MME 2015

33rd international conference

Mathematical Methods in Economics

September 9-11, 2015, Cheb, Czech Republic

CONFERENCE PROCEEDINGS

33rd International Conference

Mathematical Methods in Economics

MME 2015

Conference Proceedings

Cheb, Czech Republic
September 9 – 11, 2015

Faculty of Economics
University of West Bohemia, Plzeň

First edition
ISBN 978-80-261-0539-8

Contents

Bayesian Estimation of Time Series Models with Change Points

Václav Adamec[1]

Abstract. Time series are frequently burdened with externally induced points of change due to political decisions, changed technology or natural causes. Change point model assumes that data were generated by random process depending on the different regimes or states. This paper applies a Bayesian approach to estimation of multiple change point model of time series. It is founded upon a definition of unobserved state variable indicating a regime for sampling the specific observation. It is assumed that one or more change-points separate the regimes. Position of the change-points is unknown and must be estimated. Markov Chain Monte Carlo method is chosen to generate samples from conditional distributions of the parameters. Comparisons of alternative models are possible with quality indicators utilizing the marginal log likelihood function, primarily the Bayes factor. Diagnostic tools check the mixing properties of the sampled Markov chains required at convergence. 95% Highest Posterior Density intervals for the parameters in MCMC sample in regimes, posterior probability that observation is in specific state and posterior probabilities of regime change are supplied for every time point. This study demonstrates use of Bayesian approach to estimation and assessment of change point models applied to Poisson count problem.

Keywords: Bayesian methods, MCMC, intervention analysis, time series, multiple change-point model, Bayes factor, convergence diagnostics.

JEL Classification: C11, C52

AMS Classification: 60J10

1 Introduction

Many theoretical statisticians and practitioners believe that discovery and application of Bayesian principles in statistical inference to be one of the most fundamental developments that occurred at the end of the last century in modern statistical science. Bayesian inference implies application of a probability model to observed data and summarizing the outcome by a probability distribution of the parameters and unobserved data, such as predictions of new observations. A typical feature of Bayesian approach to statistical analysis is use of probability statements to account for the uncertainty about conclusions formulated from the available data (Gelman et al. [3]). In Bayesian statistics, all unknowns, including model parameters are treated as random variables. Bayes formula is then used to update beliefs about parameter likely value, as more data become available [13].

Despite being considered challenging, Bayesian statistical inference essentially provides simplicity to the analyst for drawing conclusions about the unobserved parameters of interest. For illustration, Bayesian probability interval for some unknown parameter can be interpreted directly as containing the unobserved parameter with high probability. It is an interpretation, which is more easily understood to and remembered by a non-professional. In the traditional "frequentist" approach, nonetheless, the confidence interval for a population parameter can be explained only in relation to repeated sampling (Gelman et al. [3]). For this reason, in Bayesian statistics, construction of high posterior probability intervals and inferences drawn from quantiles of the simulated posterior distributions became very popular and currently, they are preferred to statistical tests about the unknowns.

Widespread use of Bayesian statistical techniques was rather scarce during recent times. The reasons could be attributed not only to lack of knowledge or training, since most statistical departments at universities consider the traditional "frequentist" approaches founded on repeated sampling more appealing, but also because existing limitations in computing power and available software. Bayesian methods require more computationally intense procedures, such as high dimensional numerical integration to describe the posterior distribution of the model

[1] Mendel University in Brno, School of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic, email: vadamec@mendelu.cz.

parameters (Martin et al. [6]). Improvements in computer technology and development of new software packages in recent years contributed greatly to rising attractiveness of these methods.

According to Gelman et al. [3] Bayesian analysis should be completed in sequence of three distinct steps:

- Formulating a full probability model, a joint probability distribution of the observed and hidden quantities of the system. The probability model must be in accordance with knowledge about the underlying scientific theory and the way data were collected.
- Conditioning on the collected data is performed, calculating the correct posterior distribution - the conditional probability distribution of the unknowns, given the observed data.
- Evaluating the model fit and properties of the posterior distribution. Analysts must check, how well the model fits the data, how much sensitive are the numerical findings to the model assumptions and whether conclusions from the model are in accordance with user expectations and the underlying theory.

If the model is not found acceptable, one or several assumptions about the model in step one must be changed and the series of steps is repeated.

An important step in Bayesian analysis is formulation of prior beliefs about values of the parameters. With shortage of prior knowledge about the model parameters, the analyst may choose to apply so called "flat" or weak non-informative prior, which provides little or no information about the parameters, compared to that contained in data. Bayesian and non-Bayesian analysis then often arrive at similar conclusions with weak prior information about parameter values, since in both analyses the conclusions were driven mainly by the available data. Further, it is of great practical advantage to choose a prior distribution in the family of so-called conjugate distributions. In such case, the posterior distribution is in the same family of distributions, as the prior, which makes the calculations of the posterior uncomplicated.

Bayesian methods, like Markov chain Monte Carlo (MCMC) allow the analyst to consider a model of probability for the observed data of often-limited size and incorporate his/her prior beliefs about values of the model coefficients or distribution of random variables in the process of estimation or statistical inference. Since its discovery (see Metropolis et al. [7]), it helped circumvent the difficult task of obtaining analytical integrals required for formulating the posterior probability distributions. MCMC generates a Markov chain, whose stationary distribution after convergence approximates the posterior. The simulated samples from the posterior distribution are later used to compute the Bayes estimators of the expected value, standard error, mode, quantiles and to perform the sensitivity analysis [13]. Several variants of MCMC algorithm have been developed in the past. Gibbs sampling, Metropolis – Hastings or Reversible jump can serve as examples.

Objective of this study is to detect presence and exact timing of structural change-points in annual time series of Poisson distributed counts via Bayesian approach described in Chib [2], Park [8] and Martin et al. [6]. Comparison of candidate models, posterior probability of change-point presence, probability of regime change and Bayesian model output shall be provided. Statistical analysis and preparation of plots were secured with R-software, version 3.2.0. [10] and special add-on packages, mainly the MCMCpack [6] and coda [9].

2 Material and Methods

In this study, we applied Bayesian change-point model by Chib [2] to the Poisson count data. The time series data are assumed to stand for Poisson distributed counts $y_t \sim Poisson\,(\lambda_m)$, $m = 1,2,...,M$, generated from numerous states or regimes separated by *M* change-points. Objective of the Bayesian analysis is to determine the number and position of the change-points and quantify the impact of the regimes on the process generating the data. In the Bayesian model, the parameters of the Poisson distribution in regimes are distributed $\lambda_m \sim Gamma(\gamma, \delta)$, where $\gamma > 0$, $\delta > 0$ are the shape and scale parameters of the Gamma distribution, respectively. Probabilities that observation is in *m*-th state are distributed $p_m \sim Beta(\alpha, \beta)$, where $\alpha > 0$, $\beta > 0$ are parameters of the Beta distribution.

Several prior and posterior distributions are considered in the Bayesian analysis following the three-step procedure of Gelman et al. [3]. The prior distribution with probability density $\pi(\boldsymbol{\theta})$ reflects the analyst beliefs about the parametric vector $\boldsymbol{\theta}$ with presumed continuous distribution. The prior disregards the distribution of the observed data. The conditional distribution $\pi(\mathbf{y} \mid \boldsymbol{\theta})$ is the marginal likelihood of the empirical data, given the parametric vector $\boldsymbol{\theta}$ of the model and effectively, it is the sampling distribution of the data. The joint probability distribution of the observed and latent quantities of the system is formulated as product of the prior and the like-

lihood of the data $\pi(\mathbf{y},\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})\pi(\mathbf{y}\,|\,\boldsymbol{\theta})$. The probability density of the data $\pi(\mathbf{y})$ is received by integral of the simultaneous density

$$\pi(\mathbf{y}) = \int \pi(\boldsymbol{\theta})\pi(\mathbf{y}\,|\,\boldsymbol{\theta})d\boldsymbol{\theta}\,. \tag{1}$$

Posterior conditional density of the vector $\boldsymbol{\theta}$ given data, is obtained by application of the Bayes formula

$$\pi(\boldsymbol{\theta}\,|\,\mathbf{y}) = \frac{\pi(\mathbf{y},\boldsymbol{\theta})}{\pi(\mathbf{y})} = \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{y}\,|\,\boldsymbol{\theta})}{\pi(\mathbf{y})}\,. \tag{2}$$

Markov chains from the target posterior distribution are generated with well-chosen "burn-in" sequence, because the initial Markov samples may not originate from the desired posterior distribution. The analyst must therefore verify convergence of the Markov chain to the equilibrium posterior distribution. The diagnostics is based primarily upon visual inspection and numerical indicators of Markov chain mixing, stationarity of the posterior distribution and detection of sizeable serial correlations or cross correlations.

The better the Markov chain mixing, the smaller sample size is required after the transition "burn-in" sequence to accurately estimate the expected value, variance, mode and quantiles of the posterior distribution. Whenever possible, several Markov chains can be simulated with different starting values. Under convergence, between-chain variation becomes negligible, while within-chain variability turns large. Numerical convergence indicator devised by Gelman and Rubin [4] utilizes the above principle to produce Scale Reduction Factor (SRF). Several other formal diagnostic tests do exist to assist in checking convergence, such as those proposed by Plummer et al. [9] or Raftery and Lewis [11].

Competing Bayesian models with unequal number of change-points can be pair-wise compared by the means of Bayes factors (BF) described in Kass and Raftery [5]. BF is defined as ratio of the marginal likelihoods related to candidate models A and B to be compared, i.e. $BF_{AB} = \pi(\mathbf{y}\,|\,\boldsymbol{\theta}_A)/\pi(\mathbf{y}\,|\,\boldsymbol{\theta}_B)$, and it indicates the amount of evidence in favor of model A over model B. Square matrix of natural log transformed Bayes factors is frequently provided to assess pair-wise superiority of the contending models under study. Kass and Raftery [5] suggested guidelines to assess evidence to prefer Bayesian model A to model B (see Tab. 1).

$\mathbf{BF_{AB}}$	$\mathbf{2lnBF_{AB}}$	**Evidence for model A over B**
(0, 1)	(-∞, 0)	Negative
[1, 3)	[0, 2)	Weak
[3, 20)	[2, 6)	Positive
[20, 150)	[6, 10)	Strong
[150, ∞)	[10, ∞)	Very strong

Table 1 Guidelines to assess evidence to favor model A over model B

3 Results and Discussion

Data in this study symbolize annual counts of commercial air traffic accidents with at least one fatality collected during years 1950-2014. Data embrace only the accidents of civil commercial aircraft with at least 19 passenger seats; accidents involving planes with fewer seats were excluded. Presumably, the observed counts were influenced not only by the sheer volume of the civil air traffic, rising exponentially since the 1950s, but also through advances in technology, air traffic control, weather forecast, airline staff training and quality maintenance. Data can be viewed in the top diagram of Fig. 1, accompanied by 5-year centered linear convolution filter depicting the over time steadily declining trend. A similar decreasing pattern was documented in annual human casualty figures or occurrences of complete hull destruction beyond economical repair (not shown).

In our study, six Bayesian linear models with variable number of change-points ranging from $M = 0$ to $M = 5$ were specified and corresponding Markov chains were generated for the parameters of the Poisson distribution in $M+1$ states. In Bayesian models, position of the change-points was unknown and had to be estimated. Hyper parameters of the Gamma distribution were specified $\gamma = 23$ and $\delta = 1$.

Tab. 2 summarizes the matrix of natural logs of the Bayes factors for pair-wise assessment of the candidate models. It appears, that model with three change-points $M = 3$ is the most superior over the remaining specifications, followed closely by model with $M = 2$. Further evidence for this conclusion is given in Tab. 3 showing posterior probabilities that a specific model is correct, given one candidate model is correct. Not only model with

$M = 3$ has the highest posterior probability $p = 0.576$ of being correct, but also the largest value of the log marginal likelihood, thus providing proof for this model supremacy. This Bayesian model shall be further evaluated and discussed.

Number of change-points	0	1	2	3	4	5
0	0.00	-62.70	-74.30	-75.18	-73.94	-70.73
1	62.70	0.00	-11.60	-12.48	-11.24	-8.04
2	74.30	11.60	0.00	-0.88	0.36	3.57
3	75.18	12.48	0.88	0.00	1.23	4.44
4	73.94	11.24	-0.36	-1.23	0.00	3.21
5	70.73	8.04	-3.57	-4.44	-3.21	0.00

Table 2 Matrix of the natural logarithms of Bayes factors

Number of change-points	0	1	2	3	4	5
Posterior probability	< 0.001	< 0.001	0.246	0.576	0.171	0.007
Log marginal likelihood	-282.3	-219.6	-208.0	-207.1	-208.3	-211.5

Table 3 Posterior probabilities that a specific model is correct, given one model under study is correct

For the winning model, 115 000 samples were generated in 4 chains excluding the initial 55 000 samples of the burn-in sequence and applying the chain thinning rate of 20. Gibbs sampler, a variant of MCMC, described in Robert and Cassella [12], was used to generate the Markov chains. The size of the burn-in sequence and the thinning rate were chosen to ease convergence to a stable posterior distribution and improve the chain mixing, since smaller values of these adjustment techniques appeared inadequate. Consequently, 3 000 good samples were retained in each chain, since effective sample size adjusted for autocorrelation suggested so many samples. It should be mentioned, that unbiased estimation of the posterior means does not require zero autocorrelations of lagged chain samples, however, posterior variance estimation can be affected and efficiency may be lost. Coefficients of 1-st order serial correlations can be found in Tab. 4. Patterns of ACF showed swift exponential decay towards zero.

Parameter	Period	Mean	SD	Mode	HPD_L	HPD_U	$SRF_{0.500}$	$SRF_{0.975}$	ρ_1
λ_1	1950 – 1957	26.20	2.84	24.59	21.14	31.32	1.00	1.00	0.315
λ_2	1958 – 1977	29.15	4.35	31.43	19.88	30.00	1.00	1.00	0.584
λ_3	1978 – 2000	21.21	1.46	21.30	18.49	24.03	1.00	1.00	0.102
λ_4	2001 – 2014	11.25	0.95	11.22	9.47	13.07	1.00	1.00	0.023

Table 4 Bayes estimators of the posterior distribution and indicators of convergence

Gelman and Rubin convergence diagnostic [4] was prepared from 5 independently generated chains with dispersed starting values. Median and 97.5 % Scale Reduction Factors < 1.1 indicate likely convergence in distribution and good chain mixing (Tab. 4). Stability of the distribution was also verified by the Kolmogorov-Smirnov test of distribution equality in the first and the second half of the Markov chain [12]. Stability of the ECDF of the posterior distribution (Markov chain) can be further illustrated in plots of sample quantiles for selected probabilities depending on the Gibbs sampler iterations (see Fig. 1). Simulated Markov chains after post-simulation adjustment were used to obtain Bayes estimators of the posterior expected value, posterior standard deviation, posterior mode and to construct 95 % Highest Posterior Density intervals (see Tab. 4). Estimated quantiles of interest from the posterior distribution of the Poisson parameters λ_m are provided in Tab. 5.

Parameter	Period	$Q_{0.025}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.975}$
λ_1	1950 – 1957	20.76	24.09	26.19	28.54	31.11
λ_2	1958 – 1977	19.77	28.50	30.56	31.89	34.95
λ_3	1978 – 2000	17.40	20.56	21.31	22.05	23.55
λ_4	2001 – 2014	9.52	10.60	11.22	11.88	13.17

Table 5 Estimated quantiles of the posterior distribution

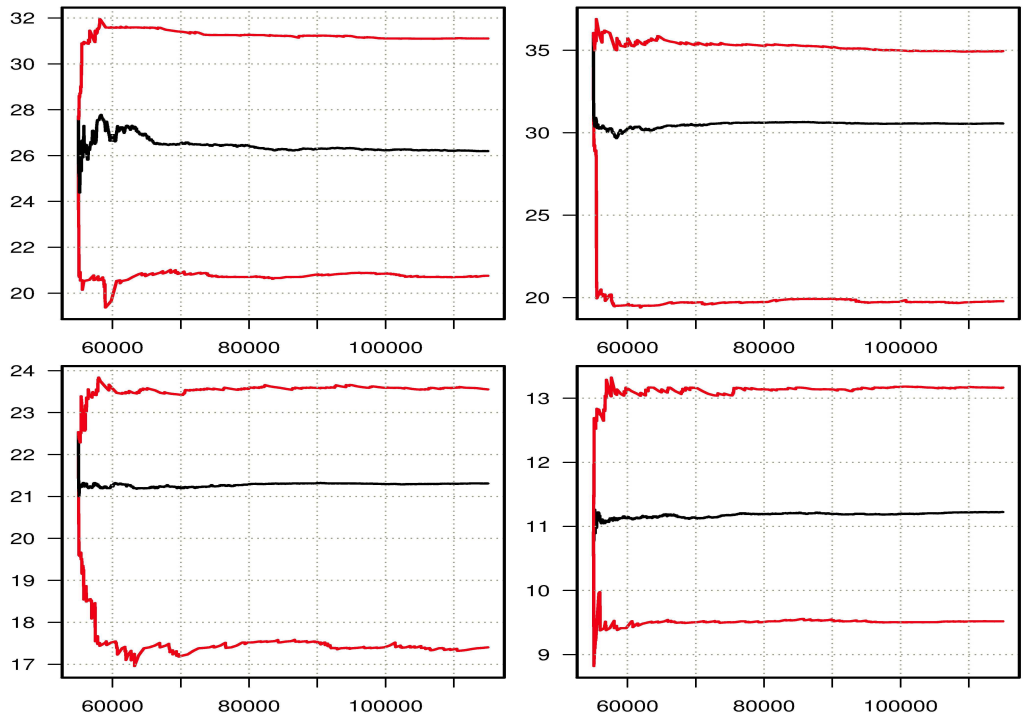


Figure 1 Plots of 2.5%, 50% and 97.5% empirical quantiles against the Gibbs sampler iterations. Quantiles in the 1-st regime are in the top left, 2-nd regime top right, 3-rd regime bottom left and 4-th regime bottom right

The model of our choice detected three change-points located in 1958, 1978 and 2001 with four separate regimes affecting the incidence of fatal airline accidents. Posterior probability that an observation changes between adjacent regimes is shown in Fig. 1 (middle diagram), where estimated change-points correspond to local peak probabilities. Posterior probability of y_t observation membership in one of four possible regimes is in the bottom graph. Posterior means, medians and modes for λ_m parameters in regimes do indicate overall decreasing tendency in fatal accidents during the period in this study (Tabs. 4 and 5). Decreasing posterior standard deviation from regime 1 to 4 points to increasing stability in the annual number of accidents, as time progresses. Timing of the change-points could be attributed to technology advancements and strict safety regulations introduced progressively in the commercial aviation.

In early stages of commercial air travel in 1950s and 1960s, the passenger aircraft, flight control and other systems of the air traffic were relatively unsophisticated and safety restrictions were relaxed. It led to comparatively high deadly accidents rates in regimes one and two in relation to the low number of flights that amounted to tens of thousand or hundreds of thousand per year, in this period. Around 1978, most flights were operated by more advanced second-generation passenger aircraft with integrated Automated Flight System, which partially contributed to lower accident rates and declining trend of accidents in the third regime. During this period, the number of passenger flights ranged from 10 to 15 million per year.

Additional substantial reduction in airliner accidents can be found in the fourth regime starting in 2001. Estimated quantiles of the posterior distribution for parameter λ_4 indicate low variations in annual counts of accidents in this period and favourable properties of this estimate. The noticeable fall was likely caused by cumulated effects of sweeping overhaul of aviation safety procedures that followed after events of September 11, 2001 and continuing penetration of fourth-generation airplanes equipped with modern Fly-By-Wire system with flight envelope protection, which prevents unusual events from happening during flight. During this period, occurrence of some types of plane accidents, as a result, was brought to near zero. It should be emphasized that the very low accident counts in regime four were achieved despite the fact that yearly numbers of commercial flights climbed to levels between 15 and 30 million.

When discussing aircraft safety, commercial aircraft manufacturers put special emphasis on the role of improvements in aviation technology. For example, Airbus [1] publicly claim that launch of the new generation of passenger aircraft causes the rate of accidents per million flights to reduce by half, compared to the previous generation. Results from the current Bayesian model do not contradict this statement. Extra focus on safety, convenience and passenger comfort contributed undoubtedly to high value and attractiveness of the services offered by commercial airlines.

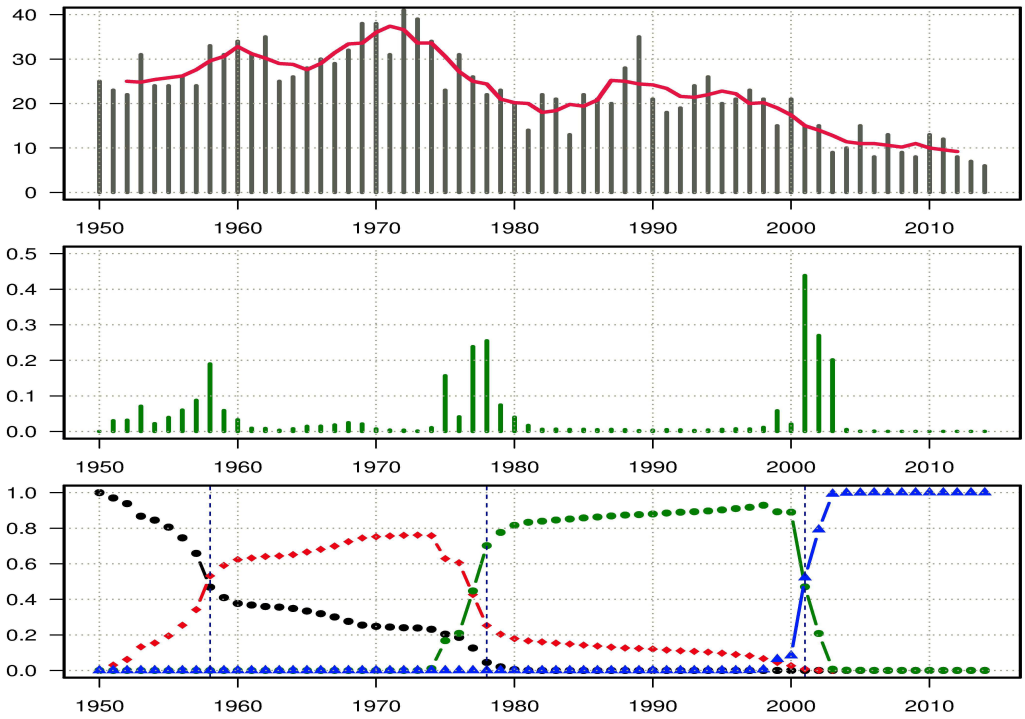


Figure 2 Observed data with 5-year moving average trend (top), posterior probability of adjacent regime change at given time (middle) and posterior probability that observation is in one of the four possible states (bottom).

4 Conclusions

In statistical practice, available data can be of insufficient size or inferior quality. In this situation, Bayesian techniques can be introduced to effectively help solve research problems, traditional approaches fail to address, as they combine the limited empirical information with analyst's prior knowledge about the unknown parameters. In this study, we fitted Bayesian change-point model to count data of air traffic accidents with fatalities and found evidence of falling accident rates, especially after 1978 and 2001. Replacement of old aircraft with those with innovative technology and added safety features reduces chances for unusual events and decreases probability of accidents, despite rising volume of air traffic.

References

[1] Airbus S.A.S.: *Commercial aviation accidents 1958-2013, a statistical analysis.* Airbus S.A.S., April 2014.
[2] Chib, S.: Estimation and comparison of multiple change-point models. *Journal of Econometrics* **86** (1998), 221–241.
[3] Gelman, A., Carlin, J. B., Stern, H.S., and Rubin, D.B.: *Bayesian data analysis*, Chapman & Hall / CRC, 1995.
[4] Gelman, A., and Rubin, D.B.: Inference from iterative simulation using multiple sequences. *Statistical Science* **7** (1992), 457–511.
[5] Kass, R., and Raftery, A.: Bayes factors. *Journal of American Statistical Association* **90** (1995), 773–795.
[6] Martin, A. D., Quinn, K. M., and Park, J. H.: MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software* **42** (2011), 1–21.
[7] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21** (1953), 1087–1092.
[8] Park, J. H.: *A change-point approach to intervention analysis using Bayesian inference.* Presented at the 2012 Annual Meeting of Korean Statistical Society.
[9] Plummer, M., Best, N., Cowles, K., and Vines, K.: CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** (2006), 7–11.
[10] R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2015, http://www.R-project.org/.
[11] Raftery, A. E., and Lewis, S. M.: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* **7** (1992), 493–497.
[12] Robert, C.P., and Cassella, G.: *Introducing Monte Carlo methods with R.* Springer, New York, 2010.
[13] Ruppert, D.: *Statistics and Data Analysis for Financial Engineering*. Springer, New York, 2011.

Equilibrium exchange rate in the Czech Republic

Emil Adámek[1]

Abstract. Lots of possible explanations of the development of the exchange rate can be found in the available literature. The aim of this paper is to find which factors influence the development of the Czech currency (measured as real effective exchange rate) using the Equilibrium Exchange Rate Approach (EER) and to assess periods with over valuated and under valuated exchange rate. This approach defines variables that cause changes in the level of the exchange rate and assess its effect on equilibrium exchange rate. In the case of empirical analysis, there exist a lot of econometric methods that can be used. This article is based on the (Johansen) cointegration analysis. Since principal of EER is to find lung run equilibrium, Vector Error Correction Model (VECM) is constructed. Economic theory defines some variables that influence equilibrium exchange rate. The most important of them are the labour productivity, interest rates and government or foreign debt. These variables are also used in this article. We use quarterly data from 2000 to 2014.
Keywords: equilibrium, exchange rate, cointegration, VECM.

JEL Classification: C22, E52, E58
AMS Classification: 62-07

1 Introduction

Komárek and Motl [10] declare four reasons why central bank should monitor estimates of equilibrium exchange rate. The first one is to gather knowledge for monetary policy implementation. The second motive is that knowledge of EER level helps central banks to set policy instruments. The third reason is that exchange rate is a key factor to evaluation competitiveness of whole economy. The last reason, which is very important especially in the case of Czech Republic, is that information about EER is crucial factor while setting central parity or conversion ratio before joining common monetary union. The aim of this paper is to find which factors influence the development of the Czech currency (measured as real effective exchange rate) using the Equilibrium Exchange Rate Approach (EER) and to assess periods with over valuated and under valuated exchange rate. Firstly, there are mentioned theoretical and empirical backgrounds of EER. Then, there are described used methods and data. In the next chapter, the results are presented. Conclusions remain last part of this paper.

2 Theoretical and empirical backgrounds

2.1 Concepts of equilibrium exchange rate

There exist a lot of concepts of EER. The classification of EER concepts is not united. One of the oldest theories is the Purchasing Power Parity Theory (PPP) which proposes that the only determinant of exchange rate development is price level (in the case of absolute version of PPP), respectively inflation rate (in the case of relative one). Nevertheless, there exist some reasons (such as the Ballasa-Samuelson effect) why this theory does not provide good explanation of exchange rate development especially in the case of transitive economies. Another traditional theory is the Uncovered Interest Parity Theory (UIP). UIP declares that expected change in the nominal exchange rate is determined by the interest rate differential. Capital Enhanced Equilibrium Exchange Rate (CHEER) is combination of PPP and UIP. It was developed by Johansen and Juselius [8]. Unlike PPP, CHEER deals with capital flows. It claims that the difference in domestic and foreign interest rates can cause misalignment between spot exchange rate and exchange rate defined by PPP. Behavior Equilibrium Exchange Rate (BEER), created by MacDonald [13], is based on above mentioned approaches but it also tries to find other determinants of exchange rate development. Škop and Vejmělek [15] call it a statistical approach because it uses statistical methods to find determinants of exchange rate development (so called *data mining*). BEER approach is the most similar approach to this paper. For other approaches to EER such as Fundamental Equilibrium Exchange Rate (FEER), Desired Equilibrium Exchange Rate (DEER), Atheoretical Permanent Equilibrium Exchange Rate (APEER), Permanent Equilibrium Exchange Rate (PEER) or Natural Real Exchange Rate (NATREX), see e.g. Driver and Westaway [5] or MacDonald [12].

[1]VŠB-Technical University of Ostrava, Faculty of Economics, Department of National Economy, Sokolská tř. 33, 701 21 Ostrava 1, Czech Republic, e-mail: emil.adamek@vsb.cz

2.2 Determinants of exchange rate

As it was mentioned above, BEER concept tries to identify variables that determine equilibrium exchange rate. First of all, it is important to discuss which kind of exchange rate is explained. Some concepts use nominal exchange rate or real exchange rate whereas others use real effective exchange rate. For further discussion see e.g. [5]. In this paper real effective exchange rate is used according to original MacDonald`s approach [13]. One of the factors that influence EER is *real interest rate differential*. According to UIP, a currency with a negative interest rate differential is expected to appreciate (because of impossibility of arbitrage opportunity). Also increasing interest rate differential will cause higher capital flows to domestic country and higher demand for its currency. Next determinant are *net foreign assets (NFA)*. Deficit of current account leads to increase of foreign debt, which is financed by foreign investors, who demand (to adjust their portfolios) a higher yield. At given interest rates, this can only be accomplished through a depreciation of the currency of the debtor country. Nevertheless, changes of NFA in reaction to exogenous shocks could be offset by adjustments in the capital stock. Since changes in the net foreign asset position and changes in the capital stock have countervailing effects, the impact on the EER may be ambiguous as Maeso-Fernandes et al. [14] claim. *Productivity differential* is very important, especially in the case of transitive economies. A higher average productivity in the domestic economy relative to the foreign leads to appreciation of domestic currency. This is known as the Ballasa-Samuelson effect. Another determinant is a *government debt*. It reflects time variant risk. Higher levels of government debt lead to depreciation of domestic currency.

2.3 Review of empirical literature

Komárek and Motl [10] compute BEER and FEER for Czech Republic using quarterly data (1996:Q1 – 2011:Q4). They estimate real (bilateral) exchange rate (CZK/EUR). As concerns BEER, the cointegration analyses are used. Depending variables are productivity difference, net foreign assets and ratio of real investments and export on GDP. The results confirm the estimates of exchange rate overvaluation by mid-1997 and in 1998, 2002 and 2008. Since 2009, the model suggests a significant slowdown in the appreciation of the equilibrium. Škop and Vejmělek [15] estimate NATREX. Quarterly data covered 1995:Q1 – 2007:Q4 periods. They constructe VECM while depending variables are output gap, time preference gap and terms of trade gap. They find that exchange rate is overvaulted in 1995-1998 periods and that NATREX appreciated in long run. Babetskii and Balázs [1] estimate BEER for Czech Republic. They use monthly data from 1993:M1 to 2004:M9. They apply several alternative cointegration techniques and identify a period of an overvaluation in 1997 and in 1999, an increasing overvaluation until 2002, an undervaluation in 2003 and a correction towards equilibrium in the second half of 2004. Komárek and Melecký [11] estimate BEER in 1994:Q1 – 2004:Q1. As determinants of the real equilibrium exchange rate they consider the productivity differential, the interest rate differential, the terms of trade, net foreign direct investment, net foreign assets, government consumption and the degree of openness. They find that the exchange rate is on average undervalued over the period 1994 to 2004 by about 7 percent with respect to the estimated BEER.

3 Methods and data

In this chapter, used methods and data are described. To evaluate EER, Vector Error Correction Model (VECM) is constructed. Quarterly data cover 2000:Q2 – 2014:Q3 periods.

3.1 Methods

In the case of cointegration analysis, it is needed that all (or at least some of them as Clark and MacDonald [3] show) variables are integrated on the same level (usually I(1)). That is why tests for stationarity are run. After the detection of integration orders, the cointegration analysis is conducted by Johansen cointegration tests. In the last step, the VECM is constructed. The theoretical backgrounds of mathematical methods follow.

Unit root tests

Dickey-Fuller test (DF) was developed by Dickey and Fuller [4]. There exist three variants of DF (so called (τ-tests):

$$\Delta y_t = \phi_1 y_{t-1} + \varepsilon_t\,, \tag{1}$$

$$\Delta y_t = \beta_0 + \phi_1 y_{t-1} + \varepsilon_t\,, \tag{2}$$

$$\Delta y_t = \beta_0 + \phi_1 y_{t-1} + \beta_2 + \varepsilon_t , \tag{3}$$

where β_0 is an intercept, β_2 is a trend and ε_t is residual component. Equation (1) includes neither intercept nor trend (it is so called model of Random Walk); equation (2) includes intercept and equation (3) includes both intercept and trend. The test statistic is defined as

$$DF = \frac{\hat{\phi}}{\hat{\sigma}(\hat{\phi})} \tag{4}$$

The null and the alternative hypothesis for all tests are H_0: $\phi = 0$ time series has a unit root and H_1: $\phi < 0$ time series has not a unit root. However this version of the DF neglects serial autocorrelation, as Heij et al [7] claim. If depended variable is conditionally autocorrelated DF cannot be used. That is why Augmented DF test (ADF) is largely used. Former equations are replaced by:

$$\Delta y_t = \phi_1 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \varepsilon_t , \tag{5}$$

while test statistic and critical values remain unchanged.

Cointegration analysis

Johansen cointegration tests (Johansen [9]) are the most common tests of cointegration[2] as Cipra [2] claims. The advantage of these tests is that they permit more than one cointegration relationship. There exist two types of Johansen cointegration tests:

1. λ_{trace} statistic:

$$\lambda_{trace}(r) = -n \sum_{i=r+1}^{m} \ln(1 - \hat{\lambda}_i) , \tag{6}$$

with hypothesis H_0: number of cointegration relationships ≤ r and H_1: number of cointegration relationships > r.

2. λ_{max} statistic:

$$\lambda_{\max}(r) = -n \cdot \ln(1 - \hat{\lambda}_{r+1}) , \tag{7}$$

with hypothesis H_0: number of cointegration relationships = r and H_1: number of cointegration relationships = r + 1.

3.2 Data

Real effective exchange rate (*REER*)

Real effective exchange rate (*REER*) can be generally expressed as:

$$RER = E_{it} = \prod_{j=1}^{n} \left(\frac{P_{it} S_{ijt}}{P_{jt}^{*}} \right)^{w_{ij}} , \tag{8}$$

where P_i denotes domestic price level, P_j^* measures price level in country *j*, S_{ij} is the nominal exchange rate and w_{ij} is the weight of country *j* in country *i*'s effective exchange rate index. In this paper *RER* is deflated by PPI. Data are used from ARAD – data series system. Data in natural logarithm are used.

Real interest rate differential (*SIR*, *LIR*)

The difference between real domestic and foreign interest rate is computed for both long run (*LIR*) and short run (*SIR*). *LIR* is measured by Maastricht criterion interest rate (10 years government bonds) - IR_L, whereas *SIR* as

[2] For another possibility – Engle-Granger test, which is based on DF, [6]

money market interest rate - IR_S. Both *LIR* and *SIR* is obtained from Eurostat database and deflated by HICP. They are computed as:

$$SIR = (IR_S - \pi) - (IR_S^* - \pi^*), \tag{9}$$

$$LIR = (IR_L - \pi) - (IR_L^* - \pi^*), \tag{10}$$

where π denotes inflation rate (computed from HICP) and * remarks foreign (euro area).

Net foreign assets (*NFA*)

Net foreign assets are measured as a ratio of Net International Investment Position to GDP in absolute value (since all values are negative) so this variable represents deficit of international investment position. This variable is in natural logarithm form. Data are obtained from ARAD – data series system.

Productivity differential (*PROD*)

Productivity differential is approximated by variable *PROD*, which is computed as a ratio between domestic (Czech) and foreign (euro area) real GDP (*RGDP*) per employee (*EMP*):

$$PROD = \frac{\frac{RGDP}{EMP}}{\frac{RGDP^*}{EMP^*}}, \tag{11}$$

where * denotes euro area. Data are gathered from Eurostat database. Time series is seasonal adjusted using Census X12 method in EViews 7.

Government debt (*DEBT*)

Variable *DEBT* is the ratio of domestic and foreign ratio of government debt to GDP. It is computed as follows:

$$DEBT = \frac{\frac{D}{GDP}}{\frac{D^*}{GDP^*}}, \tag{12}$$

where *D* is government debt, *GDP* is nominal GDP and * remains foreign country (euro area in this case). Data are obtained from Eurostat database. Missing values for euro area in period from 2001 to 2005 are computed by linear interpolation using EViews 7 software.

Therefore, real effective exchange rate can be written as:

$$REER = f(SIR, LIR, NFA, PROD, DEBT). \tag{13}$$

4 Results

The results of mathematical model are presented in this chapter. There are constructed several models while two of them are presented here (Model 1 and Model 2). Model 2 is used to compute EER for Czech Republic.

As it was mentioned above, while constructing VECM, all variables have to by integrated on the same level. ADF test results are depicted in Table 1. It can be seen that short run interest rate differential is stationary at level. Other variables are I (1).

Because *SIR* is stationary in level, this variable is not included into a model. Model 1 based on equation (13) is constructed as:

$$\ln REER_t = \alpha_0 + \alpha_1 \cdot LIR_t + \alpha_2 \cdot \ln(NFA)_t + \alpha_3 \cdot PROD_t + \alpha_4 \cdot DEBT_t + \varepsilon_t. \tag{14}$$

Based on VAR Lag Order Selection Criteria, lag 1 is chosen to run Johansen cointegration test. Both Trace Test and Maximum Eigenvalue test confirm that there is 1 cointegration relationship. Results of Model 1 are depicted in Table 2. Error Correction Term is negative and significant at 5 %.

	levels		**1st differences**	
	statistics	***p*-value**	**statistics**	***p*-value**
lnREER	-2,1587	0,2234	-7,6163	0,0000
SIR	-3,3275	0,0182	-4,6146	0,0004
LIR	-1,6826	0,4345	-6,2424	0,0000
lnNFA	-2,5435	0,1109	-8,0760	0,0000
PROD	-2,6628	0,0868	-5,8234	0,0000
DEBT	-1,9237	0,3194	-9,0558	0,0000

Source: own calculation

Table 1 Results of ADF test

	constant	**LIR**	**lnNFA**	**PROD**	**DEBT**
statistics	5,093611	-0,002741	0,284954***	-0,49394	-0,823125***
st. deviation		0,01019	0,04241	0,5221	0,1618
ECT = -0,3051**	(-0,134)	R^2 = 0,1465		Durbin-Watson = 1,9794	

Source: own calculation

Table 2 Output of Model 1

Nevertheless, two variables (*LIR* and *PROD*) are not significant. Therefore elimination methods are used. The combination of determinants which provides best results is depicted in Model 2:

$$\ln REER_t = \alpha_0 + \alpha_1 \cdot \ln(NFA)_t + \alpha_3 \cdot DEBT_t + \varepsilon_t . \tag{15}$$

	constant	**lnNFA**	**DEBT**
statistics	4,825236	0,243285***	-0,704534***
st. deviation		0,02586	0,15492
ECT = -0,4161*** (-0,1406) R^2 = 0,1818, DW = 2,05			

Source: own calculation

Table 3 Output of Model 2

The output of Model 2 is depicted in Table 3. ECT is negative and significant at 1 %. From Table 3, it is obvious that higher level of deficit in NFA[3] leads to the appreciation of currency and higher government debt leads to depreciation of currency. In table 4, there are expected and measured signs. All variables either have a right sign or are not significant.

	sign	
	expected	**model**
LIR	+	
lnNFA	+/-	+
PROD	+	
DEBT	-	-

Source: own calculations

Table 4 Expected and model signs

[3] Once again higher values of variable NFA represents higher deficit of Net International Investment Position.

5 Conclusions

The first aim of this paper is to find determinants of exchange rate development. Net foreign assets and government debt are found as the significant determinants. Both variables are largely used in other papers. The second aim is to find periods with under valuated and over valuated real effective exchange rate.

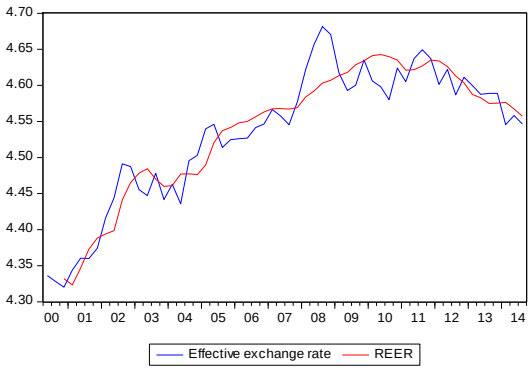


Chart 1 Effective exchange rate and REER (in logarithm)

Development of both real effective and equilibrium exchange rate is depicted in Chart 1. Real effective exchange rate is overvaulted in periods 2001-2002, 2004-2005 and 2007 to mid-2009. There is appreciation trend from 2000 to 2010 of equilibrium. Since the 2009-2011, EER has depreciated. This depreciation trend is also consistent with others paper, for example Komárek and Motl [10].

Acknowledgements

This paper was financially supported within the VŠB - Technical University SGS grant project No. SP2015/115 "Institutional and Monetary Context of Economic Integration of European Countries Today"

References

[1] Babetskii, J., and Balázs, É.: Equilibrium exchange rate in the Czech Republic: How good is the Czech BEER? *CERGE-EI, Working Papers*. 2005, **267**.
[2] Cipra, T.: *Finanční ekonometrie*. Praha: Ekopress, 2013.
[3] Clark, P., and MacDonald, R.: Exchange Rates and Economic Fundamentals: A Methodological Comparison of BEERs and FEERs. *International Monetary Fund.* 1998.
[4] Dickey, D., and Fuller, W.: Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*. 1979, **74**(366): 427-431.
[5] Driver, R., and Westaway, P.: Concepts of equilibrium exchange rates. *Bank of England, Woriking Paper*. 2003, **248**.
[6] Engle, R., and Granger C. W. J.: Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*. 1987, **55**(2): 251-276.
[7] Heij, Ch., Boer, P., Franses, P. H., Kloek, T., and Dijk, H.: *Econometric Methods with Applications in Business and Economics*. New York: Oxford University Press Inc., 2004.
[8] Johansen, S., and Juselius, K.: Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Jurnal of econometrics*. 1992, **53**: 211-244.
[9] Johansen, S.: Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*. 1991, **59**(6): 1551-1580.
[10] Komárek L., and Motl, M.: Behaviorální a fundamentální rovnovážný měnový kurz české koruny. 2012, **60**(2): 147-166.
[11] Komárek, L., and Melecký L.: The Behavioural Equilibrium Exchange Rate of the Czech Koruna. *ČNB Working Paper Series.* 2005, **5**.
[12] MacDonald, R.: Concepts of equilibrium exchange rates: An Overview. Economic Research Group of the Deutsche Bundesbank, *Discussion Paper 3/00*. 2000.
[13] MacDonald, R.: What determines real exchange rates? The long and the short of it. *Journal of International Financial Markets, Institutions and Money*. 1997, **8**: 117-153.
[14] Maeso-Fernandez, F., Osbat, Ch., Schnatz, B.: Determinants of the Euro Real Effective Exchange Rate: A BEER/PEER Approach. *European Central Bank – Working Paper **85***.2001.
[15] Škop, J., and Vejmělek, J.: Od parity kupní síly k NATREXU – případ české koruny. *Politická ekonomie.* 2009, **57**(3): 323 – 343.

Evaluation of the state of e-government in European Union countries

Eva Ardielli[1]

Abstract. E-government belongs to important trends of modernization efforts of public administration and it is also the subject of various international comparisons. The paper is focused on the comparison of the current state of e-government in the European Union countries and evaluation of the position of the Czech Republic by applying the methods of multi-criteria evaluation of alternatives. The method used in the comparison is TOPSIS (The Technique for Order Preference by Similarity to Ideal Solution). It is one of the methods of multi-criteria evaluation of alternatives, which in the result determines the overall order of alternatives. The method is based on the assumption of the existence of the matrix, including the final list of alternatives, ranked according to the final number of criteria.
In the research, there was selected the final list of alternatives (countries EU-28) and criteria (9 e-government indicators). The research is based on data set across multiple data sources. These are mainly: "eGovernment Benchmark" study from 2014, data processed by Eurostat and data managed by the United Nations. Evaluated data describe the state of e-government in the year 2013.

Keywords: e-governmet, evaluation, TOPSIS, European Union.

JEL Classification: H11
AMS Classification: 91B10, 91B06, 90B50, 62C86

1 Introduction

E-government is one of important current trends of public administration modernization and it is also the subject of international comparisons, as discussed by Khosrow-Pour [10], West [16] or Bannister [1]. Interpretation of the term "e-government" is quite broad and divergent. The general definition describes e-government as the use of information and communication technologies (ICT) in a way of government transformation in order to increase the availability, effectiveness and accountability.

The importance of e-government in the context of the modernization of public administration is dealt also with domestic authors such Hendrych [7] or Grospič [6]. E-government here does not simply represent the direction of modernization of public administration, but it is also discussed as a tool for modernizing public administration.

The aim of the paper is the comparison of the current state of e-government in the European Union countries and evaluation of the position of the Czech Republic (CR). The state of e-government is evaluated in countries EU-28 in the year 2013 based on the indicators of e-government monitored by international institutions. The structure of the article is devoted to introduction and theoretical statement, where attention is paid to the importance of information and communication policy and e-government as one of the important tools for the development of the information society. The results of empirical research then evaluate e-government in the CR and in EU countries by selected criteria, using TOPSIS method.

1.1 E-government in European Information Policy

Fundamentals of the political decision to invest in e-government services have been established in the Lisbon Strategy, which was approved by EU member states in the year 2000. As stated by Pekárek [12], the contemporary concept of e-government in the EU is based on the original eEurope initiative (An Information Society For All), which has been promoted in the EU since 1999.

E-government is in European information policies considered as one of the tools for building an information society. Prospectively, it is assumed here the certain standardization of supply range of electronic public services and of the way of services providing, which should ensure interconnectivity of some e-government solutions among EU member states. It is spoken about "Pan-European e-services", see [14]. In 2002 there was introduced

[1] VSB-Technical University of Ostrava/Faculty of Economics, department of Public Economics, Sokolská Street 33, Ostrava, Czech Republic, eva.ardielli@vsb.cz.

continuing "eEurope 2005" initiative, where e-government took also significant position. Currently is the essential European strategic framework of e-government called "Digital Agenda for Europe". The document was adopted in 2010 and it is the part of the "Europa 2020" initiative. Strategy primarily highlights the current variability in the provision of e-government services across the EU, the lack of cross-border coverage and low utilization by citizens. In order to implement the European policy in the field of e-government and its enforcement in practice across EU member states there are created so called "Action plans". These documents contain specific measures and recommendations to successful implementation of e-government. The current action plan is called „The European eGovernment Action Plan 2011-2015", see [3] and it was adopted in 2010. The European Commission has specified here the objectives of the Digital Agenda strategy in the field of e-government.

1.2 Evaluation of e-government at supranational level

E-government has been monitored as a part of the activities of many organizations. Approaches to monitoring of e-government differ considerably across organizations. For example Eurostat [4] processes and evaluates data in the field of e-government with usage of indicator measuring the interaction of citizens and businesses with public administration. The OECD has been involved in monitoring of the use of ICT in EU member countries, but e-government as a specific area is not measured here. European Commission's approach to evaluating e-government is based on evaluation of the effectiveness of European Information Policy, see [2]. This activity is based on the obligations of the European institutions. The United Nations [15] deals with the evaluation of e-government on the basis of the annual evaluation of the composite indexes of e-government and e-participation.

Most approaches aimed at assessing the overall and general state of e-government, and therefore assess performance of government at all levels of the country: federal, regional and local. Only some approaches focus on the regional or local level. However, the monitored e-government data of international organizations are not consistent with each other, as there have been monitored different time periods using different methodologies of data collecting and data processing. The organizations have been also focused on description of different subareas of e-government services, which are defined based on specific needs and purpose of the organizations. The paper focuses on the synthesis of these approaches. The synthesis output can contribute to achieving comprehensive information on the state of e-government in the EU countries based on e-government indicators monitored by Eurostat, European Commission and United Nations as well.

2 Methods

The method used in the comparison is TOPSIS (The Technique for Order Preference by Similarity to Ideal Solution). It is one of the methods of multi-criteria evaluation of alternatives. The aim of the methods of multi-criteria evaluation of alternatives is to determine the ranking of individual alternatives in terms of selected criteria, wherein the alternative with the best ranking represents the best compromise alternative, see [8]. TOPSIS method is based on the selection of alternative that is closest to the ideal alternative and furthest from basal alternative. In [5] is assumed the maximizing character of criteria. Application of TOPSIS method is as follows:

- creation of the criterial matrix that is according to [11] representing the ranking of alternatives according to respective criteria:

$$D = \begin{pmatrix} & Y_1 & Y_2 \dots & Y_j \dots & Y_r \\ A_1 & y_{11} & y_{12} \dots & y_{1j} \dots & y_{1r} \\ A_2 & y_{21} & y_{22} \dots & y_{2j} \dots & y_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_i & y_{i1} & y_{i2} \dots & y_{ij} \dots & y_{ir} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_m & y_{m1} & y_{m2} \dots & y_{mj} \dots & y_{mr} \end{pmatrix}$$

- creation of normalized criterial matrix *R* according to following formula:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{m} y_{ij}^2}}$$

where r_{ij} are elements of matrix *R*, where $i = 1,2, \dots m$; $j = 1,2, \dots r$; y_{ij} are the original input data for the alternative *i* and criterion *j*; *m* is the number of alternatives;

- calculation of weighted criterion matrix W by following equation:

$$w_{ij} = v_i * r_{ij}$$

where w_{ij} is weight normalized value and v_{ij} is weight of criterion. The acquired data matrix is multiplied by weights of respective criteria according to [13];

- determination of the ideal and basal alternative relative to the matrix values *W*, see following formulas:

$$H_j = \max(w_{ij})$$

$$D_j = \min(w_{ij})$$

for $i = 1,2, \ldots m$ and $j = 1,2, \ldots r$;

- distance calculation of alternatives to ideal alternative or from basal alternative by formula (1) and (2):

$$d_i^+ = \sqrt{\sum_{j=1}^{r} (w_{ij} - H_j)^2} \quad (1)$$

$$d_i^- = \sqrt{\sum_{j=1}^{r} (w_{ij} - D_j)^2} \quad (2)$$

for all $i = 1, 2, \ldots m$; and $j = 1,2, \ldots r$;

- calculation of the relative distance indicator of alternatives from basal alternative, see formula (3):

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (3)$$

where $i = 1,2, \ldots m$;

- arrangement of alternatives by non-growing values of c_i.

In the research, there was selected the final list of alternatives (EU-28 countries) and criteria (9 indicators of various international organizations.) The research is based on data set across multiple data sources; see [2], [4] and [15]. The summarization of monitored e-government indicators and their characteristics are shown in Tab. 1.

indicator	organization	characteristic
User Centric Government	European Commission	Shows the extent to which the service is provided online and how is the delivery perceived by the user.
Transparent Government	European Commission	Shows the extent to which governments are transparent in terms of their own responsibilities and performance.
Citizen Mobility	European Commission	Indicates the extent to which EU citizens can use online services abroad.
Business Mobility	European Commission	Indicates the extent to which businesses can use online services abroad.
Key Enablers	European Commission	Indicates the extent of online technical requirements availability as: eID, e-Documents, authentic sources, eSafe and SSO.
Online Service Index	United Nations	Describes the range and quality of online services.
E-Participation Index	United Nations	Monitors the online services and information provided by governments to citizens, the interaction among stake-holders and involvement in decision-making processes.
Internet Use – Individuals	Eurostat	Describes the percentage of individuals using the Inter-net in relation to public administration.
Enterprises Using Internet	Eurostat	Describes the percentage of enterprises using the Internet in relation to public administration.

Table 1 Monitored indicators of e-government

The evaluated data describe the state of e-government in the year 2013. To facilitate the calculations according to TOPSIS method was used the SANNA software specified in [9].

3 Results

On the basis of the TOPSIS method there was performed distance calculation from ideal and basal alternative. The coefficient of total distance of alternative i to the ideal alternative d_i+ was calculated according to formula (1). Coefficient of total distance of alternative i from basal alternative d_i- was calculated according to formula (2). Subsequently there was calculated the relative distance indicator of alternatives from basal alternative c_i. The relative distance of alternative i from the basal alternative is given by the formula (3). Values for individual alternatives are summarized in Tab. 2. The values of the calculated indicator range between 1 and 0. Value 0 corresponds to the basal alternative; value 1 corresponds to the ideal alternative, as shown in e.g. [18].

Country	d_i+	d_i-	c_i
Austria	0,03755	0,05317	0,58607
Belgium	0,04600	0,04067	0,46923
Bulgaria	0,06917	0,02068	0,23020
Croatia	0,06846	0,02001	0,22614
Cyprus	0,05405	0,03736	0,40871
Czech Republic	0,06683	0,02133	0,24195
Denmark	0,03642	0,05266	0,59116
Estonia	0,02037	0,06739	0,76787
Finland	0,02637	0,06221	0,70235
France	0,04007	0,05556	0,58097
Germany	0,05197	0,03886	0,42785
Greece	0,07189	0,02578	0,26392
Hungary	0,07086	0,01936	0,21456
Ireland	0,04574	0,04869	0,51562
Italy	0,05218	0,03769	0,41941
Latvia	0,04454	0,04591	0,50760
Lithuania	0,03309	0,05389	0,61959
Luxembourg	0,05080	0,03765	0,42562
Malta	0,03175	0,07397	0,69970
Netherlands	0,04065	0,05526	0,57616
Poland	0,05755	0,03262	0,36173
Portugal	0,03855	0,05468	0,58648
Romania	0,07772	0,01173	0,13110
Slovakia	0,07041	0,02404	0,25453
Slovenia	0,05223	0,03532	0,40343
Spain	0,04571	0,05464	0,54449
Sweden	0,03266	0,05420	0,62400
United Kingdom	0,04611	0,05168	0,52848

Table 2 Values of coefficient of total distance and relative distance indicator of alternatives

Based on the result, it is possible to determine the order of the European Union countries from the best to the worst in terms of the e-government functioning, as shown in Fig. 1.

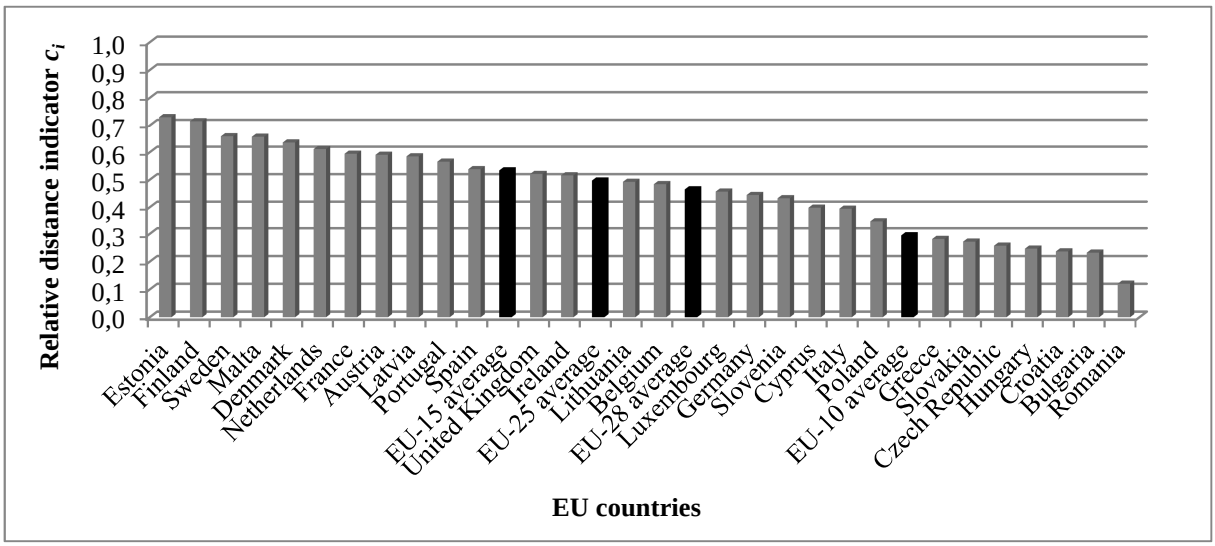


Figure 1 Evaluation of e-government state by TOPSIS method in the EU (2013)

Assessment of the state of e-government in the EU countries in 2013 showed that on the best place are Estonia and the Nordic countries - Finland and Sweden, while the worst e-government state is in Croatia, Bulgaria and Romania. The chart also captures EU-countries averages, based on which can be deduced about the difference in e-government between original EU-15 countries and the accessing countries EU-10 in 2004. The values of relative distance indicator c_i are compared in Tab. 3.

EU countries	Total relative indicator c_i
EU-15	0,53
EU-10	0,29
EU-25	0,50
EU-28	0,46

Table 3 Comparison of the relative distance indicator in EU countries

Based on the comparison of relative distance indicator c_i in EU countries, it is possible to conclude that in terms of the e-government state the EU-15 countries reach significantly better results on average (the value of c_i indicator reaches 0.53), while the EU-10 countries are lagging behind. The value of indicator c_i in the EU-10 countries reaches 0.29, which is 45 % lower value than in the case of the EU-15 countries. Here is possible to note the exceptional status of Estonia, which, though also belongs among the countries of the eastern EU enlargement in 2004, showed the best state of the e-government across the whole EU-28.

The EU countries were based on the e-government evaluation divided into three groups: countries with above-average, average and below-average e-government state see Fig. 2.

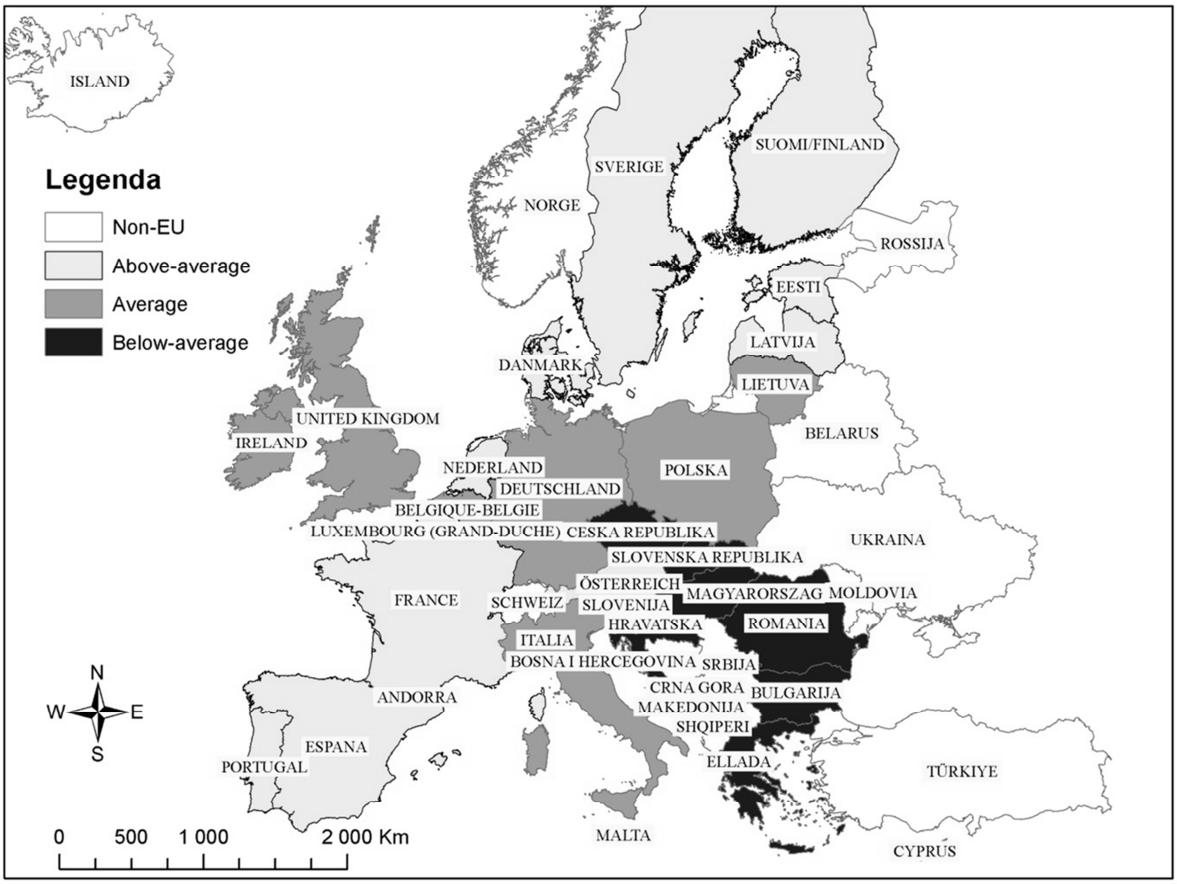


Figure 2 Breakdown of EU countries according to the e-government state (2013)

To the group of countries with above-average e-government state in EU were included the EU countries with c_i value higher than the EU-15 countries average. Among the countries with the average level of e-government were included the EU countries where the c_i value ranged between the average of EU-15 and EU-10 countries. To the group of countries with below-average e-government state were included the EU countries with c_i value lower than the EU-10 countries average. The Czech Republic is also included in this group of countries.

4 Conclusion

Based on the usage of TOPSIS method there was done the evaluation of the current state of e-government in the EU-28 and the verification of the position of the CR in international comparison for the year 2013. Results of this own research dealing with evaluation of the state of e-government in the CR reflect to a considerable extent the current results obtained on the basis of international benchmarking activities of major international institutions such as World Forum [17] or European Union [2]. The results of the evaluation of the EU-28 countries in terms of the state of e-government by TOPSIS method in the 2013 acknowledged, that the best ranking obtained Estonia, then Finland, Sweden and Malta. The worst state of e-government was reported in Romania, Bulgaria and Croatia. Based on the ranking of EU countries by the selected criteria from the best result to the worst was showed that in the average the EU-10 countries are in the field of e-government significantly lagging behind in comparison with the EU-15 countries. The value of the relative distance indicator in the EU-10 countries is only 55 % of the value of this indicator in the EU-10 countries. This indicates significantly worse state of e-government in the EU-10 countries in contrast to the EU-15 countries. One part of the research was also the evaluation of e-government in the CR in comparison with other EU countries. Based on the evaluation of the e-government state in the EU-28 in 2013 was found highly unsatisfactory position of the CR in the field of e-government. The CR ranked among the five worst countries in the EU-28, thus belongs among below-average EU countries in terms of the e-government state. The cause of the inadequacy of e-government state in the CR is mainly lack of basic concept and long-term lack of interest by Czech government. In the country there are serious shortcomings, particularly on the side of public digital services providers. Changing the attitude of government officials in this field is therefore required. E-government is a useful tool for reducing the cost of public administration and it is also the benefit for the residents in terms of time savings. This area remains therefore the main challenge to the future in the CR.

Acknowledgements

This contribution was supported by the project SP 2012/163 at VSB-TU Ostrava, Faculty of Economics.

References

[1] Bannister, F.: The curse of the benchmark: an assessment of the validity and value of e-government comparisons. *International Review of Administrative Sciences* **73** (2007), 171–188.
[2] European Commission: *E-Government Benchmark: Insight report.* Publications Office of the EU. 2014. [cit. 2015-02-20]. Available from WWW: <https://ec.europa.eu/digital-agenda/en/news/eu-egovernment-report-2014-shows-usability-online-public-services-improving-not-fast>.
[3] European Commission: *The European eGovernment Action Plan 2011-2015*. 2010 [cit. 2015-02-26]. Available from WWW: < http://ec.europa.eu/digital-agenda/en/european-egovernment-action-plan-2011-2015>.
[4] Eurostat: *Statistical database.* 2013. [cit. 2015-02-20]. Available from WWW: <http://epp.eurostat.ec. europa.eu/portal/page/portal/information_society/introduction>.
[5] Fiala, P.: *Modely a metody rozhodování.* Oeconomica, Praha, 2008.
[6] Grospič, J.: Reforma, transformace či modernizace veřejné správy? In: *Úvod do regionálních věd a veřejné správy* (Kadeřábková, J., Kouřilová, J., and Mates, P., eds.). IFEC, Praha, 2001, pp. 237-250.
[7] Hendrych, D.: *Správní věda – Teorie veřejné správy*. Wolters Kluwer ČR, a. s., Praha, 2009.
[8] Hwang, C., Yoon, K.: *Multiple Attribute Decision Making: Methods and Applications.* Springer, NY, 1981.
[9] Jablonský, J.: Software Support for Multiple Criteria Decision Making Problems. *Management Information Systems* **4** (2009), 29–34.
[10] Khosrow-Pour, M. *Practicing E-Government: A Global Perspective.* Idea Group Publishing, Hershey,2005.
[11] Milani, A.S., Shanian, A. and El-Lahham, C.: A decision-based approach for measuring human behavioral resistance to organizational change in strategic planning. *Mathematical and Computer Model-ling* **45** (2008), 1765-1774.
[12] Pekárek, A.: Analýza současných evropských trendů v oblasti e-governmentu. *Knihovna* **23** (2012), 51-62.
[13] Peng, Y., Zhang, Y., Kou, G. and Shi, Y.: A Multicriteria Decision Making Approach for Estimating the Number of Clusters in a Data Set. *PLoS one* **7** (2012), 1-9.
[14] Pomahač, R. at al.: *Veřejná správa*. C. H. Beck, Praha, 2013.
[15] UNPACS. *Data Center - Government Development Index.* 2014. [cit. 2014- 08-02]. Available from WWW: <http://unpan3.un.org/egovkb/en-us/Data-Center>.
[16] West, D. M.: E-Government and the Transformation of Service Delivery and Citizen Attitudes. *Public Administration Review* **64** (2004), 15-27.
[17] World Forum. *Global competitiveness report 2013.* [cit. 2015- 03-23]. Available from WWW: <http://www.weforum.org/reports/global-competitiveness-report-2013-2014>.
[18] Yoon, K. P., Hwang, Ch.: *Multiple Attribute Decision Making: An Introduciton.* Sage, California, 1995.

Modification of the BCWS by Work Effort

Jan Bartoška [1], Petr Kučera [2], Tomáš Šubrt [3]

Abstract. The paper deals with BCWS (Budgeted Cost for Work Performed) estimation by work effort in context of the Student Syndrome. Every existing activity in any project is, to a higher or lower extent, determined by the effect of the human agent, which commonly takes a Student Syndrome form. If Student Syndrome phenomenon is completely expressed by means of a mathematical model, it would be possible to improve prediction of its impact on work effort in project activities. Successful duration of project activities is measured by the Earned value management (EVM) tools, where the BCWS is a key parameter. The BCWS is created by work effort in a project activity. Therefore, work effort has a hidden influence on EVM. Consequently, in the real world of human resources, the BCWS is not usually linear. This paper proposes new formulas for computing the BCWS in the project for Earned value management purposes. This proposal can be used for a more precise description of the real distribution of the work effort.

Keywords: Project management, Earned Value Management, Student Syndrome, Parkinson‘s law, Mathematical model, BCWS, Planned Value, Work effort.

JEL Classification: C61
AMS Classification: 90B99

1 Introduction

Earned Value Management [8, 9 or 10] is based on comparing the Baseline and Actual plan of the project realization. Baseline and Actual plan is determined by the partial work of single resources in the particular project activities. Baseline is always based on expected resource work contours. The impact of human agent is usually not included here. The impact of human agent is usually expected only in the actual course of the project. The versatility of the human agent in projects can be described also by the first “Parkinson’s law” [4]. It is natural for people to distribute work effort irregularly to the whole time sequence which was determined by the deadline of the task termination. The questions of “Parkinson’s first law” in project management are further dealt with in e.g. [5].

Work effort of an allocated resource has very often been researched in projects from the area of informational technologies and software development, as these projects contain a high level of indefiniteness, and even common and routine tasks are unique. At the same time, it concerns the area where it is possible to find a great number of approaches to estimate how laborious the project will be or how long the tasks will take, and also case studies. The proposal for mathematical apparatus for planning the course of tasks within a case study is dealt with for instance in [3], or Barry et al. [1]. The authors Özdamar and Alanya [3] propose a particular pseudo-heuristic approach to estimate the tasks course where the indefiniteness in the project is expressed by fuzzy sets. Barry et al. [1] concentrate on the existence and expression of the relation between project duration and total effort and in their theoretical starting points they point out the dynamics of the relation between the effort and project duration when a self-strengthening loop can be expected. The work effort can be described also using system dynamic models as presented e.g. in a study from project management teaching by Svirakova [7]. The others who research the project complexity and work effort are for instance Clift and Vandenbosh [2], who point out a connection between the length of life cycle and project management structure where a key factor is again a human agent.

The aim of the paper is to propose a new procedure for the computation of BCWS (Budgeted Cost for Work Performed) for Earned Value Management including resource work contours and “Student Syndrome” impact. The authors build on the results of their previous works and use mathematical models already defined by themselves for “Student Syndrome” expression.

[1] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, Czech Republic, e-mail: bartoska@pef.czu.cz.
[2] 2 Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, Czech Republic, e-mail: kucera@pef.czu.cz.
[3] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Praha 6, Czech Republic, e-mail: subrt@pef.czu.cz.

2 Materials and methods

2.1 Earned Value Management (EVM)

Earned value management (EVM), also called deviation analysis or cost analysis, is a project management technique for measuring project performance and progress in an objective perspective. Within EVM the budgeted cost for work performed (*BCWP* or *EV* (Earned value)), actual cost for work performed (*ACWP* or *AC* (Actual cost)), and budgeted cost for work scheduled (*BCWS* or *PV* (Planned value)) are measured and calculated:

- *BCWS* is defined as the permissible budgeted cost for accomplish project plan workload at some stage during project implementation. It mainly reflected the regulation workload of plan, not the regulation cost. The calculation formula is *BCWS = Plan Workload ×Quota Budget Price.*
- *BCWP* is defined as the cost calculating the accomplishment of work and quota budget price, which could be also called EV (earned value). It quantifies the accomplishment of the project. The calculation formula is *BCWS = Accomplishment Workload ×Quota Budget Price.*
- *ACWP* is defined as actual cost at some stage during project implementation. ACWP is mainly used to reflect the values of actual consumption.

BCWS, *BCWP* and *ACWP* parameters are used to calculate other parameters [8, 9], in particular schedule variance (*SV*), cost variance (*CV*), cost performed index (*CPI*), schedule performance index (*SPI*).

For each inspection day during the project realization, the values of the parameters of BCWS, BCWP and ACWP are written down. The values of these parameters form S-Curve (see Figure 1) and consequently document the development and state of the project from in terms of cost and time [10]. The project is assessed according to the differences between these parameters, whereas the underlying assumption is a steady increase in the work done as planned, namely in the value of the BCWS parameter.

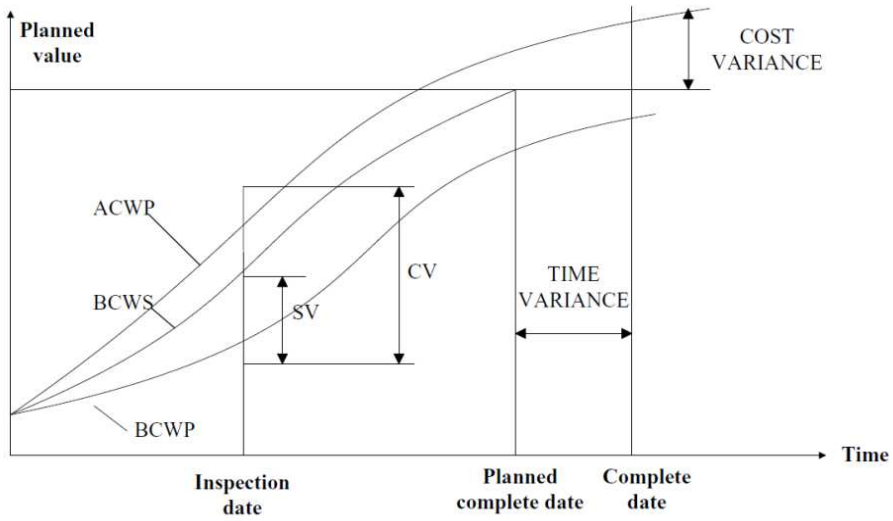


Figure 1 S-Curve [9]

2.2 "Student Syndrome" phenomenon

If there is a deadline determined for the completion of a task and a resource is a human agent, the source makes its effort during the activity realisation unevenly and with a variable intensity. Delay during activity realisation with human source participation leads to stress or to tension aimed at the resource or the tension of the resource him/herself. The development and growth of the tension evokes the increase in the work effort of the human agent allocated as a resource. Figure 2 demonstrates possible behaviour of the human resource known as the "Student Syndrome".

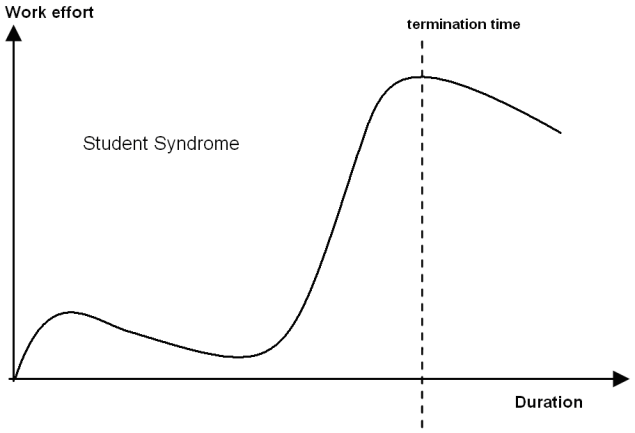


Figure 2 The variability of work effort during the "Student Syndrome"

This phenomenon has its equivalent in psychology, where it is known as „procrastination". Having been assigned a task, the human factor is under the influence of his or her natural behaviour and after a short time-interval s/he tries to create reserves in of his or her work effort. At the moment of discovery that the task is not been fulfilled in the time framework, there occurs a breakthrough in the resource's behaviour. An increasing stress of the source leads to tension and a rapid growth of work effort to enormous values.

2.3 Mathematical model of the "Student Syndrome"

In [6] the following mathematical expression of the "Student Syndrome" is proposed:

First, a function expressing the proper "Student Syndrome" is introduced. Let this function be denoted by p_1. The rate of resources utilization in the allocation of resources during the implementation of project tasks cannot exceed 100%. We cannot expect resources utilization outside of their power. Therefore, to express the "Student Syndrome", it is necessary to find such a function p_1 so that

$$\int_0^1 p_1(t)\,dt = 1 \tag{1}$$

Function p_1 should have three minima $p_1(t) = 0$ in $t = 0$, $t = 0.5$, and $t = 1$; and two maxima: former one close to the begin and latter one close to the end of the task realization. The description of the "Student Syndrome" which requires these conditions can be realized using the 4th degree polynomial:

$$p_1 = -120t^4+240t^3-150t^2+30t \tag{2}$$

Beside this, functions denoted by p_2 expressing the resource allocation according to single standard work contours of flat, back loaded, front loaded, double peak, bell and turtle are proposed in [6]. Similarly, for all these functions p_2:

$$\int_0^1 p_2(t)\,dt = 1 \tag{3}$$

Similarly, all these functions are in the form of 4th degree polynomial. For more details see [6].

The expression of the "Student Syndrome" during the realization of a task can be variously strong. Therefore, we introduce the rate r of the "Student Syndrome" which acquires values between 0 and 1. The case of $r = 0$ will represent a situation when the "Student Syndrome" does not occur at all and the resource keep the work contour exactly. On the other hand, the case of $r = 1$ means that the "Student Syndrome" manifests in its all strength and the resource absolutely ignore the work contour. Of course, both the previous cases are hypothetical. As a result, we can model the resource work effort p during a real task realization in the following way:

$$p = rp_1+(1-r)p_2 \tag{4}$$

From (1) and (3) the following holds again:

$$\int_0^1 p(t)\,dt = 1 \tag{5}$$

3 Results and discussion

This approach can be applied when computing the BCWS (Budgeted Cost for Work Performed) of a activity in the project. It is computed in the classical way of using the formula

$$BCWS = Percentage\ according\ to\ the\ calendar\ .\ BAC \tag{6}$$

where *BAC* is Budget at completion of the project. Let us denote by *a* the part of the task duration for which we want to express the BCWS (e.g. for determining the BCWS for the first quarter of the task duration, $a = 0.25$ is used). The share of the whole work effort which this part of the task duration requires can be calculated for a single resource as

$$\int_0^a p(t)\,dt = 1 \tag{7}$$

Let there are *n* resources, indexed by 1, 2, …, *n*, allocated at the task. Let r_k, p_{1k}, p_{2k} denotes r, p_1, p_2 for *k*-th resource. Then the BCWS for *a* can be computed

$$BCWS = \left(\sum_{k=1}^{n} \int_0^a (r_k\, p_{1k}(t) + (1 - r_k)\, p_{2k}(t))\,dt \right) \cdot BAC \tag{8}$$

3.1 Case study

Let us consider a project task with three resources allocated. Their work contours are back loaded, turtle and double peek, respectively. Their "Student Syndrome" rates are $r_1 = 0.3$, $r_2 = r_3 = 0.2$. This approach enables to demonstrate how much part of the work needed for implementation the task has been worked to various terms, both of single sources and in common. These terms were chosen as first quarter ($a = 0.25$), first half ($a = 0.5$) and first three quarters ($a = 0.75$) of the task duration. The results are summarized in Table 1.

a (*ratio of the task duration*)	**0.25**	**0.5**	**0.75**
Resource 1 (*front loaded*)	0.12	0.3	0.58
Resource 2 (*turtle*)	0.22	0.5	0.78
Resource 3 (*double peak*)	0.30	0.5	0.70
Total work effort	**0.22**	**0.43**	**0.68**

Table 1 Computing values of the BCWS with $BAC = 1$

The BCWS growth for single resources is presented in Figure 3. In case of none of these sources the work effort grows uniformly due to the "Student Syndrome" impact and thus BCWS also grows nonlinearly up to highly variably. Especially in case of Resource 1, where the work contour *Front loaded* is expected, we can see a decline of BCWS during the first half of the task duration and an enormous growth of BCWS at the end. As a result, the expected course of the Front Loaded in BCWS does not occur. In case of Resource 3, where the work contour *Font Double peak* is expected, "sway" of the work effort occurs and BCWS grows incredibly irregularly. At the beginning the resource spends effort higher than expected. Such an increased effort at the site where it is not planned may be inefficient for the project.

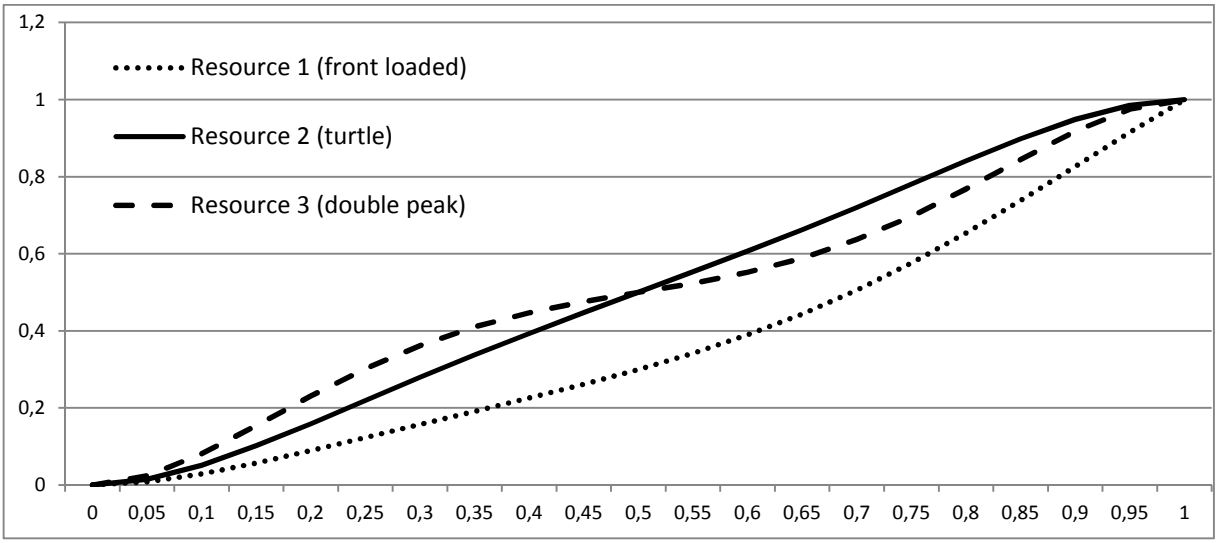


Figure 3 BCWS for Resource 1, Resource 2 and Resource 3

We can assume the involvement of a higher number of resources in an Aggregate activity in the project. BCWS in the Aggregate activity is given by the sum of planned work by single resources. If we sum BCWS of Resource 1, Resource 2 and Resource 3, we obtain BCWS for the Aggregate activity. The resulting BCWS is determined by the work contours of single resources and by the impact of the "Student Syndrome" on these work contours.

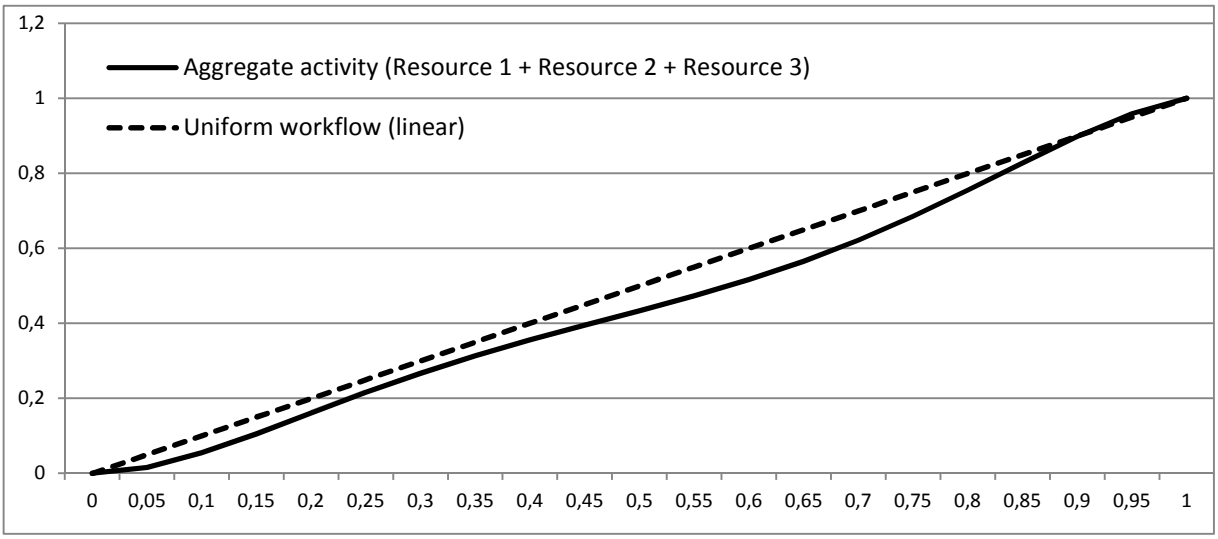


Figure 4 BCWS for Agregate activity (Resource 1 + Resource 2 + Resource 3)

The overall work effort for the Aggregate activity, despite the planned and expected work contours, is negatively affected due to the "Student Syndrome". The work effort decreases during the task realization and grows enormously at the end. BCWS may incredibly differ from the theoretically expected uniform increase in time as presumed in the classical EVM approach [9].

4 Conclusion

The paper deals with a modification of the BCWS (Budgeted Cost for Work Performed) computation including the "Student Syndrome" impact on resource work contour. The resource work effort affects the growth of BCWS. BCWS may grow unevenly up to highly variably in real environment. It is not possible to expect uniform increase of BCWS always and in case of all project tasks. In case of changing BCWS, Earned Value Management may provide with fundamentally different results for the assessment of the state and development of the project. The "Student Syndrome" impact manifests in resources work effort, changes the work contour shape and transfers itself into the whole project and EVM parameters. The work contours of the resources and the "Student Syndrome" impact may have a significant effect on both the actual and planned course of the project. As far as computations in EVM are based on unrealistically expected BCWS, EVM may be unsuccessful.

References

[1] Barry, E., Mukhopadhyay, T., Slaughter, A. S.: Software project duration and effort: An empirical study. *Information Technology and Management* **3** (2002), 113–136.

[2] Clift, T., Vandenbosh, M.: Project complexity and efforts to reduce product development cycle time. *Journal of Business Research* **45** (1999), 187–198.

[3] Özdamar, L., Alanya, E.: Uncertainty modelling in software development projects (with case study). *Annals of Operations Research* **10** (2001), 157–178.

[4] Parkinson, C.N.: *Parkinson's Law and Other Selected Writings On Management*. 1st ed., Singapure, Federal Publications (S) Pte Ltd 1991.

[5] Gutierrez, G.J., and Kouvelis, P.: Parkinson's law and its implications for project management. *Management Science* **37** (8) (1991), 990-1001.

[6] Kučera, P., Bartoška, J., Šubrt, T.: Mathematical models of the work contour in project management. In: *Proceedings of the 32nd International conference on Mathematical Methods in Economics* (Talašová, J., Stoklasa, J., Talášek, T., eds.). Palacký University, Olomouc, 2014, 518-523.

[7] Svirakova, E.: System dynamics methodology: application in project management education. In: *Proceedings of the 11th International Conference on Efficiency and Responsibility in Education* (Houška, M., Krejčí, I., Flégl, M., eds.). Czech University of Life Sciences, Prague, 2014, 813-822.

[8] Vandevoorde, S., Vanhoucke, M.: A comparison of different project duration forecasting methods using earned value metrics. *International Journal of Project Management* **24** (4) (2005), 289–302.

[9] *A Guide to the Project management body of knowledge*. Fifth edition. The Project Management Institute, Inc., Newtown Square, Pennsylvania, 2013.

[10] Kim, B., C., Kim, H., J.: Sensitivity of earned value schedule forecasting to S-curve patterns. *Journal of Construction Engineering and Management* **140** (7) (2014), 14-23.

Mathematical approximation of confusion index

Jitka Bartošová[1], Vladislav Bína[2]

Abstract. The well-known technique of cluster analysis ranks among the tools used for analyzing and understanding large datasets. Non-hierarchical clustering is also employed in EM-algorithm that uses complex empirical conditional or joint distributions for so-called "model-based" clustering. Non-hierarchical clustering methods demand preliminary knowledge of the number of clusters. Penalized information criteria (AIC, BIC) are used for preliminarily estimating the number of components for EM-algorithm. However, clustering with mixture models does not provide the possibility to achieve perfect separation; we obtain only probabilities of objects falling into the particular clusters. We have already discussed the issues concerning whether the resulting mixture components are separated and to what extent the classification is unambiguous in Bartošová and Bína (2014), and Longford and Bartošová (2014). The confusion index was used here for measuring separation and clustering. Its analytical formulation does not exist for any non-trivial pair of distributions, but under certain conditions it can be approximated. The paper presented is devoted to the derivation of approximation of the confusion index under the assumption of normality.

Keywords: confusion index; finite mixture of densities; mathematical approximation.

JEL Classification: C38, C44
AMS Classification: 62H30

1 Object clustering

A clustering serves as a tool for analyzing and understanding large data files. We can distinguish two types of clustering methods – hierarchical and non-hierarchical. Hierarchical clustering does not require any preliminary knowledge about the number of clusters and provides the possibility of creating and presenting a detailed structure of data (using a dendrogram). On the contrary, non-hierarchical clustering approaches require setting the number of clusters preliminarily. The success of this decomposition is conditioned by a good initial solution (initial decomposition) where objects are sorted according to their distance from the centroids of each cluster. Two basic categories of clustering methods correspond to this classification – distance-based methods and probabilistic approaches. Distance based methods assign objects into the classes unambiguously. Probabilistic methods do not classify unambiguously, only calculate the probabilities of membership in particular (latent) classes.

In the course of time, many methods differing in the approach to clustering have been developed – for more see the papers by Löster (2014), Řezanková (2013), Řezanková and Húsek (2012), Stankovičová and Vojtková (2007), Šulc and Řezanková (2015) among many others. Clustering methods have a wide range of applications in many different branches. Economical applications can be found, for example, in papers by Pavelka and Löster (2014), and Řezanková and Löster (2013) as well as by other authors.

1.1 Issues of Bayesian clustering

Bayesian clustering ranks among non-hierarchical clustering methods that require preliminarily setting the number of clusters (number of mixture components). The well-known EM-algorithm is based on Bayesian clustering (see Dempster, Laird and Rubin, 1977) and is based on the assumption of the mixture character of the data files. The algorithm employs the well-known fact that each continuous distribution can be approximated with sufficient accuracy using a mixture of normal curves. The number of mixture components is not known in advance – it must be estimated expertly or by using some penalized information criterion. Penalization is essential, since each additional component increases the accuracy of the model. Penalization introduces a "penalty" for the higher number of parameters in the model, i.e., it penalizes the growth of complexity.

Akaike information criterion (AIC), (see Akaike, 1973), measures the relative quality of a statistical model. AIC penalizes the number of parameters less strongly than another well-known criterion, the Bayesian information criterion (BIC) that was developed independently by Schwarz (1978). These criteria do not serve as a test of statistical significance; they provide the possibility of choosing the most suitable model from the given set.

[1,2] University of Economics in Prague, Faculty of Management, Department of Exact Methods, Jarošovská 1117/II, 37701 Jindřichův Hradec, {bartosov,bina}@fm.vse.cz..

Decomposition of the mixture into components (clusters) can be implemented using *mclust* (see Fraley and Raftery, 2003), an R package for mixture modeling. This algorithm embeds a well-known EM-algorithm for multivariate normal mixtures. The optimal number of clusters is chosen according to the Bayesian information criterion.

However, the "optimal" number of clusters does not always correspond to reality. The estimated number of clusters often exceeds the real one and it is necessary to find a method for suitably merging some of the clusters. This process is also realized by different approaches, for example, Tantrum, Murua and Stuetzle (2003) propose some graphical diagnostics and recommend employing hierarchical methods based on a dip test of unimodality (see Hartigan and Hartigan, 1985).

Non-hierarchical methods also exist, e.g., mixtures of normal components and some of these methods apply during aggregation of well-known method of *k*-means, etc. Hennig (2010) also focused on methods for aggregating components of normal mixtures into the form of (well separable) clusters, where he analyzed several different methods, many of them with some objections. Therefore, he proposed new aggregation concepts different from ones based on unimodality and he proposed the so-called "DEMP method" using an estimate of probability of a wrong classification (for more detail, see Hennig, 2010). An explicit expression also exists for determining the distance of two multivariate normal distributions (Fukunaga, 1990) based on the determinant and the inverse of covariance matrix of components. It is problematic to estimate its value with sufficient precision and the results can be biased (in the case of nearly singular matrices). An extensive amount of literature also concerns methods of cluster determination.

2 Confusion index

Whether all mixture components (estimated, e.g., by the *mclust* procedure) will be considered to be clusters can be decided using measure of their separation. In case of high separation the probability of assignment to the "right" cluster is also high, although we do not know for sure to which cluster it belongs. For a quantitative evaluation we can use the confusion index presented in the papers by Longford and Bartošová (2014), and Bartošová and Bína (2014). The confusion index of components A and B is here given by the probability that a randomly chosen observation x from component B is incorrectly classified to A. The classification rule is a conditional probability that any observation x from the domain of densities f_A (resp. f_B) is from the component A, resp. B

$$r_{x,A|AB} = \frac{p_A f_A(x)}{p_A f_A(x) + p_B f_B(x)}, \text{ resp. } r_{x,B|AB} = 1 - r_{x,A|AB}. \tag{1}$$

Confusion index $r_{A|B}$ is given by the expected probability that component A will have assigned a random sample from the distribution of the component B,

$$r_{A|B} = \int \frac{p_A f_A(x) f_B(x)}{p_A f_A(x) + p_B f_B(x)} dx \tag{2}$$

For more details, again see Longford and Bartošová (2014) or Bartošová and Bína (2014).

2.1 Mathematical approximation

Analytical expression of the confusion index does not exist for any nontrivial pair of distributions. If we need to analyze the relation of a pair of mixture components we can use an approximation of the confusion index for small deviations, i.e., for similar densities f_A and f_B. Generally, function $f(x)$ analytic in x_0 can be approximated in the neighborhood of the point using a Taylor series. From the Taylor theorem we can write

$$f(x) = T_n(x) + R_{n+1}(x) \tag{3}$$

where $\sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$ is the Taylor polynomial and $R_{n+1}(x)$ is a remainder. It can be shown that a necessary and sufficient condition for a Taylor polynomial convert for $n \to \infty$ into the Taylor series is the convergence of the remainder $R_{n+1}(x)$. Usually the Lagrange form of remainder is used $R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$, where $\xi \epsilon (x_0; x)$. For more detail see, e.g., Abramowitz and Stegun (1972).

We will assume that f_A and f_B belong in the same class of densities (e.g., normal) and differs only in the value of parameter θ, which gets values of θ_A, resp. θ_B for f_A, resp. f_B. It means that $f_A(x) = f(x; \theta_A)$ and $f_B(x) =$

$f(x;\theta_B)$ where $\theta_B = \theta_A + \Delta$. The classification rule $r_{x,A|AB}$ can be now expressed as a function of difference in parameter of component density $\Delta = \theta_B - \theta_A$, i.e., in the form

$$\varphi(\Delta) = \frac{p_A f(x;\theta_A) f(x;\theta_A+\Delta)}{p_A f(x;\theta) + p_B f(x;\theta_A+\Delta)}. \tag{4}$$

Using an expansion of function $\varphi(\Delta)$ to the second order in the neighborhood of $\Delta = 0$ we obtain

$$\varphi(\Delta) = \varphi(0) + \left.\frac{\partial\varphi(\Delta)}{\partial\Delta}\right|_{\Delta=0} \Delta + \left.\frac{\partial^2\varphi(\Delta)}{\partial\Delta^2}\right|_{\Delta=0} \frac{\Delta^2}{2} \tag{5}$$

where $p_A + p_B = 1$

$$\varphi(0) = \frac{p_1 f(x;\theta) f(x;\theta)}{p_1 f(x;\theta) + p_2 f(x;\theta)} = p_1 f(x;\theta)$$

$$\left.\frac{\partial\varphi(\Delta)}{\partial\Delta}\right|_{\Delta=0} = \left.\frac{\partial}{\partial\Delta}\left(\frac{p_1 f(x;\theta) f(x;\theta+\Delta)}{p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)}\right)\right|_{\Delta=0}$$

$$= \left.\frac{p_1 f(x;\theta) f'(x;\theta+\Delta)\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right] - p_1 f(x;\theta) f(x;\theta+\Delta) p_2 f'(x;\theta+\Delta)}{\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta\right]^2}\right|_{\Delta=0}$$

$$= \left.\frac{p_1^2 f^2(x;\theta) f'(x;\theta+\Delta)}{\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right]^2}\right|_{\Delta=0} = p_1^2 f'(x;\theta)$$

$$\left.\frac{\partial^2\varphi(\Delta)}{\partial\Delta^2}\right|_{\Delta=0} = \left.\frac{\partial}{\partial\Delta}\left(\frac{p_1^2 f^2(x;\theta) f'(x;\theta+\Delta)}{\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right]^2}\right)\right|_{\Delta=0}$$

$$= \left.\frac{p_1^2 f^2(x;\theta) f''(x;\theta+\Delta)\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right]^2}{\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right]^4}\right|_{\Delta=0}$$

$$- \left.\frac{p_1^2 f^2(x;\theta) f'(x;\theta+\Delta) 2\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right] p_2 f'(x;\theta+\Delta)}{\left[p_1 f(x;\theta) + p_2 f(x;\theta+\Delta)\right]^4}\right|_{\Delta=0}$$

$$= p_1^2 f''(x;\theta) - \frac{2p_1^2 p_2\left[f'(x;\theta)\right]^2}{f(x;\theta)} = p_1^2 \frac{f''(x;\theta) f(x;\theta) - 2p_2\left[f'(x;\theta)\right]^2}{f(x;\theta)}$$

Substituting back we obtain an approximation of the classification rule in the form

$$\frac{p_1 f_1(x) f_2(x)}{p_1 f_1(x) + p_2 f_2(x)} = p_1 f_1(x) + p_1^2 f_1'(x)\Delta + p_1^2 \frac{f_1''(x) f_1(x) - 2p_2\left[f_1'(x)\right]^2}{f_1(x)} \frac{\Delta^2}{2}. \tag{6}$$

3 Application

Sensitivity of the confusion index $r_{A|B}$ as a measure of separation, resp. interchangeability of components on the difference of parameters of component distribution $\Delta = \theta_B - \theta_A$ will be illustrated using a simulation.

3.1 Empirical study of sensitivity of the confusion index on the change of difference in component parameters

For the sake of simplicity we will assume that the mixture can be approximated by a model with two normally distributed components. Now let $f(x;\theta_A)$, $f(x;\theta_B)$ be densities of normal distribution with parameters $\{\mu_A;\sigma_A^2\}$ and $\{\mu_B;\sigma_B^2\}$. We will separately observe the influence of change in

- mean values ($\Delta = \mu_B - \mu_A$) in Figure 1,
- standard deviations ($\Delta = \sigma_B - \sigma_A$) in Figure 2.

Obviously, the values of the confusion index decrease with the growing distance of means of both distributions. It is apparent that the confusion index is asymmetrical, i.e., the behavior of index $r_{A|B}$ differs from the behavior of index $r_{B|A}$. For example, in the top of Figure 1 we can observe that for the rather extreme case of $p_A = 0.01$ and $p_B = 0.99$ with increasing distance of means the variant $r_{A|B}$ decreases much more rapidly than index $r_{B|A}$ and for value $\Delta\mu = 2.5$ the probability of assignment of random sample from B component to the A component is as low as 20 percent whereas the opposite probability assignment of a random sample from A component to the B component is more than 72 percent.

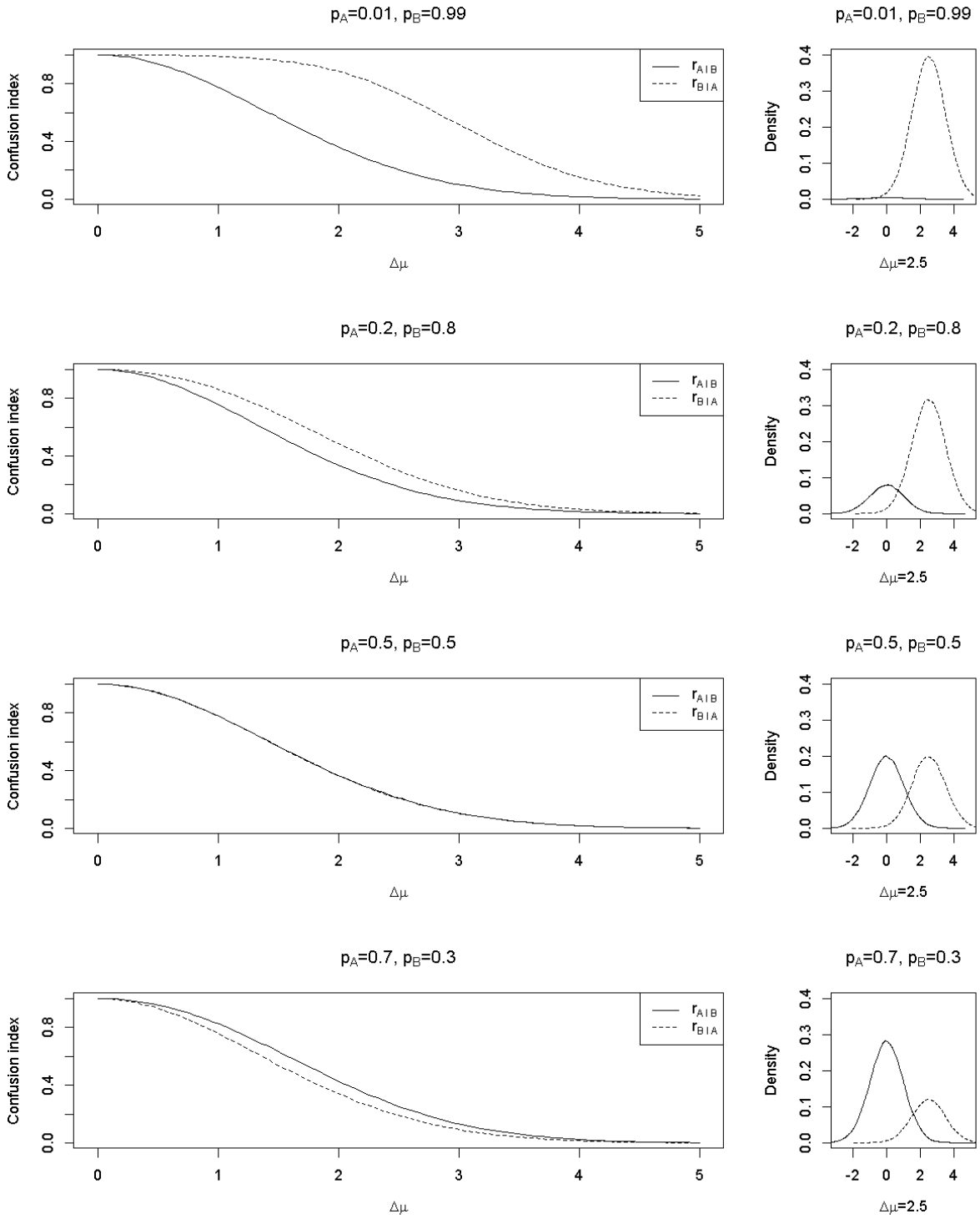


Figure 1 Influence of difference in mean values of mixture components distribution on the confusion index.

If we now observe values of the confusion index of a standard normal distribution and a distribution with a different standard deviation (see Figure 2) we can see that generally a very small standard deviation of component B leads to small values of the confusion index. The values naturally grow as variability of both distributions becomes similar, while for higher values of standard deviation of component B values of both variants begin to decrease. Roughly speaking, it appears that the values of the confusion index are higher for the components with lower variability (but it also depends on the values of p_A and p_B).

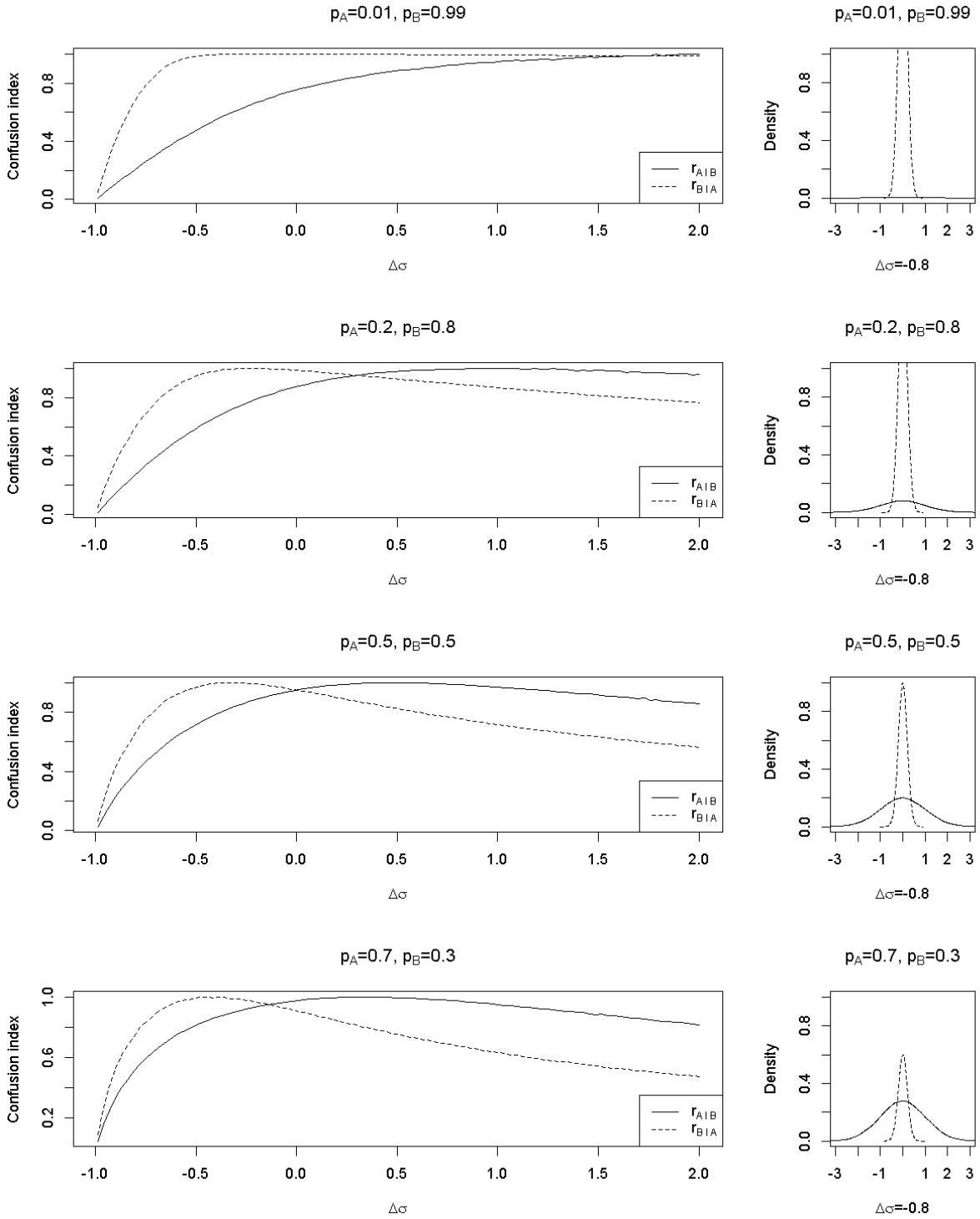


Figure 2 Influence of difference in standard deviations of mixture components on the confusion index.

4 Conclusion

In the paper presented we derived the approximation of the confusion index that can be used for measuring cluster separation. In the next part of the paper we showed (using a simulation) the influence of a difference in parameters μ and σ of normally distributed components of a mixture. According to the fact that the confusion index depends on the proportion (weight) of components in the mixture, the influence was observed for different values of marginal probabilities p_A and p_B. The comparison of results of approximation $r_{A|B}$ for small values of difference in μ and σ shows some interesting conclusions. For instance, we can observe that two similar distributions are more separated if we slightly change their standard deviations than after a similar change of their mean values. And simultaneously, we obtain an opposite result in the case of large differences. Consequently, we can once more stress that the presented approximation above is valid for only small differences.

Acknowledgements

The research was supported by GA ČR under grant no. 15-00215S.

References

[1] Abramowitz, M., and Stegun, I.A. (eds.): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, 1972.

[2] Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: *Proceedings International Symposium in Information Theory* (Petrov, and Csaki, eds.). Akademiai Kiado, Budapest, 1973, 267–281.

[3] Bartošová, J., and Bína, V.: The measure of separation of components in final mixtures of probability densities. In: Talašová, J., Stoklasa, J., and Talášek, T. (eds.) *32nd International Conference Mathematical Methods in Economics, MME 2014, Conference Proceedings*, Palacký University, Olomouc, 2014, 13–18

[4] Dempster, A., Laird, N., and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39,** 1 (1977), 1–38.

[5] Fraley, C., and Raftery, A. E.: Enhanced Software for Model-Based Clustering. Density Estimation and Discriminant Analysis. *Journal of Classification*, **20** (2003), 263–286.

[6] Fukunaga, K.: *Introduction to statistical pattern recognition*. Academic Press, Boston, 1990.

[7] Hartigan, J.A., and Hartigan, P.M.: The dip test of unimodality. *Annals of Statistics*, **13** (1985), 70–84.

[8] Hennig, C. H.: Methods for merging Gaussian mixture components. *Advances in Data Analysis and Classification* **4** (2010), 3–34.

[9] Longford, N.T., and Bartošová, J.: A confusion index for measuring separation and clustering. *Statistical Modelling*, **14**, 3 (2014), 229–255.

[10] Löster, T.: The Evaluation of CHF coefficient in determining the number of clusters using Euclidean distance measure. In: *The 8th International Days of Statistics and Economics* (Pavelka, and Löster, eds.). [online] Melandrium, Slaný, 2014, 858–869.

[11] Pavelka, T., and Löster, T.: Využití shlukové analýzy ke srovnání dopadů celosvětové finanční krize na trh práce zemí Evropské unie. *Forum statisticum slovacum*, **10**, 6 (2014), 125–130.

[12] Řezanková, H.: Klasické a současné postupy ve shlukové analýze. In: *Analýza dat 2013 Statistické metody pro technologii a výzkum*. TriloByte Statistical Software, Pardubice. 2013, 109–118.

[13] Řezanková, H., and Húsek, D.: Fuzzy Clustering: Determining the Number of Clusters. In: *Computational Aspects of Social Networks (CASoN) 2012* [CD-ROM]. Research Publishing Services, Sao Carlos, 2012, 277–282.

[14] Řezanková, H., and Löster, T.: Shluková analýza domácností charakterizovaných kategoriálními ukazateli. *E+M. Ekonomie a Management*, **16**, 3 (2013), 139–147.

[15] Schwarz, G. E. Estimating the dimension of a model. *The Annals of Statistics*, **6**, 2 (1978), 461–464.

[16] Stankovičová, I., and Vojtková, M.: *Viacrozmerné štatistické metódy s aplikáciami*. Iura Edition, Bratislava, 2007.

[17] Šulc, Z., and Řezanková, H.: Evaluation of selected approaches to clustering categorical variables. *An International Journal of the Polish Statistical Association-STATISTICS IN TRANSITION new series* [online], **15**, 4 (2014), 591–610.

[18] Tantrum, J., Murua, A., and Stuetzle, W.: Assessment and Pruning of Hierarchical Model Based Clustering. In: *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*. Washington, D.C., 2003, 197–205.

The Kitchen Sink Problem: Monte Carlo study of stochastic frontiers

Matěj Bělín[1]

Abstract. This article presents a Monte Carlo study of the problem of identification that arises when superfluous variables are added into the regression that are correlated with other regressors. The results show that the bias in coefficients that occurs is substantial even in large samples and that the standard tests for favourability of the nested models are actually likely to reject the restricted model in cases when the bias is particularly large. It is also noted that in the context of frontier analysis, this problem can be partially mitigated when attention is directed specifically to the characteristics of the production process rather than the coefficients of the frontier themselves.

Keywords: Monte Carlo Simulation, problem of identification, problem of collinearity, stochastic frontier analysis, nested models, model selection

JEL Classification: C510
AMS Classification: 62P20

1 Introduction: The problem

The term "kitchen sink models" is used to denote regressions where a large number of covariates is employed in an attempt to capture the totality of the circumstances underlying the problem under investigation and hence to obtain estimates that are free from potential omitted variable bias. For instance in macroeconomic growth models it has become customary to include several variables into the regression as determinants of the steady state following Barro [1], so with the increased availability of data, these models can feature staggering amounts of regressors, cf. Durlauf, Kourtellos and Tan [2].

As Schrodt [3] explains, the hope is that including additional explanatory variables is "at worst harmless, and at best will prevent erroneous inferences" adding however that "[n]othing could be further from the truth." The hope of benignity of superfluous variables is misplaced in situations with high correlations among regressors, as two highly precarious situations can arise:

1. Due to near collinearity, coefficients may appear insignificant, which can lead to Type II error of inappropriate failure to reject null hypothesis.
2. Due to problem of identification, the resulting coefficients may be extremely poorly estimated, showing perverse signs or extremes in magnitude.

Both of these pitfalls are manifestations of what is here called the "kitchen sink problem", where the effort to eliminate omitted variable bias results in inclusion of superfluous variables into the model, which, ironically, can be to the detriment of final results. This is the reason why Achen [4] uses the term "garbage can" models, rather than kitchen sink models. More importantly, these problems may not be immediately apparent and thus can evade the attention of the researcher. For instance, it is quite possible that regressions by Easterly and Levine [5], which attempt to distinguish the effects of investment in physical and human capital on the average income per capita noting that the coefficient on physical capital is insignificant. However, this study defines these two variables as $[\ln s - \ln(x + \delta + n)]$ for investment in physical capital and $[\ln s_h - \ln(x + \delta + n)]$ for investment in human capital. It is quite possible that the common term $\ln(x + \delta + n)$ causes collinearity between these two variables, which results in inflated standard error in one of the coefficients (and the authors do not report the correlation coefficient between these two covariates to dispel this suspicion).

Consider also the second possibility, namely that the kitchen sink model can suffer from the problem of identification. To give a specific example, consider a linear model of the relationship between variables x and y such that $y = \beta_0 + \beta_1 x + \varepsilon$ where β_0 and β_1 are the unknown coefficients and ε is the random, normally distributed disturbance. An obvious choice for fitting this model would be an OLS regression of the observed values of y on x. However, concern about the potential non-linearity can lead to specification of an augmented model, where $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon^*$ and subsequent F-test for the suitability of the nested model, which would reveal whether the augmented model is warranted. Since there is likely to be a reasonably strong correlation between x

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, <matej.belin@gmail.com>

and its square, this can potentially lead to kitchen sink problem if the true relationship is linear but the quadratic term is added nonetheless. As an illustration, let us run a small preliminary Monte Carlo simulation to show the extent of the problem: assume that the true relationship is linear such that $y = \frac{1}{3} + \frac{1}{3}x + \varepsilon$, where $\varepsilon \sim \mathcal{N}\left(0, \frac{1}{2}\right)$. We generated a population of one million individuals from which we drew five thousand subsamples of five thousand observations and performed OLS regression with linear and quadratic specifications on each subsample. The correlations between x and x^2 is very high, on average 0.99, so the problem of collinearity is likely to be serious.

Predictably, the quadratic specification yielded much wider variability of the estimated coefficient $\hat{\beta}_1$, which ranges quite wildly around the true value of 1/3 with about 14% of cases where the estimated $\hat{\beta}_1$ biased by 50% or more. Since 14% is a non-trivial proportion of all the attempts, and since this percentage will increase in smaller samples,[2] it is well worth considering potential methods of detecting these unsuitable models. It is possible to posit that the models with the most egregiously biased results would not be favoured by the F-test (or likelihood ratio test, or a functional equivalent thereof). Unfortunately, it is precisely to the contrary: the models where the bias is most serious are the ones that are favoured by F-test. Plotting the p-values of the test for equivalence of linear and quadratic models against the percentage bias in the coefficient $\hat{\beta}_1$ gives the following picture:

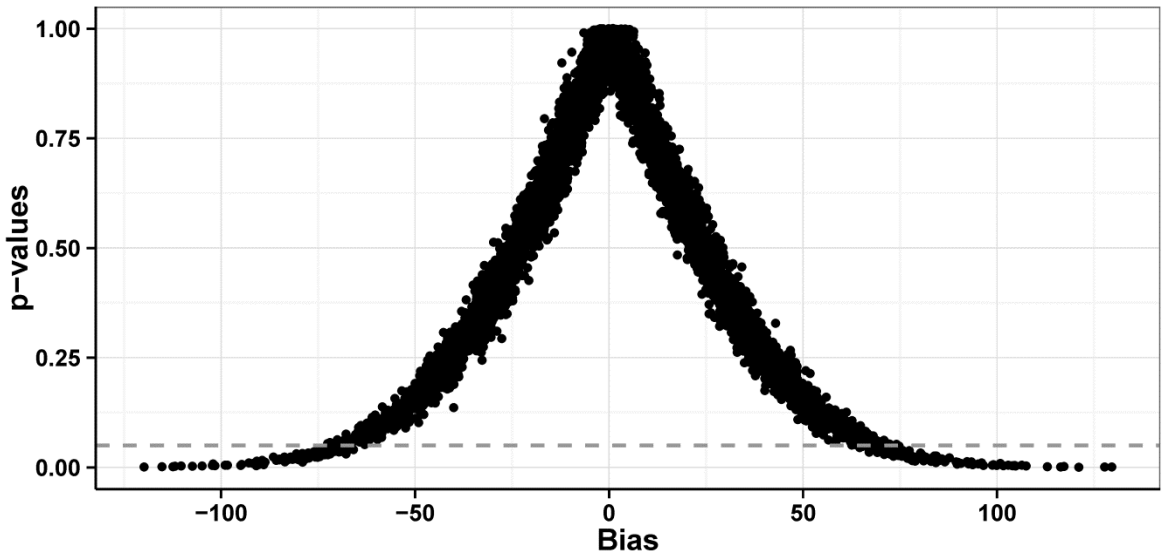


Figure 1 p-values of the F-test whether the linear and quadratic specifications are equivalent against the percentage bias in $\hat{\beta}_1$. Significance of 5% is indicated by dashed line

The A-shaped curve showing the relationship between bias and the apparent favourability of the misspecified model indicates that in the unbiased models, F-test appropriately fails to reject equivalence of the two specifications (the top of the curve), but with increasing bias in either direction, the tests start to prefer the model with the superfluous quadratic terms.

Observe that even in quite sizeable samples, the addition of a parvenu variable that is highly correlated to the other covariate results in serious problem. Therefore, this study will attempt to identify the impact of the kitchen sink problem with changing number of observations. Our strategy is to illustrate the problem on a microeconomic topic (production frontiers), since it presents an easily tractable opportunity to explore the problem, but the issue is manifestly relevant to other fields of inquiry, such as macroeconomic modelling, which has been discussed already and other fields of social sciences.

The remainder of this paper is organised as follows: first we present the details of the Monte Carlo study with results reported in the following section and we conclude in short summary of findings in the end.

[2] In samples with 25 observations, we obtain 92% of instances with bias exceeding 50% and 87% of instances with bias of 100% or more (5000 replications were performed).

2 Monte Carlo Design

Consider a Cobb-Douglas production function $\ln y = \frac{1}{3} + \frac{1}{3}\ln K + \frac{2}{3}\ln L + v - u$. A population of one million individuals is generated where K is drawn from uniform distribution $\mathcal{U}(300,\ 30000)$ and since profit maximisation under Cobb-Douglas production function implies a stable ratio between the input quantities varying only with the ratio of input prices, the values for L are generated as $L = r \times K \times \exp(q)$, where $r \sim \mathcal{U}(0.25, 0.35)$ and $q \sim \mathcal{N}(0,1)$ to give more-or-less stable relationship between the two inputs but with some disturbances arising from variability of the relative prices captured by the term r and noise caused by errors in measurement and unpredictable circumstances indicated by the term q. The random disturbance v is drawn from $\mathcal{N}\left(0, \frac{1}{2}\right)$ and the term u representing the inefficiency of a given firm is drawn from half normal distribution $\left|\mathcal{N}\left(0, \frac{1}{2}\right)\right|$, which gives us the average efficiency levels at roughly 70%.

With this basic setup, a scenario is considered, where the only problem is that translog production possibility frontier (PPF) was specified instead of the Cobb-Douglas formulation. Hence, SFA regressions were run on a subsample of the simulated population consisting of 25, 50, 100, 250, 500, 1000, and 5000 observations using the *frontier* package for *R* by Coelli and Henningsen [6]. For each sample size, 500 replications were performed.

3 Results

The estimated coefficients recorded after each replication are tallied below. For each sample size, the average value of the estimated coefficients is presented along with the numerically smallest estimate (presented under the heading Minima) and the largest estimate of all replications (under the heading Maxima). In addition, the estimated efficiency levels are compared to the true values.

The estimated coefficients are reported in Table 1 below. Observe that while the mean value (the average of all the 500 coefficients) converges to the true values in both Cobb-Douglas (CD) and translog (TL) specifications, the minimal and maximal values reveal that the ability of the translog model to recover the original coefficients is severely impeded by the kitchen sink problem. For instance the CD specification places the coefficient $\hat{\beta}_K$ in the interval 0.30 and 0.38 with 5000 observations, but the translog case gives a much wider interval between -0.22 and 0.88 for the same sample size, which is quite considerable even by modern standards.

Table 1 Estimated coefficients: means, minima, and maxima after 500 replications

Means	$\beta_0 = 1/3$		$\beta_K = 1/3$		$\beta_L = 2/3$	
N	CD	Translog	CD	Translog	CD	Translog
25	0.132	0.766	0.352	0.246	0.668	0.619
50	0.180	-0.044	0.346	0.350	0.663	0.728
100	0.255	0.213	0.339	0.304	0.661	0.718
250	0.267	0.137	0.335	0.366	0.665	0.666
500	0.300	0.144	0.333	0.382	0.667	0.652
1000	0.325	0.298	0.332	0.336	0.666	0.669
5000	0.327	0.404	0.334	0.312	0.666	0.671
Minima	$\beta_0 = 1/3$		$\beta_K = 1/3$		$\beta_L = 2/3$	
N	CD	Translog	CD	Translog	CD	Translog
25	-5.378	-109.800	-0.216	-20.980	0.231	-9.007
50	-3.215	-33.720	-0.037	-7.164	0.370	-3.263
100	-1.597	-15.420	0.107	-5.388	0.432	-2.624
250	-0.889	-10.620	0.137	-2.105	0.522	-0.638
500	-0.551	-5.442	0.209	-1.280	0.592	-0.190
1000	-0.389	-4.440	0.248	-0.615	0.613	-0.057
5000	-0.045	-1.489	0.299	-0.220	0.637	0.377
Maxima	$\beta_0 = 1/3$		$\beta_K = 1/3$		$\beta_L = 2/3$	
N	CD	Translog	CD	Translog	CD	Translog
25	4.354	97.520	1.077	28.280	1.070	9.109
50	3.508	32.530	0.870	9.322	0.981	5.711
100	2.083	17.940	0.598	4.329	0.832	4.258
250	1.677	9.426	0.497	3.232	0.790	1.775
500	1.237	5.478	0.480	2.048	0.728	1.607
1000	0.946	3.949	0.410	1.670	0.718	1.313
5000	0.558	1.958	0.376	0.881	0.697	0.926

The next step is to see whether the poor ability to estimate the coefficients impacts the estimates of efficiency as well. We estimated the inefficiency terms after each regression via the Jondrow, Lovell, Materov and Schmidt [7] estimator (JLMS estimator for short) and then compared these values with the true values of the disturbance term u. Three tests were then applied: first Spearman correlations were calculated to see whether the models identify the true ranking of different observations in terms of efficiency, then the Kolmogorov-Smirnov test for equality of distributions was calculated to see whether the shape of the estimated distribution of the inefficiency term matches the true one. Since the p-values of the KS test were consistently extremely low, only the D-statistics are reported, which should be zero if the distributions match perfectly. Finally, we calculated the average bias as of the inefficiency terms calculated as $b_u = (\hat{u} - u)/u$ to provide a guide to the magnitude of the discrepancy.

Table 2 Estimates of efficiency

	Spearman Correlations			KS D-statistic		Average Bias	
N	CD:True	TL:True	CD:TL	TL:True	CD:True	TL	CD
25	0.4408	0.4016	0.8860	0.5790	0.6069	0.1436	0.1452
50	0.4665	0.4435	0.9515	0.5570	0.5896	0.1538	0.1867
100	0.4788	0.4702	0.9789	0.4816	0.5058	0.1499	0.1730
250	0.4797	0.4768	0.9928	0.4157	0.4364	0.1351	0.1504
500	0.4820	0.4807	0.9966	0.3654	0.3723	0.1175	0.1228
1000	0.4820	0.4814	0.9983	0.3145	0.3180	0.0962	0.0981
5000	0.4829	0.4897	0.9997	0.2568	0.2569	0.0728	0.0728

The results show that the estimates of efficiency perform comparably between the two models, although the correctly specified CD version outperforms the translog slightly in smaller sample sizes. The reason for this behaviour is that the estimated regression lines do not differ too much, despite the potential divergences in the estimated coefficients. The slow evaporation of the mismatch between the true values of u and their estimates from the JLMS estimator is due to the fact that JLMS estimator gives only $E(u|\varepsilon)$, but of course there are many possible arrangements of u and v that produce the same $\varepsilon = u + v$. Thus, JLMS estimator is consistent only in the sense that it improves with sample size due to better estimation of the coefficients, which leads to better estimates of the residuals ε [8].

To test the similarity in the characteristics of the estimated regression lines, we calculate the output elasticities for the translog production function at input means and the scale elasticity (for the translog case also at input means). Let ξ denote the elasticity instead of the more customary ε to avoid confusion with the error terms.

Table 3: Output elasticities

Means	$\xi_K = 1/3$		$\xi_L = 2/3$		$\xi_{Scale} = 1$	
N	CD	Translog	CD	Translog	CD	Translog
25	0.352	0.381	0.668	0.659	1.020	1.040
50	0.346	0.342	0.663	0.664	1.009	1.006
100	0.339	0.341	0.661	0.660	1.000	1.000
250	0.335	0.334	0.665	0.665	1.000	0.999
500	0.333	0.331	0.667	0.667	0.999	0.997
1000	0.332	0.332	0.666	0.666	0.999	0.998
5000	0.334	0.335	0.666	0.666	1.000	1.001
Minima	$\xi_K = 1/3$		$\xi_L = 2/3$		$\xi_{Scale} = 1$	
N	CD	Translog	CD	Translog	CD	Translog
25	-0.216	-0.558	0.231	0.238	0.633	0.210
50	-0.037	-0.218	0.370	0.340	0.679	0.479
100	0.107	-0.023	0.432	0.433	0.806	0.640
250	0.137	0.080	0.522	0.523	0.869	0.800
500	0.209	0.164	0.592	0.591	0.906	0.857
1000	0.248	0.206	0.613	0.613	0.937	0.887
5000	0.299	0.284	0.637	0.637	0.977	0.967
Maxima	$\xi_K = 1/3$		$\xi_L = 2/3$		$\xi_{Scale} = 1$	
N	CD	Translog	CD	Translog	CD	Translog
25	1.077	1.386	1.070	1.109	1.625	1.729
50	0.870	0.912	0.981	0.899	1.340	1.540

100	0.598	0.706	0.832	0.838	1.203	1.342
250	0.497	0.560	0.790	0.798	1.116	1.192
500	0.480	0.507	0.728	0.732	1.072	1.144
1000	0.410	0.415	0.718	0.717	1.072	1.074
5000	0.376	0.399	0.697	0.697	1.027	1.049

The results make it very clear that the performance is comparable between the two models, this time it is not only the mean values but also the minima and maxima that match across the two specifications. Hence it seems that if the object of study is either the efficiency or the characteristics of the production frontier, the kitchen sink problem can be avoided.

In addition, it is worth mentioning that a similar A-shaped curve that is formed in the OLS case when the p-values for the null hypothesis of equivalence of models against bias can be obtained here as well. Plotting the p-values of the likelihood ratio test between the Cobb-Douglas and translog specification against the bias in the coefficient β_K in the model with 5000 observations gives:

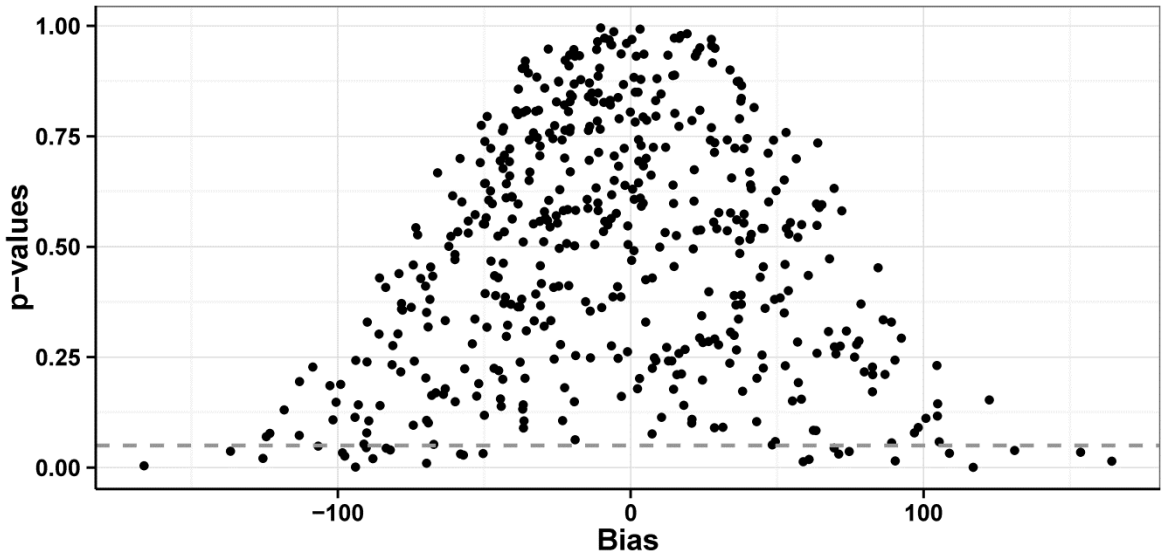


Figure 2 p-values of the LR-test whether the CD and translog specifications are equivalent against the percentage bias in coefficient β_K. Significance of 5% is indicated by dashed line.

Even though here it is not so clear that the most biased models are in fact favoured by the LR test, the result does not even begin to resemble V, which would be the expected shape if the LR test was able to weed out the most biased models. Hence, the choice of model will have to be informed by economic theory rather than the formal test.

4 Conclusion

This article shows that the attempts to prevent omitted variable bias by including extra "control" variables can easily result in problem of identification that persists even in large sample sizes. Adding covariates that are in fact irrelevant to the problem under investigation, but are nevertheless correlated with other regressors yields heavily biased results of the estimated coefficients. The impediment on the convergence of the estimated values to their true values is particularly visible on the extreme values of the coefficients, where the difference between the minimum and maximum can be massively wider in the case of misspecified model compared to the correct model.

Moreover, it has been shown that the standard tests for equivalence between the nested and the augmented model may well favour the inappropriately augmented models precisely in cases when the augmented model suffers from the most serious bias. Hence, these tests cannot be relied upon to identify the most untrustworthy models, but rather the choice of an appropriate specification has to be informed first and foremost by the economic theory.

These are problems that can affect any regression analysis. Every time there is a temptation to estimate a larger model "just to be safe," there is a danger of trading omitted variable bias for problem of identification. In the specific case discussed here, a partial remedy is available, because in the context of frontier analysis, we are not interested in the coefficients of the frontiers *per se*, but rather in the characteristics of the underlying production process, such as the output elasticities and the efficiency scores. The present study shows that these characteristics can be recovered with comparable levels of success between the larger and nested models.

Hence, the results strongly advocate the scrupulous preference for more parsimonious models. The addition of new variables is by no means harmless, especially when a more complicated functional form is imposed, such as the transcendental logarithmic production frontier. Even though models that take into account a wider range of variables may appear more realistic, in cases when their conclusions diverge significantly from the ones obtained by restricted models, the larger models should not be favoured automatically. Since one cannot rely on the sheer sample size to abjure the problem of identification, an appropriate technique has to be used to extract the information from the data correctly. Principal component analysis can be used to reduce the number of covariates, Bayesian model averaging can be used to prevent erroneous inference, even non-parametric techniques might be considered, but ultimately, the best approach is the proverbial grain of salt that is to be added while interpreting the results.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1235/2014. This support is gratefully acknowledged.

References

[1] Barro, R.J.: Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics* **106**(1991), 407-443.

[2] Durlauf, S.N., A. Kourtellos, and C.M. Tan: Are Any Growth Theories Robust? *The Economic Journal* **118**(2008), 329–346.

[3] Schrodt, P.A.: *Seven Deadly Sins of Contemporary Quantitative Political Analysis*. 2010; Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1661045.

[4] Achen, C.H.: Toward a New Political Methodology: Microfoundations and ART. *Annual Review of Political Science* **5**(2002), 423-450.

[5] Easterly, W. and R. Levine: What have we learned from a decade of empirical research on growth? It's Not Factor Accumulation: Stylized Facts and Growth Models. *World Bank Economic Review* **15**(2001), 177-219.

[6] Coelli, T. and A. Henningsen: *frontier: Stochastic Frontier Analysis. R package version 1.0*. 2013; Available from: http://CRAN.R-Project.org/package=frontier.

[7] Jondrow, J., et al.: On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *Journal of Econometrics* **19**(1982), 233-238.

[8] Greene, W.H., The Econometric Approach to Efficiency Analysis. In *The Measurement of Productive Efficiency and Productivity Growth* (H.O. Fried, C.A.K. Lovel, and S.S. Schmidt, eds.). Oxford University Press, New York, 2008.

Cost Transformation in Business Management

Beran Theodor[1], Findová Šárka[2]

Abstract. The aim of this paper is a description of a possible solution unit and performance aspects of the transformation process of the value in the internal management.
The transformation process of the value is in this paper means by two groups of internal management tools. The first set of tools is an accounting methodology solution; the second is the computational procedure using calculations. The synthesis solution is implied in the construction of the so-called conversion matrix.
The importance of corporate accounting is in recognition of the initial breakdown of costs in terms of business as a whole. This system, however, in addition to the automatic control allows extension to lower economic units in the form of so-called internal accounting circuit, which is more flexible due to the extent of its richness of detail and cost breakdown, apparently rendered to internal registers. Internal accounting circuit cannot work without linkages to other economic management tools. One of them is a preliminary calculation of direct costs. The second tool is budget in overhead costs. These links need to cover the mentioned two aspects, namely the aspect of power and formations. Transferability problem is basically solved by using the appropriate calculation method and apparatus matrix identifies the interrelationships cost performance, but also the organizational responsibility of interrelationships. The usefulness of this approach is mainly in the area of tactical control, which is not possible without the application of mathematical methods supported by computer technology.

Keywords: cost, transformation, management tools, calculation, matrix.

JEL Classification: C02
AMS Classification: 91B32

1 Introduction

Enterprise can be generally expressed as a transformation process, whose quantification has two aspects - material page and page value. The paper describes possible solutions unit and the power aspect of the value of internal transformation process. Arrangement priced production factors and evidence of their utility consumption should provide information to support decision making in internal management. The primary source of quantitative information on targeted consumption of factors of production is the corporate balance system - accounting, respectively financial accounting, whose information base is not sufficient for management purposes.

Enterprise management shall obtain information about the organizational and performance aspects of the transformation process value. The paper outlines a path through the so-called transformation matrix of transferability costs.

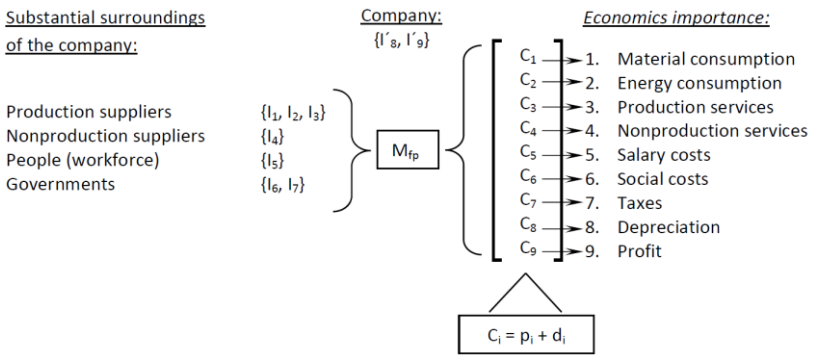


Figure 1 The emergence of cost kinds

[1] ČVUT v Praze, Fakulta strojní, Ústav řízení a ekonomiky podniku, Karlovo nám. 13, Praha 2, theodor.beran@fs.cvut.cz.
[2] ČVUT v Praze, Fakulta strojní, Ústav řízení a ekonomiky podniku, Karlovo nám. 13, Praha 2, sarka.findova@fs.cvut.cz.

where:

$I_1, I_2, \ldots I_n$	… Inputs from the surroundings of the company
Mfp	… market of factors of production
I´	… „company“ inputs
$C_1, C_2, \ldots, C_9$	… basic nine folders of division cost kinds
p_i, d_i	… cost breakdown after transformation, where:
p_i	… indirect folder of cost kind
d_i	… direct folder of cost kind

Enterprise units, cost kinds

$U_1, U_2, \ldots, U_m$ $\quad \sum_{j=1}^{m} C_{ij} = C_{im} \quad \sum_{i=1}^{g} C_{im} = C_P$

N_P … total costs (cost kinds for the whole company)

$$\begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1m} \\ C_{21} & C_{22} & \ldots & C_{2m} \\ C_{31} & C_{32} & \ldots & C_{3m} \\ C_{41} & C_{42} & \ldots & C_{4m} \\ C_{51} & C_{52} & \ldots & C_{5m} \\ C_{61} & C_{62} & \ldots & C_{6m} \\ C_{71} & C_{72} & \ldots & C_{7m} \\ C_{81} & C_{82} & \ldots & C_{8m} \\ C_{91} & C_{92} & \ldots & C_{9m} \end{bmatrix}$$

Figure 2 The sum of cost kinds according to business units - matrix

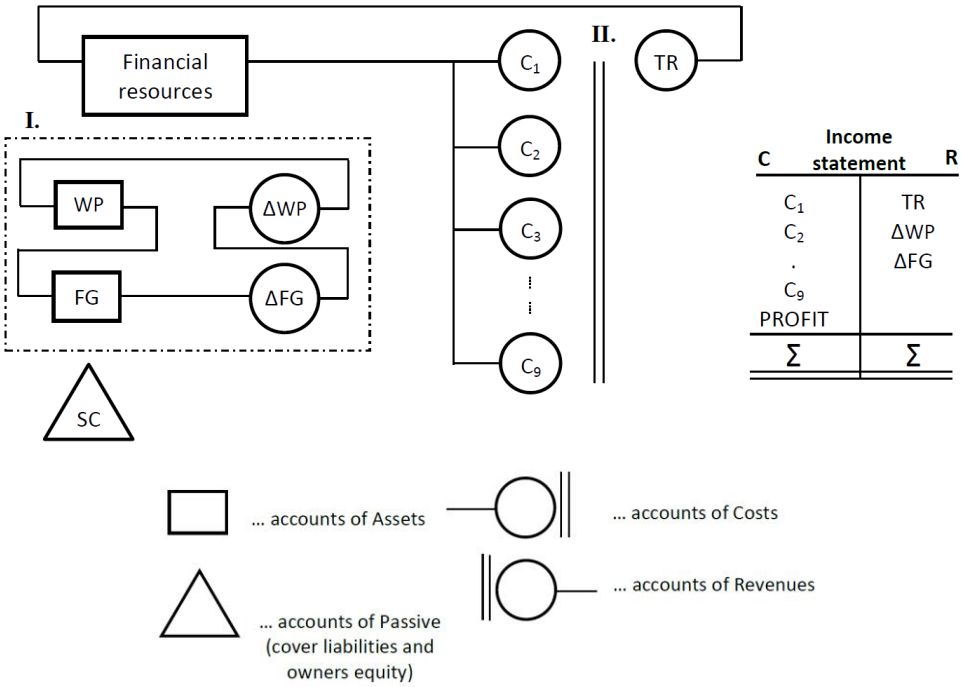


Figure 3 Cost kinds in financial accounting

Legend:

I. … „company-wide“ level awards work in progress (WP) and finished goods (FG)
II. … interruption of individual accounting lines
SC … share capital
TR … total revenues

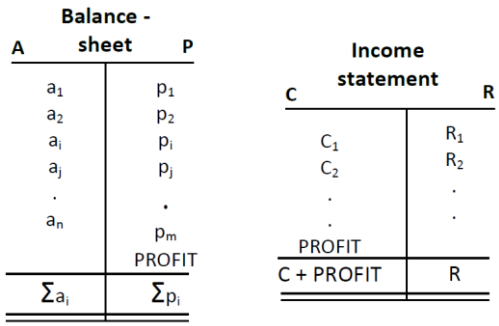


Figure 4 Statements of company

Internal accounting provides data for costing of own costs and outputs (products or services, i. e. product) and thus information about their profitability, provides data to monitor variances (controlling) and is used for control of the activities of departments and centres.

2 Function of internal accounting

Internal accounting is capable of objectification of certain information for internal control in financial terms.

Internal accounting monitors the financial information on production and non-production flows, records performance transfers between centers, that financial accounting does not record. Internal accounting provides information about formation of balance sheet profit, which is reported in the balance sheet and the income statement in the financial and internal accounting. There are evident differences between actual and planned costs in internal income statement. Each center can have savings, compliance, and cost overruns and can compile a detailed income statement. Overviews and reports of actual and projected costs used for the analysis of the causes of emergence variations (deviations).

Original and coherent functioning system composed of elements of financial management and controls (audit) and financial accounting is supplemented by internal accounting. This system is based on the integration of selected elements in the management system, but also represents a "firming" and the "strengthening" of the whole system. The advantages of this new structure stems from the above-mentioned benefits of internal accounting.

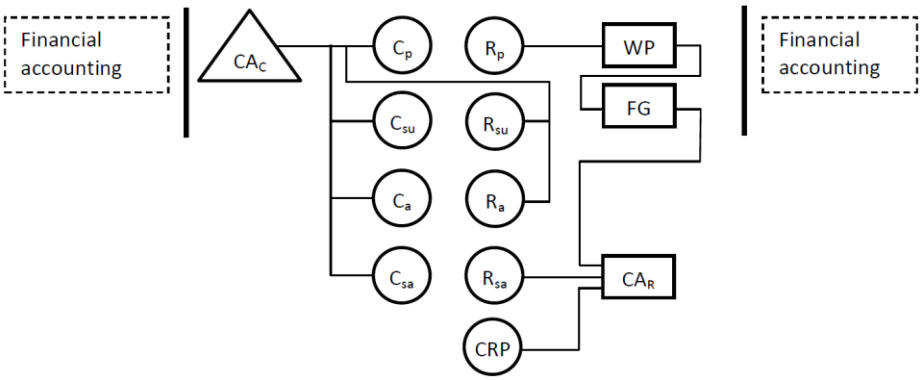


Figure 5 Internal accounting circle

Legend:

CA_C	… linking account of costs
CA_R	… linking account of revenues
C	… actual costs
R	… planning costs = revenue of centres
p	… production centre
su	… supply centre
a	… administration centre
sa	… sales centre
PP	… work in progress
FP	… finished goods
CRP	… calculated realised profit

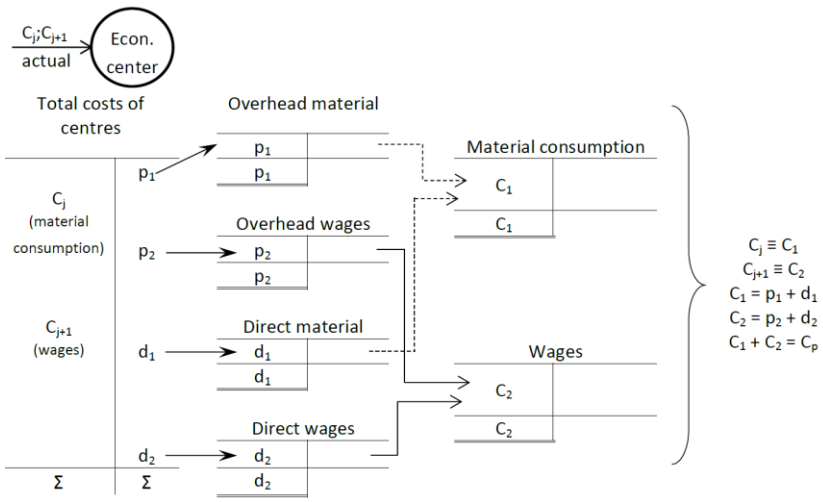


Figure 6 Costs in the economic center – Accounting expresion

3 The generation of cost kinds

We understand the kind of homogenous cost item of production factor that enters into the company of significant surroundings of the company. In the diagram (Fig. 1) is formed and the formation of the basic derivation nine-folder breakdown of costs required for other purposes.

In short, the cost kind is inferred from the surroundings in the company. In Fig. 2 is a matrix of relationships between business units and cost kinds. The consequences of this division are shown in Fig. 3. Financial Accounting captures just kind of costs, after which the posting is transferred to the billing revenues (marked ‖). There is no obvious value process, thus valuing of production. According to I. designation in Fig. 3 it is aggregate quantity WP without any explanation of its creation.

This fact can be considered as the initial reason for the creation of the second accounting circuit internal accounting. The second of the circuits is one of the instruments of mutual cost transferability and valuation tool for WP and FG. The basic layout of internal accounting is depicted in Fig. 4.

Costing model (costing breakdown of costs)

Performance	Dir. wages	Dir. mater.	Oth. dir.cos.	PC_T	PC_G	SuC	AC	SaC	Σ
	d_{11}	d_{21}	d_{31}	r_{41}	r_{51}	r_{61}	r_{71}	r_{81}	U_1
	d_{12}	d_{22}	d_{32}	r_{42}	r_{52}	r_{62}	r_{72}	r_{82}	U_2
	d_{13}	d_{23}	d_{33}	r_{43}	r_{53}	r_{63}	r_{73}	r_{83}	U_3
	.	.	.	.	.	.	.	.	U_4
	.	.	.	.	.	.	.	.	U_5
	.	.	.	.	.	.	.	.	U_6
	.	.	.	.	.	.	.	.	U_7
	.	.	.	.	.	.	.	.	U_8
	d_{1q}	d_{2q}	d_{3q}	r_{4q}	r_{5q}	r_{6q}	r_{7q}	r_{8q}	U_9
	d_1	**d_2**	**d_3**	**r_4**	**r_5**	**r_6**	**r_7**	**r_8**	**U_P**

Figure 7 Costing breakdown of cost per outputs

[d_1, d_2, d_3, r_4, r_5, r_6, r_7, r_8] … costs broken down in costing breakdown for the company

Costing breakdown → Kind breakdown ↓	Direct wages	Direct material	Other direct costs	PO_T	PO_G	SuO	AO	SaO	C_i
1. Material consumption									
2. Energy consumption									
3. Production services									
4. Nonprod. services									
5. Salary costs									
6. Social costs									
7. Taxes									
8. Depreciation									
Cost total									

Figure 8 Costing performance expressed by cost kinds

Evidence support for the filling matrix in overhead costs

- Energy
- Transport
- Maintenace
- Tool
- Main production
- Secondary production

PO_T, PO_G

- SuC – supply centre
- AC – construction, design, administrative, technology, social activity, environmental centre
- SaC – sales centre

where:

PO_T … production overheads technological
PO_G … production overheads general
SuO … supply overheads
AO … administrative overheads
SaO … sales overheads

If we denote the cost for the total production volume of C, formally written:

$$C = \sum_{i=1}^{w} C_f \quad (1)$$[3]

Formal expression of overhead rates costing method:

$$C = \sum_{k=1}^{w}\left[\sum_{i=1}^{s} d_{if} + \sum_{j=1}^{m} \Phi_j d_{1f}\right] \quad (2)$$

Checking the correctness and then perform steps:

$$C = D_C + P_C = \sum_{i=1}^{sn} C_i \quad (3)$$

The economic importance (meaning) of relations

Cf … cost types of product, marked f, product groups is intended w
Φj … designation "operator of transformation" that is the appropriate rate regime, which identifies the index j and total transformation centers, thus overhead expresses index m

Practical importance is comparison and found that in achieving the transformation usually begin to occur inequality in relation below, as do the number of direct and indirect items of cost due to the number of kinds.

$$C_i = d_i + p_i \quad (4)$$

i = 1, 2,…,n
s < n, where the index s indicates the number of direct cost items.

3.1 Groups mistakes of overhead rates costing

1. Tendency to generalize everything, and what could directly identify the product, to directing, it becomes indecipherable set of all potential costs and does not respect the principle of economy.
2. Enormously high overhead is allocate as a whole without distinguishing kinds of overhead - by one base, which are e.g. direct wages.

3.2 Recommendations for practice

1) Performing a systematic analysis of overhead costs that were excluded items that can be converted into direct costs.
2) Perform differentiated selection of costing allocation base in order to affect such a base on which depends the overhead (e.g. production area, mechanical hours, weight, volume,…).

Analysis of overhead is grounded in the gradual uncovering of so-called the costs induced.

[3] K. Macík, 2007, p. 38 – where the original formal and mathematical costing technique worked the overhead rates costing.

4 Conclusion for practise

Practical application is associated with three major process steps:

1. Detailed analysis of individual cost elements (cost kinds). Objective: The purpose of accurate incurred costs kinds so that we can cut back on items that are among the direct costs.
2. Refinement of costing methods. Objective: To accurate the overall enterprise transformation matrix in the range of overhead. It is necessary to detect relationships (dependencies) between overheads and performance (output).
3. Establishing the overall matrix for the company. Objective: finding linkages between departmental and costing breakdown of costs.

Resulting transition matrix for the whole company is a certain approximation. Matrix brings significant knowledge about the composition of the overhead that triggers corrective action regarding the first and second step. After completing these steps, we carry gear assembly matrixes of outputs. The construction of the transmission matrixes is a tool of internal management and management of whole company. More accurate of transmission matrixes lead to fair idea about the value flows in the company.

References

[1] Alnajjar, A. I., Siam, W. Z. The ability of application activity-based costing system on the air line companies: The case of the jordan aviation company. European journal of economics, finance, and administrative sciences. 2011, vol. 38, p. 138-154, ISSN 14502275.

[2] Anderson, S. W., Sedatole, K. L.. Evidence on the cost hierarchy: The association between resource consumption and production activities. Journal of management accounting research. American Accounting Association, 2013, vol. 25, no. 1, p. 119-141. DOI: 10.2308/jmar-50293, ISSN: 10492127.

[3] Beran, Th. *Oceňování výkonů v praxi vnitropodnikového řízení*. 1. vyd. V Praze: ČVUT, 2010, 216 s. ISBN 978-80-01-04998-3.

[4] Blocher, E. Cost management: a strategic emphasis. 6th ed. New York, NY: McGraw-Hill/Irwin, c2013, xxvi, 929 p. ISBN 0078025532.

[5] Buchanan, J. M.,Thirlby, G. L.S.E. essays on cost. Studies in economic theory, New York: New York University Press, 1981, c1973., 290 p.,ISBN 0814710352.

[6] Carroll, P. Overhead cost control. New York: McGraw-Hill, [1964], xii, 314p.

[7] Horngren, Ch. T. Cost accounting: a managerial emphasis. 13th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009, xxvii, 870 p. ISBN 0136126634.

[8] Kaplan, R. S., Atkinson A. Advanced management accounting. 3rd ed. Upper Saddle River, N. J.: Prentice Hall, c1998, xviii, 798 p. ISBN 0132622882.

[9] Lewis, William Arthur. Overhead costs: Some essays in economic analysis. United Kingdom: University of Manchester, 2013, 200 p. ISBN 978-020370873-6; DOI: 10.4324/9780203708736.

[10] Lucey, T. Costing. 3rd ed. London: DP Publications Ltd., 1991, 609 s. ISBN 1870941349.

[11] Macík, Karel. *Kalkulace a rozpočetnictví*. Vyd. 3., přeprac. Praha: Nakladatelství ČVUT, 2008, 213 s. ISBN 978-80-01-03926-7.

[12] Romanovsky, E. A. Hancock, N. N. Application of the device of slices for automation of matrix calculations. IOP conference series. 2014, Bristol: IOP Pub., p. 123-138. DOI: 10.1007/springerreference_18585. ISSN 1757-8981.

[13] Solomons, D. Divisional performance. Measurement and control. Homewood, Ill.: R. D. Irwin, [1968, c1965], xi, 307 p.

[14] Strumickas, M., Valanciene, L. 2010. Development of Modern Management Accounting System: Inzinerine ekonomika-engineering economics. Pragae: Sumptibus Academiae scientiarum Bohemoslovacae, p. 377-386, WOS:000283840200005, ISSN 1392-2785.

[15] Vancil, R. F.,Buddrus, L. E. Decentralization, managerial ambiguity by design: a research study and report. Homewood, Ill.: Dow Jones-Irwin, c1979, xiv, 393 p. ISBN 0910586330.

Context-Aware Recommendation in E-commerce

Ladislav Beránek[1]

Abstract. Over the last decade, most of researchers have studied new approaches of recommender systems in connection of the two main approaches, collaborative filtering and content-based filtering. The aim is to gain a system for real world situations. Specifically, applying data mining techniques to recommender systems has been effective in providing personalized information to the user by analyzing his or her preferences.
This paper proposes a context-based information filtering mechanism to enhance context aware approach. It specifically proposes a strategy for recommending based on context which is defined by various ways. Dempster-Shafer evidential theory for handling uncertainty and ignorance is exploited in order (i) to model user context acquisition process, (ii) to map user context and behaviors to recommend and (iii) to quantify behaviors usage patterns. A contextual update strategy is also proposed in order to dynamically adapt the recommendation offering according how the users consume those behaviors. Some preliminary validation tests were performed.

Keywords: E-commerce, Dempster-Shafer theory, context-awareness, recommendation systems.

JEL Classification: C49
AMS Classification: 28E15

1 Introduction

Recommendation systems have become an integral part of e-business applications. Their goal is to capture users' preferences and offer them relevant products and services. Users leave a mark within their interactions with e-business applications, weather consciously (evaluation items, filling in the questionnaire or user profile information such as age, residence, education or gender) or unconsciously (movement on the page, the length of stay, etc.). This mark is used as a feedback by e-business application operators. Based on this information it is then possible to draw conclusions or assumptions about the user's preference to a particular object. However, these functions do not provide benefits only shopkeepers, but also for users who gain a better overview of the offered products and services is complex and may look for them if they were interested.
At the present time, users access to e-business systems especially using smart phones and similar devices. These devices contain also a large variety of sensors, including a GPS sensor and other sensors that assess the environment and provide information about the context of the surroundings (location, weather information on the place and others). This context can be evaluated from a user's mobile device as a condition of some event action rules (e.g., "If I'm near certain places in the city, then Indian restaurants XY can be recommended to me, because my preference is Indian cuisine". This potential awareness contextual supports the creation of new techniques for filtering the available information and to adapt it to the specific needs of the user.
This paper proposes a context-based information filtering mechanism with the participation of contextual information. Specifically, it proposes a new strategy and recommendations strategy of increasing importance contextual information. Dempster-Shafer theory of evidence capabilities for handling uncertainty and ignorance are so used to (i) the model of the user context of the acquisition, (ii) to map the user's context and recommend behavior and (iii) to quantify the behavior and use patterns. Contextual strategy update also proposes to dynamically adapt the recommendations offer according to how users consume behavior. Overview of Dempster-Shafer theory and how it can be used for the purpose of the recommendations is the focus of section 2. Section 3 examines related works with focus on the relationship between Dempster-Shafer theory and contextual recommendation systems. The suggested model is recommendations described in Chapter 4. Finally, Section 6 analyzes some preliminary validation tests and Section 7 offers some conclusions.

2 Dempster-Shafer Theory

Information related to decision making about cyber situation is often uncertain and incomplete. Therefore, it is of vital importance to find a feasible way to make decisions about the appropriate response the situation under this uncertainty. Our model is a particular application of the Dempster-Shafer theory. The Dempster-Shafer theory [16] is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool

[1] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, beranek@ef.jcu.cz.

for combining evidence and changing prior knowledge in the presence of new evidence. The Dempster-Shafer theory can be considered as a generalization of the Bayesian theory of subjective probability [8]. In this paper, we propose a unique trust model based on the Dempster-Shafer theory which combines evidence concerning reputation with evidence concerning possible illegal behavior on an Internet auction.
In the following paragraphs, we give a brief introduction to the basic notions of the Dempster-Shafer theory (frequently called theory of belief functions or theory of evidence).

2.1 Basic Notions

Considering a finite set referred to as *the frame of discernment* Θ, a *basic belief assignment (BBA)* is a function m: $2^{\Theta} \rightarrow [0,1]$ so that

$$\sum_{A \subseteq \Theta} m(A) = 1, \tag{1}$$

where $m(\emptyset) = 0$, see [16]. The subsets of 2^{Θ} which are associated with non-zero values of m are known as *focal elements* and the union of the focal elements is called *the core*. The value of $m(A)$ expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Θ belongs to the set A but not to a particular subset of A. This value pertains only to the set A and makes no additional claims about any subsets of A. We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).
Shafer further defined the concepts of *belief* and *plausibility* [16] as two measures over the subsets of Θ as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \tag{2}$$

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \tag{3}$$

A *BBA* can also be viewed as determining a set of probability distributions P over Θ so that $Bel(A) \leq P(A) \leq Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(\neg A)$. *Moreover both of them are equivalent to m.* Thus one needs to know only one of the three functions *m*, *Bel*, or *Pl* to derive the other two. Hence we can speak about belief function using corresponding *BBA*s in fact.
Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *BBA*s m_1 and m_2 is defined as follows (the symbol $\oplus$ is used to denote this operation):

$$(m_1 \otimes m_2)(A) = \frac{1}{1-k} \sum_{B \cap C \acute{A}} m_1(B) \cdot m_2(C), \tag{4}$$

where

$$k = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \tag{5}$$

Here k is frequently considered to be a *conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 [16]. Unfortunately this interpretation of k is not correct, as it includes also internal conflict of individual belief functions m_1 and m_2 [6]. Demspter's rule is not defined when $k = 1$, i.e. when cores of m_1 and m_2 are disjoint. This rule is commutative and associative; as the rule serves for the cumulation of beliefs, it is not idempotent.
When calculating contextual discounting we also use the un-normalized (conjunctive) combination rule in the form [16] (we use the symbol $\oplus$ to denote this operation):

$$(m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C). \tag{6}$$

2.2 Operations in Product Frames

In many applications, we need to express uncertain information about several variables taking values in different domains. Let A or B be two elements (hypotheses) belonging to frames of discernments Ω_A and Ω_B. We can then define the product frame $\Omega_{AB} = \Omega_A \times \Omega_B$. Mass function $m^{\Omega_A \times \Omega_B}$ on Ω_{AB} can be seen as an uncertain relation between elements A and B. The basic operations on product space are the following [14]:

Marginalization

A mass function defined on a product space $\Omega \times \Theta$ may be marginalized in Ω by transferring each mass $m^{\Omega \times \Theta}(B)$ for $B \subseteq \Omega \times \Theta$ to its projection into Ω:

$$m^{\Omega\times\Theta\downarrow\Omega}(A) = \sum_{\substack{B\subseteq\Omega\times\Theta \\ Proj(B\downarrow\Omega)=A}} m^{\Omega\times\Theta}(B) \tag{7}$$

for all $A \subseteq \Omega$. Here $Proj(B\downarrow\Omega)$ denotes the projection of B into Ω.

Conditioning on a Product Space

Conditional beliefs represent knowledge that is valid provided that a hypothesis is satisfied. Let m be a mass function and $B \subseteq \Omega$ a hypothesis (with $m(B) = 1$). The conditional belief function $m(A|B)$ is given by (we use the un-normalized conditioning here):

$$m(A \mid B) = \sum_{\{C|C\cap B=A\}} m(C)\,. \tag{9}$$

This equation is often written for practical reasons in the form: $m^A[B] = m \otimes m_B$.

Let $m^{\Omega\times\Theta}$ be defined on the product space $\Omega\times\Theta$, and θ_0 is an element of Θ, then the conditional *bba* $m^{\Omega}[\theta_0]$ is defined by combining $m^{\Omega\times\Theta}$ with $m_{\theta_0}^{\Theta\uparrow\Omega\times\Theta}$ (with $m_{\theta_0}^{\Theta}(\theta) = 1$), and marginalizing the result on Ω:

$$m^{\Omega}[\theta_0] = \left(m^{\Omega\times\Theta} \otimes m_{\theta_0}^{\Theta\uparrow\Omega\times\Theta}\right)^{\downarrow\Omega}. \tag{10}$$

De-conditioning on a Product Space (Ballooning Extension)

Assume now that $m^{\Omega}[\theta_0]$ represents beliefs conditional on θ_0, i.e., in a context where θ_0 holds. There are usually many *bba*s on $\Omega\times\Theta$, whose conditioning of θ_0 yields $m^{\Omega}[\theta_0]$. Among these, the least committed one is defined for all $A \subseteq \Omega$ by:

$$m^{\Omega}[\theta_0]^{\hat{\uparrow}\Omega\times\Theta}(A\times\theta_0 \cup \Omega\times\overline{\theta_0}) = m^{\Omega}[\theta_0](A)\,. \tag{11}$$

This operation is referred to as the de-conditioning or ballooning extension [18] of $m^{\Omega}[\theta_0]$ on $\Omega\times\Theta$.

3 Related Work

Context-aware recommender systems are relatively new field of study [1]. The ability to provide some respective recommendation from these systems depends on the activities of the various sensors and the correct way to evaluate data from these sensors. The determination of the occurrence of situation, in which a user occurs, is an essential function for context-aware systems [2]. Evaluation of data from the sensors, because of their nature [12], is not easy. Many reasoning techniques are used to evaluate and infer the current situation [10], [11], [18]. Bayesian methods are quite popular, e.g., [15], [20]. Further techniques as fuzzy logic [7], also in combination with semantic web [5] or ontologies [12] or hidden Markov models [10], are used as well. However, these models usually require some preliminary information. Preferably, there is also belief functions theory used, see for example [13], [15].

Context-aware recommender systems are described in detail in [1], where several techniques are mentioned for implementing model-based recommendations, i.e., predictive models for calculating the probability with which the user chooses a certain type of item in a given context (e.g., support vector machines or Bayesian classifiers). As a generalization of the Bayesian probability theory, Dempster-Shafer theory extends uncertainty support, e.g., by explicitly representing ignorance in the absence of information, by offering a simple mechanism for evidence propagation or by a limited reliance on training data [9],[13]. However, it is difficult to find in the literature references to systems implementing Dempster-Shafer theory mechanisms for supporting recommendations. It is necessary to search within the decision making area in order to find researches implementing Dempster-Shafer theory based intelligent selections mechanisms. Most of them are based on payoff matrices, built by experts, linking several states of nature to different alternatives and where the knowledge of the state of nature is captured in terms of a belief mass function [4]. Based on this idea, our work also proposes to model the quantification of the context and the alternatives, e.g., items to recommend.

4 Basic Concepts of Our Model

The model for situation awareness proposed in this paper is based on the processing of the data and tabs from various sensors. Data from the sensors are processed by means of filtration and data integration. Then, mass functions are derived, alternative frames of discernment are constructed, and the comparisons with adequate frames of discernment stored in the database are accomplished. Recommendations on the basis of users' preferences are provided.

4.1 Processing data from sensors and situation inference

Data sources used for situation detection are very different, derived from various sensors (time sensor, position sensor, access logs, etc.) and other resources (for example data from human resource department - personal characteristics including personality and capability). Therefore, in the first phase, it is necessary to convert all messages received on the situation observed in a standard format. In addition, these standardized records are filtered and integrated. The aim is to simplify and eliminate redundant records, to remove records that do not meet certain requirements. These requirements may be stored in the knowledge base in the form of attribute rules and be used according to the status of the situation. The record can be removed; for example, in the absence of a key attribute of the described events or when its value is out of range and thus not relevant for the analysis of the situation.

Belief functions (basic belief assignments, mass functions) are calculated at first. For example at they can be derived from sensor reliability or can also be quantified on the basis of inference rules. For example, we can trace from various sensors that user carries out the following activities: being on some place near the restaurant with favorite cuisine, being at home, jogging or other activities which sensors are able to distinguish and infer situation in which respective user occurs.

		Recommendation hypotheses 2^{Θ}							
		ϕ	r	t	s	rt	rs	ts	rst
Context hyptheses 2^{Ω}	ϕ	1.0	0	0	0	0	0	0	0
	a	0	0.6	0	0	0	0	0	0.4
	b	0	0.1	0.6	0	0.1	0	0	0.2
	c	0	0	0	0.7	0	0.1	0.1	0.1
	ab	0	0.1	0.1	0	0.8	0	0	0
	ac	0	0.2	0.1	0.1	0	0.6	0	0
	bc	0	0	0	0	0	0	0.9	0.1
	abc	0	0	0	0	0	0	0	1.0

Table 1. Example of a context-resource recommendation (scheme of the relevant product frame)

Now, we have some uncertainty about these different aspects of a current situation. This information is expressed using established belief functions. We have no assumption that atomic elements are sets of elements of the same framework as it may relate to various aspects of the same phenomenon (situation). Instead, we believe that they can be part of various homogeneous parts of frames whose Cartesian product will be a framework that represents all the possibilities of the problem. Even further, this may be revised whenever there is new information and framework may need to be expanded to include the possible outcomes that were not previously known [19].

Since there may be several different alternative frameworks for each time point, we find the most suitable framework for resolution. We define the fitness resolution framework to meet two different aspects simultaneously. Hence we construct the frames of discernment over the data obtained from the first phase. We then chose the most appropriate frame, which has the lowest internal conflict. The Dempster's rule of combination is one of the basic notions of belief function theory [16]. It can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information [2], [3]. In this work sensors are considered to be evidential, i.e., they estimate reality in the form of a belief mass function $m_i(\cdot)$, that can be used in order to calculate belief functions concerning recommendation. Table 1.a exemplifies modelling the estimates obtained from a location system with 3 possible symbolic locations $\{a, b, c\}$. For instance, it states that *'if the location sensor estimates that the user is located at "c", then the user is actually located at "b" or "c" with an evidence of 0.1 and ...'*. It is worth noting that this sensor modelling includes ignorance modelling at evidence level in the form of the belief mass assigned to combinations of the singletons within Ω_i .

4.2 Recommendation construction

We use operations defined in Section 2.2 Operations in Product Frames for an evidential decision making process concerning recommendation. We have a the product frame on the power sets Ω and Θ. The evidential mapping links each context (elements of Ω) modelled according a belief mass function obtained from the above mentioned sensor modelling process with a common frame of discernment $\Theta = \{r_1, r_2, ..., r_N\}$ representing resources to recommend. Once again, equations (10), (11) have been used in order to calculate $m_{Ri}(\cdot)$, i.e., the

partial belief mass function representing evidences regarding how to prioritize the respective recommendation. Table 1 demonstrate the relation between for example the possible locations of a user and the recommendation to be prioritized (only 3 recommendation strategies are considered in this example: {*r,s,t*}). For instance, it states that *'if the user is located at "c", then the resources "r" or "s" should be recommended with an evidence of 0.6 and ...'*.

5 EXPERIMENTAL RESULTS

To verify the proposed framework for the situation inference, we created simulated home environment with six various sensors. In the first phase, we focused on the kitchen activities. We selected consecutive time frames describing the "breakfast preparation" activity (see Table 2). The sensor indicates for example that a toaster is in operation. Here, the constructed alternative frames of discernment from the belief functions m_1 and m_2. From these alternative frames of discernment, we chose the one that is most appropriate. After processing possible activities included in this frame of discernment together with data from database obtaining recommendations (frame of discernment) corresponding to the respective activities, we obtain the specification of the recommendation with the highest value of belief function. We are describing this process very briefly here and on a simple example. The process of constructing and evaluating all possible frames of discernment can be briefly described in the following way: at first we generate the possible frames using different partitions of the set of all cores $\{C_i\}$ from the belief functions. Then we generate the set of all partitions $\{\Omega_k\}$ of $\{C_i\}$ and we use $\{\Omega_k\}$ to generating the set of all possible cross products $\{\Theta_k\}$. Then all possible frames of corresponding recommendations $\{\Theta_k\}$ are evaluated using the equations $(9-11)$.

Sensors events (evidences)	Belief functions (situation)	Belief functions (recommendation)	Resulting recommendation
fridge, freezer, cook-stove, coffee maker, toaster motion sensor	m_1:{[{food preparation , coffee preparation}, m_1({food preparation , coffee preparation }) = 0.6] {[{food preparation , toast preparation}, m_1({food preparation , toast preparation }) = 0.4] m_2: {[{Movement in kitchen, Fast movement in kitchen}, m_2({Movement in kitchen, Fast movement in kitchen }) = 0.2] [{Movement in kitchen}, m_2({Movement in kitchen }) = 0.8] }	m_1: {[{recommendation food, recommendation coffee}, m_1({recommendation food , recommendation coffee }) = 0.8] m_2: {[{recommendation coffee maker, recommendation kitchen furniture}, m_2({recommendation coffee maker, recommendation kitchen furniture}) = 0.6]	Recommendation: Coffee
fridge, freezer, cook-stove, coffee maker, toaster, motion sensor	m_1: {[{food preparation , coffee preparation}, m_1({food preparation , coffee preparation }) = 0.3] {[{food preparation , toast preparation}, m_1({food preparation , toast preparation }) = 0.6] [{food preparation}, m_1({food preparation}) = 0.1]} m_2: {[{Movement in kitchen, Fast movement in kitchen}, m_2({Movement in kitchen, Fast movement }) = 0.8] [{Movement in kitchen}, m_2({Movement in kitchen }) = 0.2] }	m_1: {[{recommendation cook-stove}, m_1({rec. toaster, rec. coffee maker}) = 0.2] {[{rec. fridge, rec. cook stove}, m_1({rec. fridge, rec. coffee maker }) = 0.8] m_2: {[{rec. freezer, rec. cook-stove}, m_2({rec. freezer, rec. cook-stove }) = 0.9] [{rec. toaster }, m_2({rec. cook-stove }) = 0.1] }	Recommendation: Freezer
coffee maker, toaster, dishwasher, motion sensor	m_1: {[{ coffee preparation, washing dishes}, m_1({coffee preparation, washing dishes }) = 0.3] {[{ toast preparation. washing dishes}, m_1({toast preparation, washing dishes }) = 0.6] [{washing dishes}, m_1({washing dishes}) = 0.1]} m_2: {[{Movement in kitchen, Fast movement in kitchen}, m_2({Movement in kitchen, Fast movement in kitchen }) = 0.8] [{Movement in kitchen}, m_2({Movement in kitchen }) = 0.2] }	m_1: {[{rec. dishwasher, rec. washing powder}, m_1({rec. dishwasher, washing powder}) = 0.6] {[{rec. detergent, washing powder}, m_1({rec. dishwasher, washing powder}) = 0.4] m_2: {[{rec. dishwasher, rec. kitchen furniture }, m_2({rec. dishwasher, rec. kitchen furniture }) = 0.8] [{rec. detergent, rec. kitchen towel}, m_2({rec. detergent, rec. kitchen towel}) = 0.2] }	Recommendation: dishwasher

Table 2. Example of data processed within the framework.

Our approach incorporates context quality information into sensor evidence, construction of alternative frames of discernment concerning situation and recommendation suggestions. We demonstrate here our approach on a simple case study.

6 CONCLUSION

In this paper, we propose a model intended for security situation identification. This model is mainly based on the use of the belief function theory which reflects better the uncertain character of the process of security situation detection. We describe here some results of our initial study. In our future activities, we want to analyze these procedures more deeply. We are preparing more experiments with the aim to especially improve the procedures concerning the resulting description of the situation, i.e. procedures pertaining to the extraction of the knowledge from processed data from sensors and from data stored in the database.

Refrence

[1] Adomavicius, G., Tuzhilin, A., Context-aware recommender systems. In: ACM Conference on Recommender Systems, 2008, pp. 335–336

[2] Beranek, L., Knizek, J., The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions. *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 7, No. 1, 2012, pp. 34-50.

[3] Beranek, L., Tlusty, P., Remes, R., The Use of Belief Functions for an Online Auction Reputation Model. In *Proceedings of the 28th International Conference on Mathematical Methods in Economics* (*MME 2010*). Ceske Budejovice: University of South Bohemia, 2010, pp. 49 - 54.

[4] Casanovas, M., Merigo, J.M.: Fuzzy aggregation operators in decision making with Dempster-Shafer belief structure. Expert Systems with Applications Vol. 39, 2012 pp. 7138–7149

[5] Ciaramella A., Cimino M., Marcelloni F., Straccia U., Combining fuzzy logic and semantic web to enable situation-awareness in service recommendation, In. *Proceedings of the 21st international conference on Database and expert systems applications*. Berlin: Springer, 2010, pp. 31-45.

[6] Daniel, M., Several Notes on Belief Combination. In *Proceedings of the Theory of Belief Functions* Workshop. Brest: ENSIETA, 2010. pp. 1-5.

[7] Furno D., Loia V., Veniero M., A fuzzy cognitive situation awareness for airport security. *Control and Cybernetics*, Vol. 39, No. 4, 2010, pp. 959-982.

[8] Haenni, R., Shedding new light on Zadeh's criticism of Dempster's rule of combination. In *Proceedings of the Eighth Int. Conf. on Information Fusion*, Philadelphia, IEEE, 2005, pp. 879–884.

[9] Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., Devlin, S.: Evidential fusion of sensor data for activity recognition in smart homes. Pervasive and Mobile Computing Vol. 5, No. 3, 2009, pp. 236–252

[10] Liao J., Bi X., Nugent Ch., Activity Recognition for Smart Homes Using Dempster-Shafer Theory of Evidence Based on a Revised Lattice Structure. In *Proceeding of the 2010 Sixth Int. Conf. on Intelligent Environments*, Washington IEEE, 2010, pp. 46-51.

[11] van Kasteren T., Noulas A., Englebienne G., Krose B., Accurate activity recognition in a home setting. In *Proceedings of 10th International Conference on Ubiquitous Computing*, South Korea, 2008, pp. 1-9.

[12] Matheus C.J., Kokar M.M., Baclawski K., A Core Ontology for Situation Awareness. In *Proceedings of Sixth International Conference on Information Fusion*. Cairns, Australia, 2003, pp. 545-552.

[13] McKeever S., Ye J., Coyle L., Dobson S., Using Dempster-Shafer Theory of Evidence for Situation Inference. In *Proceedingd of the Fourth European Conference Smart Sensing and Context*, Lecture Notes in Computer Science 5741, 2009, pp. 149-162.

[14] Mercier, D., Lefèvre, E., Delmotte, F., Belief functions contextual discounting and canonical decompositions, *International Journal of Approximate Reasoning*, Vol. 53, No. 2, 2012, pp. 146-158.

[15] Ranganathan, A. et al, Reasoning about uncertain contexts in pervasive computing environments. *Pervasive Computing*, IEEE , Vol. 3, No 2, 2004, 62-70.

[16] Shafer, G., *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, 1975.

[17] Shubert, J., 2012. Constructing and reasoning about frames of discernment, *International Journal of Approximate Reasoning*, Vol. 53, No. 1, 2012, pp. 176-189.

[18] Skorpil, V., Stastny, J. Comparison Methods for Object Recognition. In *Proceedings of the 13th WSEAS International Conference on Systems*. Rhodos, Greece, 2009, pp. 607-610.

[19] Smets, P., Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, Vol. 9, No. 1, 1993, pp. 1-35.

[20] Ulicny, B. et al, Augmenting the Analyst via Situation-Dependent Reasoning with Trust-Annotated Facts. In *Int. Conf. on Cognitive Methods in Situation Awareness*. Miami Beach: IEEE, 2011, pp. 17-24.

L-moments and TL-moments

Diana Bílková[1]

Abstract. Application of the method of moments for the parametric distribution is common in the construction of a suitable parametric distribution. However, moment method of parameter estimation does not produce good results. An alternative approach when constructing an appropriate parametric distribution for the considered data file is to use the so-called order statistics. This paper deals with the use of order statistics as the methods of L-moments and TL-moments of parameter estimation. L-moments have some theoretical advantages over conventional moments. L-moments have been introduced as a robust alternative to classical moments of probability distributions. However, L-moments and their estimations lack some robust features that belong to the TL-moments. TL-moments represent an alternative robust version of L-moments, which are called trimmed L-moments. This paper deals with the use of L-moments and TL-moments in the construction of models of wage distribution. Three-parametric lognormal curves represent the basic theoretical distribution whose parameters were simultaneously estimated by three methods of point parameter estimation and accuracy of these methods was then evaluated. There are method of TL-moments, method of L-moments and maximum likelihood method in combination with Cohen's method. A total of 328 wage distribution has been the subject of research.

Keywords: order statistics, L-moments, TL-moments, maximum likelihood method, probability density function, distribution function, quantile function, lognormal curves, models of wage distribution.

JEL Classification: C46, C89, D31
AMS Classification: 60E05, 62E99, 62H12, 62F10

1 Introduction

The advantages of L-moments and TL-moments methods are obvious when applied to small data sets, predominantly in the fields of hydrology, meteorology and climatology, considering extreme precipitation in particular. The main aim of this paper is to utilize the two methods of parameter estimation in large data sets and compare their accuracy to that of the maximum likelihood method in combination with Cohen's method (farther only maximum likelihood method).

Data concerning the wage distribution (in the form of an interval frequency distribution) for the period 2003–2010 were downloaded from the official CSO website. Three-parameter lognormal curves represent the basic theoretical probability distribution. For all analyzed wage distributions, the model distribution parameters were estimated using the methods of L-moments, TL-moments and maximum likelihood simultaneously, their accuracy having been subsequently compared.

L-moments are an alternative system describing the shape of the probability distribution. They have certain theoretical advantages over conventional moments resting in the ability to characterize a wider range of distribution. They are more resistant to outliers compared with conventional moments and less prone to estimation bias, the approximation by asymptotic normal distribution being more accurate in finite samples. L-moments are analogous to conventional moments. They can be estimated based on linear combinations of sample order statistics, i.e. L-statistics. L-moments and their estimations, however, lack some robust features that belong to TL-moments, the latter (the trimmed L-moments) representing an alternative robust version of the former.

Different approaches to the parameter estimation and various types of estimators have been currently under discussion in some areas of statistics and applied mathematics, see, e.g. [9], the application of the L-moments method of the parameter estimation being studied, e.g. in [1], [3] and [5], the application of the TL-moments method of the parameter estimation being studied, e.g. in [2]. The wages and incomes and especially the development and modeling of wage and income distributions have been also dealt with extensively in the statistical literature; see, e.g. [6]–[8]. Three-parametric lognormal curves represent the basic theoretical distribution, see [4].

All calculations were performed using Statgraphics and SAS statistical software packages, the Microsoft Excel spreadsheet and mathematical software R.

[1] University of Economics, Prague; Faculty of Informatics and Statistics; Department of Statistics and Probability; Sq. W. Churchill 1938/4; 130 67 Prague 3; Czech Republic; e-mail: bilkova@vse.cz.

2 Theory and methods

2.1 L-moments of probability distribution

Let X be a continuous random variable that has a distribution with the distribution function $F(x)$ and quantile function $x(F)$. Let $X_{1:n} \le X_{2:n} \le ... \le X_{n:n}$ be the order statistics of a random sample of the sample size n, coming from the distribution of the random variable X. L-moment of the r-th order of the random variable X is defined as

$$\lambda_r = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot E(X_{r-j:r}), \quad r = 1, 2, \ldots . \tag{1}$$

The expected value of the r-th order statistic of a random sample of size n has the form

$$E(X_{r:n}) = \frac{n!}{(r-1)! \cdot (n-r)!} \cdot \int_0^1 x(F) \cdot [F(x)]^{r-1} \cdot [1-F(x)]^{n-r} \mathrm{d}\, F(x). \tag{2}$$

The letter "L" in "L-moments" indicates that the r-th L-moment λ_r is a linear function of the expected value of a certain linear combination of the order statistics. The actual estimation of the r-th L-moment λ_r based on the obtained data sample is then a linear combination of order data values, i.e. L-statistics. The first four L-moments of the probability distribution are now defined as

$$\lambda_1 = E(X_{1:1}) = \int_0^1 x(F)\, \mathrm{d}\, F(x), \tag{3}$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 x(F) \cdot [2\, F(x) - 1]\, \mathrm{d}\, F(x), \tag{4}$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F) \cdot \{6[F(x)]^2 - 6\, F(x) + 1\}\, \mathrm{d}\, F(x), \tag{5}$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 x(F) \cdot \{20[F(x)]^3 - 30[F(x)]^2 + 12[F(x)] - 1\}\, \mathrm{d}\, F(x). \tag{6}$$

The probability distribution can be specified by its L-moments, even if some of its conventional moments do not exist, the opposite, however, not being true. It can be proved that the first L-moment λ_1 is a characteristic of the location and the second L-moment λ_2 is that of variability. It is often desirable to standardize higher L-moments λ_r, $r \ge 3$, so that they can be independent on specific units of the random variable X. The ratio of L-moments of the r-th order of the random variable X is defined as

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \ldots . \tag{7}$$

We can also define a function of L-moments which is analogous to the classical coefficient of variation, i.e. the so called L-coefficient of variation

$$\tau = \frac{\lambda_2}{\lambda_1}. \tag{8}$$

The ratio of L-moments τ_3 is the skewness characteristic, the ratio of L-moments τ_4 being the kurtosis characteristic of the respective probability distribution. The main probability distribution properties are very well summarized by the following four characteristics: L-location λ_1, L-variability λ_2, L-skewness τ_3 and L-kurtosis τ_4. L-moments λ_1 and λ_2, L-coefficient of variation τ and ratios of L-moments τ_3 and τ_4 are the most useful characteristics allowing us to summarize the probability distribution. Their main properties are existence (if the expected value of the distribution exists, then all its L-moments exist) and uniqueness (if the expected value of the distribution exists, then L-moments define the only one distribution, i.e. no two distributions have the same L-moments).

2.2 Sample L-moments

L-moments are usually estimated from a random sample drawn from the unknown distribution. Since the r-th L-moment λ_r is a function of order statistics expected values of the r-sized random sample, it is naturally estimated using the so-called U-statistic, i.e. the corresponding function of the sample order statistics (averaged over all subsets of the sample size r that may be formed from the obtained random sample of size n).

Let $x_1, x_2, \ldots, x_n$ be a sample and $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$ an order sample. Then the r-th sample L-moment can be written as

$$l_r = \binom{n}{r}^{-1} \sum_{1 \le i1 < i2 < \ldots < ir \le n} \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot x_{i\,r-j:n}, \quad r = 1, 2, \ldots, n. \tag{9}$$

Hence the first four sample L-moments have the form

$$l_1 = \frac{1}{n} \cdot \sum_i x_i, \tag{10}$$

$$l_2 = \frac{1}{2} \cdot \binom{n}{2}^{-1} \cdot \sum\sum_{i>j} (x_{i:n} - x_{j:n}), \tag{11}$$

$$l_3 = \frac{1}{3} \cdot \binom{n}{3}^{-1} \cdot \sum\sum\sum_{i>j>k} (x_{i:n} - 2\,x_{j:n} + x_{k:n}), \tag{12}$$

$$l_4 = \frac{1}{4} \cdot \binom{n}{4}^{-1} \cdot \sum\sum\sum\sum_{i>j>k>l} (x_{i:n} - 3\,x_{j:n} + 3\,x_{k:n} - x_{l:n}). \tag{13}$$

The method of L-moments allows us to obtain parameter estimations by setting the first p sample L-moments (p = 3 in this case) equal to the corresponding L-moments of the basic probability distribution. Using the following equations, we get parameter estimates by the method of L-moments for the case of the three-parameter lognormal distribution ("L" meaning L-moments estimation method)

$$z = \sqrt{\frac{8}{3}} \cdot \Phi^{-1}\left(\frac{1+t_3}{2}\right), \tag{14}$$

$$\sigma^{L} \approx 0{,}999\,281\,z - 0{,}006\,118\,z^3 + 0{,}000\,127\,z^5, \tag{15}$$

$$\mu^{L} = \ln\left[\frac{l_2}{\operatorname{erf}\left(\frac{\sigma^{L}}{2}\right)}\right] - \frac{\sigma^{2^{L}}}{2}, \tag{16}$$

$$\theta^{L} = l_1 - \exp\left(\mu^{L} + \frac{\sigma^{2^{L}}}{2}\right). \tag{17}$$

2.3 TL-moments of probability distribution

An alternative robust version of L-moments will be introduced now. This modification of L-moments is called the "trimmed L-moments" and is noted TL-moments. In this modification of L-moments, the expected values of order statistics of a random sample (in L-moments definition of probability distributions) are replaced by the expected values of order statistics of a larger random sample, the sample size growing in such a way that it corresponds to the total size of the adjustment, as shown below.

TL-moments have certain advantages over conventional L-moments and central moments. TL-moment of probability distribution may exist even if the corresponding L-moment or central moment of this probability distribution does not exist, as it is the case of the Cauchy distribution. Sample TL-moments are more resistant to outliers in the data. The method of TL-moments is not intended to replace the existing robust methods, but rather as their supplement, particularly in situations with outliers in the data.

In this robust modification of L-moments, the expected value $E(X_{r-j:r})$ is replaced by that of $E(X_{r+t1-j:r+t1+t2})$. For each r, we increase the size of a random sample from the original r to $r + t_1 + t_2$, working only with the expected values of these r modified order statistics $X_{t1+1:r+t1+t2}, X_{t1+2:r+t1+t2}, \ldots, X_{t1+r:r+t1+t2}$ by trimming t_1 and t_2 (the lowest and highest value, respectively, from a conceptual sample). This modification is called the r-th trimmed L-moment (TL-moment) and marked $\lambda_r^{(t1,t2)}$. Thus, TL-moment of the r-th order of a random variable X is defined as

$$\lambda_r^{(t_1,t_2)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot E(X_{r+t_1-j:r+t_1+t_2}), \quad r = 1, 2, ... \tag{18}$$

It is evident from the expressions (18) and (1) that TL-moments are reduced to L-moments when $t_1 = t_2 = 0$. Although we can also consider applications where the adjustment values are not equal, i.e. $t_1 \neq t_2$, we focus only on the symmetry of $t_1 = t_2 = t$. Then the expression (18) can be rewritten

$$\lambda_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot E(X_{r+t-j:r+2t}), \quad r = 1, 2, \ldots. \tag{19}$$

For $t = 1$, the first four TL-moments have the form

$$\lambda_1^{(1)} = E(X_{2:3}), \tag{20}$$

$$\lambda_2^{(1)} = \frac{1}{2} E(X_{3:4} - X_{2:4}), \tag{21}$$

$$\lambda_3^{(1)} = \frac{1}{3} E(X_{4:5} - 2X_{3:5} + X_{2:5}), \tag{22}$$

$$\lambda_4^{(1)} = \frac{1}{4} E(X_{5:6} - 3X_{4:6} + 3X_{3:6} - X_{2:6}). \tag{23}$$

Measurements of location, variability, skewness and kurtosis of a probability distribution analogous to conventional L-moments (3)–(6) are based on $\lambda_1^{(1)}$, $\lambda_2^{(1)}$, $\lambda_3^{(1)}$ a $\lambda_4^{(1)}$.

The expected value $E(X_{r:n})$ can be written using the formula (2). Applying the equation (2), we can re-express the right side of the equation (19)

$$\lambda_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \frac{(r+2t)!}{(r+t-j-1)! \cdot (t+j)!} \cdot \int_0^1 x(F) \cdot [F(x)]^{r+t-j-1} \cdot [1-F(x)]^{t+j} \, \mathrm{d}F(x), \; r = 1, 2, \ldots. \tag{24}$$

It is necessary to bear in mind that $\lambda_r^{(0)} = \lambda_r$ normally represents the r-th L-moment with no adjustment.

The expressions (20)–(23) for the first four TL-moments ($t = 1$) may be written in an alternative manner

$$\lambda_1^{(1)} = 6 \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \, \mathrm{d}F(x), \tag{25}$$

$$\lambda_2^{(1)} = 6 \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot [2F(x) - 1] \, \mathrm{d}F(x), \tag{26}$$

$$\lambda_3^{(1)} = \frac{20}{3} \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot \{5[F(x)]^2 - 5F(x) + 1\} \, \mathrm{d}F(x), \tag{27}$$

$$\lambda_4^{(1)} = \frac{15}{2} \cdot \int_0^1 x(F) \cdot [F(x)] \cdot [1 - F(x)] \cdot \{14[F(x)]^3 - 21[F(x)]^2 + 9[F(x)] - 1] \, \mathrm{d}F(x). \tag{28}$$

The distribution can be identified by its TL-moments, although some of its L-moments and conventional moments do not exit.

TL-skewness $\tau_3^{(t)}$ and TL-kurtosis $\tau_4^{(t)}$ can be defined analogously as L-skewness τ_3 and L-kurtosis τ_4

$$\tau_3^{(t)} = \frac{\lambda_3^{(t)}}{\lambda_2^{(t)}}, \tag{29}$$

$$\tau_4^{(t)} = \frac{\lambda_4^{(t)}}{\lambda_2^{(t)}}. \tag{30}$$

2.4 Sample TL-moments

Let $x_1, x_2, \ldots, x_n$ be the sample and $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$ an order sample. The expression

$$\hat{E}(X_{j+1:j+l+1}) = \frac{1}{\binom{n}{j+l+1}} \cdot \sum_{i=1}^{n} \binom{i-1}{j} \cdot \binom{n-i}{l} \cdot x_{i:n} \tag{31}$$

is considered to be an unbiased estimation of the expected value of the $(j + 1)$-th order statistic $X_{j+1:j+l+1}$ in the conceptual random sample of the sample size $(j + l + 1)$. Now we assume that in the definition of the TL-moment $\lambda_r^{(t)}$ in (19) the expression $E(X_{r+t-j:r+2t})$ is replaced by its unbiased estimation

$$\hat{E}(X_{r+t-j:r+2t}) = \frac{1}{\binom{n}{r+2t}} \cdot \sum_{i=1}^{n} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j} \cdot x_{i:n}, \tag{32}$$

which is obtained by substituting $j \to r + t - j - 1$ and $l \to t + j$ in (31). Now we get the r-th sample TL-moment

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \hat{E}(X_{r+t-j:r+2t}), \quad r = 1, 2, ..., n - 2t, \tag{33}$$

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \frac{1}{\binom{n}{r+2t}} \cdot \sum_{i=1}^{n} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j} \cdot x_{i:n}, \quad r = 1, 2, ..., n - 2t, \tag{34}$$

which is an unbiased estimation of the r-th TL-moment $\lambda_r^{(t)}$. Let us note that for each $j = 0, 1, \ldots, r - 1$, the values $x_{i:n}$ in (34) are not equal to zero only for $r + t - j \le i \le n - t - j$ in relation to combination numbers. Simple adjustment of the equation (34) provides an alternative linear form

$$l_r^{(t)} = \frac{1}{r} \cdot \sum_{i=r+t}^{n-t} \left[\frac{\sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \binom{i-1}{r+t-j-1} \cdot \binom{n-i}{t+j}}{\binom{n}{r+2t}} \right] \cdot x_{i:n}. \tag{35}$$

The above results can be used to estimate TL-skewness $\tau_3^{(t)}$ and TL-kurtosis $\tau_4^{(t)}$ by simple ratios

$$t_3^{(t)} = \frac{l_3^{(t)}}{l_2^{(t)}}, \tag{36}$$

$$t_4^{(t)} = \frac{l_4^{(t)}}{l_2^{(t)}}. \tag{37}$$

3 Results and discussion

Table 1 shows parameter estimations obtained using the three methods and the value of the *S* criterion (the sum of absolute deviations of the observed and theoretical frequencies for all intervals) for the total wage distribution in the Czech Republic, giving an approximate description of research outcomes for all 328 wage distributions. We found out that the method of TL-moments provided the most accurate results in almost all, with minor exceptions, wage distribution cases, the deviations having occurred mainly at both ends of the wage distribution due to extreme open intervals of an interval frequency distribution. Table 1 indicates that for the total wage distribution set for the whole Czech Republic in 2003–2010, the method of TL-moments always yields the most accurate output in terms of the *S* criterion. As for the research of all 328 wage distributions, the second most accurate results were produced by the method of L-moments, the deviations having occurred again at both ends of the distribution in particular. The latter method brought the second most accurate results in terms of all total wage distribution data sets over the period 2003–2010. In the majority of cases, the maximum likelihood method was the third most accurate approach. (For all cases, see Table 1.) Figure 1 represents the development of probability density function of three-parametric lognormal curves with parameters estimated using the method of TL-moments (related to the results in Table 1).

4 Conclusion

A relatively new class of moment characteristics of the probability distribution has been introduced in this paper. The probability distribution characteristics of the location (level), variability, skewness and kurtosis have been constructed using L-moments and their robust extension – TL-moments method, the former (as an alternative to classical moments of probability distributions) lacking some robust features that are typical for the latter.

The accuracy of TL-moments method was compared to that of L-moments and the maximum likelihood method. Higher accuracy of the former approach in comparison to that of the latter two methods has been proved by examining 328 wage distribution data sets. Advantages of L-moments over the maximum likelihood method have been demonstrated by the present study as well. The criterion of the sum of all absolute deviations of the observed and theoretical frequencies for all intervals for tackling wage distributions, has been employed.

Disadvantages of used methods are problems of computational complexity. Said methods of estimating parameters may be used in any field of economics, in which we model the empirical distribution using some theoretical continuous probability distribution, for example, when modelling amounts of damage in non-life insurance.

	Method of TL-moments			Method of L-moments			Maximum likelihood method		
Year	**μ**	**σ^2**	**θ**	**μ**	**σ^2**	**θ**	**μ**	**σ^2**	**θ**
2003	9.060	0.631	9,066	9.018	0.608	7,664	9.741	0.197	2.071
2004	9.215	0.581	8,552	9.241	0.508	6,541	9.780	0.232	0.222
2005	9.277	0.573	8,873	9.283	0.515	6,977	9.834	0.229	0.270
2006	9.314	0.578	9,383	9.284	0.543	7,868	9.891	0.211	0.591
2007	9.382	0.681	10,028	9.388	0.601	7,903	9.950	0.268	0.162
2008	9.439	0.689	10,898	9.423	0.624	8,755	10.017	0.264	0.190
2009	9.444	0.704	10,641	9.431	0.631	8,685	10.020	0.269	0.200
2010	9.482	0.681	10,617	9.453	0.621	8,746	10.034	0.270	0.201

Year	**Criterion *S***	**Criterion *S***	**Criterion *S***
2003	108,437.01	133,320.79	248,331.74
2004	146,509.34	248,438.78	281,541.41
2005	137,422.05	231,978.79	311,008.23
2006	149,144.68	216,373.24	325,055.67
2007	198,670.74	366,202.87	370,373.62
2008	206,698.93	357,668.48	391,346.02
2009	193,559.55	335,999.20	359,736.37
2010	210,434.01	235,483.68	389,551.44

Table 1 Parameter estimations obtained using the three methods of point parameter estimation and the value of *S* criterion for total wage distribution in the Czech Republic

Source: Own research

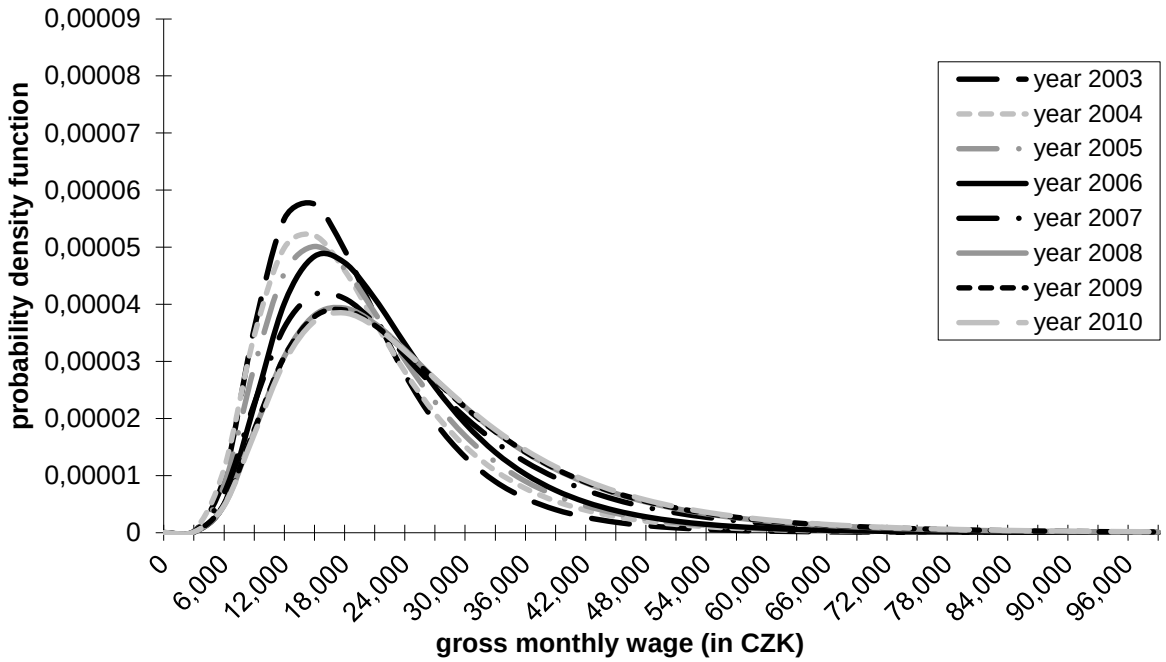


Figure 1 Development of probability density function of three-parametric lognormal curves with parameters estimated using the method of TL-moments

Source: Own research

Acknowledgements

This paper was processed with contribution of long term institutional support of research activities number IP400040 by Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic.

References

[1] Bílková, D.: Lognormal Distribution and Using L-Moment Method for Estimating Its Parameters. *International Journal of Mathematical Models and Methods in Applied Sciences [online]* **6** (2012), 30–44.

[2] Elamir, E. A. H., and Seheult, A. H.: Trimmed L-moments. *Computational Statistics & Data Analysis* **43** (2003), 299–314.

[3] Hosking, J. R. M.: L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society (Series B)* **52** (1990), 105–124.

[4] Johnson, N. L., Kotz, S., and Balakrishnan, N.: *Continuous Univariate Distributions*. John Wiley & Sons, New York, 1994.

[5] Kyselý, J., and Picek, J.: Regional Growth Curves and Improved Design Value Estimates of Extreme Precipitation Events in the Czech Republic. *Climate Research* **33** (2007), 243−255.

[6] Mallick, S. K.: Income Distribution and Consumption Deprivation: An Analytical Link. *Journal of Income Distribution* **17** (2008), 25–36.

[7] Marek, L., and Vrabec, M.: Probability Models for Wage Distributions. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics 2013*. The College of Polytechnics Jihlava, Jihlava, 2013, 575–581.

[8] Monti, M., and Santoro, A.: A Note on Between-Group Inequality with an Application to Households. *Journal of Income Distribution* **18** (2009), 49–62.

[9] Šimpach, O.: Faster Convergence for Estimates of Parameters of Gompertz-Makeham Function Using Available Methods in Solver MS Excel 2010. In: *Proceedings of the 30th International Conference on Mathematical Methods in Economics 2012*. Silesian University in Opava, School of Business Administration in Karviná, Karviná, 2012, 870–874.

Impact of German Economic Growth on Czech Tourism Demand

Sára Bíza Bisová[1]

Abstract. In the Czech tourism analyses statistics summaries and graphical representations are frequently used. Nevertheless, more complex and sophisticated methods such as multivariate macroeconometric models are poorly applied in this area in the Czech Republic. Most macroeconometric analyzes usually concentrates on stabilization policies, or a special part of the macro-system, using the main macroeconomic indicators.
Our aim is to analyze the Czech incoming tourism and the influence of the German economic situation on the Czech inbound tourism demand using VAR models approach. In this context, Granger causality testing in bivariate VARs and impulse response functions analysis in more-dimensional VARs are included. Three representations of Czech tourism demand are defined where the emphasis is placed on incoming tourism to the Czech Republic. The outputs indicate relevance of all the three definitions of Czech tourism demand and significant relation between Czech prices, German product and Czech tourism demand. According to the results, the economic situation in Germany can be considered to be an indicator of incoming tourism in the Czech Republic and the prices influence the tourism in general.

Keywords: Granger causality, impulse response functions, macroeconomic analysis, tourism demand, vector autoregression, Granger causality.

JEL Classification: C32, C51, E10, F59
AMS Classification: 62M10, 91B84

1 Introduction

Tourism is very important source of income and employment in many countries. Therefore, tourism data collection and analyses become more relevant. Also in the Czech Republic income from tourism represent important part of output and is very important for employment. The Czech Republic is among the major European tourist destinations. In the Czech Republic tourism comes under the Ministry of regional development which is the coordinating and methodical authority for tourism. CzechTourism is organization of the Ministry engaged in the development and progress of tourism in the Czech Republic and closely cooperating with the Ministry. A fundamental objective of CzechTourism is to promote the Czech Republic as a tourist destination abroad and also in the Czech Republic. CzechTourism activities include marketing, education, consulting, international cooperation etc. For tourism data collection there is also other important institution - Czech Statistical Office. In recent years, Ministry and CzechTourism try to implement projects to expand and improve the statistical databases and analyses of tourism. Unfortunately, currently are not available rich historical data from tourism and therefore statistical and econometric analyses are limited. There are only small set of variables available and many time series are not as long as desirable. Most of the analyses concentrate on the progress of chosen indicators or on comparison of selected territorial units. They are usually based on interpreting of descriptive statistics while econometric analyzes are very uncommonly used in this area.

We can distinguish three basic forms of tourism as defined in EU Manual [9] – domestic, inbound and outbound tourism. Domestic tourism contains activities of residents within a country of reference as a part of a domestic or an outbound trip. Inbound tourism includes activities of non-residents within the country of reference on an inbound trip. Outbound tourism is then represented as activities of residents outside the country of reference as a part of an outbound or a domestic trip. Domestic and inbound tourism together form internal tourism. This study concentrates on the Czech internal tourism with a focus on the Czech incoming tourism.

In recent studies many indicators measuring tourism demand are defined and applied - tourist arrivals (departures), tourist expenditures, nights spent in collective accommodation establishments etc. Next to the indicators mentioned above we use also occupancy rate of hotels and similar accommodations. Song, Li in [7] specify a model for tourism demand and construct ex post forecast for evaluation the quality of the model. They used

[1] University of Economics Prague, Dept. Of Econometrics, W. Churchill Sq. 4, Prague 3, xbiss00@vse.cz.

real tourist expenditure and arrivals in their aggregate and per capita form in Hong Kong by visitors from Australia, the UK and the USA (key source markets) as demand for tourism. Next to it they put other macroeconomic indicators in their macroeconometric model of demand for tourism – e.g. real GDP, CPI and exchange rate. They defined ADLM (autoregressive distributed lag model) with mentioned versions of demand for tourism for each of three incoming countries and their findings suggest, that using tourist arrivals as demand for tourism is more influenced by origin country income and habit effects, while tourist expenditure is connected with destination prices relative to those in the origin country. They also state, that in their study, model with aggregate expenditure was predicted most precisely followed by total arrivals, per capita arrivals and per capita expenditure was the poorest one. Wong et al. in [8] used more methodologies for modelling and forecasting tourism demand for Hong Kong with aim to examine the efficiency of combining forecasts using tourist arrivals from key source markets (ten countries) – ARIMA, ADLM, ECM and VAR. They conclude that although the combined forecasts do not always outperform the single model forecasts, combined forecasts can reduce risk of forecasting failure and should be preferred to single model forecasts in many practical situations. Biagi, Pulina in [1], [2] used VAR models and Granger causality in empirical analysis of Sardinia. Authors defined several models and conclude, that tourism supply is demand-driven, tourism demand is quality-driven and there is simultaneous Granger causality between tourism demand and capacity (number of beds). They define a proxy variable for quality of existing accommodation by the number of accommodation and number of beds within 3-5 stars category.

In this paper we employ indicators defining the demand for tourism in multidimensional econometric analysis using VAR models and impulse response functions. We estimate three VAR models and construct impulse response functions (IRFs) using the recursive identification technique. We also use testing for causality based on prediction principle presented by Granger. The concepts of VAR models and impulse response functions are frequently used for the econometric modeling of the macroeconomic stabilization policies, but using tourism indicators as occupancy rate of bed places and bedrooms in hotels and similar accommodation, arrivals of non-residents and nights spent at tourist accommodation establishments are less frequented. Our goal is to check the relevancy of the three definitions of tourism demand and to find out the influence of change in Czech prices and German economic growth (represented by German product) on the tourism demand in Czech Republic under the assumption that Germany is the main incoming country to Czech Republic - Germany is strongly economically connected with the Czech Republic through import and export and also through the tourism.

The paper is organized as follows: section 2 states theoretical background of applied econometric approach, data are presented in section 3 and empirical analysis in section 4. Conclusions are summarized in section 5.

2 VAR models

The structural form of VAR model (SVAR) without intercepts can be defined as follows [4], [5]

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{B}\mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t$ is a structural disturbance vector generated by vector white noise process, with identity covariance matrix $E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}) = \mathbf{\Lambda}$, where diagonal elements are variances of structural shocks, $\mathbf{\Pi}(L)$ is a polynomial matrix in the lag operator and $\mathbf{y}_t$ is a vector of *m* endogenous variables. The elements outside the diagonal in matrix $\mathbf{B}$ may be non-zero, therefore, some of the shocks can influence more endogenous variables of the system [4].

We can estimate the reduced form

$$\mathbf{y}_t = \mathbf{A}^{-1}\mathbf{\Pi}(L)\mathbf{y}_{t-1} + \mathbf{v}_t, \tag{2}$$

where $\mathbf{v}_t$ is a disturbance vector of reduced form with a covariance matrix $E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{\Sigma}$. For the disturbance vectors $\mathbf{u}_t$ and $\mathbf{v}_t$ holds the following formulae [4]

$$\mathbf{v}_t = \mathbf{A}^{-1}\mathbf{B}\mathbf{u}_t \quad \text{or} \quad \mathbf{A}\mathbf{v}_t = \mathbf{B}\mathbf{u}_t \tag{3}$$

and

$$E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}}) = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}E(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}})\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1} \quad \text{or} \quad \mathbf{\Sigma} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}. \tag{4}$$

The matrix $\mathbf{A}$ identifies the relation between the structural disturbances $\mathbf{u}_t$ and the reduced form shocks $\mathbf{v}_t$. We employ the **recursive identification** scheme, so-called **Choleski decomposition**, to get a just identified VAR, where the matrix $\mathbf{A}$ is lower triangular and the matrix $\mathbf{B}$ is diagonal, see [6].

The VAR models are introduced for the macroeconomic analyses in [6]. The estimation of VAR models is usually followed by construction of IRFs. IRFs are substantial in the macroeconomic policy analyses. For deriving and interpretation of IRF see [5].

2.1 Granger causality

Within the VAR models Granger causality (GC) presented in [6] for testing causality in econometric models is usually applied. The outputs of Granger concept of causality for stationary time series cannot be interpreted as causality in terms of cause and effect. The principle of GC is based on predictive properties of an analyzed set of variables. Concretely, if the past values of one variable (or the subset of variables) can help in predicting the remaining variable. In two-variable (X, Y) system we can simply test, if X can be better predicted using past values of Y compared to using only historical data of X. In this most simple situation the test is a partial F-test on regression parameters of lagged Y in the regression for X. We can write

$$E(X \mid X_{t-1}, X_{t-2}, ..., X_{t-k}, Y_{t-1}, Y_{t-2}, ..., Y_{t-k}) \neq E(X \mid X_{t-1}, X_{t-2}, ..., X_{t-k}), \tag{5}$$

where maximal lag k is selected on different basis, for example using AIC criterion.

In case of simultaneous GC we talk about the feedback system. In more dimensional VAR is possible to test GC of a subset of variables on a last one of the system. More about GC test in [3], [4].

3 Data

For the analysis five Czech and German macroeconomic indicators were applied, as we assume, that incoming tourists traveling from Germany to Czech Republic represent about 20% of arrivals of non-residents to the Czech Republic when working with number of guests in collective accommodation establishments. German data are real output (*GDPr_de*, real gross domestic product in constant prices (2005=100)) measured in million units of national currency[2]. Czech data contain consumer price index (*CPI_cz*), with basis in 2005 (2005=100) [3] and three definitions of tourism demand - net occupancy rate of bed-places and bedrooms in hotels and similar accommodation in Czech Republic (*OCCUPANCY_cz*) [4], total nights spent by non-residents at tourist accommodation establishments (*NIGHTS_cz*) [5] and arrivals of non-residents at tourist accommodation (*ARRIVALS_cz*)[6]. All the time series are measured at quarterly frequencies from Q1 2002 to Q4 2014 (52 observations) because earlier data of Czech indicators for tourism demand – namely *OCCUPANCY _cz*, *ARRIVALS_cz* and *NIGHTS_cz* - are not available. Data series were obtained from Eurostat, Czech Statistical Office and Deutsche Bundesbank.

For the purpose of causality testing, time series were not seasonally adjusted for the GC analysis, because seasonality is crucial information for causality testing in this case. For IRF analysis of more-dimensional models seasonal dummies for tourism demand variables and X12 ARIMA method for German output were used to reach better interpretable IRFs that are quicker returning to equilibrium. For the estimation EViews 7 and GRETL 1.9.5 software were used.

Table 1 shows the ADF tests of all the above mentioned variables in levels. All the variables, except for occupancy rate[7], are non-stationary at 5 % significance level and therefore should be transformed to reach the stationarity for the application in VAR. The occupancy rate was also transformed due to the interpretation. In order to eliminate the non-stationarity, differences between logarithms were computed (for their useful interpretation as growth rates)

$$\Delta \ln(y_t) = \ln y_t - \ln y_{t-1} = \ln \frac{y_t}{y_{t-1}}, \tag{6}$$

where y_t represents each of the analysed variables in period t. All the time series are stationary in first differences – see table 1. Variable *OCCUPANCY_cz* was only differentiated, because is represented in percentage terms and therefore does not need to be logarithmized.

[2] Deutsche Bundesbank – www.bundesbank.de.
[3] Czech statistical office - www.czso.cz.
[4] Eurostat - http://ec.europa.eu/eurostat.
[5] Eurostat - http://ec.europa.eu/eurostat.
[6] Eurostat - http://ec.europa.eu/eurostat.
[7] In Figure 2 we can see, that occupancy rate does not seem clearly stationary during the monitored time period due to higher levels of the variable in 2002. In longer view the occupancy rate probably become a non-stationary time series.

LEVEL	t-statistic	Prob.*	D-LOG	t-statistic	Prob.*
GDPr_de	-1.055576	0.7264	***GDPr_de***	-2.459771	0.0148
CPI_cz	-0.440897	0.8939	***CPI_cz***	-4.400840	0.0009
ER	-2.001260	0.2855	***ER***	-5.039293	0.0001
OCCUPANCY_cz	-3.746667	0.0063	***OCCUPANCY_cz***	-5.398486	0.0000
ARRIVALS_cz	-1.836504	0.3588	***ARRIVALS_cz***	-3.565863	0.0103
NIGHTS_cz	-2.515502	0.1183	***NIGHTS_cz***	-3.300190	0.0209

*MacKinnon (1996) one-sided p-values

Notes: Test critical values -3.58 (1% level), -2.93 (5% level), -2.60 (10% level), IR not in logs

Table 1 ADF tests for variables in levels

Figure 1 shows graphs of time series described above in levels. We can clearly see the non-stationarity of all variables in their level form.

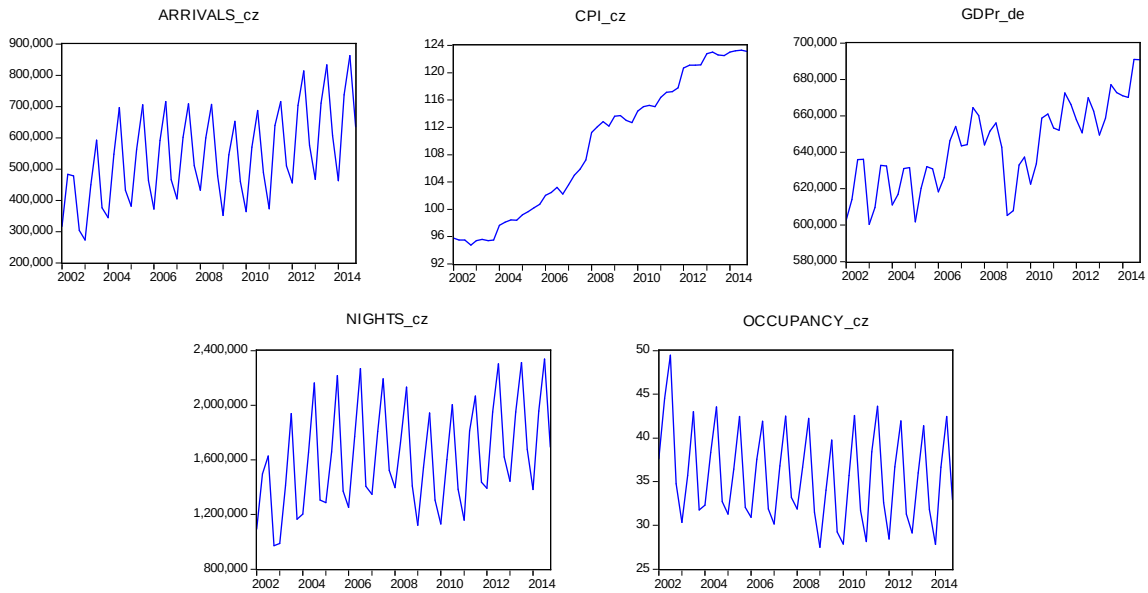


Figure 1 - Time series plots

4 Application

The application part includes causality testing, specification of more-dimensional VAR model and chosen impulse response functions. For the analysis were used abbreviations of variables and data transformations according to the description in section 3 (Data).

4.1 Granger causality testing

The Granger causality was analyzed between German output and each of three variables representing tourism demand in Czech Republic. The same was performed also for Czech price level and tourism demand. The results of all the GC tests in bivariate VAR models are available in Table 2. The first and fourth columns contain tested pair of variables in the given order, the second and fifth columns inform about the number of included lags in the appropriate bivariate VAR and the third and sixth columns include the p-value for the test.

There is evident relationship between German product and tourism demand in the Czech Republic. All the pairs of variables have uni- or bi-directional GC at 5 % significance level. The same applies to Czech CPI. In all the three cases of tourism demand definition there is relationship running from GDP to tourism demand. In case of *ARRIVALS_cz* the results show even simultaneous causality. It could be interpreted to mean that more arrivals to Czech Republic entail less traveling of German tourists within Germany and concurrently less foreign tourists coming to Germany because they travel to Czech Republic instead of Germany (therefore decrease in GDP)[8].

[8] Under the assumption of ceteris paribus, there could be also general increase in tourism in EU

But we must keep in mind, that GC test interpretation is based on principle of predictions and not in the economic sense of cause and effect, although we would like it so.

The case of Czech price level is similar to German GDP, there are also significant GC tests in all the three pairs. There is relationship in sense of GC running from Czech CPI to tourism demand. In case of occupational rate there is even bi-directional GC – one possible interpretation can be increasing prices in hospitality management projected into CPI as a result of higher demand for accommodation capacity and for the related services eventually.

GC test	df*	prob.	GC-test	df*	prob.
GDPr_de › ARRIVALS_cz	7	0.0448	***ARRIVALS_cz › GDPr_de***	7	0.0128
GDPr_de › NIGHTS_cz	7	0.0429	***NIGHTS_cz › GDPr_de***	7	0.1009
GDPr_de › OCCUPATION_cz	7	0.0010	***OCCUPATION_cz › GDPr_de***	7	0.2459
CPI_cz › ARRIVALS_cz	4	0.0057	***ARRIVALS_cz › CPI_cz***	4	0.9758
CPI_cz › NIGHTS_cz	4	0.0063	***NIGHTS_cz › CPI_cz***	4	0.9535
CPI_cz › OCCUPATION_cz	7	0.0000	***OCCUPATION_cz › CPI_cz***	7	0.0028

*Selection according to AIC criteria.

Table 2 – Granger causality test

4.2 The three-variable VAR

The Granger causality does not bring any information about the sign of relation between two variables. IRFs inform us about the response of all variables of the VAR system to a unit exogenous shock in each endogenous variable. These responses include the signs of the effects and therefore can be also used for causality analysis. Three versions of three-variable VAR models were estimated and impulse response functions were constructed. Figure 2 shows IRFs representing responses of the tourism demand variables to a unit exogenous shock in *CPI_cz* and *GDPr_de_sa*.

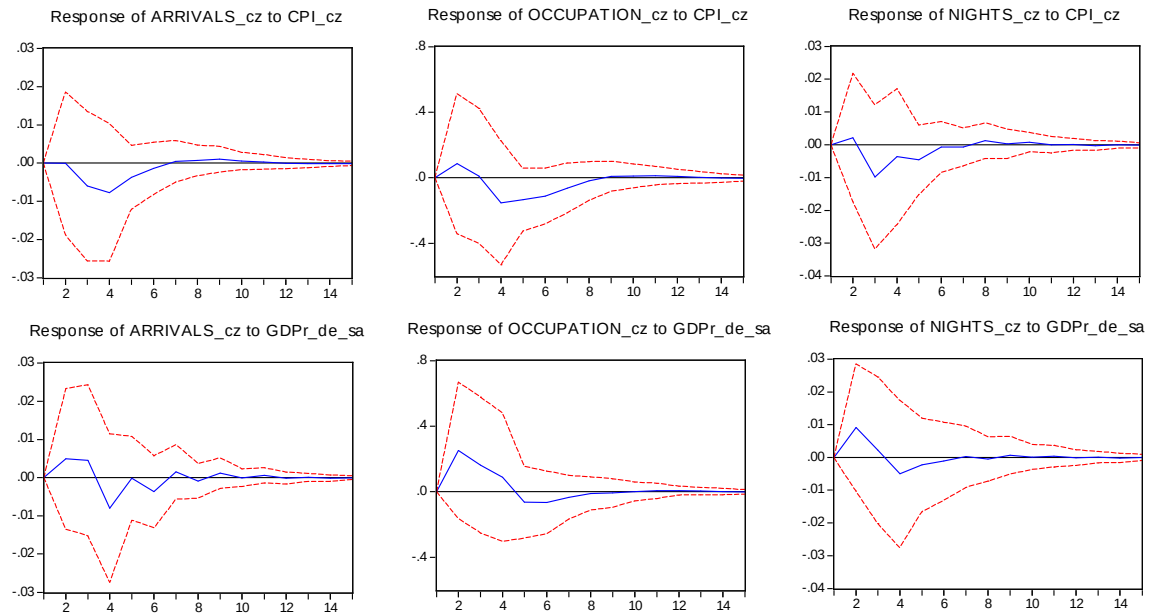


Figure 2 – Impulse response functions (response to Cholesky one S.D. innovations ± 2 S.E.)

In all the VAR models Czech consumer price index and German gross domestic product appear. The third variable is one of the three definitions of tourism demand in the Czech Republic, (therefore three versions of VAR). For the purpose of better interpretation of IRFs, variable *GDPr_de* was seasonally adjusted by the X12 ARIMA method (we use the abbreviation *GDPr_de_sa*). Seasonality in variables representing Czech tourism demand appear to be deterministic and thus we put seasonal dummies into the three-variable VARs. IRFs using seasonally adjusted variable *GDPr_de_sa* provide similar outputs (the same signs) with smoother responses to German output in comparison to models using *GDPr_de* without X12 ARIMA transformation. All the VAR

models for IRF analysis were estimated with 3 lags although more than 3 lags were recommended by AIC criterion. IRFs for three-variable VAR(4) and higher lag length VAR models contained worse interpretable outputs with responses changing sign in some cases.

The first row of the matrix of IRFs includes the responses of tourism demand variables to an unexpected unit exogenous shock in Czech prices. In the second row appear reactions of tourism variables to an unexpected unit exogenous shock in German output. It must be noted that the transformation of all the variables change their interpretation into the growth rates. The results of the three models suggest similar IRFs. We can see similar responses of all the three tourism demand variables to an exogenous shock in *CPI_cz* and also to a shock *GDPr_de_sa*. The unit positive shock in *CPI_cz* is followed by decrease in Czech tourism demand in all the three versions with a slight persistence at the beginning of the simulated horizon – the fall starts after about one or two quarters. One possible interpretation can lead to delayed reaction of foreign tourists to the change in Czech prices. On the other hand, the confidence intervals are relative wide and therefore bring some uncertainty to our interpretation. In the second row of Figure 2 demand variables increase immediately as a response to a positive shock in German GDP with a peak around two quarters after the shock. It suggests positive impact of German economic growth on the tourism demand in the Czech Republic.

5 Conclusion

For the analysis were used three possible definitions of Czech tourism demand with focus on incoming tourism, Czech price level and German economic growth derived from German output. Ganger causality testing in bivariate VAR models was applied and IRFs for the three three-variable VAR models were constructed. The outputs from Granger causality testing and IRF analysis suggest that an intervention or shock influencing German economic situation or Czech price level can significantly affect the demand for tourism in the Czech Republic. Also the impulse response function analysis corresponds with our assumptions about the modelling system. Tourism demand decrease in response to increase in prices and increase in reaction to positive impulse in German economic growth. Our analysis suggests that we can deduce the changes in the development of Czech tourism demand through the economic situation in Germany, the key source market for Czech tourism. The Czech tourism demand is also determined by the Czech prices influencing the decision making of potential visitors about their traveling. Our other conclusion is that all the three definitions of tourism demand are relevant and applicable for the Czech tourism demand analysis. More dimensional VAR models expanded to include exchange rate, other definitions of Czech tourism demand representing tourist expenditures and frequency domain causality tests are the subject of our upcoming analyses.

Acknowledgements

The research project was supported by Grant No. IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Biagi, B., Pulina, M.: Tourism, environmental quality and economic growth: empirical evidence and policy imlications. *CRENOS Working papers* **9** (2006).
[2] Biagi, B., Pulina, M.: Bivariate VAR models to test Granger causality between tourist demand and supply: Implications for regional sustainable growth. *Paper in Regional Science*, 2009, Vol. 88, No. 1, pp. 231-244.
[3] Granger, C.W.J.,: Investigating Causal Relations by Econometric Models and Cross-spetral Methods. *Econometrica*, 1969, Vol. 37, No. 3, pp. 424-438.
[4] Hušek, R.: *Ekonometrická analýza.* Oeconomica, Praha, 2007.
[5] Lütkepohl, H.: *New introduction to multiple time series analysis.* Springer, 2005.
[6] Methodological manual for tourism statistics - Version 3.1, *Eurostat Manuals and Guidelines*, European Union, 2014 edition, 2014. Available online [1.5.2015]: http://ec.europa.eu/eurostat/.
[7] Sims, Ch. A.: Macroeconomics and reality. *Econometrica*, 1980, Vol. 48, No. 1, pp. 1-48.
[8] Song, H., Li, G.: Tourism demand modelling and forecasting: how should demand be measured? *Tourism Economics*, 2010, Vol. 16, No. 1, pp. 63-81.
[9] Wong, K., Song, H., Witt, S. F. & Wu, D. C.: Tourism forecasting: to combine or not to combine? *Tourism Management* **28** (2007), 1068-1078.

Investment portfolio making under uncertainty by a two-stage decision making procedure

Adam Borovička[1]

Abstract. The main aim of this paper is to solve the problem of investment portfolio making under uncertainty. For this purpose, a two-stage decision making procedure is proposed. In the first phase, the approach provides aggregate valuation of investment alternatives which are divided into "bad" and good" ones. Then an investment portfolio is made from the selected alternatives during the second phase. The proposed procedure offers the complex view of the investment in contrast to the current approaches solving rather only the partial fragments of an investment process. The concept accepts the elements of uncertainty, as inaccurate input data, approximate preferences about a relevance of criteria or a final value of particular portfolio characteristics. These vague elements are expressed by the (triangular) fuzzy numbers. For the first phase, the fuzzy multi-criteria evaluation method is proposed. The algorithm is based on modifications of the current concepts ELECTRE I and III. In terms of the second phase, the fuzzy multiple objective programming method is introduced which is educed from the fuzzy method KSU-STEM. The whole proposed two-stage decision making procedure is applied to making portfolio of the open unit trusts offered by the Česká spořitelna investment company. The investment situation is introduced, the portfolio making process is described step by step and the results are analyzed.

Keywords: Fuzzy number, multi-criteria evaluation method, multiple objective programming method, portfolio making, two-stage decision making procedure, uncertainty, unit trust.

JEL Classification: C44
AMS Classification: 90B50, 90C70

1 Introduction

Many people decide where to invest their free financial resources. The important investment decision is usually the result of a complex decision making process. We have several approaches solving only the fragments of the entire investment decision making process, e. g. determination of investment policy, decision making criteria and their importance, choice of suitable investment alternatives, or portfolio making. So the intention of this paper is to propose such a decision making procedure containing all mentioned activities in order to solve the particular investment problem. The procedure is defined as two-stage using the principles and approaches of the decision making theory. At the beginning, a potential investor sets his/her investment policy. After this primary phase of the process, which is not included in this text in more detail, a selection of suitable investment instruments from the preselected set (defined on the basis of the investment policy) can be started in terms of the first stage. Then an investment portfolio of the selected instruments is made in the second stage. There can be many elements of uncertainty in terms of these two processes. It means that the input data can be in inaccurate form, the investor preferences about a relevance of criteria or a final value of particular portfolio characteristics can be expressed only approximately. These elements are specified via the triangular fuzzy numbers. Then to select the set of "good" investment alternatives, the fuzzy multi-criteria evaluation method is proposed. To make the final investment portfolio, the proposed fuzzy multiple objective programming method is applied. The algorithms of both concepts are affected by current approaches, but are adapted to the investment decision making. In the practical part, the portfolio of open unit trusts offered by the Česká spořitelna investment company is made. In order to fulfill this main aim of the paper, the proposed two-stage decision making procedure is applied. The investment situation is introduced, the investment process is described. The results are analyzed and commented.

After an introduction to the problems, both phases of the proposed two-stage decision making procedure are briefly described. Next two chapters are devoted to the proposed fuzzy multi-criteria evaluation method and fuzzy multiple objective programming method. The algorithms of these methods are shortly described with an emphasis on special improvements and modifications compared to the current concepts. Then the two-stage

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, adam.borovicka@vse.cz.

procedure is applied to making portfolio of the open unit trusts. In summary, some ideas for future research are specified.

2 Two-stage decision making procedure

The two-stage decision making procedure is proposed in order to comprise the whole investment decision making process. At the beginning of the first stage, the potential investor usually forms his/her investment policy with some investment counsel. It means to determine a purpose (goal) of the investment, to specify evaluative criteria (risk, return, cost, etc.) and their importance, to set passive or active work with portfolio, to determine investment horizon. Then a range of the potential investment instruments is specified. In the second part of the first phase, the *fuzzy multi-criteria evaluation method* is applied in order to reduce sometimes wide group of investment alternatives that are divided into "bad" and "good" ones. The first phase helps to orient in the particular investment situation, to analyze preselected investment alternatives by the choice criteria. The investor knows which alternatives are "bad" and which would be suitable for a final investment portfolio. A reduction of the extensive set of possible investment instruments can also serve to computational simplification in the second phase when the mathematical model is formulated and solved. In the second phase, the portfolio is made from the selected "good" investment instruments by the *fuzzy multiple objective programming method*. The investor can invest his/her free financial resources in the chosen investment instruments in the stated shares.

2.1 Fuzzy multi-criteria evaluation method

Many current multi-criteria evaluation methods are "fuzzified". They accept criteria values or weights in the approximate form. We know fuzzy WSA, AHP, fuzzy conjunctive and disjunctive method, fuzzy modifications of TOPSIS method, and method based on the (fuzzy) preference relation scoring AGREPREF, ELECTRE I, ELECTRE III, or PROMETHEE.

The proposed fuzzy multi-criteria evaluation method eliminates some drawbacks and unfitnesses regarding a practical application of a portfolio making of the current concepts [4]. The method accepts the input data in the strict and the approximate form which the most methods do not enable. Any additional information (threshold values or specification of utility function) from a decision maker is not required because it could be very problematic for a decision maker. The vague elements are expressed as triangular fuzzy numbers. The alternatives are scored by a fuzzy preference relation in order to divide the alternatives into "good" (effective) and "bad" (ineffective). The algorithm takes into account the differences in criteria values compared to other methods (e.g. ELECTRE III, conjunctive or disjunctive method). A comparison of the triangular fuzzy numbers is made by McCahone's approach [7]. The distance between fuzzy numbers is measured by the vertex method [5] that is modified. In the end, the proposed method is user-friendly, the algorithm is comprehensible.

The algorithm of the developed method can be briefly described in the following steps:

Step 1: The criteria matrix $\mathbf{Y} = (y_{ij})$ is specified, where y_{ij} $(i = 1,2,...,n;\ j = 1,2,...,k)$ is an evaluation of the *i*-th alternative by the *j*-th criterion. The criteria values can be in the strict form, in the vague form as triangular fuzzy numbers as well. An importance of the *j*-th criterion is specified in the strict form as weight v_j.

Step 2: The strict and fuzzy criteria values are compared. As in ELECTRE III [9], the following sets of criteria indices are specified, where the set $I^{\min}$, or $I^{\max}$ contains the indices of minimizing, or maximizing criteria

$$I_{iPj} = \left\{ r \wedge s \mid y_{ir} > y_{jr},\ y_{is} < y_{js};\ r \in I^{\max}, s \in I^{\min} \right\} \qquad i, j = 1,2,...,n;\ i \neq j$$
$$I_{jPi} = \left\{ r \wedge s \mid y_{jr} > y_{ir},\ y_{js} < y_{is};\ r \in I^{\max}, s \in I^{\min} \right\} \qquad i, j = 1,2,...,n;\ i \neq j \ .$$

Step 3: The matrices **S** and **R**, whose elements express the grades of preference, are determined. The matrix **S** is defined as in the ELECTRE III technique. A concept of the matrix **R** is derived from ELECTRE I method [8], however it takes into account the criteria importance and works with the standardized criteria values. The element of the matrix **R** is formulated for each couple of alternatives *i* and *j* $(i, j = 1,2,...,n)$ as follows

$$r_{ij} = \frac{\sum_{h \in I_{iPj}} \left(v_h \mid y'_{ih} - y'_{jh} \mid \right)}{\sum_{h=1}^{k} v_h \mid y'_{ih} - y'_{jh} \mid} \qquad I_{iPj} \neq \emptyset, \qquad r_{ij} = - \qquad i = j, \qquad r_{ij} = 0 \qquad \text{else},$$

where y'_{ih}, or. y'_{jh} $(i, j = 1,2,...,n;\ h = 1,2,...,k)$ is a standardized criteria value. The difference (distance) between standardized fuzzy criteria values is measured by the modified vertex method [4].

Step 4: The aggregate preference of the *i*-th alternative in face of the *j*-th alternative is set by the proposed rule

$$s_{ij} > s_{ji} \wedge r_{ij} > r_{ji} .$$

The rule is a modification of ELECTRE III technique, the thresholds are eliminated. The *effective alternative* is such an alternative that embodies the highest discrepancy between number of alternatives in face of which is preferred and number of alternatives that are preferred in face of it. The concept combines an approach ELECTRE I and III approaches to eliminate the ELECTRE I drawback, when the effective alternative does not have to exist.

2.2 Fuzzy multiple objective programming method

To make investment portfolio actively, I focused on the group of interactive multiple objective programming methods. The oldest interactive methods are GDF and STEM. Then the methods are "fuzzified", e. g. fuzzy STEM, which is called KSU-STEM. Other methods working with the different types of fuzzy numbers are proposed. Fuzzy goal technique with interactive procedure was also proposed. Furthermore, several concepts based on the α-cut principle are known.

The proposed fuzzy objective programming method also eliminates some drawbacks and unfitnesses regarding a practical application of a portfolio making of the current concepts [4]. As in the fuzzy multi-criteria evaluation method, the data can be expressed as triangular fuzzy numbers or real values. The hardly determinable additional information is not required from a decision maker (α-cut, goal values etc.). The interactive procedure is included in the algorithm, so the decision maker can actively partake of solution making. His/her preferences of objective function values can be expressed only approximately by means of the triangular fuzzy numbers.

The algorithm of the developed method can be briefly described in the following several steps:

Step 1: Define *k* objective functions. Denote the objective functions with strict coefficients as $f_i(\mathbf{x}) = \mathbf{c_i^T x}\ (i \in I_{min}^{striktní} \cup I_{\max}^{striktní})$, where I_{min}^{strict}, or I_{max}^{strict} is a set of indices of minimizing, or maximizing objective functions. $\mathbf{c_i^T} = (c_{i1}, c_{i2}, ..., c_{in})$ is a vector of strict coefficients and $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ denotes a vector of variables. The objective functions with fuzzy coefficients are denoted as $\tilde{f}_i(\mathbf{x}) = \tilde{\mathbf{c}}_\mathbf{i}^\mathbf{T}\mathbf{x}\ (i \in I_{min}^{fuzzy} \cup I_{max}^{fuzzy})$, where I_{min}^{fuzzy}, or I_{max}^{fuzzy} is a set of indices of minimizing, or maximizing fuzzy objective functions. $\tilde{\mathbf{c}}_\mathbf{i}^\mathbf{T} = (\tilde{c}_{i2}, \tilde{c}_{i2}, ..., \tilde{c}_{in})$ is a vector of fuzzy coefficients. The *i-th* fuzzy objective function is defined as following triangular fuzzy number

$$\tilde{f}_i(\mathbf{x}) = (\mathbf{c_i^{l^T}x}, \mathbf{c_i^{m^T}x}, \mathbf{c_i^{u^T}x}) = \left(f_i^{\,l}(\mathbf{x}),\ f_i^{\,m}(\mathbf{x}),\ f_i^{\,u}(\mathbf{x}) \right) \qquad i \in I_{min}^{fuzzy} \cup I_{\max}^{fuzzy},$$

where $\mathbf{c_i^{l^T}} = (c_{i1}^l, c_{i2}^l, ..., c_{i2}^l)$, $\mathbf{c_i^{u^T}} = (c_{i1}^u, c_{i2}^u, ..., c_{i2}^u)$ and $\mathbf{c_i^{m^T}} = (c_{i1}^m, c_{i2}^m, ..., c_{i2}^m)$ are vectors of lower, upper and middle parameters. The vector of weights $\mathbf{v} = (v_1, v_2, ..., v_k)^T$ is known.

Step 2: The ideal and basal values of the objective functions are computed. Similar to STEM method [2], the ideal value is stated by means of one objective model, where the extreme of particular objective function is searched on the set of the strict conditions *X*. The ideal value of the objective functions with the fuzzy coefficients is stated similarly; however three separate one objective models must be solved. The basal value of each objective function is calculated in connection with the ideal values of all objective functions. This principle is derived from the approach of STEM method.

Step 3: As in [1], so-called fuzzy goals are formulated. This approach takes into account an acquisition of the values of particular objective function. The minimizing, or maximizing objective function with strict coefficients $f_i(\mathbf{x})$ $(i \in I_{min}^{strict}$, or $i \in I_{max}^{strict})$ is converted to the fuzzy goal expressed by the right-side, or the left-side triangular fuzzy number $\tilde{F}_{f_i(\mathbf{x})}$ with the following membership function

$$\mu_{\tilde{F}_{f_i(\mathbf{x})}}(\mathbf{c_i^T x}) = \begin{cases} 1 & \mathbf{c_i^T x} \le f_i^{\,I} \\ \frac{f_i^{\,B} - \mathbf{c_i^T x}}{f_i^{\,B} - f_i^{\,I}} & f_i^{\,I} \le \mathbf{c_i^T x} \le f_i^{\,B} \\ 0 & \mathbf{c_i^T x} \ge f_i^{\,B} \end{cases}, \text{ or } \qquad \mu_{\tilde{F}_{f_i(\mathbf{x})}}(\mathbf{c_i^T x}) = \begin{cases} 1 & \mathbf{c_i^T x} \ge f_i^{\,I} \\ \frac{\mathbf{c_i^T x} - f_i^{\,B}}{f_i^{\,I} - f_i^{\,B}} & f_i^{\,B} \le \mathbf{c_i^T x} \le f_i^{\,I} \\ 0 & \mathbf{c_i^T x} \le f_i^{\,B} \end{cases},$$

where $f_i^{\,B}$, or $f_i^{\,I}$ is the basal, or the ideal value of the *i*-th objective function with the strict coefficients. The basal and ideal value of the objective functions with the fuzzy coefficients is in the vague form (as triangular fuzzy number), so the fuzzy goal is created via three triangular fuzzy numbers.

Step 4: The model of fuzzy linear programming is formulated. This model can be transformed to the following strict form via the concept of fuzzy goals (Bellman optimality principle)

$$
\begin{aligned}
& z = \alpha \to \max \\
& v_i \frac{f_i^B - \mathbf{c_i^T x}}{f_i^B - f_i^I} \geq \alpha \qquad i \in I_{min}^{strict}, \qquad v_i \frac{f_i^{lB} - \mathbf{c_i^{lT} x}}{f_i^{lB} - f_i^{ll}} \geq \alpha \qquad v_i \frac{f_i^{mB} - \mathbf{c_i^{mT} x}}{f_i^{mB} - f_i^{ml}} \geq \alpha \qquad v_i \frac{f_i^{uB} - \mathbf{c_i^{uT} x}}{f_i^{uB} - f_i^{ul}} \geq \alpha \qquad i \in I_{min}^{fuzzy} \\
& v_i \frac{\mathbf{c_i^T x} - f_i^B}{f_i^I - f_i^B} \geq \alpha \qquad i \in I_{max}^{strict}, \qquad v_i \frac{\mathbf{c_i^{lT} x} - f_i^{lB}}{f_i^{ll} - f_i^{lB}} \geq \alpha \qquad v_i \frac{\mathbf{c_i^{mT} x} - f_i^{mB}}{f_i^{ml} - f_i^{mB}} \geq \alpha \qquad v_i \frac{\mathbf{c_i^{uT} x} - f_i^{uB}}{f_i^{ul} - f_i^{uB}} \geq \alpha \qquad i \in I_{max}^{fuzzy}, \qquad (1) \\
& \mathbf{x} \in X \\
& 0 \leq \alpha \leq 1
\end{aligned}
$$

where f_i^{lB}, f_i^{mB}, f_i^{uB}, and f_i^{ll}, f_i^{ml}, f_i^{ul} are the basal and the ideal values of parameters of the triangular fuzzy numbers describing the *i-th* objective function with the fuzzy coefficients, α is the grade of membership of a solution which should be the greatest as possible.

Step 5: If a solution of the previous model (1) is not acceptable for a decision maker, the interactive procedure is started. A decision maker divides the objective functions into 3 groups – objective functions whose values are acceptable and do not have to get worse, objective functions whose values must be improved and objective functions whose values can be got worse. The preferences about a possible downgrade of the values can be expressed in the approximate form by a decision maker. The vague preferences are formulated by the triangular fuzzy numbers. Model (1) is modified, some conditions are added, some conditions are eliminated. The interactive procedure is inspired by the methods STEM and KSU-STEM [6]. But these concepts are improved by a possibility of the preference expressed in the vague form which is better applicable in a real decision making situations. The criteria relevance is also conceived differently that in the mentioned concepts. Thus, an importance of the objective functions, whose values can be got worse, are explicitly conserved. This process is repeated until the solution is acceptable by a decision maker.

3 Making portfolio of the open unit trusts

The real application describes a situation of investment decision making for a long period in order to save money for pension age. It results from my personal situation. Moreover, I do think that it is a typical situation for many younger people in a productive age. Besides "standard" products as retirement income insurance, life insurance or building savings, an investment counsel presents open unit trusts to the client. The potential investor will play active role in the process of portfolio making. He/She specifies the evaluative criteria in cooperation with an investment counsel. He/She sets their importance linguistically.

3.1 Criteria

As long-standing client of Česká spořitelna, I chose the open unit trusts managed and offered by the Česká spořitelna investment company. The offer consists of 9 *mixed* funds (*Akciový Mix FF*, *Dynamický Mix FF*, *Fond řízených výnosů OPF*, *Fond životního cyklu 2020 FF*, *Fond životního cyklu 2030 FF*, *Konzervativní Mix FF*, *Osobní portfolio 4*, *PLUS Otevřený podílový fond*, *Vyvážený Mix FF*), 5 *bond* funds (*High Yield dluhopisový OPF*, *Korporátní dluhopisový OPF*, *Sporobond OPF*, *Sporoinvest*, *Trendbond*) and 3 *stock* funds (*Global Stocks FF*, *Sporotrend OPF*, *Top Stocks*). These 17 investment alternatives are evaluated by the most important 3 criteria – *return*, *risk* and *cost*. Other possible criteria, e. g. portfolio diversification or market mood, are included in the model via special conditions. The return is measured monthly in a period from January 2010 to July 2014. This period can be considered as "more calm", then it could properly represent a development in the long-term investment horizon. The return is expressed by the triangular fuzzy number. Its middle parameter is determined as average monthly return. Lower, or upper parameter is computed as an average monthly return reduced, or topped by three standard deviation of monthly returns [4]. This operation is based on a presumption of the normally distributed returns. The triangular fuzzy number tries to cover return sufficiently in the vague form. The risk is measured by the average absolute negative deviation proposed in [4]. This concept takes into account only values less than average return. The costs include all fees connected with an investment to the open unit trusts, namely entry, supervisor, manager or license fees.

3.2 Two-stage investment portfolio making

At the beginning of the portfolio making process, a potential investor expresses an importance of the criteria linguistically in the scale of 5 terms – *very low*, *low*, *middle*, *high*, *very high* importance. The importance is: return – *high*, risk – *very high*, cost – *very low*. The investor is risk-averse which results from a purpose of the investment. The linguistically expressed criteria importance is transformed to the criteria weights via the fuzzy weights estimation method proposed in [3] and [4]. The weights are: return – 0,4, risk – 0,547, cost – 0,053.

In terms of the first phase, "the best" investment alternatives from each group of the open unit trusts are selected by the proposed fuzzy multi-criteria evaluation method. Only one investment alternative is evaluated as effective ("the best") in each group, namely mixed unit trust *Fond řízených výnosů OPF*, bond fund *Sporoinvest* and stock fund *Global Stocks FF*. Fond řízených výnosů OPF has not so high return compared to the most other mixed funds. But the differences between the return values of the mixed funds are not considerable. Moreover, the value of the most important criterion risk and also cost is significantly the lowest. The bond fund Sporoinvest is selected because its risk and cost is in the lowest level and its return is the second greatest. In the group of stock funds, Global Stocks FF has the lowest risk and Top Stocks has the greatest return. Although Global Stocks FF has greater cost than Top Stocks, the lowest level of the most important criterion makes it a winner.

In terms of the second phase, a portfolio of the selected unit trusts is made by the proposed fuzzy multiple objective programming method. Three objective functions are determined. The objective function expressing the portfolio return is specified as a weighted sum of returns of the selected unit trusts

$$\tilde{f}_1(x_1, x_2, x_3) = \sum_{i=1}^{3} \tilde{v}_i x_i = (\sum_{i=1}^{3} v_i^l x_i, \sum_{i=1}^{3} v_i^m x_i, \sum_{i=1}^{3} v_i^u x_i),$$

where $\tilde{v}_i$ $(i = 1,2,3)$ is the return of the *i-th* unit trusts specified as triangular fuzzy number in the form (v_i^l, v_i^m, v_i^u), x_i is a share of the *i-th* unit trust ($i = 1$ for Fond řízených výnosů OPF, $i = 2$ for Sporoinvest, $i = 3$ for Global Stocks FF). This objective function is maximizing. The objective function expressing the portfolio risk is specified as a weighted sum of risks of the selected unit trusts

$$f_2(x_1, x_2, x_3) = \sum_{i=1}^{3} r_i x_i,$$

where r_i $(i = 1,2,3)$ is the risk of the *i-th* unit trust. This objective function is minimizing. The objective function expressing the portfolio cost is formulated as a weighted sum of cost of the selected unit trusts

$$f_3(x_1, x_2, x_3) = \sum_{i=1}^{3} n_i x_i,$$

where n_i $(i = 1,2,3)$ is the cost of the *i-th* unit trust. This objective function is also minimizing. The set of strict conditions *X* is specified in the following form

$$X = \left\{ \sum_{i=1}^{3} x_i = 1; 0.2 \le x_i \le 0.5, i = 1,2,3 \right\}.$$

$\sum_{i=1}^{3} x_i = 1$ ensures making a portfolio. The conditions $0.2 \le x_i \le 0.5, i = 1,2,3$ represent a diversification tool which ensures that any unit trust will not create the most of portfolio, insignificant part as well. The basal and ideal values of three above-mentioned objective functions are calculated via finding their extremes on the set *X*. After formulation of the fuzzy goals, a strict mathematical model of linear programming can be formulated as follows

$$\begin{aligned}
&z = \alpha \to \max \\
&0.4\frac{\sum_{i=1}^{3} v_i^l x_i - 4.122}{2.547} \ge \alpha \qquad 0.4\frac{\sum_{i=1}^{3} v_i^m x_i - 9.102}{0.254} \ge \alpha \qquad 0.4\frac{\sum_{i=1}^{3} v_i^u x_i - 11.535}{3.051} \ge \alpha \\
&0.547\frac{1.21 - \sum_{i=1}^{3} r_i x_i}{0.609} \ge \alpha \qquad 0.053\frac{4.152 - \sum_{i=1}^{3} n_i x_i}{1.584} \ge \alpha \\
&0.2 \le x_i \le 0.5 \qquad i = 1,2,3 \\
&\sum_{i=1}^{3} x_i = 1 \\
&0 \le \alpha \le 1
\end{aligned} \quad , \tag{2}$$

A solution of this model can be obtained by the simplex method.

The primary investment portfolio is in the following table (Table 1).

Fund	Share	Criterion	Value
Fond řízených výnosů	26.24 %	Return	(6.3652, 91343, 11,9035) %
Sporoinvest	50 %	Risk	0.6754 %
Global Stocks FF	23.76 %	Cost	2.7086 %

Table 1 Primary investment portfolio

The unit trust Sporoinvest has the greatest possible share because it has significantly the lowest level of risk. The highest importance of risk criterion causes greater share of the fund Fond řízených výnosů with lower risk than the fund Global Stocks FF with greater return. The investor obtains the values of portfolio characteristics recounted to the annual base. The return is "defuzzified" by the average value from three parameters of the triangular fuzzy number. Then the risk is 3.05 %, return 8.11 % and cost 2.71 %. The investor would like to improve the value of risk at the expense of the values of return and cost that are very good for him/her. A possible downgrade of these values is expressed approximately via the triangular fuzzy numbers. The modified model (2) is solved. Because a value of risk is close to the ideal value, other small positive change of risk (hundredths percentage points) implicates a greater decrease return (tenths percentage points). On the basis of this analysis, the investor accepts a primary solution (portfolio in Table 1).

4 Conclusion

The paper deals with the problem of a portfolio making under uncertainty. To cover the entire investment process, the complex two-stage decision making procedure is proposed. In the theoretical part, the emphasis is put on the analytic apparatus used in the proposed procedure. The principle and algorithm of the proposed fuzzy multi-criteria evaluation method and fuzzy multiple objective programming method are briefly described. In the practical part, the portfolio of open unit trusts is made by the proposed procedure. Of course, the decision making procedure can be also used for making portfolio of another investment instruments or even for other decision making situation. For future research, using other types of fuzzy numbers could be discussed, another mechanism for a comparison of the fuzzy numbers or an involvement the vague structural coefficients and right sides of constraints in a mathematical model could be also studied in the fuzzy multiple objective programming method. These modifications should lead to the possibility of other applications.

Acknowledgements

The research project was supported by Grant No. IGA F4/54/2015 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague and Grant No. 13-07350S of the Grant Agency of the Czech Republic.

References

[1] Bellman, R. E., and Zadeh, L. A.: Decision making in a fuzzy environment. *Management Science* **17** (1970), 141–164.

[2] Benayoun, R., de Montgolfier, J., Tergny, J., and Laritchev, O. I.: Linear programming with multiple objective functions: STEP Method (STEM). *Mathematical Programming* **1** (1971), 366–375.

[3] Borovička, A.: Fuzzy weights estimation method based on the linguistic expression of criteria relevance. *Ekonomická revue* **17** (2014), 13–23.

[4] Borovička, A.: *Vytváření investičního portfolia podílových fondů pomocí fuzzy metod vícekriteriálního rozhodování*. Ph.D. Thesis, Department of Econometrics, University of Economics, Prague, 2015.

[5] Chen, C. H.: Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems* **114** (2000), 1–9.

[6] Lai, Y. J., and Hwang, C. L.: *Fuzzy Multiple Objective Decision Making: Methods and Applications*. Springer, Berlin, 1996.

[7] McCahone, C.: *Fuzzy Set Theory Applied to Production and Inventory Control*. Ph.D. Thesis, Department of Industrial Engineering, Kansas State University, 1987.

[8] Roy, B.: Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *La Revue d'Informatique et de Recherche Opérationelle (RIRO)* **2** (1968), 57–75.

[9] Roy, B.: ELECTRE III: algorithme de classement basé sur une représentation floue des préférences en présence des critères multiples. *Cahiers du CERO* **20** (1978), 3–24.

DEA-risk models with Value at Risk inputs

Martin Branda[1]

Abstract. DEA-risk models with diversification are suitable for accessing efficiency of investment opportunities available on financial markets. One of the most important properties of these models – compared with the traditional ones – is that the optimal solutions correspond to efficient investment opportunities, thus can be used by investors to revise their inefficient portfolios. In this paper, we focus on DEA models where Value at Risk serves as an input and we enable short sales under margin requirements. We discuss possible reformulations necessary to solve the resulting model. The approach is demonstrated on real data from US stock market.

Keywords: Data envelopment analysis, diversification, value at risk, portfolio selection

JEL classification: C44
AMS classification: 90C15

1 Introduction

Since the seminar work of Markowitz [13], the portfolio selection problem has deserved much attention by researchers as well as practitioners. Various models, formulations, risk measures have been introduced and investigated. In this paper, we focus on diversification-consistent data envelopment analysis (DEA) models with inputs which quantify risk. These models are useful not only for classifying of portfolios as efficient and inefficient, but the optimal values can be used by investors to revise their inefficient portfolios to get efficient ones. This important property does not hold for the traditional DEA models, see [3] for a deep discussion. A convergence of the traditional DEA efficient frontier to the diversification-consistent one was investigated by [12].

Diversification-consistent models were investigated by several papers. In [6, 7, 10], the author focused on mean–variance and mean–variance–skewness efficiency. Positive parts of coherent risk measures were employed by [11] as the inputs. In [1], the author avoided cutting the risk measures by using general deviation measures, which are positive for all nondegenerated random variables. Paper [3] returned to the coherent risk measures and proposed DEA models based on directional distance measures which enable to use inputs which can take positive and negative values. Under particular choices of the inputs and output, it can be even shown that DEA-risk models are equivalent to stochastic dominance efficiency tests, see [4, 5].

In this paper, we focus on mean-risk efficiency of investment opportunities, in particular we use Value at Risk (VaR) to quantify the risk. VaR does not belong to any of the above mentioned classes of risk functionals, thus we cannot rely on their properties, e.g., on subadditivity and monotonicity, and propertied of the DEA-VaR model have to be investigated carefully. Moreover, we consider the set of investment opportunities with short sales allowed under margin requirements. We formulate a general DEA model and then propose its reformulation under finite discrete distribution, which is then used in the empirical part. We investigate efficiency of 48 US stock representative portfolios. We propose their ranking under various choices of the margin parameters and also report several dominating portfolios.

The paper is organized as follows. Section 2 contains basic notation and definitions. DEA-risk model with value at risk inputs is formulated in Section 3. Efficiency of US stock representative portfolios is investigated in Section 4. Section 5 concludes the paper.

[1]Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, Prague 186 75, Czech Republic
tel.: +420 221 913 404, fax: +420 222 323 316, branda@karlin.mff.cuni.cz

2 Preliminaries and notation

Let R be a random variable representing the rate of return. For arbitrary $\alpha \in (0,1)$, Value at Risk (VaR) can be defined as the optimal solution of the following chance-constrained problem

$$VaR_\alpha(R) = \min\{\xi : P(-R \le \xi) \ge \alpha\},$$

i.e. the random loss is lower than VaR_α with high probability α, cf. [14]. Note that value at risk represents α-quantile of random variable $-R$.

We consider the following set of attainable investment opportunities based on n investment opportunities with random rates of return R_i:

$$\mathcal{R} = \left\{ \sum_{i=1}^{n} R_i x_i + \left(1 - \sum_{i=1}^{n} x_i\right) r : \quad 1 - \sum_{i=1}^{n} x_i + \gamma \sum_{i=1}^{n} x_i^+ \ge \beta \sum_{i=1}^{n} x_i^-, \right. \tag{1}$$

$$\left. x_i = x_i^+ - x_i^-,\ x_i^+ \ge 0, x_i^- \ge 0,\ i = 1, \dots, n \right\}, \tag{2}$$

where $\beta \ge 1$ denotes the margin requirement level on short position for risky assets and $0 \le \gamma < 1$ the nonnegative discount rate for the long position of the owned risky assets. If γ is equal to zero, then only cash can be used for margins. Moreover, the not-invested cash can be deposit for rate r and the cash invested over budget has to be borrowed for the same rate. Analogous constraints were considered by [9] in their mean-variance DEA-risk model.

Under the finite discrete distribution of $R \in \mathcal{R}$ with S realizations r_s and probabilities $p_s > 0$, we can obtain the value at risk on the level α as the optimal value of the following mixed-integer linear programming problem

$$\begin{aligned} VaR_\alpha(R) = \min_{\xi, y_s}\ & \xi \\ \text{s.t.}\ & \\ \sum_{s=1}^{S} p_s y_s &\ge \alpha, \\ -r_s - \xi &\le M(1 - y_s),\ s = 1, \dots, S, \\ y_s &\in \{0,1\},\ s = 1, \dots, S, \end{aligned} \tag{3}$$

where $M > 0$ is a sufficiently large constant. Similar reformulation is then used in the following Section.

3 A DEA-risk model with value at risk inputs

Let $R_0 \in \mathcal{R}$ be our benchmark that is compared with all $R \in \mathcal{R}$ based on the values of considered VaRs on levels $\alpha_1, \dots, \alpha_K$ and expected return. The following program projects the benchmark investment opportunity to the weak Pareto-Koopmans efficient frontier in directions e, d_k:

$$\begin{aligned} \theta^{\mathrm{DC}}(R_0) = \max_{\theta, x_i}\ & \theta \\ \text{s.t.}\ & \\ \mathbb{E}\left[\sum_{i=1}^{n} R_i x_i\right] + \left(1 - \sum_{i=1}^{n} x_i\right) r &\ge \mathbb{E}[R_0] + \theta \cdot e(R_0), \\ VaR_{\alpha_k}\left(\sum_{i=1}^{n} R_i x_i + \left(1 - \sum_{i=1}^{n} x_i\right) r\right) &\le VaR_{\alpha_k}(R_0) - \theta \cdot d_k(R_0),\ k = 1, \dots, K, \\ 1 - \sum_{i=1}^{n} x_i + \gamma \sum_{i=1}^{n} x_i^+ &\ge \beta \sum_{i=1}^{n} x_i^-, \\ x_i^+ - x_i^- &= x_i,\ x_i^+, x_i^- \ge 0. \end{aligned} \tag{4}$$

We sat that the benchmark R_0 is efficient if the optimal value is equal to zero, i.e. the corresponding mean-VaR image lies on the efficient frontier in the multivariate mean-VaR space. Based on suggestion of [3], we employ the following nonnegative directions

$$\begin{aligned} e(R_0) &= \max_{R\in\mathcal{R}} \mathbb{E}[R] - \mathbb{E}[R_0], \\ d_k(R_0) &= VaR_{\alpha_k}(R_0) - \min_{R\in\mathcal{R}} VaR_{\alpha_k}(R). \end{aligned}$$

Under finite discrete distribution, these directions can be evaluated using similar reformulation as for the whole DEA model.

Let the distribution of random returns be finite discrete with realizations $r_{i,s}$, $s = 1,\dots,S$ and corresponding probabilities $p_s > 0$, $\sum_{s=1}^{S} p_s = 1$. Then, problem (4) can be formulated and solved as a (large) mixed-integer linear programming problem

$$\theta(R_0) = \max_{\theta,\xi_k,x_i y_{k,s}} \theta \tag{5}$$

s.t.

$$\sum_{s=1}^{S} p_s \sum_{i=1}^{n} r_{i,s}x_i + \left(1 - \sum_{i=1}^{n} x_i\right) r \geq \mathbb{E}^S[R_0] + \theta \cdot \tilde{e}^S(R_0), \tag{6}$$

$$\sum_{s=1}^{S} p_s y_{k,s} \geq \alpha_k,\ k = 1,\dots,K, \tag{7}$$

$$-\sum_{i=1}^{n} r_{i,s}x_i - \Big(1 - \sum_{i=1}^{n} x_i\Big)r - VaR^S_{\alpha_k}(R_0) + \theta \cdot \tilde{d}^S_k(R_0) \leq M(1 - y_{k,s}),$$

$$k = 1,\dots,K,\ s = 1,\dots,S, \tag{8}$$

$$1 - \sum_{i=1}^{n} x_i + \gamma \sum_{i=1}^{n} x_i^+ \geq \beta \sum_{i=1}^{n} x_i^-, \tag{9}$$

$$x_i^+ - x_i^- = x_i, \tag{10}$$

$$y_{k,s} \in \{0,1\},$$

where $M > 0$ is a sufficiently large constant. Moreover, $\mathbb{E}^S$, $VaR^S_{\alpha_k}$, $\tilde{e}^S$, $\tilde{d}^S_k$ are computed using the discrete distribution and before solving the problem. Now, let us describe the constraints:

(6) ensures that expected return of portfolio is higher or equal to the expected return of benchmark and measures possible improvement (increase),

(7) models value at risk on level α_k using binary variables $y_{k,s}$

(8) identified if random loss is lower than $VaR^S_{\alpha_k}(R_0) - \theta \cdot \tilde{d}^S_k(R_0)$; if not, than the corresponding $y_{k,s} = 0$,

(9) short positions have to be covered by available cash and discounted assets in long position,

(10) split the portfolio weights into positive and negative parts.

4 Numerical study

In this section, we investigate efficiency of 48 US industry representative portfolios observed monthly from July 2004 to June 2014, i.e. we consider $S = 120$ observations of random returns. The industry portfolios are based on four-digit SIC codes and are listed in freely available Kenneth French library[1]. We used the observations as realizations of a discrete distribution of returns r_{is}, $s = 1,\dots,S$ and set the probabilities to $p_s = 1/120$. We consider $K = 4$ VaRs on the most commonly used levels $\alpha_k \in \{0.75, 0.9, 0.95, 0.99\}$. We use $r = 2\%$ p.y. as a rate of return of the riskless asset. The resulting diversification-consistent DEA models were solved using solver CPlex connected to the modelling system GAMS.

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french

We consider four different combinations of the parameters $\beta \in \{1, 1.5\}$ and $\gamma \in \{0, 0.2\}$. The results can be found in Table 1. We can observe that the rankings are very similar for all considered values of the parameters. Portfolio Smoke has the highest ranking for models where cash as well as assets in long positions can be used to cover the short positions, whenever portfolio Guns is ranked as first in the models where only cash can be used. However, only portfolio Guns lies on the efficient frontier under $\gamma = 0$. In Table 2, we report the portfolio weights which correspond to the investment opportunities which dominate Smoke under various choices of the parameters and – according to the theory – are efficient. Note that only nonzero weights are reported.

γ	β	Agric	Food	Soda	Beer	Smoke	Toys	Fun	Books
0	1	0.62 (19)	0.55 (5)	0.62 (17)	0.55 (6)	0.39 (2)	0.69 (36)	0.69 (34)	0.75 (45)
0.2	1	0.61 (22)	0.54 (4)	0.61 (18)	0.55 (6)	0.45 (1)	0.67 (34)	0.68 (36)	0.72 (45)
0	1.5	0.45 (15)	0.41 (5)	0.48 (19)	0.43 (7)	0.11 (2)	0.59 (35)	0.56 (32)	0.66 (45)
0.2	1.5	0.45 (15)	0.39 (5)	0.47 (21)	0.41 (6)	0.23 (1)	0.55 (35)	0.54 (33)	0.62 (45)
		Hshld	Clths	Hlth	MedEq	Drugs	Chems	Rubbr	Txtls
0	1	0.63 (22)	0.6 (11)	0.66 (27)	0.67 (29)	0.57 (7)	0.61 (14)	0.63 (24)	0.67 (32)
0.2	1	0.61 (19)	0.59 (10)	0.64 (25)	0.65 (27)	0.56 (7)	0.6 (15)	0.62 (23)	0.67 (35)
0	1.5	0.53 (28)	0.43 (10)	0.54 (30)	0.57 (34)	0.43 (12)	0.43 (11)	0.49 (24)	0.52 (26)
0.2	1.5	0.49 (25)	0.43 (10)	0.52 (29)	0.54 (32)	0.42 (7)	0.45 (14)	0.48 (24)	0.55 (34)
		BldMt	Cnstr	Steel	FabPr	Mach	ElcEq	Autos	Aero
0	1	0.69 (33)	0.73 (41)	0.74 (44)	0.65 (25)	0.63 (23)	0.67 (31)	0.71 (38)	0.6 (12)
0.2	1	0.67 (32)	0.7 (41)	0.72 (44)	0.64 (26)	0.63 (24)	0.66 (31)	0.69 (38)	0.59 (11)
0	1.5	0.55 (31)	0.62 (41)	0.63 (42)	0.5 (25)	0.48 (20)	0.53 (27)	0.59 (36)	0.45 (14)
0.2	1.5	0.53 (31)	0.59 (41)	0.61 (43)	0.49 (26)	0.47 (22)	0.52 (27)	0.56 (36)	0.45 (12)
		Ships	Guns	Gold	Mines	Coal	Oil	Util	Telcm
0	1	0 (1)	0.53 (4)	0.77 (47)	0.65 (26)	0.77 (48)	0.59 (9)	0.58 (8)	0.6 (10)
0.2	1	0.48 (2)	0.54 (5)	0.75 (47)	0.65 (28)	0.76 (48)	0.6 (12)	0.57 (8)	0.58 (9)
0	1.5	0 (1)	0.34 (4)	0.68 (48)	0.43 (8)	0.66 (46)	0.41 (6)	0.43 (9)	0.46 (17)
0.2	1.5	0.23 (2)	0.36 (4)	0.64 (47)	0.48 (23)	0.65 (48)	0.42 (8)	0.43 (9)	0.45 (11)
		PerSv	BusSv	Comps	Chips	LabEq	Paper	Boxes	Trans
0	1	0.73 (42)	0.62 (21)	0.67 (28)	0.71 (37)	0.62 (20)	0.67 (30)	0.61 (13)	0.62 (16)
0.2	1	0.71 (42)	0.61 (20)	0.65 (30)	0.68 (37)	0.61 (21)	0.65 (29)	0.6 (14)	0.61 (17)
0	1.5	0.63 (43)	0.48 (22)	0.53 (29)	0.61 (39)	0.46 (18)	0.56 (33)	0.44 (13)	0.46 (16)
0.2	1.5	0.6 (42)	0.47 (20)	0.52 (28)	0.57 (39)	0.46 (17)	0.53 (30)	0.45 (13)	0.45 (16)
		Whlsl	Rtail	Meals	Banks	Insur	RlEst	Fin	Other
0	1	0.62 (18)	0.62 (15)	0.48 (3)	0.76 (46)	0.69 (35)	0.72 (40)	0.71 (39)	0.74 (43)
0.2	1	0.61 (16)	0.6 (13)	0.5 (3)	0.73 (46)	0.67 (33)	0.7 (40)	0.69 (39)	0.71 (43)
0	1.5	0.48 (21)	0.49 (23)	0.28 (3)	0.68 (47)	0.6 (38)	0.61 (40)	0.59 (37)	0.65 (44)
0.2	1.5	0.47 (19)	0.46 (18)	0.31 (3)	0.64 (46)	0.57 (38)	0.58 (40)	0.56 (37)	0.62 (44)

Table 1 Efficiency scores and rankings (in brackets)

5 Conclusions

In this paper, we have proposed a new DEA-risk model with value at risk inputs and short sales allowed under margin requirements. We have shown that under finite discrete distribution of the rates of return, the model can be solved as a large mixed-integer linear programming problem. The numerical results have shown that the ranking of the investment opportunities is not influenced significantly under various choices of the margin parameters. We have also reported several dominating portfolios. Future research will be devoted to stability and robustness of the proposed model. An important question is also the proper choice of the number of inputs and outputs, cf. [15].

γ	0	0.2	0	0.2
β	1	1	1.5	1.5
Smoke	0.344	0.374	0.680	0.819
Books	-0.532	-0.438	-0.090	-0.146
Clths	0.000	0.029		
Txtls	0.008	0.093		
Steel		-0.008		-0.021
FabPr		0.042		
Ships	0.623	0.712	0.269	0.267
Gold				-0.030
Mines	0.025			
Coal	-0.019	-0.166		-0.001
PerSv		-0.085		
Banks		-0.225	-0.011	-0.064
Other	-0.195			
Riskless	0.746	0.672	0.152	0.177

Table 2 Weights of portfolios dominating to Smoke

Acknowledgements

This work was supported by the Czech Science Foundation under the Grant GP13-03749P.

References

[1] Branda, M.: Diversification-consistent data envelopment analysis with general deviation measures. *European Journal of Operational Research* **226** (3) (2013), 626–635.

[2] Branda, M.: Reformulations of input-output oriented DEA tests with diversification. *Operations Research Letters* **41** (5) (2013), 516–520.

[3] Branda, M.: Diversification-consistent data envelopment analysis based on directional-distance measures. *Omega: International Journal of Management Science* **52** (2015), 65–76.

[4] Branda, M., Kopa, M.: DEA-risk efficiency and stochastic dominance efficiency of stock indices. *Czech Journal of Economics and Finance* **62** (2) (2012), 106–124.

[5] Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22** (1) (2014), 13–35.

[6] Briec, W., Kerstens, K., Lesourd, J.-B.: Single period Markowitz portfolio selection, performance gauging and duality: a variation on the Luenberger shortage function. *Journal of Optimization Theory and Applications* **120** (1) (2004), 1–27.

[7] Briec, W., Kerstens, K., Jokung, O.: Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach. *Management Science* **53** (2007), 135–149.

[8] Dempster, M.A.H. (Ed.): Risk Management: Value at Risk and Beyond. Cambridge University Press, 2010.

[9] Ding, H., Zhou, Z., Xiao, H., Ma, C., Liu, W.: Performance Evaluation of Portfolios with Margin Requirements, *Mathematical Problems in Engineering*, vol. 2014, Article ID 618706, 8 pages, 2014. doi:10.1155/2014/618706

[10] Joro, T., Na, P.: Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research* **175** (2006), 446–461.

[11] Lamb, J.D., Tee, K-H.: Data envelopment analysis models of investment funds. *European Journal of Operational Research* **216** (3) (2012), 687–696.

[12] Liu, W., Zhou, Z., Liu, D., Xiao, H.: Estimation of portfolio efficiency via DEA. *Omega: International Journal of Management Science* **52** (2015), 107–118.

[13] Markowitz, H.M.: Portfolio selection. *Journal of Finance* **7** (1) (1952), 77–91.

[14] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26 (2002), 1443–1471.

[15] Toloo, M., Tichý, T.: Two alternative approaches for selecting performance measures in data envelopment analysis. Measurement 65 (2015), 29–40.

Security risk factors: ANP model for risk management decision making

Helena Brožová[1], Jan Rydval[1], Libor Šup[2],Moufida Sadok[3], Peter Bednar[4]

Abstract. Information is a valuable asset supporting management decisions and business operations within the enterprise. Consequently, securing the company critical information assets from sophisticated insider threats and outsider attacks is essential to ensure business continuity and compliance with regulatory frameworks. Security risk management is the process that identifies threats and vulnerabilities of an enterprise information system, evaluates the likelihood of their occurrence and estimates their potential business impact. It is a continuous process that allows cost effectiveness of implemented security controls and provides a dynamic set of tools to monitor the security level of the information system. Given the uncertainty and complexity of security risks analyses, the identification of risk factors as well as the estimation of their business impact require tools for assessment of risk with multi-value scales according to different stakeholders' point of view. Therefore, the purpose of this paper is to model risk factors using semantic network to develop the decision network and the Analytical Network Process (ANP) to evaluate factors of complex problems taking into consideration quantitative and qualitative data. As a decision support technique ANP also measures the dependency among risk factors related to the elicitation of individual judgement.

Keywords: Information security, Risk factors, Semantic networks, Analytical network process, Multi-criteria decision making.

JEL Classification: C44 Operations Research • Statistical Decision Theory
AMS Classification: 90B50 Management decision making, including multiple objectives
90B90 Case-oriented studies

1 Introduction

This paper seeks to explore the potential relevance of the ANP use in information systems security (ISS) context to support the exploration of individual understanding of security risks leading to richer elaboration of problem spaces. In fact, the objectives of ISS process are to help to identify new threats and vulnerabilities, to estimate their business impact and to provide a dynamic set of tools to control the security level of the information system. The identification of a number of risk factors requires a classification according to their severity and impact on the information system activity. However, this classification should pay attention to the influence of contextual variables such as the exploration of multiple perspectives of contextual understanding of security risk factors. The involvement of organizational stakeholders to assess security risks with multi-value scales would result in a better understanding of the role and application of security functions in situated practices and an achievement of contextually relevant risk analysis [1], [13]. Consequently, we argue in this paper that the ANP could potentially provide an approach to evaluate security risk factors taking into consideration quantitative and qualitative data. As a decision support technique the ANP also measures the dependency among risk factors related to the elicitation of individual judgement. A case study is used to illustrate this concept.

The remainder of this paper is organized as follows. In the section 2, a short review of ANP and semantic network concepts found in literature are provided. In Section 3 a case study is given to illustrate the use of ANP and semantic network techniques. The empirical findings of the case study are also discussed in this section and in the section 4. Finally, concluding remarks are presented in section 5.

[1]Dept. of Systems Engineering, Faculty of Agricultural Economics and Management, Czech University of Life Sciences, Prague.
[2]Dept. of Systems Engineering, Faculty of Agricultural Economics and Management, Czech University of Life Sciences and Dept. of Security, Czech University of Life Sciences, Prague.
[3]Higher Institute of Technological Studies in Communication in Tunis, Tunisia and School of Computing, University of Portsmouth, UK
[4]School of Computing, University of Portsmouth, UK, and Dept. of Informatics, Lund University, Sweden

2 Materials and Methods

The ANP is a multiple criteria decision method based on the network representation of a decision problem which considers the dependence across elements and levels of a decision problem [11], [12]. The crucial step of the ANP is the pairwise comparison of all pairs of elements related to the same element from higher level or different cluster from decision network. This judgement has to be consistent, in case it is not we use simple role showing how to improve judgement consistency. The steps of the ANP method for this study are as follows:

1. The first step – the semantic network describing the elements of the ISS decision problem and its relations is constructed.
2. The second step - the network is created based on the semantic network to describe inner dependence within a set (clusters) of decision elements, and outer dependence among different sets (clusters) of the decision elements.
3. The third step - the pairwise comparisons of the elements within and across the clusters are made. The consistency of these comparisons is also checked.
4. The fourth step - if the comparison is not consistent, the decision maker will see how to change and adjust the comparison [4].
5. The fifth step - the normalized supermatrix with the preferences derived from the previous pairwise comparisons is calculated.
6. The sixth step - the limiting supermatrix is computed using program SuperDecision [18] and global preferences of decision elements are obtained [11] .

The semantic network of the decision problem can be used as started point for creation of the ANP decision network. The basic advantage of the semantic network is that it contains information similarly as information is stored in human memory, and it is machine-understandable. This means that it can be machine-processed. It is possible to analyse facts and information included in the semantic network and to acquire new knowledge about represented facts [17], [19].

Such semantic network illustrates different points of view as well as the relationships between different relevant elements within a decision context. In effect, semantic networks were originally used to express meanings of various expressions in natural language. According to Sowa [16] semantic networks are used namely because of their ability to easily provide usable system to represent information and to focus mainly on the organization of a large number of information sources. They also support the description of complex processes and offer a tool to represent the understanding of a problem space.

Semantic (associative) network is defined as a directed graph consisting of nodes and edges [16]. Nodes represent items of described problem and edges connecting these nodes represent relationships between these items. Fundamental types of these relations are as follows:

- IS-AN-INSTANCE-OF (IIO) relationship is used to express the fact that a particular object (for instance of a particular class) belongs to the specified class.
- IS-A-KIND-OF (IKO) relationship can express that a class is a subclass of another class.
- IS-A-PART-OF (IPO) relationship serves to express that a certain class of objects is composed of some parts.

Pairwise comparison of items of the created decision network is then made by IT expert. This judgement is filled in special form in MS Excel (**Table 1**) which helps through the entire process.

A - most important					**Equally**					**B - most important**
A	9	7	5	3	1	3	5	7	9	**B**
Poor password practices	x									Poor reporting procedures
Poor password practices						x				RAT installed
Poor reporting procedures								x		RAT installed

Table 1 MS Excel form for pairwise comparisons

After the user files this table the consistency index is computed. If the comparisons are not consistent the automatic calculation shows suitable third value of each individual preference. These ideal values can serve as a guideline for adjusting the initial evaluation (**Table 2**).

				Weights						
Poor reporting procedures	1.000	9.000	0.333	**0.324**	*-2.206*	*9.000*	*0.333*	*Lambda*	*3.2056*	**Consistency index**
RAT installed	0.111	1.000	0.143	**0.056**	*0.111*	*-2.206*	*0.143*	*Determinant*	*-6E-05*	**0.10**
RAT installed	3.000	7.000	1.000	**0.62**	*3.000*	*7.000*	*-2.206*			
	Ideal values			1						
	1.167	**0.643**								
		0.019								

Table 2 Ideal values for pairwise comparisons

3 Case Study – ISS management

In the particular context of ISS it is necessary to construct different scenarios of security risks reflecting different points of view. The comparison and integration of security risk factors within a semantic network provide an overview of the relative importance and impact of a particular security factor. This leads to identify the most important security risk factors and to assess their impact on the efficiency and effectiveness of an information system. Consequently, appropriate mitigation decisions to cope with and reduce security risks could efficiently be made.

In this section we describe three diagrams as produced by three stakeholders (end user, network administrator, security expert) respectively following a security breach. These diagrams contain the most important risk security factors and their relations.

The view of security expert about possible security risk factors explaining the data breach is captured in Figure 1. The view of the network administrator about possible security risk factors explaining the data breach is captured in Figure 2. The end user's view about possible security risk factors explaining the data breach is captured in Figure 3.

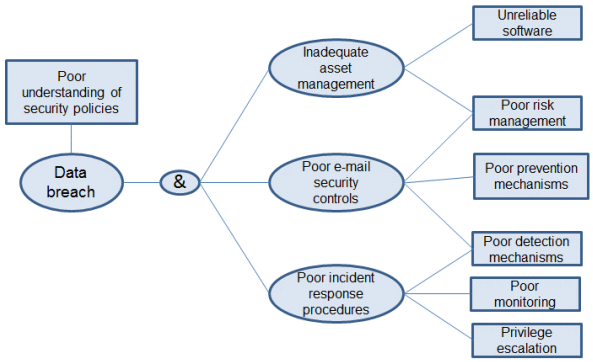


Figure 1 View of the security expert

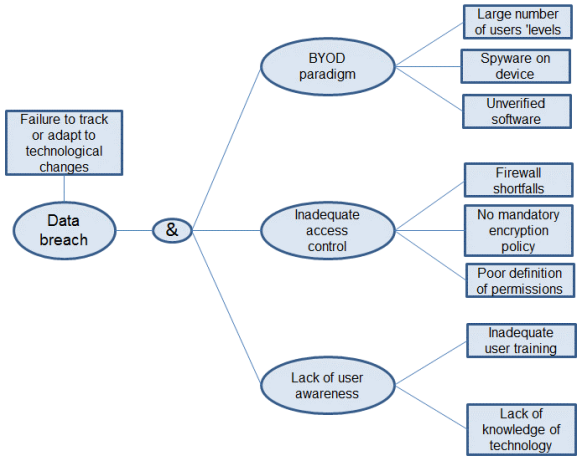


Figure 2 View of the network administrator

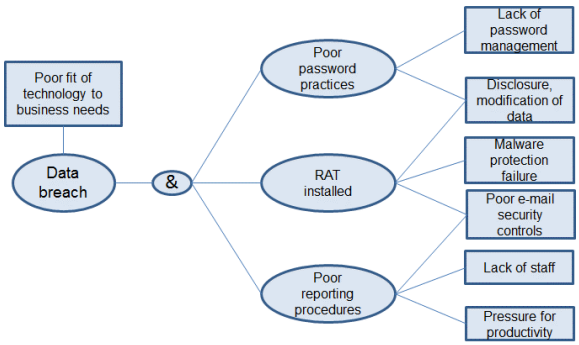


Figure 3 View of the end user

3.1 Semantic Network of Security Risk Factors

Semantic network can display individual elements influencing issues of data breach. It provides us with information about relationships in the network between individual factors of the ISS and how they can influence the threatened data. However, it does not give us the quantitative information about the importance of these factors and how much they influence the data breach; however it can be used as started point for creation of the ANP decision network.

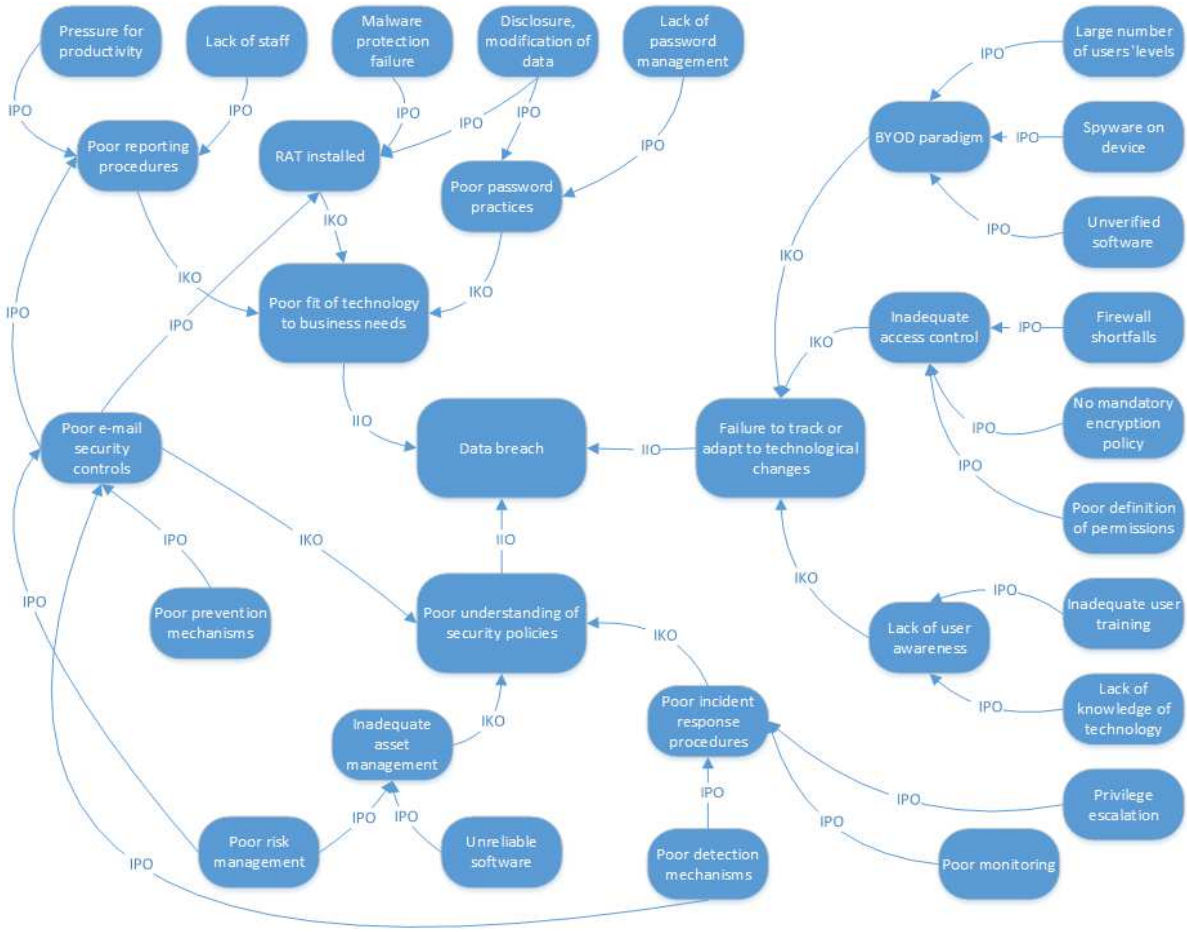


Figure 4 Semantic Network of Data Breach Factors

The figure above shows that the issue of data breach consists of three main instances corresponding to the views of individual stakeholders: *Poor understanding of security policies*, *Failure to track or adapt to technological changes*, *Poor fit of technology to business needs*.

Each of these three major instances of the *Data Breach* incorporates various elements expressing different possibilities of the instances formation. These elements could be broken down into a different amount of sub elements expressing the component parts of the element responsible for particular instance of the *Data Breach*.

Some elements, respectively sub-elements can play multiple roles within the semantic network, which are caused by a different point of view from which the element is observed. The role of the element and its affiliation to the class of elements or sub-elements is displayed in the Figure 4 using evaluated connection between the elements (IIO, IKO or IPO).

For example, the element *Poor e-mail security controls* plays two kinds of roles. At first it is a kind of instance of item *Poor understanding of security policies* (oriented connection IKO) and at second it is a part (sub-element) of the element *RAT installed* (connection IPO).

3.2 ANP Evaluation of Security Risk Factors

Semantic network (**Figure 4**) is used as a decision network for the ANP method. The importance of each factor is made using Saaty pairwise comparison of all factors related to the factor or cluster on the higher level. This evaluation is made by an IT expert. The consistency of these judgements was checked. If the comparison was not

consistent, IT expert got a feedback and an advice on how to adjust the comparison. After this process we received the consistent Saaty matrices and the ANP model was calculated by SuperDecision software.

More important area in ISS management according to the results is *Poor understanding of security policies* (**Table 3**). Our analysis shows that the *Poor risk management*, the *Privilege escalation*, the *Poor detection mechanisms* and the *Poor monitoring* are the most important factors explaining the possible *Data breach*. These results are coherent with previous studies as many of the existing risk analysis models and frameworks focus mainly on the technical modules related to the development of security mitigation and prevention and do not pay much attention to the influence of contextual variables affecting the reliability of the provided solutions [14], [15].

1 Poor understanding of security policies	
1.1 Inadequate asset management	
1.1.1 Unreliable software	0.005
1.1.2 Poor risk management	0.375
1.2 Poor e-mail security controls	
1.2.1 Poor prevention mechanisms	0.038
1.2.2 Poor detection mechanisms	0.121
1.3 Poor incident response procedures	
1.3.1 Poor monitoring	0.121
1.3.2 Privilege escalation	0.174

2 Failure to track or adapt to technological changes	
2.1 BYOD paradigm	
2.1.1 Large number of users 'levels	0.022
2.1.2 Spyware on device	0.004
2.1.3 Unverified software	0.005
2.2 Inadequate access control	
2.2.1 Firewall shortfalls	0.015
2.2.2 No mandatory encryption policy	0.015
2.2.3 Poor definition of permissions	0.015
2.3 Lack of user awareness	
2.3.1 Inadequate user training	0.019
2.3.2 Lack of knowledge of technology	0.002

3 Poor fit of technology to business needs	
3.1 Poor password practices	
3.1.1 Lack of password management	0.031
3.1.2 Disclosure, modification of data	0.007
3.2 RAT installed	
3.2.1 Malware protection failure	0.031
3.3 Poor reporting procedures	
3.3.1 Lack of staff	0.001
3.3.2 Pressure for productivity	0.001

Table 3 Importance of security risk factors related to data breach

4 Discussion

Methods or approach of risk assessment of the ISS factors can be divided into i) deterministic methods, ii) probability methods, iii) methods using analogies and iv) multi-criteria evaluation methods [5]. From the last-mentioned approach Delphi method is often used for the multi-criteria evaluation of risk [2], [7]. The ANP method used in this paper for the evaluation of the importance of defined indicators (elements) is not commonly used in the Czech Republic as stated Procházková [6] although the potential of the ANP in the risk assessment is considerable. Both the Delphi method the ANP method use the group of experts, and the reliability of the experts is a critical factor of the quality of risk assessment and evaluation. Procházková [6] states that reliability of experts assessment can be expressed as $r = 1 - \frac{1}{2^n}$ where r is the confidence level and n is the number of experts. The confidence level can reach values from the interval [0.5, 1.0), the more experts are included into evaluation process the final assessment is more relevant.

In our approach the decision network of risk assessment is based on semantic network. However, an easier way to capture the interdependence of the elements of risk analysis is used in mind map by Buzan and Buzan [3]. They state that the mind map is a graphic layout of text (accompanied by pictures) showing the connections. This approach (resembling the structure of decision trees) of graphical representations of mind processes using structured texts supplemented by pictures does not work with oriented connections and nodes; however oriented connections and nodes can be used in the semantic network to express the relationships of seemingly unrelated ideas and information and therefore the importance of elements of semantic network can be calculated.

Oriented connections and nodes in form of directed graphs can be used to determine the value of risks using Bayesian networks [5] which are based on the method of conditional probability. Oriented connections and nodes in the ANP decision network are used in many decision problems which consider the dependence across elements and levels of decision problem, for instance for assessment of the importance of individual elements of the consumer's behaviour or of the framing effect [8], [9], [10].

5 Conclusion

The ANP is used for evaluation of the ISS risk factors in the process that identifies threats and vulnerabilities of an enterprise information system, evaluates the likelihood of their occurrence and estimates their potential business impact. This approach is going through developing multiple contextual understandings of information sys-

tem security risks using semantic network technique which is used by stakeholders in a dialogue for the discovery and description of relationships between the different problem elements as understood by each stakeholder, creation of the decision network and evaluation of individual relations. The automatic control of the judgement consistency is incorporated into the calculation. The aim of this process is to classify the importance of security risk factors and to implement effective security controls and procedures. Future research aiming to collect data related to the risk analysis for various types of information technologies projects would help in gaining better adjustment and estimation of security risk factors.

Acknowledgements

This research is supported by the Internal Grant Agency of the University of Life Sciences Prague – project IGA PEF 20151027 - The use of the Analytic Network Process to identify and to assess security risks of Information Technologies projects.

References

[1] Bednar, P., Katos, V.: Digital forensic investigations: a new frontier for Informing Systems, in D'Atri, A. and Sacca, D. (Ed.) *Information Systems: People, Organizations, Institutions and Technologies*, Springer Physica-Verlag, Berlin Heidelberg, ISBN: 978-3-7908-2147-5, 2010.

[2] Briš, R.: *Reliability, Risk and Safety: Theory and Applications*. CRC Press / Balkema, Leiden, 2009, ISBN 978-0-415-55509-8.

[3] Buzan, T., Buzan, B.: *Das Mind-Map-Buch. Die besteMethodezurSteigerungihresgeistigen Potentials*. Moderne Verlagsgesellschaft Mvg, 2002.

[4] Hlavatý, R.: Saaty`s matrix revisited: Securing the consistency of pairwise comparisons, In: *Proceedings of the 32st International conference on Mathematical Methods in Economics*. Palacký University, Olomouc, 2014, 83–88.

[5] IEC/ISO. Risk management – Risk assessment techniques IEC/ISO 31010. 1st ed. Geneva: ISO, 2009. ISBN 2-8318-1068-2.

[6] Procházková, D.: *Analýza a řízení rizik*. České vysoké učení technické v Praze, Praha, 2011, ISBN 978-80-01-04841-2.

[7] Procházková, D.: *Metody, nástroje a techniky pro rizikové inženýrství*. Karolinum, Praha, 2011, ISBN 978-80-01-04842-9.

[8] Rydval, J.: Quantification of Framing Effect using ANP and AHP. *In Mathematical Methods in Economics 2011*, JanskaDolina, Slovakia, Professional Publishing, 2011, ISBN 978-80-7431-058-4.

[9] Rydval, J., Bartoška, J.: Quantification of Framing Effect in the Meat Distribution by ANP, *Mathematical Methods in Economics 2013*, College of Polytechnics Jihlava, 2013, ISBN 978-80-87035-76-4.

[10] Rydval, J., Brožová, H.: Quantification of Framing effect in education Process using ANP. In: *Proceedings of Efficiency and Responsibility in Education International Conference 2011*, Prague, CULS, 2011, ISBN 978-80-213-2183-0.

[11] Saaty, T. L.: *Decision Making with Dependence and Feedback: The Analytic Network Process*, The Analytic Hierarchy Process Series. Pittsburgh: 2001, IX, RWS Publications.

[12] Saaty, T.L.: *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback*. Creative Decisions Foundation, 2003.

[13] Sadok, M., Katos, V., Bednar, P.: Developing contextual understanding of information security risks, *International Symposium on Human Aspects of Information Security & Assurance (HAISA 2014)*, Plymouth University, 2014.

[14] Salmela, H.: Analysing business losses caused by information systems risk: a business process analysis approach, *Journal of Information Technology*, Vol. 23, 2008, pp185-202.

[15] Siponen, M. and Willison, R.: Information security management standards: Problems and solutions, *Information & Management*, Vol. 46, 2009, pp267-270.

[16] Sowa, J. F.: *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole Publishing Co., Pacific Grove, CA, 2000.

[17] Steyvers, M., Tenenbaum, J. B.: The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth, *Cognitive Science*, **29** (2005), 41–78.

[18] SuperDecisions: Software for Decision-Making. URL <http://www.superdecisions.com/>

[19] Xia, Z. Y., Bu, Z.: Community detection based on a semantic network, *Knowledge-Based Systems*, **26**(2012), 30-39.

Economizing logistic costs of the manufacturing system using mathematical modeling to aid decision-making

Robert Bucki[1], Bronislav Chramcov[2]

Abstract. The paper highlights the problem of mathematical modelling of manufacturing processes taking place in the highly sophisticated logistic system. Its real state is represented by a manufacturing hall equipped with robotic centers which carry out pre-programmed operations. The charge material is supplied in the form of liquid aluminum and stored in base tanks which are supplied in time by an outsourcing company. Charge material is transported from the base tanks to robotic centers by trolleys with built-in tanks. Ready products (e.g. engine blocks) are moved to the storing area located next to the adequate centre. Customers' orders are to be made at optimized costs in time. Therefore, the main criterion remains to minimize production costs within the existing manufacturing and technological constraints. The main goal of mathematical modeling is to support decision making as the course of production cannot be alternated once it is set. The system is controlled by heuristic algorithms.

Keywords: manufacturing system optimization, heuristic algorithms, modeling and simulation.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Modeling complex logistic systems is an important issue in contemporary manufacturing environment [1]. Such systems require implementing discrete programming methods [2] as well as considering decision making while modeling and controlling complex manufacturing systems [3]. This all leads to optimization of such systems [4,6]. Moreover, lean manufacturing lets us minimize manufacturing costs [5,7]. The automotive industry has recorded continuous growth in recent years. This trend is set to continue: new markets and technologies are opening up completely new perspectives. However, this growth places high demands on the productivity of both existing and newly planned production lines and facilities. Robots have been present in the automotive industry for the last three decades which means that the cost-intensive models of the past have meanwhile made way for versatile and reliable high-tech robots which can be employed almost universally. They can be integrated flexibly and virtually without risk into almost any work process. At the same time, they create the ideal conditions for future production expansions or changeovers. Robots manage short-term increase in productivity or secure the future long term manufacturing activity [8].

The paper focuses on implementing robotic centers designed to make orders consisting of various types of engine blocks ordered by various clients. As the engine block is the main housing of the engine it has to include a certain number of requirements such as the wear resistance, long lasting, maintenance, and withstand the pressure created when combustion takes place. Based on the above features the most widely used material are cast iron and aluminum alloys to manufacture the cylinder block. Cast iron alloys are used because they contain good mechanical properties, low cost, and availability compared with other metals. But certain aluminum alloys contain most of the characteristics of cast iron but with low weight. And also, an aluminum alloy casted engine block gives a good surface finish and high machinability compared with cast iron alloys. Aluminum alloys main feature for its popularity is its low weight reducing the weight of the engine as well as the whole vehicle. But the main disadvantage is their cost compared with grey cast iron. Aluminum alloy has a good machinability properties compared with grey cast iron. Moreover, aluminum is used to form the major parts of the engine. Another benefit resulting from implementing aluminum in car manufacturing is its noise reduction possibility. The more aluminum is used in the production of a vehicle, the less the weight of the vehicle is and the less fuel it consumes, thereby reducing the amount of harmful emissions into the atmosphere.

[1] Institute of Management and Information Technology, ul. Legionów 81, 43-300 Bielsko-Biała, Poland, rbucki@wsi.net.pl
[2] Tomas Bata University in Zlín, Faculty of Applied Informatics, Nad Stráněmi 4511, 760 05 Zlín, The Czech Republic, chramcov@fai.utb.cz.

2 The analysis of the assumed manufacturing system

To control the manufacturing process certain data is required and must be precisely defined. First of all, the whole system is subject to thorough analysis. The manufacturing system consists of robotic centers and the melting stove for liquid aluminum. Each robotic centre is equipped with a dedicated heating stove (robotic centre tank), a casting machine, a saw, a cooling tunnel and transit store for cast iron sleeves. Robotic centers can produce various products, which means their times of manufacturing cycles differ and there are different demands for aluminum. Aluminum bars are melted in the melting stove however; liquid aluminum can also be delivered in some solutions of manufacturing systems (it is brought in special tanks just in time by trucks). The melting stove has to have the minimum required steady level of liquid aluminum. If there is need for it, another melting stove is activated in the manufacturing system. A trolley with a tank transports liquid aluminum along determined routes. There can be more trolley tanks if necessary. After emptying the tank it returns back to the tank filling station where it is filled with liquid aluminum and degassed before it is back on the route. All trolley tanks must undergo the process of degassing at the station. The tank filling station receives signals from all robotic centers about the state of aluminum in their own heating stoves. The state of these stoves is defined and they must be emptied throughout the manufacturing process. Liquid aluminum is poured into tanks from the melting stove within 3 minutes if the full stove contains 22000 kg of liquid aluminum. The tanks contain 1000 kg of liquid aluminum each. If the stove contains 10000 kg of liquid aluminum, the pouring operation takes up to 6 minutes. This leads to the conclusion that the coefficient of pouring time should be taken into account. The tank filling process takes place at the tank filling station and is subsequently transported to the tank degassing station which takes 1,5 minutes. The degassing process takes approximately 6 minutes. After this operation tanks of aluminum are transported to the robotic centers by a forklift. The forklift with a tank of liquid aluminum covers different distances to the robotic centers in the determined time. Moreover, an additional time is needed to transport the tank to the melting stove from the last robotic centre. The transport route is one way only because of safety regulations and as the width of the route enables only this kind of solution. Too long a time taken for the transport operation could lead to freezing aluminum after 30 minutes and subsequent losses and extra tank regeneration costs. After 35-40 cycles of filling a tank, a need for cleaning arises. It is performed at the tank cleaning station within the time of 30 minutes. After 230 cycles of filling the tank, it is regenerated within 2 days. This operation is carried out by a specialist company.

Each robotic centre is equipped with its heating stove (robotic centre tank) of a maximum capacity of 3000 kg which keeps the temperature within the given tolerance range. The minimum level of liquid aluminum in the heating stove is 2000 kg. Filling the heating stove with 1000 kg of aluminum from the tank lasts 60s. The casting machine in a robotic centre is equipped with the injection system with a portal which takes metal from the heating stove by a ceramic tank and delivers it to the injection chamber and moreover, four cast iron sleeves are inserted by the robot. If the amount of metal in the heating stove is too small, the tank takes too little aluminum and this does not meet the proper product requirements. After placing sleeves from the transit store for cast iron sleeves into the special form, it is closed and filled with liquid metal. In this way, an injection dose of aluminum is ready and after casting solidification in the form, the casting machine is opened and the robot takes the ready cast (an engine block) out of the form. A ready product leaves the form every 100 seconds. Then the robot takes the cast on the saw station where the gating system weighing 4 kg is cut during 30 seconds and flashes are removed. Then the cast is taken to the cooling tunnel where there are 9 casts. Subsequently, the operator takes the cast out of the cooling tunnel and controls it visually during 90 seconds and deposits into the container for finished products which can contain 56 products.

The ready containers are stored in front of the thermal stove where the heat treatment process takes place. The stove houses 18 containers. If the stove is empty the process lasts approximately 4 hours however; if the stove is full, the baskets with casts leave it every 1080 seconds. The baskets are stored in the storing field before the finishing line. There is a place for one basket in front of the finishing line. Engine blocks are put separately on the input transport roller conveyor and they are subject to operations in the press stations, two times in the buffing spindle and in the shot blasting machine which takes 23s. The output transport roller conveyor delivers them to the lines of visual inspection where they are packed and distributed. The press station removes remains of casting. A cast is put on the press by the input portal and taken out of it by the first robot in the finishing line which puts the cast further on the buffing spindle station. Two buffing spindle stations are in fact a spindle with an end milling cutter used to remove burrs which cannot be removed at the press station. There is an intermediate table between two buffing spindle stations whose goal is to transport the cast to the subsequent operation. Throughout the milling operation cycle the second robot holds a cast while burrs are removed. The shot blasting machine is responsible for blast cleaning in order to achieve adequate roughness of the surface structure as well as the required quality. The third robot puts casts on 20 adequate hooks which are placed in a line and moved through the shot blasting machine. The same robot pulls casts, pours shot out of the cast and employs the output transport roller conveyor.

The trolley tanks need a determined timescale of delivering metal to ensure continuity of the manufacturing process. The number of melting stoves and transport tanks is determined however, it should result from the order to be made and a standstill time or replacement times of machines. There is also no need to deliver liquid metal to centers which are characterized by a minimum level of liquid aluminum.

The mathematical model is to form the basis for the subsequent simulation process imitating delivering liquid aluminum to robotic centers in order not to stop the whole manufacturing activity at the smallest number of tanks (costs of tanks and their regeneration) and transport trolleys (costs of forklifts and their rent). Additionally, showing the work time of trolleys is another vital issue which results in the need of estimating the number of workers necessary for the company to employ in the shift.

3 The mathematical description of the system

3.1. Definition of the system

The manufacturing hall is equipped with robotic centers which carry out pre-programmed operations on charge material. The charge material is delivered in the form of liquid aluminum and stored in base tanks. It is assumed that base tanks are supplied in time from outside company sources. Charge material is transported from base tanks to robotic centers by means of trolleys (forklifts) with built-in tanks. Ready products are moved to the storing place which is usually located next to the robotic centre. Customers' orders are presented in the vector $Z = [z_n]$, where z_n is the nth order. The state of orders is modified after each manufacturing decision according to the form (1) where $x_{n,k}$ is the number of the product of the nth order made at the kth stage. The stage k, $k=1,...,K$ is the moment of making the manufacturing decision.

$$z_{n,k} = \begin{cases} z_{n,k} - x_{n,k} & \text{if there is a manufacturing decision,} \\ z_{n,k-1} & \text{otherwise.} \end{cases} \tag{1}$$

Let the matrix G^{tank} represent the state of base tanks for liquid aluminum in the manufacturing hall and consequently variable $g_{l,k}^{\text{tank}}$ is the capacity of the lth base tank at the kth stage expressed in liters. This variable changes according to the form (2) where $x_{l,k}^{\text{tank}}$ represents the amount of the liquid aluminum removed from the lth base tank at the kth stage expressed in liters.

$$g_{l,k}^{\text{tank}} = \begin{cases} g_{l,k}^{\text{tank}} - x_{l,k}^{\text{tank}} & \text{if there is a manufacturing decision,} \\ g_{l,k-1}^{\text{tank}} & \text{otherwise.} \end{cases} \tag{2}$$

It is possible to define the maximum capacity of base tanks by vector $G^{\text{tank(max)}}$ where the variable $g_l^{\text{tank(max)}}$ represents the maximum capacity of the lth base tank expressed in litres. Consequently, base tanks are characterized by their minimum capacity. It is possible to define them by vector $G^{\text{tank(min)}}$ where the variable $g_l^{\text{tank(min)}}$ represents the minimum capacity of the lth base tank expressed in litres.

If the equation in the form (3) is valid, then the lth base tank is performing its function in the manufacturing process. Otherwise, the lth base tank is excluded from the manufacturing process.

$$\left(g_l^{\text{tank(min)}} \le g_{l,k}^{\text{tank}} \le g_l^{\text{tank(max)}}\right) \quad \forall l \in \{1,2,...,L\} \ \wedge \ \forall k \in \{1,2,...K\} \tag{3}$$

Customers' orders are made in R robotic centers placed in the manufacturing hall. The matrix $\Omega = [\omega_{n,r}]$ represents adjustment of robotic centers responsible for making customers' orders where $\omega_{n,r}$ represents adjustment of the rth robotic center to the nth order. Moreover, $\omega_{n,r} = 1$ if the nth order can be made in the rth robotic center, $\omega_{n,r}$=0 otherwise. Each rth robotic centre is supplied with liquid aluminum transported by the uth trolley. It is possible to define by matrix $A = [\alpha_{r,u}]$ where $\alpha_{r,u}$ represents the adjustment of the uth trolley to the rth robotic center. Moreover, $\alpha_{r,u} = 1$ if the uth trolley is entitled to supply the rth robotic center, $\alpha_{r,u} = 0$ otherwise. Let the matrix $B = [\beta_{l,u}]$ represents the adjustment of trolleys equipped with tanks to the base tanks from which molten aluminum is poured where $\beta_{l,u}$ represents the adjustment of the lth trolley to the uth base tank. Moreover, $\beta_{l,u} = 1$ if the uth trolley is entitled to be supplied by the lth base tank, $\beta_{l,u}$=0 otherwise.

It is possible to specify the state of trolley tanks by means of the matrix G^{trolley} where the variable $g_{u,k}^{\text{trolley}}$ represents the capacity of the *u*th tank at the *k*th stage expressed in liters and it is changed according to the form (4) where $x_{u,k}^{\text{trolley}}$ is the amount of the liquid aluminum removed from the *u*th trolley tank at the *k*th stage expressed in liters.

$$g_{u,k}^{\text{trolley}} = \begin{cases} g_{u,k}^{\text{trolley}} - x_{u,k}^{\text{trolley}} & \text{if there is a manufacturing decision,} \\ g_{u,k-1}^{\text{trolley}} & \text{otherwise.} \end{cases} \tag{4}$$

The maximum capacity of trolley tanks can be defined by the vector $G^{\text{trolley(max)}}$ where variable $g_u^{\text{trolley(max)}}$ represents the maximum capacity of the *u*th tank (in liters). Consequently, trolley tanks are characterized by their minimum capacity and are defined by the vector $G^{\text{trolley(min)}}$ where the variable $g_u^{\text{trolley(min)}}$ represents the minimum capacity of the *u*th tank (in liters).

Similarly, the state of the *r*th robotic center tank can be specified in the matrix G^{centre} where the variable $g_{r,k}^{\text{centre}}$ represents the capacity of the *r*th robotic center tank (heating stove) at the *k*th stage expressed in liters. This variable changes according to the form (5) where $x_{r,k}^{\text{centre}}$ is the amount of the liquid aluminum removed from the *r*th tank of the robotic centre due to the manufacturing decision at the *k*th stage expressed in liters.

$$g_{r,k}^{\text{centre}} = \begin{cases} g_{r,k}^{\text{centre}} - x_{r,k}^{\text{centre}} & \text{if there is a manufacturing decision,} \\ g_{r,k-1}^{\text{centre}} & \text{otherwise.} \end{cases} \tag{5}$$

The maximum capacity of robotic centre tanks is defined by the vector $G^{\text{centre(max)}}$ where the variable $g_r^{\text{centre(max)}}$ represents the maximum capacity of the *r*th robotic centre tank (in liters). Consequently, the minimum capacity of robotic centre tanks is defined by the vector $G^{\text{centre(min)}}$ where the variable $g_r^{\text{centre(min)}}$ represents the minimum capacity of the *r*th robotic centre tank (in liters).

3.2. Transport times

The filling times for trolley tanks can be defined in the matrix T^{trolley} where the time $t_{l,r}^{\text{trolley}}$ represents the filling time of the *u*th trolley tank from the *l*th base tank. Similarly, the filling times for robotic centre tanks are defined by the matrix T^{centre} where the time $t_{u,r}^{\text{centre}}$ represents the filling time of the *r*th robotic centre tank from the *u*th trolley tank along the certain route.

In case there is another transportation tank which is to carry out a transport duty, it is obliged to wait for the tank occupying this route to free it. The awaiting place is determined by the place of the already completed operation. Transportation routes allow only one trolley tank to carry out transport operations.

The transport times of trolleys are defined also by the matrices. The transport times from base tanks to robotic centers is defined by the matrix $T(u)^{\text{tr}}$ where the time $t(u)_{l,r}^{\text{tr}}$ represents the transport time of the load of aluminum from the *l*th base tank to the *r*th robotic centre tank by means of the *u*th trolley tank. Consequently, the transport times from robotic centers to base tanks take the values according the matrix $T(u)^{\text{trback}}$ where the time $t(u)_{r,l}^{\text{trback}}$ represents the transport time of the *u*th trolley tank from the *r*th robotic centre tank to the *l*th base tank.

3.3. Production times

The production times in robotic centers take the values according the matrix T^{pr} where the time $t_{n,r}^{\text{pr}}$ represents the production time of the *n*th order in the *r*th robotic centre. The robot programming times in robotic centers are defined by the matrix T^{inf} where the time $t_{n,r}^{\text{inf}}$ is the robot programming time for the *n*th order in the *r*th robotic centre. A robot operating in the working centre is always allowed by the time which lets us collect a product (casting) and this is the reason why there is no possibility for it not to collect the product. Then the matrix T^{pr} meets the current needs of the manufacturing system.

The maintenance is scheduled for each robotic centre. The times of maintenance in robotic centers are defined by the vector T^{main} where the time t_r^{main} represents the maintenance time for the *r*th robotic centre. The

times for scheduling maintenance operations are assumed in the vector T^{sch} where the time t_r^{sch} is the time of scheduling maintenance operations for the *r*th robotic centre.

The temperature of liquid aluminum is maintained automatically in base tanks within defined values. The temperature of liquid aluminum in trolley tanks is not maintained so time limits are specified for the time of storing the charge material. The allowable storing times of charge material in trolley tanks can be defined by the vector $T^{\max}$ where the time $t_u^{\max}$ represents the maximal time of storing the charge material in the *u*th trolley tank according to its current state.

The scale is divided by each conventional unit for every trolley tank capacity. The trolley tank must be filled from the base tank if the given amount in the *u*th trolley tank at the *k*th stage $g_{u,k}^{\text{trolley}}$ cannot be placed in the *r*th robotic centre within the given time. Then the dedicated trolley tank waits for the subsequent possibility to be filled i.e. for the moment when the value $g_r^{\text{centre(min)}}$ is reached.

3.4. Available capacity

According to the above, the available capacity of the base tanks, trolleys and robotic centre tanks can be calculated. The available capacity of *l*th base tank can be calculated according to the form (6). Consequently, the available capacity of the *u*th trolley tank is obtained from the equation (7) and the available capacity of the *r*th tank of the robotic centre can be calculated according to the equation (8).

$$p_{l,k}^{\text{tank}} = g_l^{\text{tank(max)}} - g_{l,k}^{\text{tank}},\ l=1, 2, \ldots, L;\ \ k=1, 2, \ldots, K; \tag{6}$$

$$p_{u,k}^{\text{trolley}} = g_u^{\text{trolley(max)}} - g_{u,k}^{\text{trolley}},\ u=1, 2, \ldots, U;\ \ k=1, 2, \ldots, K; \tag{7}$$

$$p_{r,k}^{\text{centre}} = g_r^{\text{centre(max)}} - g_{r,k}^{\text{centre}},\ r=1, 2, \ldots, R;\ \ k=1, 2, \ldots, K; \tag{8}$$

If the condition in the form (9) is valid then it is possible to fill the tank of the *r*th robotic centre. Otherwise, the *u*th trolley tank must be filled.

$$p_{r,k}^{\text{centre}} \geq g_{u,k}^{\text{trolley}} \quad \vee \quad p_{r,k}^{\text{centre}} \geq g_{u,k}^{\text{trolley}} - g_u^{\text{trolley(min)}} \tag{9}$$

4 Heuristic control algorithms

Control algorithms are responsible for determining the order to be manufactured and the trolley tank to be used for transporting aluminum. This part of article presents the division of proposed heuristic algorithms to control the discussed manufacturing process.

In order to carry out practical simulation procedures the most vital thing becomes the need for implementing strategies on the basis of which the following heuristic algorithms are based:

1. *The strategy with automatic adjustment* - During making orders in casting centers there is a possibility to adjust a machine to another product if the preceding order is fully made. The choice of this strategy enables us to implement all machines for each of three product types.
2. *The strategy without automatic adjustment* - During making orders and after completing them in casting centers there is no possibility to adjust a machine to another product. The choice of this strategy defines certain machines to adequate products.

The created simulation tool has been built on the basis of the specification assumptions included in this article and implements the following practical algorithms:

- *The heuristic of the minimum state of liquid aluminum in the heating stove* - This heuristic chooses a centre to be served with the given state of the aluminum level which should be safe for the manufacturing process. The safe minimum state prevents manufacturing faulty products and reduces costs of failures which can result from avoiding this kind of action.
- *The heuristic of the fixed state of aluminum in the heating stove* - This heuristic chooses a centre to serve on condition it meets criteria set by the user however; the minimum state of aluminum in the heating stove cannot fall below 2500 kg.
- *The heuristic of the furthest center* - This heuristic chooses a machine whose delivery time is the longest. While servicing centers, first of all, the minimum state of the heating stove and, only then, the delivery time to the center is verified as well.

- *The heuristic of the nearest center* - This heuristic chooses a machine whose delivery time is the shortest. While servicing centers, first of all, the minimum state of the heating stove is verified and, only then, the delivery time to the center is verified as well.
- *The random choice of centers* - The casting centers are chosen at random on condition that the minimum state of aluminum in the heating stove is kept.
- *Random choice of heuristics* - This course of action chooses casting centers by means of all heuristics at random excluding the heuristic of the fixed state of aluminum in the heating stove on condition that the minimum state of aluminum in the heating stove is kept.
- *The heuristic of the maximum order* - First, this heuristic chooses centers characterized by the biggest order without taking into account the minimum state of aluminum in the stove. As a result, manufacturing centers with smaller orders wait for completing manufacturing bigger orders in other centers.
- *The heuristic of the minimum order* - First, this heuristic chooses centers characterized by the smallest order without taking into account the minimum state of aluminum in the stove. As a result, manufacturing centers with bigger orders wait for completing manufacturing smaller orders in other centers.
- *Random choice of orders* - Centers for manufacturing orders are chosen at random without taking into account the minimum state of aluminum in the stove.

5 Conclusions

The paper highlights the problem of mathematical modeling of manufacturing processes taking place in the highly sophisticated logistic system. Its real state is represented by a manufacturing hall equipped with robotic centers which carry out pre-programmed operations. The charge material is supplied in the form of liquid aluminum and stored in base tanks which are supplied in time by an outsourcing company. Charge material is transported from the base tanks to robotic centers by trolleys with built-in tanks. Ready products (e.g. engine blocks) are moved to the storing area located next to the adequate centre. Customers' orders are to be made at optimized costs in time. Therefore, the main criterion remains to minimize production costs within the existing manufacturing and technological constraints. The main goal of mathematical modeling is to support decision making as the course of production cannot be alternated once it is set. The real model is brought to the form of the synthetic environment for which necessary specification assumptions are made. The system is controlled by heuristic algorithms. Equations of state illustrate changes throughout the manufacturing process.

Acknowledgements

This work was supported in part by the European Regional Development Fund within the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

References

[1] Bucki, R. Chramcov, B., Jašek, R. Modelling the Complex Logistic System with Regeneration Plants. *In: Mathematical Modelling in Logistics - Decision Making Processes.* Faculty of Management at the University of Prešov in Prešov, Grafotlač Prešov, 2013, pp. 36-69, ISBN 978-80-555-0824-5.

[2] Bucki, R., and Marecki, F. Multi-Stage Discrete Programming of the Production Line. *Communications* **1** (2013), 62-67.

[3] Bucki, R. Logistic Decision-Making in Modelling and Control of the Complex Production System. In: *Antonyová, A. et al. Mathematical Modelling in Logistics – Optimum Control.* Faculty of Management, University of Prešov in Prešov, Grafotlač Prešov, 2012, pp. 51-62, ISBN 978-80-555-0602-9.

[4] Bucki, R., and Suchánek, P. The Method of Logistic Optimization in E-commerce. *Journal of Universal Computer Science 18*, (2012), 1238-1258.

[5] Chramcov, B., Bucki, R., and Beran, P. Lean Manufacturing System Design Based on Computer Simulation: Case Study for Manufacturing of Automotive Engine Control Units. In: *Modrák, V., Semančo, P. Design and Management of Lean Production Systems*, IGI Global, 2014.

[6] Suchánek, P., Vymětal, D., and Bucki, R. Decision-Making in the Context of E-Commerce Systems. In: *Laura Ştefănescu. Intelligent Decision Support Systems for Managerial Decision Making.* ASERS Publishing House, 2011, 49-88.

[7] Modrák, V., and Semančo, P.: *Design and Management of Lean Production System*. IGI Global, Hershey, 2014.

[8] Kuka industrial robots. *Handling of die cast parts* [online]. ©2015 [cit. 2015-05-02]. Available from: http://www.kuka-robotics.com/en/solutions/solutions_search/L_R247_Handling_of_die_cast_parts.htm

Factors of Czech market share growth in European Union

Petra Čekmeová[1]

Abstract. As the Czech Republic is continuously improving its market position in European Union, an interesting question arises. What are the forces behind its raising market share? This paper provides a decomposition of Czech market share growth to competitiveness and structural effects with aim to quantify the contribution of individual effects to its increasing market share in European Union. We use Constant Market Shares analysis, using both analytical and econometric approach. The analysis is executed on 25 member states of European Union. Time series consist of yearly observation on country-level imports from Czech Republic and European Union, from 1999 to 2014. We follow a model using Törnqvist index systems with symmetric weights. The results suggest that the Czech Republic's increasing portion in European Union's imports can be attributed mainly to improvements in its competitiveness and partially to geographical distribution of its exports. Commodity composition of the exports has generally negative effect. These outcomes are verified by econometric version of the analysis applying weighted OLS with fixed effects.

Keywords: Constant Market Share Analysis, Czech Republic, Competitiveness.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Export performance of the country always attracts a lot of attention among economists and policy makers. Particularly, the country's competitive position is often highlighted in political debates. If the country is able to maintain or improve its market share in foreign markets, it will positively affect its export performance. And in turn, improved export performance has positive impact on the nation's economic prosperity, mainly in the case of small open economies such as the Czech Republic.

Market share of the Czech Republic in the European Union follows rising trend. Even during the Global Financial Crisis, its export performance has not been easing up. Its share in imports of 25 member states of the European Union (EU25)[2] has increased by almost 2.5 percentage points over a period of 2000-2014. Notice that only three member states enjoyed an increase in their market shares by more than 2 percentage points during this period. As the Czech Republic is continuously improving its market position, an interesting question arises: "What are the forces behind this continuous improvement of the Czech market share in the EU25?"

Generally, a country can improve its position in foreign markets thanks to favourable structure of its export flows as well as rising competitiveness. One of the popular methods of applied international economics, called Constant Market Shares analysis (CMSA), is used to identify forces behind the country's export growth. It isolates an effect of pure competitiveness from structural factors. There exist a variety of models to conduct an analytical version of CMSA, but even less studies dealing with an econometric version of this analysis. The main idea of the decomposition is that the country which exports to markets with increasing share in world imports, specialises to highly demanded commodity groups or enjoys high level of external competitiveness tends to have rising market share. Moreover, we can investigate the sources of external competitiveness as the analysis provides detailed structural decomposition of individual effects. Despite this, there has not been any academic study which would deal with the decomposition of the Czech export performance using CMSA until now.

The purpose of this paper is to *find out the factors of increasing market share of the Czech Republic in the European Union.* The paper provides a decomposition of Czech market share growth to competitiveness and structural effects with aim to quantify the contribution of individual effects to its increasing market share, applying both analytical and econometric version of CMSA.

The paper is organized as follows. After a short introduction, the second section introduces a methodology applied in this study, including description of data used in the empirical analysis. The third section presents the

[1] Masaryk University, Department of Economics, Lipová 41a, Brno, Czech Republic, e-mail: 441117@mail.muni.cz.
[2] Bulgaria, Croatia and Romania were excluded from the analysis regarding their relatively short membership in EU and availability of data required for the analysis.

results of the CMSA carried out on the Czech Republic's market share growth in EU25. The last section contains concluding remarks summarizing the main results of our analysis.

2 Methodology and data

2.1 Analytical version of the Constant market shares analysis

Since a pioneering study of Tyszynski (1951) serious methodological debate has occurred over time[3]. Thus, there is a variety of methods to conduct CMSA. In the analytical approach, we follow a model proposed by Milana (1988) standing for its solution for an index problem. It is an adjusted version of that used by Fagerberg and Sollie (1987). Applying index number theory, Milana came to conclusion that the use of superlative index numbers had to be preferred to Paasche or Laspayers indices. The weights used in decomposition are symmetric based on the assumption that the export structure of the initial (final) year should not be dominant over the whole period. The decomposition is executed by application of Törnqvist-type indexes, where elementary components with higher proportional variations are aggregated by direct superlative indices and the rest of the components are derived implicitly.

The model is as follows:

$$\Delta s = \sum_m \frac{1}{2}[M_m^{t-1} + M_m^t]\Delta c_m + \sum_c \sum_m \frac{1}{2}[s_{cm}^{t-1} + s_{cm}^t]\Delta \frac{Q_{cm}}{\sum_c Q_{cm}} \frac{1}{2}[c_m^{t-1} + c_m^t] + \sum_c \sum_m \Delta s_{cm} \frac{1}{2}\left[\frac{Q_{cm}^{t-1}}{\sum_c Q_{cm}^{t-1}} + \frac{Q_{cm}^t}{\sum_c Q_{cm}^t}\right]\frac{1}{2}[c_m^{t-1} + c_m^t] \quad (1)$$

where

$$M_m = \sum_c s_{cm}\left[\frac{Q_{cm}}{\sum_c Q_{cm}}\right],\ c_m = \frac{\sum_c Q_{cm}}{\sum_c \sum_m Q_{cm}},\ s_{cm} = \frac{q_{cm}}{Q_{cm}}.$$

In the model, $q_{c,m}$ denotes Czech exports of the c-th commodity to the m-th market, $Q_{c,m}$ denotes EU25 export of c-th commodity to m-th market, M_m denotes market share of the Czech Republic in m-th market, C_m denotes share of m-th market's imports in total world import and $s_{c,m}$ denotes export share of Czech Republic in m-th market's imports of c-th commodity.

The scheme (1) allows us to decompose a change of the focus country's market share to "market-distribution" (first row), "commodity-composition" (second row) and "competitiveness" (third row) effect, respectively. The interpretation of these effect is the following: the first component describes the growth in focus country's market share thanks to increasing shares of its exports markets in world import; the second measures a contribution of increase in shares of given commodity group in each markets and the competitiveness effect, reflects the pure performance of the focus country (market gains beside favourable structural patterns).

In comparison with the majority of the works concentrating on the export growth between initial and final year, we present a model where all consecutive years of the given time period are analysed. Shorter the time period analysed more accurate is the Törnqvist discrete time approximation (Milana, 1988). It is an additional reason why we apply the model (1).

The empirical analysis is executed on 25 member states of the European Union, based on data from the Eurostat. The Czech Republic is selected as a focus country and the other members are treated as standard in the analysis. The choice of standard follows from the export structure of Czech Republic which is highly focused on European markets as well as its membership in this regional trading block. Time series consist of yearly observation on imports of individual member countries from Czech Republic and European Union from 1999 to 2014. Only import flows entered to analysis to avoid inconsistency between aggregated import and export flows. In all, 7840 observations were analysed. The data had been collected for one-digit level of SITC (rev.3) comprising 10 commodity groups for each country and then aggregated according to scheme (1).

[3] For more details see f.e. Svennilson (1954), Baldwin (1958) , Spiegelglas (1959), Leamer and Stern (1970), Richardson (1971a, 1971b), Fagerberg and Sollie (1987), Merkies and van der Meer (1988), Milana (1988), Ahmadi-Esfahani (2006),

2.2 Econometric version of constant market share analysis

To decompose the growth rate of Czech market share in EU25 into structural and competitiveness effects, we apply a methodology of Fontagné (2010) which relies on econometric model proposed by Chaptea (2005).[4] The empirical analysis, as before, is performed on 25 member states of the EU (Czech Republic as focus country and other members as standards). The data contain yearly observation on country-level exports of 10 commodity groups (SITC, rev.3) from 1999 to 2014.

To capture variations over time we focus on annual growth rates of each trade flow rather than on the change between initial and final year of the analysed period. The analysis is provided in three steps. At first, growth rate of Czech exports to each importer for a given commodity group and year are explained. Then, the annual results are aggregated at the level of exporter. Finally, we transpose the decomposition of export growth into decomposition of market share growth. As for the software, we use GRETL.

The contribution of competitiveness and structural factors are estimated via weighted (by w_{lmp}^t) OLS with fixed effects where the growth rates of each individual trade flow are regressed on dummies for exporter (l), importer (m) and commodity groups (p). The regression model is as follows:

$$ln\left(\frac{X_{lmp}^t}{X_{lmp}^{t-1}}\right) = \alpha_l^t + \beta_m^t + \gamma_p^t + \varepsilon_{lmp}^t \tag{2}$$

where X_{lmp}^t denotes the value of Czech exports of given commodity group to given importer in time t, α_l^t represents the amount of growth which can be attributed to the competitiveness of the Czech Republic in time t, β_m^t and γ_p^t catch up the contribution of average market and commodity structure of European trade to the annual growth of Czech exports in time t. The weight, $w_{lmp}^t = 0{,}5\left(X_{lmp}^{t-1}/X^{t-1} + X_{lmp}^t/X^t\right)$, denotes the average weight of Czech export flow in the EU25 trade. As the estimation is executed for each consecutive year of the period 2000-2014, we get 15 annual effects for the Czech Republic, each commodity group and for almost each importer (one country is excluded due to exact multicollinearity). Comparing to the analytical version of the CMSA, an econometric analysis can be used to predict statistical significance of the individual effects by estimating the corresponding standard errors.

The growth rate of Czech exports is given by the logarithm of the Törnqvist index of its exports of each commodity group (p) to each importer (m):

$$dlnX_l^t = ln\left(\frac{X_l^t}{X_l^{t-1}}\right) \approx \sum_{mp} \frac{w_{lmp}^t}{w_l^t} ln\left(\frac{X_{lmp}^t}{X_{lmp}^{t-1}}\right) \tag{3}$$

Using (2) and (3) and normalizing the estimated effects, $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$, we can express the contribution of competitiveness and structural factors to the overall growth rate of Czech exports in the following form:

$$ln\left(\frac{X_l^t}{X_l^{t-1}}\right) = \bar{\alpha}_l^t + \sum_m \frac{w_{lm}^t}{w_l^t} \bar{\beta}_m^t + \sum_p \frac{w_{lp}^t}{w_l^t} \bar{y}_p^t \tag{4}$$

where $\bar{\alpha}_l^t = \hat{\alpha}_l^t + \sum_m w_m^t \hat{\beta}_m^t + \sum_p w_p^t \hat{\gamma}_p^t$, $\bar{\beta}_m^t = \hat{\beta}_m^t - \sum_m w_m^t \hat{\beta}_m^t$ and $\bar{y}_p^t = \hat{\gamma}_p^t - \sum_p w_p^t \hat{\gamma}_p^t$. The right-hand side elements of equation (4) represents (successively): competitiveness effect, market-distribution effect and commodity-composition effect. Normalization of estimated effects is required because $\hat{\alpha}_l^t$ measures the fraction of Czech export growth relative to the importer omitted in the regression (7). To obtain competitiveness effect, independent on the choice of omitted country, we calculate least square mean ($\bar{\alpha}_l^t$) and normalize the remaining ($\bar{\beta}_m^t$, $\bar{y}_p^t$), too.

To transpose the decomposition of annual export growth into decomposition of market share growth, we subtract from both left and right-hand side of the equation (4) the logarithmic change in EU25 exports in time t ($\mathrm{dln}X^t = lnX^t/X^{t-1}$) and take the exponentials of the resulting expressions, resulting in:

$$g_i^t \equiv \exp(dlnX_l^t - dlnX^t) - 1 = (1 + CE'_l) * (1 + MDE'_l) * (1 + CCE'_l) - 1 \tag{5}$$

where $CE'_l = \exp(\bar{\alpha}_l^t - dlnX^t) - 1$, $MDE'_l = \exp\left(\sum_m \frac{w_{lm}^t}{w_l^t} \bar{\beta}_m^t\right) - 1$, $CCE'_l = \exp\left(\sum_p \frac{w_{lp}^t}{w_l^t} \bar{y}_p^t\right) - 1$.

[4] Chaptea´s model is an improvement of the weighted variance analysis of growth rates of Jayet (1993).

3 Empirical results

3.1 Decomposition of Czech market share growth – analytical approach

The results of CMSA, using analytical approach, are reported in the Table 1. The annual values of three effects (CCE, MDE, CE) represent their contribution to year-on-year changes of the Czech market share in EU25 (Δ s) in percentage points.

year	CCE	MDE	CE	Δ s
2000	0,003	-0,017	**0,121**	0,107
2001	0,005	0,0282	**0,176**	0,2092
2002	-0,008	-0,014	**0,153**	0,131
2003	-0,021	0,048	**0,101**	0,128
2004	-0,005	0,031	**-0,035**	-0,009
2005	-0,036	0,017	**0,216**	0,197
2006	-0,016	0,036	**0,184**	0,204
2007	0,022	0,062	**0,239**	0,323
2008	-0,063	0,055	**0,277**	0,269
2009	-0,020	0,021	**0,124**	0,125
2010	-0,010	0,031	**0,155**	0,176
2011	-0,029	0,040	**0,146**	0,157
2012	-0,048	0,004	**0,069**	0,025
2013	-0,012	0,029	**-0,007**	0,010
2014	0,038	0,019	**0,155**	0,212

Table 1 Decomposition of Czech market share growth in the EU25 [in % point]

As reported in the first column of the Table 1, commodity-composition effect had negative sign almost in the whole period and consequently decreased the Czech market share by reported percentage points. It implies that the Czech Republic is oriented on commodities which share in European imports stagnate or decline having in turn negative effect on its market share growth. The two latest years project from this long-term trend, but we except that it is only short run deviations as the structure of its exports has not changed significantly. Substantially different is the situation in the case of second structural effect. Market distribution of Czech exports has caused partial decline in its market share only in two years (2000, 2002) and in rest of the analyzed period positively contributed to year-on-year changes. It reflects that the portions of Czech republic´s trading partners in world import have developed favorably. As reported in the fourth column of the Table 1, the most significant contribution to the export share growth of Czech Republic can be assign to sources beside structural factors with the highest contribution in 2008. Competitiveness effect has shown slightly negative growth only in 2004 and 2013. Both years were accompanied by negative or stagnant market share growth which indicates a strong positive correlation between competitiveness and export performance of Czech Republic.

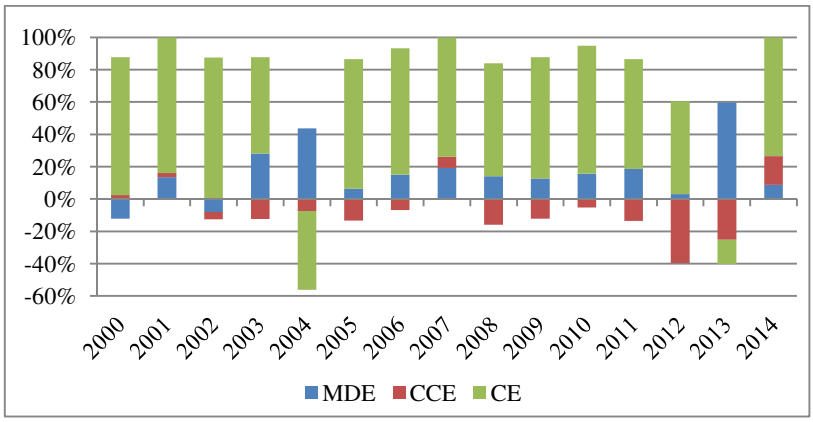


Figure 1 Contribution of individual effects to the market share growth of the CR in the EU25

Graph 1 summarizes the contribution of individual effect to market share gains over the analysed period. It is obvious that almost in the whole period; positive competitiveness effect caught up the largest portion of the market share growth. Its average contribution from 2000 to 2014 was 114.10%, while the market-demand effect contributed in average by only 8.73%. The commodity-composition effect influenced Czech export performance negatively with average contribution of -22.4%. Thus we can conclude that the driving force of Czech market share growth lies mainly in its external competitiveness.

3.2 Decomposition of Czech market share growth – econometric approach

The outcomes of econometric version of CMS analysis are reported in the Table 2 and Figure 2 where we can see the contribution of competitiveness, market-distribution and commodity composition effects to the growth rate of the Czech market share in the European Union (EU25).

year	CCE´	MDE´	CE´	g
2000	-0,65%	-3,60%	**10,83%**	6,15%
2001	4,49%	-6,52%	**16,88%**	14,16%
2002	0,33%	-2,49%	**10,10%**	7,70%
2003	0,45%	1,39%	**4,94%**	6,88%
2004	-5,24%	-0,80%	**27,67%**	20,02%
2005	-1,89%	-2,88%	**8,65%**	3,53%
2006	1,41%	-2,21%	**8,22%**	7,32%
2007	0,75%	-0,56%	**9,98%**	10,19%
2008	1,51%	0,63%	**5,21%**	7,47%
2009	-2,87%	0,86%	**4,20%**	2,08%
2010	0,59%	0,67%	**4,95%**	6,29%
2011	0,09%	0,55%	**3,46%**	4,13%
2012	4,26%	1,22%	**-8,27%**	-3,20%
2013	0,40%	0,79%	**-1,86%**	-0,69%
2014	0,61%	0,30%	**5,10%**	6,05%

Table 2 Decomposition of growth rate of the Czech market share in the EU25

The last column of the Table 2 presents annual growth rates of Czech market share in EU25 calculated for each year of the period 2000-2014, in a multiplicative form (based on equation (5)). The first and second columns of the Table 2 display the growth rates of Czech market share due to commodity and market structure of its exports. Compared to results of the previous analysis, the impact of commodity-composition of the Czech exports is not predominantly negative. On the contrary, the effect of market distribution is mostly negative in the first part of the analysed period and strictly positive since 2008. However, we can see that the contribution of these structural factors are only marginal compared to the competitiveness effect, which is displayed in the fourth column. Moreover, it is obvious that the effect of pure competitiveness on annual growth rates of the Czech market share in the EU25 is positive during almost the whole analysed period. Only in 2012 and 2013 we notice negative competitiveness effects which in turn resulted in negative growth rate of Czech market share. Thus, we can sum up that the main factor of the market growth rate of the Czech Republic can be contributed to its external competitiveness.

4 Conclusion

Since the beginning of the 21st century, the Czech Republic has considerably improved its market position in the European Union. Over the period of 15 years (1999-2014), only three member states (including the Czech Republic) enjoyed an increase in their market shares by more than 2 percentage points. The purpose of this paper was to find out the factors behind the improving market position of the Czech Republic in the European Union.

The paper provided a decomposition of the Czech market share growth in 25 member states of the European Union over the period of 1999-2014 to competitiveness and structural effects. The decomposition, with aim to quantify the contribution of structural factors and pure competitiveness to annual market share changes, was executed via Constant Market Shares analysis. We applied both analytical and econometric versions of the method. Thus, in addition, the paper provided an empirical application of two different approaches of the Constant Market Shares analysis for the case of the Czech Republic.

The results of the Constant market share analysis suggest that the competitiveness effect had significantly positive effect on the Czech market share in the European Union's imports, while the geographical distribution of its exports contributed to the overall change only partially. On the other side, the commodity composition effect had negative impact on the development of the Czech market share almost in the whole period. The effects of structural factors little bit differ using an econometric approach, but their contribution to the growth rate of the Czech market share in the European Union rest marginally.

The general conclusion from this study is that the improvement in the relative position of the Czech Republic in the European market is driven mainly by its external competitiveness and not by structural shifts in its export flows. Moreover, it implicitly leads to a conclusion that the driving force behind the favourable development of the Czech market share in the European Union lies in factors which determine its external competitiveness.

Acknowledgements

Supported by the specific research project No. MUNI/A/1223/2014 at Masaryk University.

References

[1] Ahmadi-Esfahani, F. Constant Market Shares Analysis: Uses, Limitations and Prospects. *The Australian Journal of Agricultural and Resource Economics* **50** (2006), 510-526.
[2] Baldwin, R.E.: The commodity composition of world trade: selected industrial countries 1900-1954. *Review of Economics and Statistics* **40** (1958), 50-71.
[3] Chaptea, A., Gaulier, G.: World trade competitiveness: A disaggregated View by Shift-Share Analysis. *CEPII Working paper* **23** (2005), 51 p.
[4] Fagerberg, J. and Sollie, G. The Method of Constant-Market-Shares Analysis Revisited. *Applied Economics* **19** (1987),1571-1584.
[5] Fontagné, L., Chaptea, A., and Zignano, S.: European Export Performance. *CEPII Working paper* **12** (2010), 38 p.
[6] Jayet, H.: *Analyse spatial quantitative: une introduction.* Economica, Paris, 1993.
[7] Leamer, E.E. and Stern: *Quantitative International Economics.* Allen & Bacon, Boston, 1970.
[8] Merkies, A.H.Q.M. and van der Meer, T.: A theoretical Foundation for Constant Market Shares Analysis. *Empirical Economics* **13** (1988), 65-80.
[9] Milana, C.: Constant-Market-Shares Analysis and Index Number Theory. *European Journal of Political Economy* **4** (1988), 453-478.
[10] Richardson, J.D.: Constant-market-shares analysis of export growth. *Journal of International Economics* **1** (1971a), 227-239.
[11] Richardson, J.D.: Some Sensitivity Tests for a "Constant-Market-Shares" analysis of Export growth. *The Review of Economics and Statistics* **53** (1971b), 300- 304.
[12] Skriner, E.: Competitiveness and Specialisation of the Austrian Export Sector – A Constant-Market-Shares Analysis. *FIW Working Paper* **32** (2009), 22 p.
[13] Spiegelglas, S.: World exports of manufactures, 1956 vs. 1937. *The Mancheste School* **27** (1959), 111-139. SVENNILSON, I.: *Growth and Stagnation in the European Economy.* U.N. Economic Commission for Europe, Geneva, 1954.
[14] Tyszynski, H.: World Trade in Manufactured Commodities, 1899-1950. In: *The Manchester School of Economics and Social Studies* **19** (1951), 272-304.

Extended Model of Tourist Routes Optimization

Anna Černá[1], Jan Černý[2], Vladimír Přibyl[3]

Abstract. The paper starts with the following problem concerning the design of tourist routes (trails) for hiking or biking or horse riding or cross-country skiing (one of them, not more): Given a "candidate" network of (mostly) field and forest unpaved roads. The network is represented by an undirected graph. Each edge of the graph is characterized by its "length" (in km or in min.) and by the design cost necessary to recondition the track. Moreover, each edge and each vertex have a list of tourist points of interests (PoI) like waterfalls, memorials etc., occurring at it. Each PoI is evaluated by its attractiveness score (AtS). The problem is to find the feasible and the most attractive route connecting two given points.
The paper shows how to define the AtS for an individual PoI and how to derive the total AtS for a given tourist route, connecting two vertices of the graph. The main purpose of the paper is to present a mathematical model enabling to find a tourist route, connecting two given vertices, not exceeding the limits of cost and length and maximizing its AtS.
The model is extended in two senses. First, comparing to the "classic" minimization of costs, it incorporates the new objective "attractiveness" of the route. Second, it extends the class of admissible objective functions. So far, only additive functions, with respect to the edges and vertices, were considered, whereas the non-additive ones are admitted in the model as well.

Keywords: Tourist route, points of tourist interest, attractiveness score, model, optimization.

JEL Classification: C61, H76, O18, R58
AMS Classification: 05C35, 90C35

1 Introduction

This paper is focused on the design of optimal tourist routes (trails) in a given "candidate" network of (mostly) field and forest unpaved roads. The results can be applied mainly to hiking, biking, horse riding or cross-country skiing routes optimization. In the sequel, the route is supposed to be simple, i.e. to pass any its internal node only once. The origin and the destination of the route, however, may be different or identical.

The network is represented by a connected undirected graph $G = (V, E, d, c)$ where the vertex set V is defined as $V = \{1, 2, \ldots, n\}$, $n > 2$, $E \subset V \times V$ is the edge set, the function d: $E \to \langle 0; \infty)$ represents length of edges, the function c: $E \to \langle 0; \infty)$ represents the design costs necessary to recondition of the track. Naturally, it is supposed that the triangle inequality holds for the function d and, therefore, the function d can be extended to the set $V \times V$ where it expresses the distance between the pairs of vertices. Moreover, it can be extended to the set R of all routes defined on G in such a sense that the length $d(e)$ is counted in $d(r)$ if e is included into r. The function c can be similarly extended to the set R as well.

The "classic" problem is to find a route connecting two given vertices v_o (the origin) and v_d (the destination) by a path r, having the properties that the length $d(r)$ of r belongs to the given interval $\langle d_1; d_2\rangle$ and the cost $c(r)$ of r is minimal.

This paper represents two extensions of the classic problem. First, it introduces new function defined as on V, as on E. The values $a(v)$ and $a(e)$ represent "attractiveness score" of the vertex v or the edge e respectively. Second, the attractiveness $a(r)$, as an objective function, need not be additive with respect to the vertices and edges belonging to r. The *extended problem* may be formulated in several versions, e.g. as follows:

EP1. Given a connected undirected graph $G = (V, E, d, c)$, vertices v_o (the origin) and v_d (the destination), a function $a(.)$ defined on the set $R(v_o, v_d)$ of all routes connecting vertices v_o and v_d and, finally, given limits d_0 and c_0 for the length and cost respectively. The problem is to find a route $r \in R(v_o, v_d)$ s.t. $d(r) \le d_0$, $c(r) \le c_0$, $a(r) \to \max$.

[1] University of Economics, Prague/Faculty of Management, Jindřichův Hradec, cerna@fm.vse.cz.
[2] University of Economics, Prague/Faculty of Management, Jindřichův Hradec, cerny@fm.vse.cz.
[3] University of Economics, Prague/Faculty of Management, Jindřichův Hradec, pribyl@fm.vse.cz.

EP2. Given a connected undirected graph $G = (V, E, d, c)$, vertices v_o (the origin) and v_d (the destination), a function $a(.)$ defined on the set $R(v_o, v_d)$ of all routes connecting vertices v_o and v_d and limits d_0 and a_0 for the length and attractiveness respectively. The problem is to find a route $r \in R(v_o, v_d)$ s.t. $d(r) \leq d_0$, $a(r) \geq a_0$, $c(r) \rightarrow$ min.

1.1 State of the Art

The theory concerning "classic" problem of shortest paths in so called "doubly weighted" graphs can be found e.g. in the subchapter 4.D of [8], where one can found that it was studied since sixties and that it was proved that minimization of length at constrained duration is NP-hard.

On the other hand, the attractiveness as a quantitative characterization of a tourist route is quite new, for the first time it was mentioned in [2] in 2012. One can see some distant relation of bicycle route attractiveness to "Bicycle Level of Service", mentioned by [6] in 1997. This quantitative measure tries to express the cyclist's feeling of safety. Similar view can be found in [7] from 2013, where one can find a list of some objectives connected with safety and a cyclist's comfort. Other papers, like [4], [10] and [13] added the travel time or the velocity) as another point and [11] added the pavement quality. Further, [9] or [5] emphasized influence of cycling on human health.

As seen, the scientific results concerning non-motorized transport were oriented almost exclusively to bicycling and specifically to its use to commuting, see e.g. [1]. In [12] the application is extended to walking as well and the factor of air quality is accented. Moreover, [10] touched also leisure cycling and spoke about a terrain for recreational use and this is the point of view of the current paper as well.

1.2 Tourist Point of Interest (PoI): Attractiveness Quantification

The attractiveness value (score) $a(p)$ of a tourist point of interest p, briefly PoI (e.g. a statue, waterfall, beautiful avenue of trees or watchtower) expresses the intensity of positive feeling of a tourist enjoying p.

1.3 Basic Attractiveness Quantification of Vertices and Edges

First the denotation $P(v)$ is defined as the set of all PoI's that belong to the vertex v and $P(e)$ is similarly defined for the edge e, for all $v \in V$ and $e \in E$. Then, for all $v \in V$, the *basic attractiveness* $ba(v)$ is defined as the sum of all $a(p)$, $p \in P(v)$ and $ba(e)$ is similarly defined for all $e \in E$.

1.4 Attractiveness Quantification of Routes

Additive Attractiveness of Routes

Basic attractiveness function can be extended to the set R of all routes in G similarly as the length function d: The attractiveness $ba(v)$ or $ba(e)$, respectively, is counted in $ba(r)$ if v or e is included in r. Consequently, the basic attractiveness function is *additive* with respect to the components of the route r, since any adding of a vertex v or an edge e to r leads to adding the values $ba(v)$ or $ba(e)$ to $ba(r)$.

Non-Additive Attractiveness of Routes

If the previous assertion is not valid for at least one route r then the attractiveness function a is said *non-additive*. One type of non-additivity is studied in [3]. There, if a route r contains e.g. a loop (which is allowed there) and passes around the same PoI p for the second time, the acquired attractiveness is smaller than in the case of the first passing.

In this paper, the approach is a bit different (and not only since the loops are not allowed here). The attractiveness is reduced not only during second passing around the same individual PoI p, but also around another PoI p' which is "very similar" to p. It is supposed, that the set P of all PoI's occurring at the vertices and edges of the graph G is divided into subsets $P_1, \ldots, P_\omega$. Two PoI's p and p' belonging to the same subset (class) P_i are considered "very similar".

For each positive integer $i \leq \omega$ a real number $a_i > 0$ is given such that the basic attractiveness of an "standard attractive" PoI $p \in P_i$ is $ba(p) = a_i$, if it is "above average attractive" then it is defined $ba(p) = 2a_i$, and if p is "unusually attractive" then $ba(p) = 3a_i$. Such an evaluation of attractiveness is equivalent to the conversion of one "above average attractive" PoI into two "standard attractive" PoI's and one "unusually attractive" PoI into three "standard attractive" PoI's. Therefore, the total "basic" attractiveness $ba(v)$ of a vertex v and $ba(e)$ of an edge e can be expressed by the sum

$$ba(v) = \sum_{i=1}^{\omega} a_i m_i(v) \quad \text{and} \quad ba(e) = \sum_{i=1}^{\omega} a_i m_i(e) \tag{1}$$

where $P_i(v) = P_i \cap P(v)$ and $m_i(v)$ is the "recalculated" number of elements in the set $P_i(v)$, where each "standard" p is counted once, each "above attractive" twice and each "unusually attractive" three times. The denotations $P_i(e)$ and $m_i(e)$ are defined analogically.

In contrast with [3], where the first passing around a PoI p is considered basic and the attractiveness of any next passing is reduced with respect to the first one, now the second and further passing around a PoI p from the given P_i is considered basic with attractiveness $ba(p)$ and the first passing around some PoI from P_i its attractiveness is increased. If $p \in P_i$ and p is "standard attractive" then the attractiveness of the first passing around it is $a1(p) = a_i + f_i$ where $f_i > 0$ is given for all classes P_i. Consequently, the attractiveness of a route r is defined by the sum

$$a(r) = \sum_{i=1}^{\omega} \left(g(r,i)\, f_i + \sum_{v \in r} a_i m_i(v) + \sum_{e \in r} a_i m_i(e) \right) \tag{2}$$

where the indicator $g(r, i) = 1$ when the route r passes at least once around a PoI $p \in P_i$, i.e. when at least one member $m_i(v)$ or $m_i(v)$ in (Ar) is positive.

2 Optimization Model and Methods for the Problem EP2

This paper is focused on the design of optimal simple tourist routes connecting given vertices v_o and v_d in the given graph $G = (V, E, d, c)$, resolving the problem E2. The problem E1 can be solved analogically.

In the sequel, a non-additive attractiveness function $a(.)$ defined by (Ar) will be used. Moreover, one natural simplification is done. Assume that V' is a set, $V' = V$ if $v_o \neq v_d$ and $V' = V \cup \{fv_d\}$ otherwise, where fv_d is fictive (dummy) vertex. Further, $E' = E$ if $v_o \neq v_d$ and $E' = E \cup \{(v, fv_d): (v, v_d)\}$ otherwise; then each added edge (v, fv_d) has the same length cost and PoI's as the original (v, v_d).

It is obvious that, in the case of $v_o = v_d$ the optimal route r' from v_o to fv_d in the graph $G' = (V', E', d, c)$ unambiguously defines the optimal route r from v_o to $v_d = v_o$ in the graph G'. Therefore, the following model will suppose that $v_o \neq v_d$ and, obviously, denotation of the vertices can be chosen in such a way that $v_o = 1$, $v_d = n$.

2.1 LP model for EP2

Besides the denotations already defined, the following denotations, parameters and variables are used in the model:

$GD = (V, ED, d, c)$ is the digraph corresponding to the graph G in the following sense:

- $e = (v, w) \in E \Leftrightarrow vw \in ED$ and $wv \in ED$
- $d(vw) = d(wv) = d(v, w)$, $c(vw) = c(wv) = c(v, w)$, and $m_i(vw) = m_i(wv) = m_i(v, w)$ for each $(v, w) \in E$ and each $i = 1, \ldots, \omega$.

M represents a "big number", i.e. a substitute of infinity. E.g. M is the sum of all values $c(v, w)$ times the sum of all values $d(v, w)$ times the sum of all values $a_i f_i$ times the sum of all values $m_i(k).m_i(j, k)$ or something like this.

The binary variable $u_j = 1$ expresses the fact that the j^{th} vertex is chosen to the route and $u_j = 0$ means the opposite for all $j = 1, \ldots, n$.

The binary variable $z_i = 1$ expresses the fact that route passes around any PoI from P_i and $z_i = 0$ means the opposite for all $i = 1, \ldots, \omega$.

The binary variable $x_{jk} = 1 \Leftrightarrow$ the k^{th} vertex is the immediate successor of the j^{th} vertex on the route for all $ij \in ED$.

The positive integer y_j is the ordinal number of the j^{th} vertex on the route.

The LP problem:

To find the binary variables z_i, $i = 1, \ldots, \omega$, u_j, $j = 1, \ldots, n$, x_{ij}, $ij \in ED$, and positive integer variables y_j, $j = 1, \ldots, n$ minimizing the objective function

$$\sum_{jk \in ED} c(jk) x_{jk} \rightarrow \min \tag{3}$$

and meeting the following constraints:

$$u_1 = u_n = 1 \text{ and } y_1 = 1 \tag{4}$$

$$x_{jk} \leq \frac{u_j + u_k}{2} \text{ for each } jk \in ED \tag{5}$$

$$u_k \leq \sum_{j \in \{1,\ldots,n\};\, jk \in ED} x_{jk} \text{ for each } k = 2, \ldots, (n-1) \tag{6}$$

$$y_k \geq y_j + 1 + M(x_{jk} - 1) \text{ and } y_k \leq y_j + 1 + M(1 - x_{jk}) \text{ for each } jk \in ED \tag{7}$$

$$u_k \leq y_k \leq n.u_k \text{ for each } k \in \{1, 2, \ldots, m\} \tag{8}$$

$$\sum_{k \in \{1,\ldots,n\};\, 1k \in ED} x_{1k} = 1 \quad \sum_{k \in \{1,\ldots,n\};\, k1 \in ED} x_{k1} = 0 \quad \sum_{k \in \{1,\ldots,n\};\, nk \in ED} x_{nk} = 0 \quad \sum_{k \in \{1,\ldots,n\};\, kn \in ED} x_{kn} = 1 \tag{9}$$

$$\sum_{j \in \{1,\ldots,n\};\, jk \in ED} x_{jk} = \sum_{j \in \{1,\ldots,n\};\, kj \in ED} x_{kj} \text{ for each } k = 2, \ldots, (n-1) \tag{10}$$

$$z_i \leq \sum_{k=2}^{n-1} m_i(k) u_k + \sum_{jk \in ED} m_i(jk) x_{jk} \leq M.z_i \text{ for each } k = 2, \ldots, (n-1) \tag{11}$$

$$\sum_{i=1}^{\omega} \left(f_i z_i + a_i \left(\sum_{k=2}^{n-1} m_i(k) u_k + \sum_{jk \in ED} m_i(jk) x_{jk} \right) \right) \geq a_0 \tag{12}$$

$$\sum_{jk \in ED} d(jk) x_{jk} \leq d_0 \tag{13}$$

Remark. It is obvious that the LP model for the problem EP1 can be obtained by a small modification of the LP model for EP2. Which one will be used depends on the budget situation.

2.2 Solution of both Problems by Depth-First-Search Technique

The proposed method (in the sequel denoted DFSM) is based on the combinatorial "depth first search" inspection of all possible routes. Clearly, its computational complexity very highly increases with the increasing number of vertices and edges. So that it cannot be practically used for optimization in large networks. On the other hand, it's clear, that the calculation can be stopped at any time the feasible solution is reached and computational time has exceeded acceptable limit. However, the solution probably is not optimal in this case, and the method should be considered heuristic.

This method works with the same digraph *GD* as in case of LP model. The building procedure starts with the initial vertex v_o and, in each step, it selects the first unused edge $e \in ED$ incident with v_o leading into the unused vertex (loops are not allowed). If exist, such edge is appended to the current route. If the sum of the length of current route and the length of the shortest path from ending vertex of current route to the destination v_d does not exceed the maximum length of route d_o, the procedure iterates resuming from the ending vertex of the added edge. Otherwise, the procedure backtracks, the last edge is discarded and the next one edge is considered. When

the current route reaches the destination vertex and the total length does not exceed d_o, the total cost $c(r)$ and the attractiveness $a(r)$ are computed, the route is added to the list of possible routes and procedure backtracks.

The optimal solutions of the problems EP1 and EP2 then can be obtained from the resulting list of possible routes by using filtering and ordering. It takes time less than 1 second. The advantage of this method is that the way of calculating the total cost and attractiveness of the route could easily be changed even after the list of possible routes is created.

The different strategies could be used for the selection of edge being added to current route. The edges incident with particular vertex could be for example ordered in accordance to edge or ending vertex attractiveness, edge length or cost. In case of problem EP2 it is possible to modify the algorithm in order to implement cutting of unpromising branches when the cost of current route exceeds c_0. These strategies influence the "heuristic" effectivity of the algorithm and improve the method ability to reach better solution in shorter time. This could be interesting in case of premature interruption of the procedure.

3 Computational Experience

The both above described methods were implemented and tested on the randomly generated network with 30 vertices and 94 directed edges. The edges cost $c(v,w) \in \langle 0;100)$ were generated randomly. LP model was implemented and solved in Gurobi 5.6.3 and combinatorial (DFSM) method in the environment of Visual Basic for Application in MS Access, which offers easy to use tools for ordering and filtering. Both methods were tested on standard hardware configuration (PC with Windows 7, Intel® Core™ i5 CPU 750@ 2,67 GHz, RAM 4 GB).

For testing purposes we assume only 3 classes of PoI. Each with basic attractiveness of standard PoI a_i = 1 and f_i =1. Each vertex and edge were assigned by randomly generated vector (m_1,m_2,m_3) of the "recalculated" numbers of "standard attractive" PoI. Test network is shown in Figure 1. The vertices are labeled in the form [m_1,m_2,m_3] and the edges in the form (edge cost| m_1,m_2,m_3).

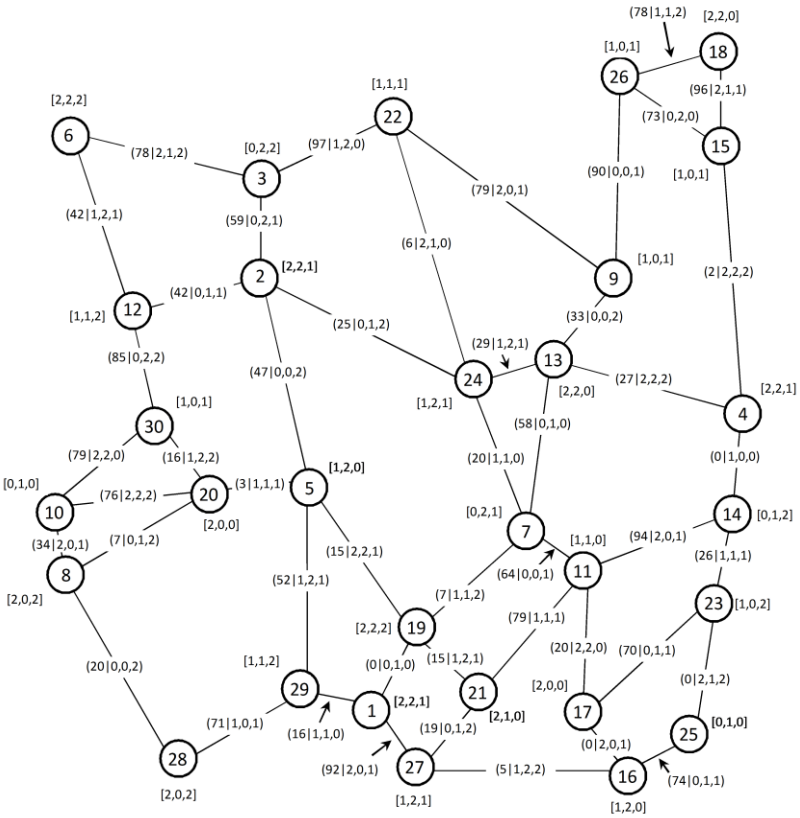


Figure 1 Test network

The tests were carried out for v_o = 6, v_d = 18 and different parameters d_o and a_o. LP model was implemented in the EP2 version described in 2.1. It is obvious from the description in 2.2, that the DFSM solves both EP1 and EP2 problems simultaneously and moreover for all possible values of a_o. This fact should be taken into account when comparing the computational times of the methods.

The solutions obtained by the both methods were exactly the same for all test cases. Table 1 brings the comparison of computational times of both methods in different test cases. The small size of test network has enabled us to find all potential routes with all reachable levels of attractiveness by DFSM method, even for d_o = 1000. Thus in the case, when d_o is greater then the length of the longest route without loops. LP model was tested with a_o set to approximately 90% of maximal potential attractiveness obtained by DFSM.

DFSM*		LP model**		
d_o	Elapsed time[s]	d_o	a_o	Elapsed time[s]
60	< 1	60	52	< 1
80	< 1	80	77	< 1
100	3	100	105	1
120	10	120	126	< 1
140	27	140	141	1
160	51	160	155	10
180	75	180	167	20
200	88	200	169	15
1000	95	1000	169	30

* DFSM solves EP1 and EP2 simultaneously

** LP model solves EP2

Table 1 Comparison of Computational Times

Table 1 shows, that in case of DFSM the computational time continuously increases with d_o. It is fully consistent with the combinatorial character of the method. In the contrary, in case of LP model, the computational times depend on particular parameters and don't increase continuously. The tests we have carried out in addition to those described in table 1 show, that the computational time of LP model depends (for the used test network structure and cost parameters of edges) more closely with the parameter a_o than with d_o.

4 Conclusion and Outline of Future Research

This paper studies a new branch of non-motorized transport management theory – the design of routes for non-motorized transport for leisure purposes, taking into account the attractiveness of routes. The main contributions of the paper are:

1. Introduction of a non-additive attractiveness function, comparing with [2] or [3] where only additive one was used.
2. Formulation of two optimization problems.
3. Presentation of two exact methods of solution.
4. Outline of computational experience.

The future research is expected in two directions: To improve the algorithms of the problems EP1 and EP2 and to extend the application aspects, e.g. from the point of view of optimal design to the view of optimal use of just designed network.

Acknowledgements

The authors would like to thank the Czech Science Foundation for the support of the project 402/12/2147 whose results are described in this paper.

References

[1] Buehler, R.: Determinants of Bicycle Commuting in the Washington, D.C. Region: The Role of Bicycle Parking, Cyclist Showers, and Free Car Parking at Work. *Transportation Research Part D: Transportation and Environment* **17** (2012), 525-531.

[2] Černá, A., and Černý, J.: Note on Optimal Paths for Non-Motorized Transport on the Network. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*, Karviná, Czech Republic, September 11-13, 2012, pp. 91-94.

[3] Černá, A., Černý, J., Malucelli, F., Nonato, M., Polena, L., and Giovannini, A." Designing Optimal Routes for Cycle-Tourists. Transportation Research Procedia **3** (2014), 856 – 865.

[4] Furth, P.: Bicycling Infrastructure for Mass Cycling. In: A Transatlantic Comparison. City Cycling (Pucher, J., and Buehler, R. eds), MIT Press, Cambridge, MA, 2112, 105-139.
[5] Goodman, A., Sahlqvist, S., and Ogilvie, D.: New Walking and Cycling Routes and Increased Physical Activity: One- and 2-Year Findings From the UK iConnect Study. *Am J Public Health* **104** (2014), e38–e46.
[6] Landis, B., Vattikuti, V., and Brannick, M.: (1997). "Real-Time Human Perceptions: Towards a Bicycle Level of Service, *Transportation Research Record* **1578** (1997), Transportation Research Board, National Research Council, Washington, DC.
[7] Lin, J.J., and Yu, C. J.: A bikeway network design model for urban areas. *Transportation: Planning, Policy, Research, Practice* **40** (2013), 45-68.
[8] Plesník, J.: *Grafové algoritmy* (Algorithms in Graphs, in Slovak). VEDA, Bratislava, 1983.
[9] Pucher, J., Buehler, R., Bassett, D., and Dannenberg, A.: Walking and Cycling to Health: Recent Evidence from City, State, and International Comparisons. *Americal Journal of Public Health* **100** (2010), 1986-1992.
[10] Sener, I.N., Eluru, N., and Bhat, C.R.: An Analysis of bicycle route choice preferences in Texas, US. *Transportation*, **36** (2009), 511-539.
[11] Stinson, M.A., and Bhat, C.R.: Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation Research Record*, **1828** (2003), Transportation Research Board, National Research Council, Washington, DC, 107–115.
[12] Thakuriah, P., Metaxatos, P., Lin, J., and Jansen, E.: Factors affecting propensity to use bicycle and pedestrian facilities in suburban location. *Transportation Research Part D : Transportation and Environment* **17** (2012), 341–348.
[13] Tranter, P.J.: Effective speed: Cycling because it's faster. In: City Cycling (Pucher, J., and Buehler, R. eds), MIT Press, Cambridge, MA, 2112, 57-74.

Interval data and sample variance: A study of an efficiently computable case

Michal Černý[1], Ondřej Sokol[2]

Abstract. Computation of the tight upper bound of the sample variance of an interval-valued dataset is known to be NP-hard. However, there are various polynomial-time computable special cases, which are quite natural from a statistical point of view. We analyze the algorithm by Ferson, Ginsburg, Kreinovich, Longpré and Aviles designed for one such special case. The algorithm can be applied to any dataset and it is exponential in the maximum number of intersecting $\frac{1}{n}$-narrowed intervals. This number corresponds to the largest clique in the associated interval intersection graph. We analyze the size of the clique for some natural probabilistic data generating processes, showing that the size of the clique grows "moderately". This implies that in spite of the theoretical NP-hardness result, the analyzed algorithm is indeed useful in many practical cases, even more than originally claimed by the authors of the algorithm.

Keywords: Interval data, sample variance, average time complexity.

JEL classification: C44
AMS classification: 90C15

1 Introduction

We deal with the problem that arises when exact data values are not known, but intervals which these values belong to are known. These situations happen when we process rounded, censored or categorized data. Similar problems are solved when we work with measurements which have known tolerances or when we work directly with intervals, e.g. interval predictions, bid-ask spread or intervals of daily minimum and maximum values in the stock market.

Sample variance is basic descriptive statistic, which is essential for many statistical techniques and methods. It includes computation of various statistics (e.g. t-statistic), data normalizations etc. When only interval data are at our disposal, our goal is to determine lower and upper bounds of selected statistics. From these lower and upper bounds we can draw conclusions over the entire dataset.

2 Problem statement

A general framework. Consider a one-dimensional dataset $x_1, \ldots, x_n$ (a random sample from a certain distribution, say), which is unobservable. What is observable is a collection of intervals $\boldsymbol{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \boldsymbol{x}_n = [\underline{x}_n, \overline{x}_n]$ such that it is guaranteed that

$$x_i \in \boldsymbol{x}_i, \quad i = 1, \ldots, n. \tag{1}$$

Assume that the only information about the distribution of $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ given the observed values $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_n)^{\mathrm{T}}$ and $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n)^{\mathrm{T}}$ is that the axiom (1) holds a.s.

Let a statistic $S(x_1, \ldots, x_n)$ be given. (Formally, $S : \mathbb{R}^n \to \mathbb{R}$ is just a continuous function.) Due to our weak assumptions, the only information we can infer about $S(x_1, \ldots, x_n)$ from the observable data

[1]University of Economics in Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, cernym@vse.cz
[2]University of Economics in Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, xsoko00@vse.cz

$\underline{x}, \overline{x}$ is the lower and upper bound, respectively, of the form

$$\underline{S} = \min\{S(\xi) : \underline{x} \le \xi \le \overline{x}\}, \quad \overline{S} = \max\{S(\xi) : \underline{x} \le \xi \le \overline{x}\},$$

where the inequality "$\le$" between two vectors in $\mathbb{R}^n$ is understood componentwise.

Examples. Bounds $\underline{S}, \overline{S}$ for many important statistics have been extensively studied in literature. Sometimes, the situation is easy: for example, where $S \equiv \widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ is the sample mean, then we immediately get

$$\underline{\widehat{\mu}} = \frac{1}{n}\sum_{i=1}^n \underline{x}_i, \quad \overline{\widehat{\mu}} = \frac{1}{n}\sum_{i=1}^n \overline{x}_i. \tag{2}$$

Another important statistic is the sample variance $S \equiv \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$, when the true mean μ is known. Then $\overline{\widehat{\sigma^2}} = \frac{1}{n}\sum_{i=1}^n \max\{(\underline{x}_i - \mu)^2, (\overline{x}_i - \mu)^2\}$ and $\underline{\widehat{\sigma^2}} = \frac{1}{n}\sum_{i=1}^n s_i^2$, where

$$s_i^2 = \begin{cases} (\overline{x}_i - \mu)^2, & \text{if } \overline{x}_i < \mu, \\ (\underline{x}_i - \mu)^2, & \text{if } \underline{x}_i > \mu, \\ 0, & \text{if } \underline{x}_i \le \mu \le \overline{x}_i. \end{cases}$$

This paper is a contribution to the case of sample variance when the true mean is unknown. This statistic will be investigated in the next section. However, other statistics have been studied too; we refer the reader to Černý and Hladík [1] (t-ratio), Kreinovich [8] and Xiang et al. [10] (entropy), Kreinovich et al. [7] (higher moments) and others. Summaries of approaches to computing various statistics under interval uncertainty can be found in [6] and [9].

3 Sample variance

From now on we will work with the sample variance

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \widehat{\mu})^2, \tag{3}$$

where the true mean μ is unknown and is replaced by the sample mean $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$. The computation of the upper and lower bounds, respectively, reduces to the optimization problems

$$\underline{\widehat{\sigma}^2} = \min_{x\in\mathbb{R}^n} \left\{ \frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j \right)^2 : \underline{x} \le x \le \overline{x} \right\}, \ \overline{\widehat{\sigma^2}} = \max_{x\in\mathbb{R}^n} \left\{ \frac{1}{n}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j \right)^2 : \underline{x} \le x \le \overline{x} \right\}.$$

It is obvious that computation of $\underline{\widehat{\sigma}^2}$ is a convex quadratic problem, which can be solved in weakly polynomial time. However, Ferson et al. [4] introduced another interesting method, yielding a strongly polynomial algorithm.

The computation of $\overline{\widehat{\sigma^2}}$ is easily seen to be NP-hard (a proof can be found in [5]). Moreover, it is known to be inapproximable with an arbitrary absolute error, see Černý and Hladík [1]. The cited paper also gives a useful positive statement: computation of $\overline{\widehat{\sigma^2}}$ can be done in pseudopolynomial time.

The computational properties of $\overline{\widehat{\sigma^2}}$ have been studied extensively, see e.g. Ferson [4], Xiang [10], Dantsin [3]. The most interesting fact is that many special cases of the problem are known that can be solved efficiently. Our question is whether these cases are "rare" or "frequent" in practice. In particular, the statistical approach usually assumes that the observed data $\underline{x}, \overline{x}$ are generated by an underlying random process. Then, the question is whether the hard instances—such as the instances resulting from NP-hardness proofs—occur with a high or low probability.

An interesting algorithm for computation of $\overline{\widehat{\sigma^2}}$ by Ferson et al. The paper by Ferson et al. [4] presents an algorithm with the following nice property. To fix notation, for an interval $\boldsymbol{x} = [\underline{x}, \overline{x}]$ we denote $x^C = \frac{1}{2}(\overline{x} + \underline{x})$ its center and $x^\Delta = \frac{1}{2}(\overline{x} - \underline{x})$ its radius. For an $\alpha > 0$, let $\alpha\boldsymbol{x}$ denote the interval $[x^C - \alpha x^\Delta, x^C + \alpha x^\Delta]$.

Theorem 1. *There exists a polynomial $p(n)$ for which the following holds true: when $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are given such that*

$$\text{every } k\text{-tuple of distinct indices } i_1, \ldots, i_k \in \{1, \ldots, n\} \text{ satisfies } [\underline{\widehat{\mu}}, \overline{\widehat{\mu}}] \cap \tfrac{1}{n}\boldsymbol{x}_{i_1} \cap \tfrac{1}{n}\boldsymbol{x}_{i_2} \cap \cdots \cap \tfrac{1}{n}\boldsymbol{x}_{i_k} = \emptyset, \tag{4}$$

then the algorithm by Ferson et al. makes at most $2^k p(n)$ steps. [Here, $\underline{\widehat{\mu}}, \overline{\widehat{\mu}}$ are given by (2).] □

Remark. The original paper presents a weaker statement, in particular the intersection property from (4) is written as $\frac{1}{n}\boldsymbol{x}_{i_1} \cap \frac{1}{n}\boldsymbol{x}_{i_2} \cap \cdots \cap \frac{1}{n}\boldsymbol{x}_{i_k} = \emptyset$.

So, we are interested in the minimum k for which the condition (4) holds. This task can be reformulated in a graph-theoretic way: consider the undirected graph $G_n = (V_n, E_n)$, where $V_n = \{1, \ldots, n\}$ and $\{i, j\} \in E_n$ iff $\frac{1}{n}\boldsymbol{x}_i \cap \frac{1}{n}\boldsymbol{x}_j \cap [\underline{\widehat{\mu}}, \overline{\widehat{\mu}}] \neq \emptyset$ $(i \neq j)$. Let ω_n be the size of the largest clique of G_n. Now we can, obviously, restate Theorem 1:

Corollary 2. *The algorithm runs in time $O(2^{\omega_n} p(n))$, where p is a polynomial.* □

4 Description of the algorithm

The algorithm proposed by Ferson et al. [4] works as follows:

- We sort all $2n$ endpoints of the narrowed intervals $\frac{1}{n}\boldsymbol{x}_1, \ldots, \frac{1}{n}\boldsymbol{x}_n$ into a sequence of ascending order $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(2n)}$. After that we divide the real line into $2n+1$ segments $\left[x_{(k)}, x_{(k+1)}\right]$, where we set $x_{(0)} := -\infty$ and $x_{(2n+1)} := \infty$.
- We compute the $\underline{\widehat{\mu}}$ and $\overline{\widehat{\mu}}$ and select all intervals $\left[x_{(k)}, x_{(k+1)}\right]$ that intersect with $[\underline{\widehat{\mu}}, \overline{\widehat{\mu}}]$.
- For each remaining interval $\left[x_{(k)}, x_{(k+1)}\right]$ and for each $i = 1, \ldots, n$ we select the value of x_i following this rule:
  - if $x_{(k+1)} < x_i^C - \frac{1}{n}x_i^\Delta$, then we select $x_i = \overline{x}_i$,
  - if $x_{(k)} > x_i^C + \frac{1}{n}x_i^\Delta$, then we select $x_i = \underline{x}_i$,
  - otherwise we consider both possible values for x_i.

  We get one or several sequences of x_i. If the average of $x_1, \ldots, x_n$ is in the interval of $\left[x_{(k)}, x_{(k+1)}\right]$, then we compute its sample variance using formula (3).
- We return the largest value of the computed sample variances as $\overline{\widehat{\sigma^2}}$.

From the description it follows the algorithm makes at most $2^k n^2$ steps, where k denotes the maximal number of narrowed intervals that have a common point in interval $[\underline{\widehat{\mu}}, \overline{\widehat{\mu}}]$. If we construct G_n, then $\omega_n = k$.

The main idea of the algorithm is based on the fact that if all values of $x_1, \ldots, x_n$ are fixed except x_i, then $\overline{\widehat{\sigma^2}}$ is a quadratic function of one variable. As the function $\overline{\widehat{\sigma^2}}$ is convex, it will attain its maximum in one of the endpoint of $[\underline{x}_i, \overline{x}_i]$. The maximum is attained at the endpoint which is the farthest from the point of minimum of this quadratic function, which, in this case, is $\widehat{\mu}$ computed from fixed x_j for all $j \neq i$.

5 Stochastic setup

Of course, the algorithm can be slow, e.g. when we meet an instance with $x_1^C = x_2^C = \cdots = x_n^C$ (such instances result e.g. from the proof of NP-hardness of computation of $\overline{\widehat{\sigma^2}}$). However, here we switch ourselves to a particular stochastic setup, where it seems that such instances are rare.

We assume that the centers $x_1^C, \ldots, x_n^C$ are sampled from a distribution Φ and that the radii $x_1^\Delta, \ldots, x_n^\Delta$ are sampled from a nonnegative distribution Ψ and the samples are independent. Observe that now G_n is a random graph and its clique number ω_n is a random variable. We also assume that Φ is a continuous distribution with finite first and second moments and its density function is limited from above and Ψ has finite first and second moments.

6 Simulations

The objective of the simulation was to estimate $\mathsf{E}\omega_n$, where E stands for the expectation.

We simulated sets of intervals with centers $x_1^C, \ldots, x_n^C$ sampled from $N(0,1)$. The radii of intervals $x_1^\Delta, \ldots, x_n^\Delta$ were sampled from $\mathrm{Exp}(\lambda)$ with $\lambda \in \{0.1,\ 0.5,\ 1,\ 5\}$ for $n = 25, \ldots, 2000$. Values of n were generated from 25 to 300 with the step of 25, the rest of n up to 2000 was generated with the step of 100. For each n we generated 500 samples from which we computed the expected value of ω_n.

Figure 1 shows the relationships between ω_n and n for various parameters of λ. For comparison we plot the function $\log n$ into the figure. Figure 2 shows the same relationships but with the logarithmic horizontal axes. Comparing the shapes of curves corresponding to $\log n$ and curves obtained by our simulations, we conclude that ω_n can be approximated by $\log n$ with sufficient accuracy. A statistical test confirms that this is correct.

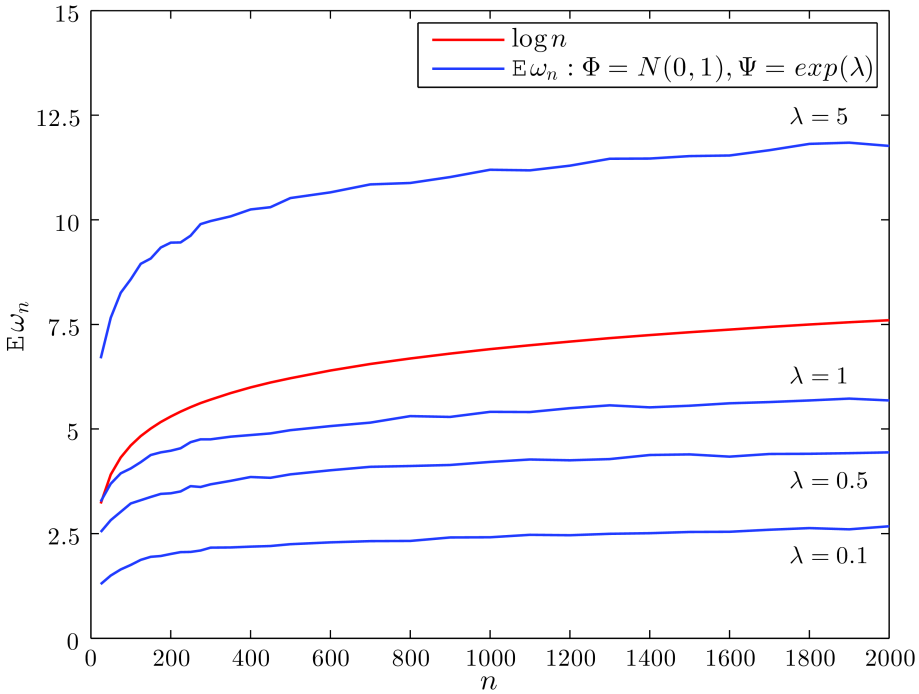


Figure 1 Results of simulations $\mathsf{E}\omega_n$ for various λ of the distribution $\Psi = \mathrm{Exp}(\lambda)$.

In Figure 3 we show that ω_n is affected only by the ratios $\frac{\mathrm{Var}(\Phi)}{\mathrm{Var}(\Psi)}$ and $\frac{\sqrt{\mathrm{Var}(\Phi)}}{\mathsf{E}(\Psi)}$. It is clearly seen that $\mathsf{E}\omega_n$ is the same for various values of σ^2 and λ from distributions $\Phi = N(0, \sigma^2)$ and $\Psi = \mathrm{Exp}(\lambda)$, when the ratios $\frac{\mathrm{Var}(\Phi)}{\mathrm{Var}(\Psi)}$ and $\frac{\sqrt{\mathrm{Var}(\Phi)}}{\mathsf{E}(\Psi)}$ remain the same.

However, the main result of this paper is that ω_n is approximable by $\log n$ for every ditributions Φ and Ψ if the conditions stated in Section 5 are true.

Results. Using simulation, we have found that the maximal clique ω_n of G_n randomly generated can be approximated by the logarithmic function of n. That yields an interesting result in terms of expected computational complexity, because maximal clique ω_n determines exponential complexity of the algorithm proposed by Ferson et al. [4], which is used for computation of $\overline{\widehat{\sigma^2}}$ under interval uncertainty. From the Corollary 2 we found that the complexity of computing $\overline{\widehat{\sigma^2}}$ from random data which are assumed to be generated from Φ and Ψ have polynomial complexity.

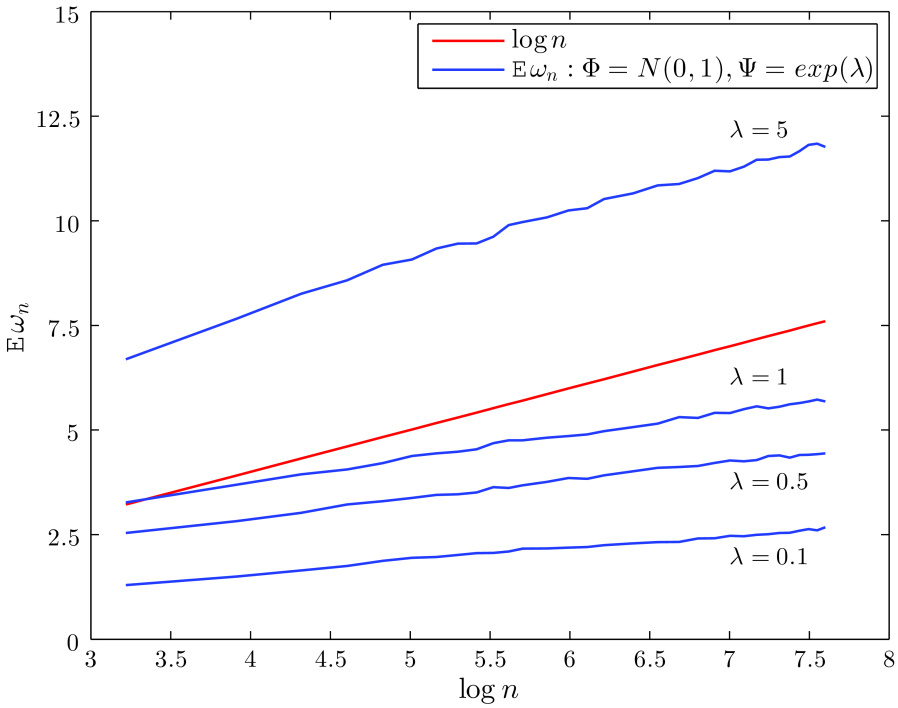


Figure 2 Results of simulations $\mathsf{E}\omega_n$ for various λ of the distribution $\Psi = \mathrm{Exp}(\lambda)$. Horizontal axis is logarithmic.

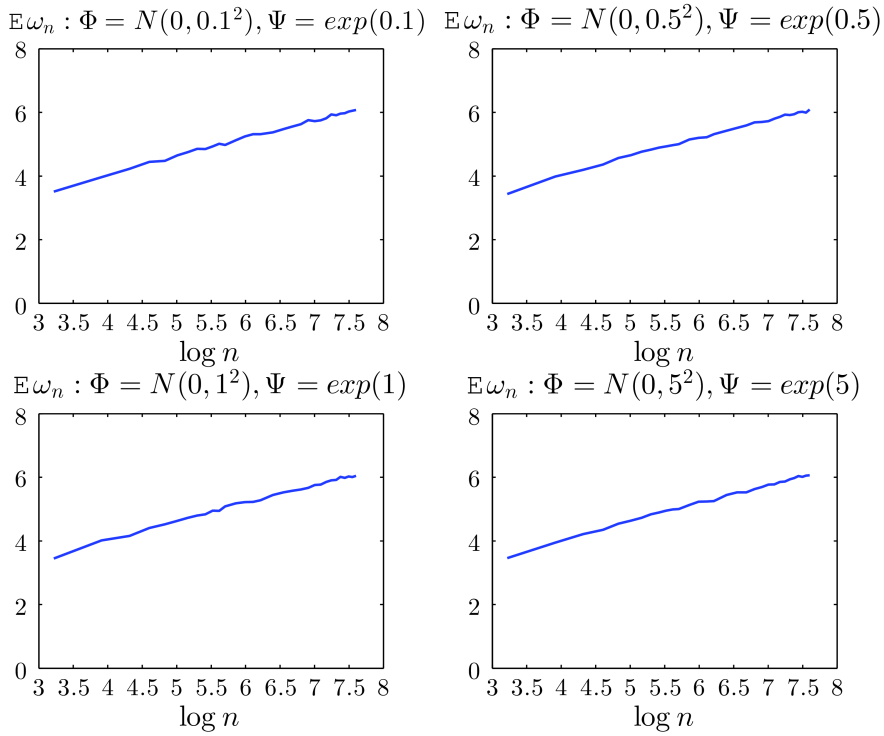


Figure 3 Results of simulation $\mathsf{E}\omega_n$ for various σ^2 and λ for distributions $\Phi = N(0, \sigma^2)$ and $\Psi = \mathrm{Exp}(\lambda)$ while $\mathrm{Var}(\Phi) = \mathrm{Var}(\Psi)$ and $\sqrt{\mathrm{Var}(\Phi)} = \mathsf{E}(\Psi)$. Horizontal axes are logarithmic.

7 Conclusions

Though the problem of computation of $\overline{\widehat{\sigma}^2}$ is generally NP-hard, we have shown that for many practical situations $\overline{\widehat{\sigma}^2}$ can be calculated in polynomial time. These results were obtained by using Monte Carlo simulations but have not been analytically confirmed so far.

Our main conjecture is: *if Φ is a continuous distribution with finite first and second moments and its density function is limited from above and Ψ has finite first and second moments, then $\mathsf{E}\omega_n = O(\log n)$.* If the conjecture is true, then Corollary 2 implies that *the algorithm runs in polynomial time on average.* And this is the main and most interesting finding; even though we currently cannot prove the conjecture, we present simulations strongly supporting it. To this end, the simulations confirm at least the fact that the algorithm of Ferson et al. is indeed useful in practice (as long as the random intervals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are generated by a "reasonable" random process, formalized here by the pair of distributions Φ and Ψ).

Further research should focus on a wider class of statistics for which Ferson's method could be efficiently applied; this class may contain e.g. the coefficient of variation or some change-point statistics, see e.g. [2].

Acknowledgment. The work of M. Černý was supported by the University of Economics in Prague under Grant IGA F4/54/2015. The work of O. Sokol was supported by the Faculty of Informatics and Statistics, University of Economics in Prague.

References

[1] ČERNÝ, Michal; HLADÍK, Milan. The Complexity of Computation and Approximation of the t-ratio over One-dimensional Interval Data. *Computational Statistics & Data Analysis.* 2014, Vol. 80, pp. 26–43.

[2] ČERNÝ, Michal. Binary segmentation and Bonferroni-type bounds. *Kybernetika.* 2011, Vol. 47, No. 1, pp. 38–49.

[3] DANTSIN, Evgeny; KREINOVICH, Vladik; WOLPERT, Alexander; XIANG, Gang. Population Variance under Interval Uncertainty: A New Algorithm. *Reliable Computing.* 2006, Vol. 12, No. 4, pp. 273–280.

[4] FERSON, Scott; GINZBURG, Lev; KREINOVICH, Vladik; LONGPRÉ, Luc; AVILES, Monica. Exact Bounds on Finite Populations of Interval Data. *Reliable Computing.* 2005, Vol. 11, No. 3, pp. 207–233.

[5] FERSON, Scott; GINZBURG, Lev; KREINOVICH, Vladik; LONGPRÉ, Luc; AVILES, Monica. Computing Variance for Interval Data is NP-Hard. *ACM SIGACT News.* 2002, Vol. 33, No. 2, pp. 108–118.

[6] FERSON, Scott; KREINOVICH, Vladik; HAJAGOS, Janos; OBERKAMPF, William; GINZBURG, Lev. *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty.* Sandia National Laboratories, Albuquerque. 2007.

[7] KREINOVICH, Vladik; LONGPRÉ, Luc; FERSON, Scott; GINZBURG, Lev. Computing Higher Central Moments for Interval Data. *Technical Report.* University of Texas at El Paso. 2004. Available from: http://digitalcommons.utep.edu/cs_techrep/374/.

[8] KREINOVICH, Vladik. Maximum Entropy and Interval Computations. *Reliable Computing.* 1996, Vol. 2, No. 1, pp. 63–79.

[9] NGUYEN, Hung T.; KREINOVICH, Vladik; WU, Berlin; XIANG, Gang. Computing Statistics under Interval and Fuzzy Uncertainty. *Studies in Computational Intelligence.* 2012.

[10] XIANG, Gang; CEBERIO, Martine; KREINOVICH, Vladik. Computing Population Variance and Entropy under Interval Uncertainty: Linear-time Algorithms. *Reliable computing.* 2007, Vol. 13, No. 6, pp. 467–488.

Routes Design Using Own and Hired Vehicles

Zuzana Čičková[1], Ivan Brezina[2], Juraj Pekár[3]

Abstract: Costs of distribution constitute a major share of the total logistic costs of almost every organization. Decision making in this field can be supported by mathematical models of various types of routing problems. Classical routing problems are aimed at satisfying customer demand with the routes originating and ending in a central depot, while the open routing problems do not consider the need for vehicles to return. The importance of open versions arises e.g. from car-hire options. This paper describes new model of vehicle routing problem that allows combining classical capacited vehicle routing problem (CVRP) modeling the routes realized by own vehicles together with the possibility to realize open routes with hired vehicles (OVRP). Although hiring vehicles could be more expensive per unit distance traveled, their use could lead to considerable saving due to reduction of the total cost of the route. The analysis is focused on proportion between the cost corresponded to using the own and hired vehicles.

Keywords: Open Vehicle Routing Problem, Mixed Integer Programming, Mathematical Model.
JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

Distribution management is one of most important field in many business areas. A lot of different routing problems that enable to model many real-life problems can be very rewarding in logistics process. Frequent decision-making in the area of distribution is based on the possibility of hiring vehicles as opposed to using own vehicles and in that way to specify the number of own vehicles relative to total cost of rented vehicles. The unit cost per km of own vehicles consist not only of fuel cost but also include e.g. amortization, taxes and driver's wage. On the other side, the rental cost depends on cost per km and also includes the fixed cost per car. The known capacitated vehicle routing problem (CVRP) can be briefly described as follows ([1], [2], [3], [5], [6], [8]): Consider a depot from which some products have to be delivered to a set of customers. It is assuming the known shortest distance between depot and each customer's location, as well as between each pairs of customer's location. The goal is to find the optimal shortest route (starting and ending at the depot) for a vehicle (vehicles) so that each customer demand is met. The capacity of vehicle (or fleet of vehicles) is (are) known (if more than one vehicle are used, the same capacity of all of them is supposed), so that all customers' demand needs to be served with the use of some vehicle (all the demands are met in full).

Open models of routing problems ([4]) are an extension of CVRP. The difference between the open vehicle routing problem OVRP and the classical CVRP model is based on the assumption that the vehicle (vehicles) does (do) not need to return to the depot after servicing the last customer. The open models allow describing different real-world applications e.g. car-hire options that result to different logistic costs. Although hiring vehicles could be more expensive per unit distance traveled, their use could lead to considerable saving due to reduction of the total cost of the route. In the practical application it could be interesting to set the rate specifying the ratio between the own and rented vehicles so that it is effective rent at least one external vehicle.

This paper describes the mathematical model that allows setting the specific indicator describing the limit value which indicates the maximal rate (calculated as cost ratio between the cost of own and rented vehicles) when it is effective renting at least one external vehicle. The empirical analysis is based on regional data in region Nitra in Slovakia.

2 Vehicle routing problem and open vehicle routing problem

The classical versions of routing problems can be described using following notation: Let $N = \{1, 2, ...n\}$ be the set of served nodes (customers) and let $N_0 = N \cup \{0\}$ be a set of nodes that represents the customers together

[1] University of Economics in Bratislava, zuzana.cickova@euba.sk.
[2] University of Economics in Bratislava, ivan.brezina@euba.sk.
[3] University of Economics in Bratislava, juraj.pekar@euba.sk.

with the initial node (depot). A shortest distance time d_{ij} is associated with pairs $i, j \in N_0$, $i \neq j$. Mathematical models of selected routing problems are possible to formulate as binary programming problems. The models involve binary variables $x_{ij}\,(i, j \in N_0\ , i \neq j\ )$ that enable to model if the node i precedes node j in a route of the vehicle $x_{ij} = 1$ and $x_{ij} = 0$ otherwise. Certain demand $g_i,\ i \in N$, which has to be met from the initial node ($i = 0$), is associated with each customer. The distribution is performed using a vehicle with a certain capacity (g). The goal is to identify the routes of vehicle so that the total traveled distance is as low as possible (the known shortest distances between all nodes $d_{ij},\ \ i, j \in N_0$ is supposed) with respect to the following restrictions: the depot represents initial node and also the final node of every route, from this node the demands $q_i,\ i \in N$ of all the other nodes are met (in full), each node (except depot) is visited exactly once and total demand on route must not exceed the capacity of the vehicle (g). The model implicitly assumes that $q_i \leq g$ for all $i \in N$, i.e. the demand of each customer does not exceed the capacity of the vehicle. Further on, the variables $u_i, i \in N$ that based on well-known Miller-Tucker- Zemlin's formulation, e.g. Miller et al. ([7]) are employed. Those variables represent cumulative demand of customers on one particular route.

Based on this assumption, the model of CVRP can be stated as follows:

$$\min f(\mathbf{X},\mathbf{u}) = \sum_{i \in N_0} \sum_{\substack{j \in N_0 \\ i \neq j}} d_{ij} x_{ij} \tag{1}$$

$$\sum_{i \in N_0} x_{ij} = 1, \quad j \in N, \quad i \neq j \tag{2}$$

$$\sum_{j \in N_0} x_{ij} = 1, \quad i \in N, \quad i \neq j \tag{3}$$

$$u_i + q_j - g(1 - x_{ij}) \leq u_j, \quad i \in N_0, \quad j \in N, \quad i \neq j \tag{4}$$

$$q_i \leq u_i \leq g, i \in N \tag{5}$$

$$u_0 = 0 \tag{6}$$

$$x_{ij} \in \{0,\ 1\}, i, j \in N_0, i \neq j \tag{7}$$

Objective function (1) determines the total distance traveled. Equations (2) and (3) ensure that each customer (except the origin) is visited exactly ones. Equations (4) are anti-cyclical conditions that prevent the formation of such sub-cycles which do not contain a initial node ($i = 0$). The set of variables $u_i,\ i \in N$ ensures the calculation of current load of vehicles in its route to i-th customer (including). Equations (5) ensure that all demands on the route must not exceed the capacity of the vehicle. Equation (6) ensures that load of the vehicle is set to zero in the depot.

Mathematical model of OVRP can be stated as follows:

$$\min f(\mathbf{X},\mathbf{u}) = \sum_{i \in N_0} \sum_{j \in N_0} d_{ij} x_{ij} \tag{8}$$

$$\sum_{i \in N_0} x_{ij} = 1, \quad j \in N, \quad i \neq j \tag{9}$$

$$\sum_{j \in N_0} x_{ij} \leq 1, \ i \in N\ , \ i \neq j \tag{10}$$

$$\sum_{j \in N} x_{0j} \geq 1 \tag{11}$$

$$u_i + q_j - g(1 - x_{ij}) \leq u_j, \quad i \in N_0, \quad j \in N, \quad i \neq j \tag{12}$$

$$q_i \leq u_i \leq g\ , \ i \in N \tag{13}$$

$$u_0 = 0 \tag{14}$$

$$x_{ij} \in \{0,\ 1\}, i, j \in N_0, i \neq j \tag{15}$$

The objective function (8) models the total distance traveled by all of the vehicles. Equations (9) ensure that only one of the vehicles enters each customer exactly once and equations (10) ensure that the corresponding vehicle does not need to depart from every customer, because the route ends after serving the last of them. Equation (11) ensures that each vehicle starts its route exactly once. Equations (12) avoid the presence of sub-tour and also calculate the real cumulative demands of customers for the next node on the route based on

previous node. Equations (13) ensure that all demands on the route must not exceed the capacity of the vehicle. The fix value of variable u_0 is set up by equation (14).

3 Mathematical model based on cost rate between the own and hired vehicles

Further on, consider the possibility to realize the closed routes (starting and ending at the depot) together with the possibility to realize open routes. Mathematical programming formulation requires two type of binary variables: the variables $x_{ij},\ i, j \in N_0$ with a following notation: $x_{ij} = 1$ if customer i precedes customer j in a route of the own vehicle and $x_{ij} = 0$ otherwise and the variables $y_{ij},\ i, j \in N_0$ with a following notation: $y_{ij} = 1$ if customer i precedes customer j in a route of the rented vehicle and $y_{ij} = 0$ otherwise. Fixed cost associated with number of rented vehicle are known and designated as f. The model deals with the additional variable r representing the cost rate between the own and hired vehicles. Moreover we suppose the known scalar designated as *cvrp* representing the total cost of the route based on CVRP formulation (1) – (7):

The mathematical model can be stated as follows:

$$\max f(\mathbf{X}, \mathbf{Y}, \mathbf{u}, r) = r \tag{16}$$

$$\sum_{i \in N_0} \sum_{\substack{j \in N_0 \\ i \neq j}} d_{ij} x_{ij} + r \sum_{i \in N_0} \sum_{\substack{j \in N_0 \\ i \neq j}} d_{ij} y_{ij} + f \sum_{j \in N_0} y_{0j} \leq cvrp \tag{17}$$

$$\sum_{i \in N_0} x_{ij} + \sum_{i \in N_0} y_{ij} = 1,\ j \in N,\ i \neq j \tag{18}$$

$$\sum_{j \in N_0} x_{ij} + \sum_{j \in N_0} y_{ij} \leq 1,\ i \in N,\ i \neq j \tag{19}$$

$$u_i + q_j - g(1 - x_{ij}) \leq u_j,\ i \in N_0,\ j \in N,\ i \neq j \tag{20}$$

$$u_i + q_j - g(1 - y_{ij}) \leq u_j,\ i \in N_0,\ j \in N,\ i \neq j \tag{21}$$

$$\sum_{i \in N_0} x_{ij} - \sum_{i \in N_0} x_{ji} = 0,\ j \in N,\ i \neq j \tag{22}$$

$$q_i \leq u_i \leq g,\ i \in N \tag{23}$$

$$\sum_{i \in N_0} \sum_{\substack{j \in N_0 \\ i \neq j}} y_{ij} \geq 1 \tag{24}$$

$$u_0 = 0 \tag{25}$$

$$x_{ij} \in \{0, 1\},\ y_{ij} \in \{0, 1\},\ i, j \in N_0,\ i \neq j \tag{26}$$

The objective function (16) enables to find maximal rate that model the proportion between the hired and own vehicles. Equations (17) represent the total cost of the route must to be not more expensive compared to the CVRP. Equations (18) ensure that only one of the vehicles (own or rented) enters each customer exactly once and equations (19) ensure that the vehicle does not need to depart from every customer, because the route of rented vehicle ends after serving the last of them. Equations (20) and (21) avoid the presence of sub-tour (for own and rented vehicle) and also calculates the real cumulative demands of customers of the next node on the route based on previous node. Equations (22) ensure preserve the connectivity of the routes. Equations (23) ensure that all demands on the route must not exceed the capacity of the vehicle. Equation (24) ensure that at least one rented vehicle is used. The fix values of variables u_0 are set up by equation (25).

4 Routes Design Using Own and Hired Vehicles in Slovak Region

Slovakia is divided into 8 regions: Region Bratislava (BA), Region Banská Bystrica (BB), Region Košice (KE), Region Nitra (NR), Region Prešov (PO), Region Trenčín (TN), Region Trnava (TT) and Region Žilina (ZA):

Region Nitra (NR) is divided into 7 districts: Nitra (0), Komárno (1), Levice (2), Nové Zámky (3), Šaľa (4), Topoľčany (5), Zlaté Moravce (6).

Input data for region Nitra: 7 districts, distribution centre (regional capital is Nitra, $i = 0$), the shortest distances between all district towns and between distribution centre and each district towns is designated as d_{ij}, vehicle capacity g was set to 12, the demand of district towns was associated with number of inhabitant (Komárno 6, Levice 7, Nové Zámky 9, Šaľa 3, Topoľčany 4, Zlaté Moravce 3). The first goal was to set the

distribution based on CVRP (using the own vehicles). The resulting value of objective function (8) is 337, which value is also the value of coefficient *cvrp* in (17). Next, the fixed coefficient of rented vehicle (*f*) was set to 15 per vehicle and problem (16) – (26) was solved in accordance to those additional parameters. Resulting value of variable *r* specifies the maximal ratio representing effective renting at least one external vehicle.

The computational experiments were provided on the base of before mentioned data. The mathematical model was implemented in software GAMS (solver Couenne 23.5.1) on PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8. The results are given below:

Results of CVRP (1) – (7):
Objective value *337*:
The routes of own vehicles: Nitra (0) - Levice (2)- Zlaté Moravce (6) - Nitra (0)
Nitra (0) - Nové Zámky (3) - Šaľa (4) - Nitra (0)
Nitra (0) - Topoľčany (5) - Nitra (0)
Nitra (0) - Komárno (1) - Nitra (0)
Implementing model (16) – (27):
Objective value: *r = 1.717*,
The routes of own vehicles: Nitra (0) - Levice (2)- Zlaté Moravce (6) - Nitra (0)
Nitra (0) - Nové Zámky (3) - Šaľa (4) - Nitra (0)
Nitra (0) - Topoľčany (5) - Nitra (0)
The route realized by rented vehicle: Nitra (0) - Komárno (1)

Based on results listed above it can be stated the following: The total cost of the routes could be decreased by car rental if the ratio between the cost per km of rented and own vehicle is less than *1.717* (supposing other conditions are the same and the fixed coefficient of rented vehicle was set to 15).

Conclusion

This paper considers the version of open vehicle routing problem (OVRP) that allows combining CVRP together with the possibility to realize open routes. The mathematical formulation was provided on the base of mixed integer programming (MIP). Presented formulation allows setting the limit value which indicates the maximal rate (calculated as cost ratio between the cost of own and rented vehicles) when it is effective renting at least one external vehicle. The computational experiments were based on regional data in Slovakia. Software implementation was realized in GAMS (solver Couenne 23.5.1).

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/0245/15 „Transportation planning focused on greenhouse gases emission reduction“.

References

[1] Babej, A., Gežík, P., Mieresova, L.: Waste collection - modelling based on vehicle routing problem. In: *Nové trendy v ekonometrii a operačním výzkumu 2013* (Jablonský, J., Reiff, M. , eds.). Vydavateľstvo EKONÓM, Bratislava, (2013).
[2] Čičková, Z., Brezina, I.: An evolutionary approach for solving vehicle routing problem. In: *Quantitative methods in economics multiple criteria decision making XIV*. IURA EDITION, Bratislava, (2008), 40–44.
[3] Čičková, Z., Brezina, I., Pekár, J.: *Solving the Real-life Vehice Routing Problem with Time Windows Using Self Organizing Migrating Algorithm*. Ekonomický časopis **61**(5/2013), 497–513.
[4] Čičková, Z., Brezina, I., Pekár, J.: *Open vehicle routing problem.* In: *Mathematical Methods in Economics 2014*. Olomouc: Faculty of Science, Palacký University, 2014 124 – 127.
[5] Desrochers, M., Desrosiers, J., Solomon, M.: A New Algorithm for the Vehicle Routing Problem with Time Windows. *Operations Research*. **40** (1992), 342-354.
[6] Fábry, J., Kořenář, V., Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Mathematical Methods in Economics 2011 [CD-ROM]*. Praha, PROFESSIONAL PUBLISHING, (2011), 140–144.
[7] Miller, C.E., Tucker, A.W., Zemlin, R.A.: *Integer programming Formulation of Traveling Salesman Problems*. Journal of the ACM, 7(4/1960), 326-329.
[8] Palúch, S., Peško, Š.: *Kvantitatívne metódy v logistike*. EDIS vydavateľstvo ŽU, Žilina, 2006.

Nonlinear model of the Eurozone labor market

Ondřej Čížek[1]

Abstract. A nonlinear model of unemployment is formulated and applied in this paper. The formulation is motivated by a Keynesian principle of weak aggregate demand. The basic idea is that the rise of unemployment causes the fall of aggregate demand for final goods. This in turn leads to the fall of the demand for labor, which leads to even higher unemployment. The model is econometrically estimated for the Eurozone and it is shown that the described mechanism gives rise to the multiplicity of the equilibrium rate of unemployment. Factors influencing the position of the equilibrium points are studied and economic implications are made from this analysis.

Keywords: nonlinear modeling, unemployment, aggregate demand, multiple equilibria.

JEL Classification: E24, J23, J64
AMS Classification: 91G70

1 Introduction

Europe today is characterized by high unemployment as well as weak aggregate output demand. For this reason, the model formulated in this article is based on the following transition mechanism: "*High unemployment leads to low consumption demand. Production is also low as no one can afford to buy the goods, which sustains high unemployment.*" This formulation principle is supported by the Nobel Prize winner in economics, Joseph Stiglitz [4], who states that Europe's problem today is a lack of aggregate demand. This article contributes to the existing literature by incorporating the Keynesian principle of weak aggregate demand into the basic model of the labor market.

The formulated model will be econometrically estimated using the data from the Eurozone. The main contribution of the paper is the finding that there are multiple equilibrium unemployment rates for the estimated model. Multiplicity of equilibriums is caused by the above mentioned transition mechanism. More specifically, there is low demand for labor in times of high unemployment. Therefore, it is hard to find a job, which sustains unemployment at high levels, preventing its return to a lower more effective equilibrium.

2 Model

2.1 Unemployment dynamics

Unemployment is modeled in a continuous time environment with data available only at discrete dates. Shimer's (2012) methodology is used to express the dynamics of unemployment in discrete time in the following form

$$U_{t+1} = \frac{s_t}{s_t + f_t} \cdot \left(1 - e^{-(s_t + f_t)}\right) \cdot L_t + e^{-(s_t + f_t)} \cdot U_t \,, \tag{1}$$

where U_t is the number of unemployed,

L_t represent the labor force,

s_t is separation rate and f_t represents job-finding rate.

The measurement of transition rates f_t, s_t is also based on Shimer's (2012) methodology. According to his evidence, there are substantial fluctuations in job finding probability during business cycle frequencies, while separation probability is nearly acyclic. This suggests that, in order to understand fluctuations in unemployment, one must understand the fluctuations in job-finding probability. The formulation of the model presented in this paper is based upon this result. The emphasis will, therefore, be given to the model of job-finding probability.

[1] University of Economics, Prague, W. Churchill Sq. 4, 130 67, Prague 3, Czech Republic.

2.2 Matching function

Standard modeling device is the aggregate matching function $M(\bullet)$. The flow of workers from unemployment to employment during a given period $[t,t+1)$ will be assumed to be related to the number of unemployed workers U_t and (unfilled) job vacancies V_t at the beginning of the time interval $[t,t+1)$, according to the Cobb-Douglas matching function with constant returns to scale

$$M_t \equiv M(V_t,U_t) = A \cdot U_t^{\alpha} \cdot V_t^{1-\alpha}, \; A>0, \; \alpha \in (0,1). \tag{2}$$

The assumption of constant returns to scale is in line with most empirical work (see survey performed by Pissarides, Petrongolo [2]).

Another common presumption is random matching, which means that all unemployed workers U_t have an equal sampling probability. In this case, the job-finding probability F_t is[2]

$$F_t \equiv \frac{M_t}{U_t} = A \cdot \theta_t^{1-\alpha}, \tag{3}$$

where $\theta_t \equiv V_t / U_t$ denotes the market tightness.

The corresponding job finding rate is given by

$$f_t = -\ln(1-F_t). \tag{4}$$

2.3 Labor demand

The model of the demand for labor is inspired by the Keynesian principle of weak aggregate demand. It is based on the following transition mechanism:

↓ unemployment rate → ↑ purchasing power → ↑ aggregate demand for final products →
↑ demand for labor → ↑ vacancies posted by firms.

This mechanism is modeled by the following linear approximation:

$$V_t = a - b \cdot U_t, \tag{5}$$

where V_t is the number of vacancies posted by firms.

The relation (5) can will be expressed in terms of the market tightness by dividing both sides by the number of unemployed:

$$\theta_t = a \cdot \frac{1}{U_t} - b. \tag{6}$$

From a practical point of view, it will be more convenient to use unemployment rate u_t instead of the number of unemployed U_t, which leads to:

$$\theta_t = a \cdot \frac{1}{u_t} - b, \tag{7}$$

where u_t is the unemployment rate.

In order to ensure that $\theta_t \geq 0$, the relation (7) is further modified as follows:

[2] The corresponding job finding rate is given by $f_t = -\ln(1-F_t)$.

$$\theta_t = \begin{cases} a \cdot \dfrac{1}{u_t} - b, & \text{for } u_t < \overline{u}, \\ \underline{\theta}, & \text{for } u_t \geq \overline{u}. \end{cases} \tag{8}$$

where $\underline{\theta} \geq 0$ is a lower bound of the market tightness.

3 Data

Data from the Eurozone[3] were used in the application part of this article. All such data is available at the database of the Eurostat, which can be found online at:

http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.

Unemployment U_t was obtained from the series „Unemployment by sex and age groups –monthly average, 1000 persons (une_nb_m)".

Job-finding probability F_t is measured by the Shimer's (2012) method. To apply this methodology, the series for the short-term unemployed persons have to be used. Specifically, the number of workers, whose unemployment has not exceeded one quarter is used in this paper, which can be found in the Eurostat database under title „Unemployment by sex, age and duration of unemployment (lfsq_ugad)".

Market tightness is defined as $\theta_t \equiv V_t / U_t$, so it remains to describe data on unfilled vacancies V_t. This variable was retrieved from a job vacancy rate v_t. The title of the corresponding time series in the Eurostat database is „Job vacancy statistics – quarterly data, NACE rev. 2 (jvs_q_nace2)".

4 Econometric estimation

Firstly, the stochastic version of the regression (3) was estimated

$$F_t = A \cdot \theta_t^{1-\alpha} \cdot e^{\varepsilon_t}, \tag{9}$$

where $\alpha \in (0,1)$ and ε_t is i.i.d. random error.

The estimation was performed by ordinary least squares (OLS) after log-linearization for the data ranging from 2005 Q1 to 2014 Q2. The results are as follows:[4]

$$\ln\left(\hat{F}_t\right) = \underset{(0.08)}{-1.03} + \left(1 - \underset{(0.05)}{0.66}\right) \cdot \ln\left(\theta_t\right), \; R^2 = 0.59. \tag{10}$$

The estimate $\hat{\alpha} = 0.66$ is in line with the results of other empirical studies, which are summarized by Pissarides and Petrongolo (2001) and according to which this parameter ranges from 0.2 to 0.8.

Secondly, the regression (8) was estimated in the following form

$$\theta_t = a \cdot \frac{1}{u_t} - b + \eta_t, \tag{11}$$

where η_t is i.i.d. random error.

The parameter $\underline{\theta}$ is not estimated econometrically. The lowest value of market tightness was attained in 2013 and was approximately equal to 0.12. We definitively conclude that $\underline{\theta} < 0.12$. However, it is impossible to

[3] The data used is expressed at fixed composition whenever possible. Fixed composition means that the whole time series contains the data for all 18 countries, even if the evolution of the Eurozone was changing during the time. The Eurozone was set up in 1 January 1999 by the following 11 countries: Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain. Another countries joined the Eurozone later on: Greece (2001), Slovenia (2007), Cyprus and Malta (2008), Slovakia (2009), Estonia (2011) and Latvia (2014).

[4] Standard errors of the estimated coefficients are indicated in parentheses below the parameters.

say anything more concerning the value of the $\underline{\theta}$ parameter on the basis of the historical data. For this reason, the value of the $\underline{\theta}$ parameter is not obtained using econometric techniques. Instead, various plausible values of this parameter will be taken into account and consequences to the model properties will be analyzed.

The regression (11) was estimated by standard OLS for $t \in \{2003\ \text{Q4}, ..., 2014\ \text{Q3}\}$ as follows:

$$\hat{\theta}_t = \underset{(0.002)}{0.03} \cdot \frac{1}{u_t} - \underset{(0.02)}{0.18}, \quad R^2 = 0.86. \tag{12}$$

5 Equilibrium unemployment rates

Under the assumption of constant labor force $L_t = L$ and separation rate $s_t = s$, the equation (1) is slightly modified as follows:

$$u_{t+1} = \frac{s}{s + f_t} \cdot \left(1 - e^{-(s+f_t)}\right) + e^{-(s+f_t)} \cdot u_t,$$

where $u_t = U_t / L$ is unemployment rate.

This equation implies that a stationary unemployment rate $u_t = u$ satisfies

$$0 = \left(\frac{s}{s + f(u)} - u \right) \cdot \left(1 - e^{-(s+f(u))}\right), \tag{13}$$

where $s = 0.02$ is the separation rate in 2014 Q3 (the last date in the dataset),
$f(u)$ indicates that a stationary value of f is a function of a stationary unemployment rate u.

In order to describe the function $f(u)$ in more detail, let's start with the equilibrium value of market tightness, which is obtained from (12)

$$\theta(u) = 0.03 \cdot \frac{1}{u_t} - 0.18. \tag{14}$$

From now on, I will return to the specification (8), which yields a stationary value of market tightness in the following form

$$\theta(u) = \max\left[\left(0.03 \cdot \frac{1}{u_t} - 0.18 \right), \underline{\theta} \right]. \tag{15}$$

Different values of the lower bound $\underline{\theta} \geq 0$ will be discussed later in this section.

The probability of finding a job in a stationary state is obtained from (10)

$$F(u) \equiv e^{-1.03} \cdot \theta(u)^{1-0.66} \tag{16}$$

and the corresponding job-finding rate is

$$f(u) \equiv -\ln(1 - F(u)), \tag{17}$$

which defines the function $f(\bullet)$ in the equation (13).

Equation (13) is nonlinear, hence multiple solutions may exist. The expression on the right hand side of (13) is a function of the variable u, which will be denoted $g(u)$

$$g(u) \equiv \left(\frac{s}{s + f(u)} - u \right) \cdot \left(1 - e^{-(s+f(u))}\right). \tag{18}$$

The function $g(u)$ calculated on the assumption that $\underline{\theta} = 0.01$ (graph (a)) and $\underline{\theta} = 0.04$ (graph (b)) is depicted at the following figure

lower bound $\underline{\theta} = 0.01$ | lower bound $\underline{\theta} = 0.04$

Figure 6a goes here | Figure 6b goes here

(a) | (b)

Figure 1: Equilibrium unemployment rates for different values of the lower boundary of market tightness $\underline{\theta}$.

In the graph (a), we can see that there are three solutions of the equation $g(u) = 0$. Therefore, three equilibrium unemployment rates $u^1 = 0.08$, $u^2 = 0.14$ and $u^3 = 0.21$ exist in this case. The equilibriums u^1 and u^3 are stable, while the equilibrium u^2 is unstable. This follows immediately from the fact that the value $g(u_t)$ can be interpreted as $u_{t+1} - u_t$. In the graph (b), it is shown that the equilibrium u^3 disappears if the lower boundary for the market tightness is at least 0.04.

6 Economic consequences of multiplicity

Market tightness in the Eurozone was 0.34 at the beginning of the economic crisis in 2007. Since then, it has been practically falling steadily and reached a minimum value of 0.13 in 2013.

From these facts, it seems to me that the market tightness could quite easily fall below 0.04. A multiple equilibrium model of the labor market is thus more suitable for describing unemployment dynamics in the Eurozone during the current economic crisis than a traditional model with unique equilibrium.

The existence of multiple equilibrium unemployment rates in my model can be explained by a less effective labor market during times of high unemployment. Firms open only a few vacancies during a recession (crisis) because demand for their output is low. For this reason, it is hard for unemployed workers to find jobs, which maintains unemployment at high levels and prevents its return to a lower equilibrium point.

Note that the unemployment rate in the Eurozone was 0.12 in 2013, which is very near to the unstable equilibrium u^2. The future dynamics of the unemployment rate are thus unclear. It is probably the case that a small negative shock would cause the unemployment rate to converge to the equilibrium u^3. On the other hand, a positive shock may lead to convergence at the point u^1.

The message for policymakers is that it seems reasonable not to wait until it is too late to make the unemployment rate converge to the lower equilibrium u^1. Some years from now, the unemployment rate might be on its way to the equilibrium u^3 well above the point u^2.

7 Conclusion

This article contributes to the existing literature by introducing the Keynesian principle of weak aggregate demand into the basic model of the labor market.

The significant finding is that incorporating the principle of weak demand gives rise to a multiplicity of equilibrium unemployment rates. It was shown in this paper that there are two stable equilibrium points, provided that the number of unfilled vacancies (market tightness) is sufficiently low in times of high unemployment. Similar results were obtained by Kaplan, Menzio (2015), who calibrated their model for the US economy and found that the feedback between employment and product market generates multiple equilibriums.

There are two stable equilibrium unemployment rates for the model econometrically estimated for the Eurozone. The lower stable equilibrium unemployment rate attains a value of 0.08. The higher stable equilibrium emerges when market tightness is allowed to fall below a value of 0.04. The lower it is allowed to fall, the higher the second stable equilibrium. Specifically, if the lower boundary for market tightness is 0.01 then a higher equilibrium would attain a value 0.21.

There is also an unstable equilibrium unemployment rate, which equals 0.14. The importance of this lies in the fact that the unemployment rate in the Eurozone was 0.12 in 2013. The labor market in the Eurozone is thus very close to unstable equilibrium these days. In such a situation, only a tiny negative shock could cause the

unemployment rate to converge to a higher stable equilibrium. Similarly, only a small positive shock may lead the convergence to the lower equilibrium point.

Acknowledgements

Financial support of VŠE IGA IG403014 and VŠE IP400040 is gratefully acknowledged by the author.

References

[1] Kaplan, G., and Menzio, G. Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. *Journal of Political Economy* (forthcoming), 2015.
[2] Pissarides, C. A., and Petrongolo, B. Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, **39** (2001), 390-431.
[3] Shimer, R. Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics*, **15(2)** (2012), 127-148.
[4] Stiglitz, J. E. Crisis: Principles and Policies: With an Application to the Eurozone Crisis. In: *Proceedings of the International Economics Association's 2012 Buenos Aires round-table*, 2013.

Revisited Zero-Crossing Method for Hurst Exponent Estimation in Time Series

Martin Dlask[1], Jaromir Kukal[1], Quang Van Tran[1]

Abstract. Fractional Gaussian noise (fGn) is a stochastic process whose properties can be used for estimating Hurst exponent (H) and fractal dimension of original integrated process. Traditional methods provide point estimation of zero-crossing probability, and therefore only single value of the exponent H. Our innovative approach is based on signal segmentation – using Bayesian analysis we can determine mean and confidence interval for Hurst exponent in each particular segment. Moreover, we are able to obtain its aggregate estimate by utilizing all statistical features calculated at the interval level. The optimal segmentation of fGn sample is performed on the basis of unimodal principle. The methodology of optimal segmentation is subsequently applied to the analysis of stock market indices and the results are compared with values obtained by using standard technique of fractal dimension estimation.

Keywords: Time series, fractional Gaussian noise, zero-crossing rate, Bayesian approach, optimal segmentation, stock market.

JEL classification: 60G22
AMS classification: C11

1 Introduction

Precise estimation of the Hurst parameter of a time series is one of the most difficult problems in fractal geometry. Since the most mathematical formulas are valid only in fully continuous world, it is extremely important to develop accurate estimation methods for analysing discrete real-valued data, such as signal samples, time series or samples of stochastic processes.

Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn), invented by Mandelbrot [7], belong to the well-known continuous random processes with fractal patterns. These two processes are both dependent on Hurst parameter H and can be converted to each other by means of differencing in the case fBm→fGn or cumulative sum in the latter case. Therefore, it is sufficient to investigate one of them at once and utilize its properties for accurate parameter estimation. However, not only accurate point estimate is essential for fractal analysis, the model error and other statistical indicators need to be taken into account as well. Since the majority of financial time series tend to have independent increments causing their Hurst parameter to be very close to 0.5 (Sang [9], Couillard [3]), it is necessary to obtain additional information regarding to the fractal dimension.

In this paper we investigate the properties of fGn and on the basis of traditional zero-crossing method we improve the point estimate of H with mentioned additional characteristics such as standard deviation and appropriate confidence intervals. Whereas the standard approach does not provide any, our improved method can recognise more fractal patterns in the time series and therefore it can be a starting point for later multi-fractal analysis.

2 Traditional zero-crossing approach

The number of zero-crossings (or level crossings) of Gaussian processes and its relevant fractal properties were published in several papers by Azais [1], Feuerverger [4] and Couerjolly [2]. The main result of these studies is the dependence of Hurst parameter on the probability of zero-crossing of random continuous

[1]FNSPE CTU, Brehova 7, Prague, Czech Republic,
dlaskma1@fjfi.cvut.cz, jaromir.kukal@fjfi.cvut.cz, quangvan.tran@fjfi.cvut.cz

nowhere differentiable process at fixed interval. Our aim is to improve the method by analysing the number of zero-crossings of fGn process and to deduce the formula another way.

Suppose $x_k = x_k(H)$ is a sampling of fGn process with parameter H, zero mean and unit variance on fixed interval for $k = 1, ..., N+1$. The appropriate autocorrelation function ρ_k [7] equals

$$\rho_k = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right). \tag{1}$$

We denote the first autocorrelation coefficient by $\rho \equiv \rho_1 = 2^{2H-1} - 1$. The probability density function of two consecutive sample elements is

$$f(x_k, x_{k+1}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x_k^2 - 2\rho x_k x_{k+1} + x_{k+1}^2\right)\right), \tag{2}$$

therefore the probability of zero-crossing equals

$$p^* = \text{prob}\left(x_k x_{k+1} < 0\right) = \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^{+\infty}\int_0^{+\infty} \exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2 + 2\rho xy + y^2\right)\right) \mathrm{d}x\mathrm{d}y. \tag{3}$$

Using the substitution $\xi = (\rho x + y)(1-\rho^2)^{-\frac{1}{2}}$ and $\eta = x$ one obtains

$$p^* = \frac{1}{\pi} \iint_{O^*} \exp\left(-\frac{\xi^2 + \eta^2}{2}\right) \mathrm{d}\xi\mathrm{d}\eta, \tag{4}$$

where $O^* = \left\{\left(\rho x + y), x\sqrt{1-\rho^2}\right) \in \mathbf{R}^2 : x, y \in \mathbf{R}_0^+\right\}$. Substituting $\xi = r\cos\varphi$ and $\eta = r\sin\varphi$ the crossing probability can be expressed as

$$p^* = \frac{1}{\pi} \int_{\varphi_{\min}}^{\varphi_{\max}} \mathrm{d}\varphi \cdot \int_0^{+\infty} \exp\left(-\frac{r^2}{2} r\right) \mathrm{d}r = \frac{\varphi_{\max} - \varphi_{\min}}{\pi}. \tag{5}$$

Due to the form of area O^* we have $\varphi_{\min} = 0$ and $\varphi_{\max}$ equals

$$\varphi_{\max} = \begin{cases} \arctan\frac{\sqrt{1-\rho^2}}{\rho} & \rho > 0 \\ \pi - \arctan\frac{\sqrt{1-\rho^2)}}{|\rho|} & \text{otherwise} \end{cases}. \tag{6}$$

Utilizing equations (5) and (6), it is possible to express the probability p^* as

$$p^* = \frac{1}{\pi}\,\text{arccot}\,\frac{\rho}{\sqrt{1-\rho^2}} \tag{7}$$

and finally using the identity $\rho = 2^{2H-1} - 1$, the point estimate of Hurst parameter H equals

$$H = 1 + \log_2 \cos\frac{\pi p^*}{2}, \tag{8}$$

which is revisited form of well-known formula published by Couerjolly [2], originally in the form

$$H = \frac{1}{2}\left(1 + \log_2(1 + \epsilon|\cos(\pi \cdot S_N)|)\right), \tag{9}$$

where the relative number of zero crossing S_N with respect to the sample length N and $\epsilon = \text{sgn}\left(H - \frac{1}{2}\right)$ are assumed to be known.

3 Bayesian estimation of crossing probability

We already know how to obtain point estimate of H with given fGn sample of length $N+1$ – if the number of zero-crossings in the sample equals Z, we can approximately express H as

$$H \approx 1 + \log_2 \cos \frac{\pi Z}{2N}. \tag{10}$$

Unfortunately, from formula (10) we do not know the distribution of Hurst exponent H neither the distribution of zero-crossings p^*. Suppose, that the probability p^* is known for investigated sample. Then the number of zero-crossings Z comes from binomial distribution $Z \sim \mathrm{Bi}(N, p^*)$ with probability density function

$$f(Z|p^*) = \binom{N}{Z} (p^*)^Z \cdot (1-p^*)^{N-Z}. \tag{11}$$

Using the Bayesian inverse rule [5], the posterior probability density equals

$$f(p^*|Z) = \frac{(p^*)^Z (1-p^*)^{N-Z}}{\mathrm{B}(N+1, N-Z+1)}, \tag{12}$$

where $\mathrm{B}(p,q)$ is the standard beta function. We can generalize the result by means of using Dirichlet prior as

$$f_{\mathrm{PRIOR}}(p^*) = \frac{(p^*)^{-\alpha}(1-p^*)^{-\alpha}}{\mathrm{B}(1-\alpha, 1-\alpha)}, \tag{13}$$

for parameter $\alpha \in [0,1)$. After the application of Bayesian rule, we get the posterior probability for p^* in the form

$$f_{\mathrm{POST}}(p^*|Z) = \frac{(p^*)^{Z-\alpha}(1-p^*)^{N-Z-\alpha}}{\mathrm{B}(Z+1-\alpha, N-Z+1-\alpha)}, \tag{14}$$

therefore the probability p^* comes from Beta distribution $p^* \sim \mathrm{B}(Z+1-\alpha, N-Z+1-\alpha)$. Different estimates of p^* can be obtained for different choices of parameter α, however it is reasonable to consider only natural Bayesian approach with $\alpha = 0$ or Jeffreys-Perks law [6],[8] with $\alpha = 1/2$.

4 Segmentation and Hurst exponent estimation

When trying to find different fractal patterns (different H values) in the investigated signal, it is necessary to divide the original sampling into disjoint segments and determine the statistical features in each interval separately. Function (14) describes the probability of zero-crossing in the fGn sample in general – considering the segmentation into L disjoint segments, each with $N+1$ elements, the probability $f_{\mathrm{L}}(p)$ of zero-crossing of the whole signal sample with $L \cdot (N+1)$ elements equals

$$f_{\mathrm{L}}(p) = \frac{1}{L} \sum_{k=1}^{L} \frac{p^{Z_k-\alpha}(1-p)^{N-\alpha-Z_k}}{\mathrm{B}(Z_k+1-\alpha, N-Z_k+1-\alpha)}, \tag{15}$$

where Z_k is the number of zero-crossings in k-th segment. The function $f_{\mathrm{L}}(p)$ should be unimodal for all L only under the fGn assumption of the whole sample with the same H in each block. Nevertheless, considering real data, where the input sample contains significantly different number of zero-crossings in each segment, the function $f_{\mathrm{L}}(p)$ is likely to have more peaks whose amount is dependent on parameter $L > 1$. For certain L^* large enough the function becomes unimodal and we denote this number as the *optimal segmentation* of the given sample defined as

$$L^* = \begin{cases} 1 & f_L(p) \text{ is unimodal for all } L \in \mathbf{N} \\ \min\{L > 1 : f_L(p) \text{ is unimodal}\} & \text{otherwise} \end{cases}$$

On the basis of the function $f_L(p)$ we can determine the boundaries $p_{\min}, p_{\max}$ of confidence interval for p with confidence level $\widetilde{\alpha}$ as

$$\int_0^{p_{\min}} f_L(p)dp = \frac{\widetilde{\alpha}}{2}, \tag{16}$$

$$\int_0^{p_{\max}} f_L(p)dp = 1 - \frac{\widetilde{\alpha}}{2}. \tag{17}$$

Equation (15) together with point estimate of H from equation (8) provide the expected value of H in the form

$$\mathrm{E}(H) = 1 + \int_0^1 f_L(p) \log_2 \cos \frac{p\pi}{2} \mathrm{d}p \tag{18}$$

and variance σ^2

$$\sigma^2 = \int_0^1 \left(1 + \log_2 \cos \frac{p\pi}{2}\right)^2 f_L(p)\mathrm{d}p - (\mathrm{E}\, H)^2. \tag{19}$$

Finally, we obtain the boundaries of confidence interval for H as

$$H_{\min} = 1 + \log_2 \cos \frac{p_{\max}\pi}{2}, \tag{20}$$

$$H_{\max} = 1 + \log_2 \cos \frac{p_{\min}\pi}{2}, \tag{21}$$

where $p_{\min}$ and $p_{\max}$ were calculated from (16) and (17) with given confidence level $\widetilde{\alpha}$.

5 Application to stock market indices

Once the principle of optimal segmentation is known, it can be applied to the development of stock market indices. We analysed the segmentation of nine time series showing the performance of each stock market. At first it is necessary to investigate the behaviour of segmentation-dependent probability density. Probability density function $f_L(p)$ for TSX stock market index is depicted in Figure 1 for $\alpha = 1/2$ and $L = 2$ (top left), 4 (top right), 7 (bottom left) and optimal segmentation $L^* = 12$ (bottom right). Results for $\alpha = 0$ show similar behaviour, therefore we consider only the case $\alpha = 1/2$ in the following calculations. Statistical calculations and visualisations were performed in the Matlab environment and therefore all the tables and figures included in this paper are results of own computations.

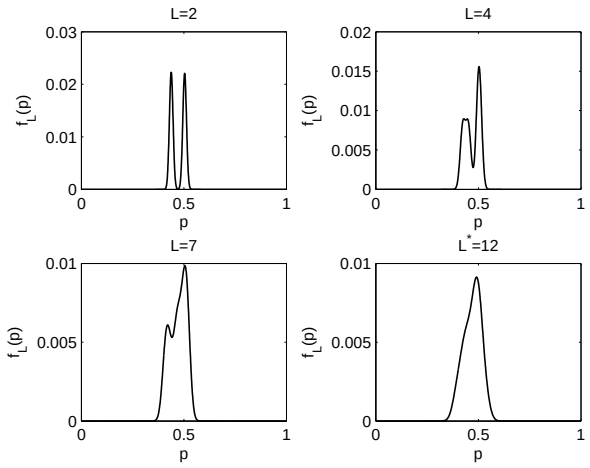


Figure 1 Bayesian density for various sizes of segmentation blocks of TSX.

With increasing number of segments L the support of function $f_L(p)$ extends while the total range decreases causing the unimodal property for large parameters L. It is generally not true that the function is unimodal for all parameters $L > L^*$. We demonstrate this fact in Figure 2 showing the 95% confidence intervals and mean values of H for all feasible divisions with unimodal $f_L(p)$. The optimal segmentation occurs for $L^* = 5$, however, the function has not unique peak for all $5 < L < 10$. Once the segmentation on L intervals is fine enough, the new additions to segmentation-dependent probability distribution are no longer significant to spoil the trend of the whole function obtained by dividing into $L - 1$ blocks.

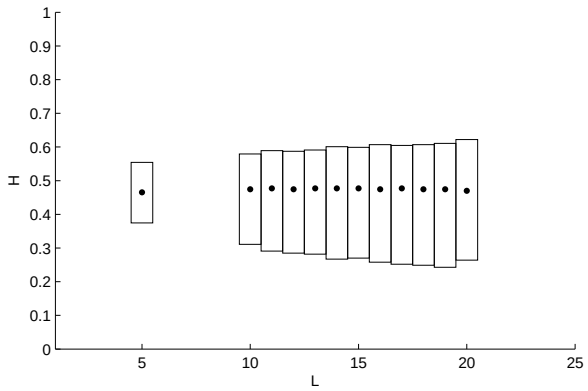


Figure 2 95% confidence intervals for feasible segmentations of CAC40.

Table 1 shows the traditional point estimate (10) of H together with expected values and 95% confidence intervals obtained by Bayesian approach without segmentation (for $L = 1$) and revisited approach showing the parameter of optimal segmentation L^* and appropriate statistical characteristics. When the parameter of optimal segmentation equals 1 (for DAX,HSI and SMI), it means the function $f_L(p)$ is unimodal from the beginning signifying more or less constant parameter H in the entirety of the signal. NASDAQ and TSX are the stock markets whose Hurst parameters varies in time a lot, because their optimal segmentation parameter equals 14 and 12, respectively.

index	traditional method	no segmentation ($L = 1$)		optimal segmentation		
	H^*	EH	95% CI	L^*	EH	95% CI
CAC40	0.4745	0.4672	(0.4435; 0.4909)	5	0.4652	(0.3746;0.5545)
DAX	0.5007	0.4790	(0.4556; 0.5023)	1	0.4790	(0.4556;0.5023)
FTSE	0.5502	0.4906	(0.4676; 0.5135)	3	0.4863	(0.4361;0.5482)
HSI	0.5664	0.4974	(0.4746; 0.5201)	1	0.4974	(0.4746;0.5201)
NASDAQ	0.5631	0.5584	(0.5375; 0.5792)	14	0.5418	(0.4008;0.7136)
NIKKEI	0.4583	0.4553	(0.4312; 0.4793)	3	0.4532	(0.4034;0.5068)
SMI	0.5303	0.5087	(0.4863; 0.5310)	1	0.5087	(0.4863;0.5310)
SP500	0.4722	0.4456	(0.4212; 0.4699)	6	0.4460	(0.3225;0.5628)
TSX	0.6107	0.5646	(0.5439; 0.5852)	12	0.5607	(0.3957;0.6992)

Table 1 Comparison of traditional and revisited method.

6 Conclusion

Knowing the probability density function of both Hurst exponent and zero-crossing rate, it is possible to determine new statistical characteristics for H. Due to the application of novel methodology, we obtain usually wider confidence intervals for the majority of stock markets. However, the optimal expected values remain almost the same in comparison with the case without segmentation. In most cases, larger confidence interval indicate that the specific index does not follow the same fractal behaviour during the entire investigated period. Therefore the method is able to find new fractal patterns by means of dividing original signal – when the optimal segmentation is reached, the estimates are more realistic and therefore useful for further analysis such as investment recommendations or predictions.

Acknowledgements

The paper was created with the support of CTU in Prague, Grant SGS14/208/OHK4/3T/14.

References

[1] Azais, J. M.: Conditions for convergence of number of crossings to the local time: Application to stable processes with independent increments and to Gaussian processes. *Probability and mathematical statistics* **11.**, 1 (1990), 19–36.

[2] Couerjolly, J. F.: Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *Journal of statistical software* **5.**, 7 (2000), 1-53.

[3] Couillard, M., Davison M.: A comment on measuring the Hurst exponent of financial time series. *Physica A* **348.**, 2005, 404–418.

[4] Feuerverger, A., Hall, P., Wood, A. T. A.: Estimation of Fractal Index and Fractal Dimension of Gaussian Process by Counting the Number of Level Crossings. *Journal of Time Series Analysis* **15.**, 6 (1994), 587-606.

[5] Gelman, A.: *Bayesian Data Analysis*, Chapman and Hall/CRC, 2003.

[6] Jeffreys, H: An invariant form for the prior probability in estimation problems. *Proc. Royal Soc. London (A)* **186.** 1007 (1945) ,453–461.

[7] Mandelbrot, B. B., Ness, J. W.: Fractional Brownian motions, Fractional Noises and Applications. *SIAM Review* **10.**, 1968, 422–437.

[8] Perks, W.: Some observations on inverse probability. *J. Inst. Actuar.* **73.** 2 (1947), 285–312.

[9] Sang, H., Ma, T., Wang, S.: Hurst exponent analysis of financial time series. *Journal of Shanghai University (English Edition)* **5.** 4(2001), 269–272.

Determinants of job and skill mismatch in the graduate labour market

Zuzana Dlouhá[1]

Abstract. The paper investigates determinants of job and skill mismatch among the university graduates from the University of Economics, Prague. The data were obtained from the international cross-sectional Flexible Professional in the Knowledge Society (REFLEX) survey that collected information regarding the labour market status of the higher education graduates five years after graduation. The method of binary logit model with vertical and horizontal mismatch as dependent variables was used. As factors of mismatches graduate's characteristics, education and early career experience were employed in the models. Several hypotheses were tested, such as advantage of male graduates and graduates with highly educated parents for a better match between education and occupation, or that type of contract at current work of the respondent can work as trade-off between tenancy and job match.

Keywords: binary logit model, labour market mismatch, REFLEX survey

JEL Classification: I23, J24
AMS Classification: 62P20

1 Introduction

Qualitative mismatch between educational level of employees and that required by the jobs available in the labour market is one of the most discussed dimensions of the job and skill mismatch that mainly affects university graduates. The discrepancies between qualifications and skills that individuals possess and those needed by the labour market negatively affect economic competitiveness and growth, increase unemployment, undermine social inclusion, and generate significant economic and social costs. From the micro-level point of view mismatches in job/education and skills lead to lower job satisfaction and lower salary than properly matched [4], but we have to be careful when interpreting these conclusions: educational requirements for a certain job can rise over time, individuals can be overqualified due to low ability for that level of qualification or they chose to work at less stressful work, etc. [2]

In the study, we partially followed the approach introduced in [3] where determinants of mismatching in four countries (Hungary, Poland, Lithuania Slovenia) were investigated such as age, higher parental education, study-related work experience, etc. The main finding of the study is that mismatch in first occupation has strong and long lasting effect on the job match even five year after the graduation. We were also inspired by the study in which factors of dissatisfaction with chosen study programme were investigated and variable horizontally mismatched at first job were confirmed as statistically significant [1].

The aim of this paper is to determine factors of job and skills mismatch among the university graduates from the University of Economics, Prague. We tested the hypotheses as follows: male graduates and graduates with highly educated parents, and graduates with work experience (both study and non-study-related) can have an advantage for a better match between education and occupation, and that type of contract can work as a trade-off between tenancy and job match.

The paper is composed of chapters as follows. Chapter 2 describes the methodology and dataset together with variables included in the models, Chapter 3 explores the determinants of educational (horizontal and vertical) and skills mismatch among university graduates. In Chapter 4 we summarize our findings.

2 Methodology and data

We briefly introduce methodology used for determining the factors of job and skill mismatch within the following section. Next we describe the analysed dataset and provide basic descriptive statistics of the variables incorporated into models.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, e-mail: figlova@vse.cz.

2.1 Binary choice model

We consider a class of binary response models of the form [5]:

$$\mathrm{P}(y = 1|\boldsymbol{x}) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(\beta_0 + \boldsymbol{x\beta}), \tag{1}$$

where G is a function taking on values strictly between 0 and 1: $0 < G(z) < 1$, for all real numbers z. For the estimation of response probabilities we use logit model, hence G is the logistic function:

$$G(z) = \frac{e^z}{1+e^z} = \Lambda(z), \tag{2}$$

which has values between 0 and 1 for all real numbers z and this is the cumulative distribution function for a standard logistic random variable. The expression P/(1 – P) presents *odds ratio* in favour of the occurrence of an event. If we take the natural log of this expression, the result is

$$L = ln\frac{\mathrm{P}}{1-\mathrm{P}} = \beta_0 + \boldsymbol{x\beta}, \tag{3}$$

and L (logit, hence the name logit model) is the natural log of the odds ratio linear in $\boldsymbol{x}$ and in parameters.

We use the maximum likelihood method for the estimation of the logit model. This method is consistent, normally distributed and efficient in large samples.

Marginal (partial) effect for continuous explanatory variable on $p(x) = \mathrm{P}(y = 1|\boldsymbol{x})$ is obtained from the partial derivative:

$$\frac{\partial p(\boldsymbol{x})}{\partial x_j} = g(\beta_0 + \boldsymbol{x\beta})\beta_j, \tag{4}$$

where $g(z) = \frac{\mathrm{d}G}{\mathrm{d}z}(z)$ is probability density function associated with G. Due to non-negativity of the density function, the partial effect of the x_j will always have the same sign as β_j.

Marginal (partial) effect for discrete explanatory variable on the probability x_k going from c_k to $c_k + 1$ is

$$G[\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k(c_k + 1)] - G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k c_k). \tag{5}$$

Standard errors of marginal effects (4) and (5) can be calculated using e.g. delta method [4].

To analyse the correctness of the model several tests will be run. For testing multiple restrictions the likelihood ratio (LR) as a twice the difference in the log-likelihoods is used:

$$LR = 2(L_{ur} - L_r), \tag{6}$$

where L_{ur} is the log-likelihood for the unrestricted model and L_r is the log-likelihood for the restricted model. We test the joint null hypothesis that all slope coefficients are zero. As a goodness of fit we calculate the following McFadden's R^2:

$$\text{McFadden's } R^2 = 1 - (L_{ur} / L_r) \tag{7}$$

Hosmer-Lemeshow and Andrews χ^2 tests as another comparison of the fitted expected values to the actual values were used.

2.2 Data and variables

The data were obtained from the international cross-sectional Flexible Professional in the Knowledge Society (REFLEX) survey that was held by the Education Policy Center, Charles University, Prague, in 2013. This project is a large scale international project that has been carried out in 15 European countries and Japan and it was financed as a Specific Targeted Research Project (STREP) of the European Union's Sixth Framework Programme. Altogether 21 public, and 15 private colleges and universities participated on this survey in the Czech Republic. We were provided only with the sub-sample of graduates from the University of Economics, Prague that consists of 1,704 respondents that graduated during the period 2008–2012 and that are currently employed. Respondents were asked to exclude jobs they left within 4 months after graduation. Data was weighted by the proportion of faculties, type of study, gender, economical status and year of graduating. Due to lack of answers in some variables we use final dataset of 1,298 graduates.

We consider three specifications of the binary logit model, each with different dependent variable with differing between males and females. Firstly, dependent dummy variable *Horizontal job mismatch* where a graduate is considered as mismatched when his/her job is not related to the field of study of the programme was constructed

(model A). Second dummy dependent variable *Vertical job mismatch* mismatches level of education and current graduate's job (model B). Third dependent variable *Skill mismatch* variable is based on the response to a question asking respondents to rate on a 1 (not at all) to 5 (very high extent) to which their skills and knowledge were utilized in their job with a response 1 or 2 deemed consistent with overskilling (model C).

We included usual demographic characteristics: gender is coded as 1 for females, age is measured in years. Parental highest earned education is distinguishing between mother's and father's education, coded as follows: 1 – elementary/basic school, 2 – secondary education with apprenticeship certificate, 3 – secondary education with school-leaving exam, 4 – bachelor's, master's and doctoral type of study. Continuous variables related to work experience, measured in months, differentiate between two options of the graduate's job experience in study-related and non-study-related activities. Left first employment is dummy variable that distinguish those respondents who still work in their first employment after five years of graduation from those who already left their jobs. We also included number of jobs the graduate had during the labour market career until the data collection. Dummy variable type of contract in the current occupation takes the value of 1 if the job is permanent in contrast to fixed-term employment or self-employment and this variable test possible trade-off between job safety and mismatch.

In Table 1 we present mean and standard deviation of all variables. We can observe that 22.03% of the respondents are horizontally mismatched at their first job, only 9.63% of them are mismatched vertically, and 51.39% of the graduates are skill mismatched. Average age of finishing the studies is almost 28 years. Mother's highest earned education is a little bit higher than father's one, comparing 3.3621 versus 3.3482. We can conclude that graduates work experience in study related field is in average 4 months longer that in non-study-related field. Most of graduates have permanent type of contract (almost 78%).

Variable	Mean	Std. dev.
Horizontal job mismatch	0.2203	0.4146
Vertical job mismatch	0.0963	0.2951
Skills mismatch	0.5139	0.5000
Gender (female = 1)	0.6048	0.4891
Age	27.9769	3.4016
Father's education	3.3482	0.8897
Mother's education	3.3621	0.7863
Work experience (study related, in months)	27.5532	279.1236
Work experience (not study related, in months)	23.1960	277.7259
Left first employment (= 1)	0.6210	0.4853
Number of jobs	1.5609	0.8584
Type of contract (permanent = 1)	0.7766	0.4167

Table 1 Means and standard deviations of variables

3 Results

The main results of the binary logit estimation of nine different models (model A, B and C, all observations, female nad male) specifications are presented below in Tables 2, 3, and 4. For all the estimations and further calculations econometric software EViews 8 was used. As odds ratios with standard errors are not included in EViews 8, we programmed own procedures. While the value of the coefficient does not give a clear interpretation of the relationship when using a binary logit model, the sign of the coefficient does have the same interpretation as with an ordinary least square model [5].

From the results of model A in the Table 2 we observe that being female have positive effect on the probability of horizontally mismatched at first job of the graduate, significant at 5% level. On the other hand, age of respondent lowers this probability, however this variable is not statistically significant. There is no gender difference in estimated model for significance of dummy variables related to working experience, both study and non-study-related. Mother's education in model A of females decreases the probability of being horizontally mismatched and is statistically significant when comparing to males (model A of males). Females are 1.359 times more likely than males to be horizontally mismatched controlling all other variables on average when interpreting the odds ratio (3) from full model. We reject the hypotheses that type of contract can work as trade-off between tenancy and job match.

The explanatory variables are also jointly highly significant in all models despite the relatively low value of McFadden's pseudo R^2. The results of Hosmer-Lemeshow (H-L statistics) and Andrews χ^2 tests indicate small difference between the fitted expected values and actual values. Therefore we cannot reject that the models are providing an insufficient fit to the data.

	Full		Female		Male	
Predictors	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**
Intercept	–0.716 (0.846)	0.489	1.300 (1.064)	3.668	–4.072 (1.480)***	0.017
Gender (female = 1)	0.307 (0.147)**	1.359				
Age	–0.038 (0.024)	0.963	–0.048 (0.031)	0.953	–0.034 (0.043)	0.966
Father's education	0.118 (0.092)	1.125	0.107 (0.107)	1.113	0.153 (0.182)	1.165
Mother's education	–0.172 (0.109)	0.842	–0.305 (0.130)**	0.737	0.053 (0.192)	1.054
Work exper. (study related)	–0.023 (0.004)***	0.978	–0.025 (0.005)***	0.975	–0.020 (0.008)**	0.980
Work exper. (non-study-rel.)	0.015 (0.004)***	1.015	0.011 (0.004)**	1.011	0.027 (0.007)***	1.027
Left first employment (= 1)	0.388 (0.265)	1.474	–0.169 (0.329)	0.845	1.645 (0.503)***	5.182
Number of jobs	0.224 (0.148)	1.251	–0.100 (0.197)	0.905	0.857 (0.254)***	2.356
Type of contract (permanent = 1)	0.0002 (0.000)	1.000	–0.001 (0.003)	0.999	0.236 (0.288)	1.267
McFadden's R^2	0.0612		0.0523		0.1118	
LR statistics	83.738***		45.680***		54.650***	
N	1 298		785		513	
H-L statistics	19.873**		23.125***		12.156	
Andrews stat.	24.351***		26.235***		56.659***	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2 Model A - determinants of horizontal job mismatch

The estimated coefficients and odds ratios of model B in which the dependent variable is vertical job mismatch are presented in Table 3. Vertical mismatch is particularly more probable for older respondents (statistically significant at 1% level). We see also strong effect for those who left first occupation. There is no proof of significant causal relationship being horizontally mismatched and having a permanent job. More jobs during early career increase the probability of the individual being vertically mismatched.

	Full		Female		Male	
Predictors	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**
Intercept	1.473 (1.303)	4.364	4.750 (1.664)***	115.57	–4.426 (2.130)**	0.012
Gender (female = 1)	0.528 (0.215)**	1.695				
Age	–0.128 (0.040)***	0.880	–0.179 (0.052)***	0.837	–0.066 (0.063)	0.936
Father's education	0.057 (0.127)	1.059	0.071 (0.142)	1.074	–0.009 (0.264)	0.991
Mother's education	–0.478 (0.148)***	0.620	–0.640 (0.170)***	0.527	–0.103 (0.284)	0.902
Work exper. (study related)	–0.033 (0.007)***	0.968	–0.035 (0.009)***	0.965	–0.033 (0.013)**	0.968
Work exper. (non-study-rel.)	–0.00006 (0.001)	1.000	0.001 (0.006)	0.999	0.00007 (0.001)	1.000
Left first employment (= 1)	0.774 (0.366)**	2.169	0.172 (0.430)	1.188	3.116 (0.963)***	22.548
Number of jobs	0.508 (0.187)***	1.661	0.198 (0.240)	1.219	1.466 (0.393)***	4.331
Type of contract (permanent = 1)	–0.0001 (0.001)	1.000	–0.001 (0.003)	0.999	0.0003 (0.001)	1.000
McFadden's R^2	0.0818		0.0897		0.1028	
LR statistics	67.326***		50.541***		25.732***	
N	1 298		785		513	
H-L statistics	14.539*		7.664		10.843	
Andrews stat.	18.362**		9.748		99.096***	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 Model B - determinants of vertical job mismatch

Table 4 reports the results of model C – skill mismatching. We observe that younger graduates increase skills mismatching in all models, and we can see also positive effect in study related working experience with no gender difference. Respondents who already left first employment decrease the probability of skills mismatched when comparing to vertical mismatch model B.

	Full		Female		Male	
Predictors	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**	**Coeff. (SE)**	**Odds**
Intercept	–0.525 (0.657)	0.592	–1.246 (0.860)	0.288	0.538 (1.078)	1.713
Gender (female = 1)	–0.126 (0.117)	0.882				
Age	0.031 (0.018)*	1.031	0.045 (0.024)*	1.046	0.008 (0.029)	1.008
Father's education	–0.016 (0.076)	0.985	–0.085 (0.094)	0.918	0.107 (0.134)	1.113
Mother's education	0.078 (0.088)	1.082	0.153 (0.112)	1.165	–0.009 (0.146)	0.991
Work exper. (study related)	0.012 (0.003)***	1.012	0.011 (0.004)***	1.011	0.015 (0.005)***	1.015
Work exper. (non study related)	–0.003 (0.003)	0.997	–0.003 (0.004)	0.997	–0.003 (0.006)	0.997
Left first employment (= 1)	–0.537 (0.215)**	0.585	–0.435 (0.272)	0.647	–0.698 (0.354)**	0.498
Number of jobs	–0.173 (0.123)	0.841	–0.070 (0.158)	0.933	–0.293 (0.197)	0.746
Type of contract (permanent = 1)	–0.00006 (0.000)	1.000	0.001 (0.005)	1.001	–0.237 (0.231)	0.789
McFadden's R^2	0.0251		0.0246		0.0359	
LR statistics	45.104***		26.727***		25.443***	
N	1 298		785		513	
H-L statistics	14.873*		13.831*		9.254	
Andrews stat.	14.787		14.621		9.413	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Model C - determinants of skills mismatch

4 Conclusion

The analysis determines the factors of being mismatched horizontally and vertically with job, and skills mismatching. We used the data of graduates from the University of Economics, Prague, within five years after graduation. The main factors of mismatching are: gender (horizontal and vertical mismatching), age (skills mismatching) of the respondent and selected early effects, such as working experience (both study and non-study-related), and left first employment. The results of the study are similar to results found by [3] in the term of the same sign and significance of estimated regression coefficients for the following variables: age of the respondent, left first employment, and number of jobs.

In the future, we plan to widen this study in terms of comparison of job and skills mismatching with other universities in the Czech Republic. A comparison with other countries participating in the REFLEX survey is also possible. One can also compare results with previous years in order to evaluate possible developments in time. We also consider estimate the impact of mismatches on wages.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics, Prague, project no. IG403015 and was prepared with the institutional support for the long-term development of science and research at the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Dlouhá, Z., Dlouhý M.: Factors of dissatisfaction with chosen study programme. In: *Mathematical Methods in Economics (MME2014)*, 151-155 (2014).

[2] McGuiness, S., and Sloan, PJ.: Labour market mismatch among UK graduates: An analysis using REFLEX data. *Economics of Education Review* **30** (1), 130-145 (2011).

[3] Robert, P.: Job mismatch in early career of graduates under post-communism. *International Journal of Manpower*, **35** (4), 500-513 (2014).

[4] Robst, J.: Education and job match: The relatedness of college major and work. *Economics of Education review* **26,** 397-407 (2007).
[5] Wooldridge, J. M.: *Introductory Econometrics: A Modern Approach*. 5th edition, Cengage Learning, Mason, (2012).

Long-Term Trends in Efficiency of National Health Systems: An Evaluation by Two-Stage Data Envelopment Analysis.

Martin Dlouhý[1]

Abstract. The objective of the paper is to evaluate efficiency of national health systems of developed countries from a long-term perspective. The data on health systems come from three sources: OECD Health Statistics 2014, European Health for All Database and Eurostat. The sample includes 32 developed countries for the period 1990-2010. Two-stage data envelopment analysis was used as a non-parametric method of efficiency evaluation. Efficiency evaluation of health systems was divided into two parts: efficiency of health resources to produce health services and efficiency of health services to produce health of population. The efficiency scores suggest some improvements in efficiency of national health systems in the long-term. These results are the same for both one- and two-stage data envelopment analysis, for both constant and variable returns to scale.

Keywords: health systems, data envelopment analysis, efficiency evaluation.

JEL Classification: C44, I10
AMS Classification: 90C90

1 Introduction

Efficiency evaluation in health care is a theoretical as well as a practical problem. Health care is an application field with specific context and characteristics: final output (health improvement) is hard to measure and hard to express in monetary values; the causality between input and output is not always certain; apart from efficiency of health system, equity and moral principles have to be borne in mind; there is usually a high variation in demand for services [6]. But this complexity is the reason why health care as an application field is so challenging for researchers.

One of the most comprehensive attempts to evaluate performance of health systems was The World Health Report 2000 [14]. The World Health Report presented an index of national health system's attainment and an index of performance relative to potential. These measures are based on five goals: the level and distribution of health, the level and distribution of responsiveness of the system to the legitimate expectations of the population, and fairness of contribution to financing the health system. An important objective is to see what one can learn from the best. The World Health Report aims to stimulate a debate about better ways of measuring health systems performance and thus finding a successful new direction for health systems to follow. Five objectives (achievements) were identified and a survey of 1006 respondents was conducted to derive a set of weights [14].

If only outcomes of health system are considered, the Report talks about *attainment*. The second question is what a health system should be able to achieve with the same level of resources - *performance*. The resources are measured as per capita health expenditure in international dollars. The World Health Report suggests that to assess *relative performance* requires a scale, one end of which establishes an upper limit or *frontier*, corresponding to the most that could be expected of a health system. This *frontier* represents the level of attainment which a health system might achieve, but which no country surpasses. The similar terms - *relative performance* or *efficiency frontier* - are used in the data envelopment analysis (DEA), which is a method of the production function approximation based on linear programming. DEA uses quantities of inputs consumed and outputs produced to calculate the relative efficiencies of units in the sample. The relative technical efficiency is defined as the ratio of total weighted input to total weighted output or vice versa. DEA permits each unit to select its own weights. It is assumed that each unit (a country in this analysis) will select the weights that maximize its own efficiency ratio. The efficient frontier represents the maximum amounts of output that can be produced by the given amounts of input or, alternatively, the minimum amounts of inputs required to produce the given amount of output.

We provide a comparison of fundamental characteristics of the World Health Report 2000 methodology and data envelopment analysis in Table 1. Both methodologies are able to deal with multiple inputs and outputs and offer the performance score and the estimation of frontier. The approaches differ in the possibility of country's individual weights allowed in DEA. The World Health Report, when analyzing the stewardship, also emphasizes

[1] University of Economics, Dept. of Econometrics, 4 W. Churchill Sq., Prague, Czech Republic, dlouhy@vse.cz.

the role of democratic government to select national preferences (i.e. weights) to the objectives [14]. Therefore the information on individual weights might be interesting. Another advantage of DEA is the possibility to offer a given country a group of peers, which consists of countries with the same preferences/the same relative performance in individual objectives.

World Health Report 2000	**Data Envelopment Analysis**
Multiple inputs and outputs	
Dealing with multiple outputs and inputs	Dealing with multiple outputs and inputs
Weights	
Weights are constant for all countries - based on a survey among experts	Weights are individual for each country - set by the DEA model with assumption of performance maximization
Production Frontier Estimation	
Econometric methods	Linear programming
Results	
Overall performance (ranking, potential performance)	Technical efficiency (= performance ranking and potential performance) Individual weights (= country's preferences) Reference set (= peer countries)

Table 1 Comparison of Methodologies: The World Health Report 2000 and DEA

Our aim is to contribute to the debate on efficiency measurement in health care. The objective of this study is to evaluate efficiency of national health systems of selected developed countries from a long-term perspective by using data envelopment analysis.

2 Data and Methods

Data

The data on national health systems come from three sources. The primary source of data was the *OECD Health Statistics 2014*. Some data and some new countries were added from the *European Health for All Database* and from the *Eurostat*. The database includes 32 developed countries for the period 1990-2010. We observe development in the countries in ten-year intervals, thus we can maximally have 3 observations for each country. Unfortunately, many cases had to be deleted from the sample due to missing data. Finally, we were able to use 10 observations from the year 1990, 25 from the year 2000, and 20 from the year 2010.

Efficiency evaluation of national health systems was divided into two steps (stages): efficiency of health resources to produce health services (Stage 1) and efficiency of health services to produce health of population (Stage 2). In Stage 1, the inputs comprise the number of physicians in full time equivalents and the total number of hospital beds, while the outputs (intermediate products) comprise the number of hospital discharges and the number of outpatient contacts. In Stage 2, the inputs comprise the outputs of Stage 1 (the number of hospital discharges and the number of outpatient contacts) and the outputs comprise the total population and the life expectancy at birth (in years).

Data Envelopment Analysis

For efficiency evaluation, we will use data envelopment analysis (DEA), which is able to deal multiple inputs and outputs and with the two-stage efficiency models. DEA uses quantities of inputs consumed and outputs produced to calculate the relative technical efficiency of a decision-making unit. The relative technical efficiency of the unit is defined as the ratio of its total weighted output to its total weighted input or, vice versa, as the ratio of its total weighted input to its total weighted output. DEA was developed by Charnes, Cooper, and Rhodes in 1978 [2]. Since 1978 a great variety of DEA models with various extensions and modifications has been developed and used for a great number of applications in the public and private sector (for example [3] [4] [5] [9]).

For each production unit DEA (a) calculates the efficiency score; (b) determines the relative weights of inputs and outputs; and (c) identifies the peers for each unit that is not technically efficient. The peers of an inefficient unit are technically efficient units with similar combinations of inputs and outputs. The peers serve as

benchmarks which show potential improvements that an inefficient unit can attain. Because the peers are real production units, the efficiency improvements are attainable.

Hollingsworth [8] reviewed 188 papers on frontier efficiency measurement in health care. The review mainly included DEA-based methods that dominated the literature; however, there was also an increasing number of parametric techniques, such as stochastic frontier analysis. DEA became quite popular as a method of efficiency evaluation in health care in the Czech Republic, too ([7] [11] [13]).

Two-Stage DEA Model

In the two-stage DEA model, we model each unit as two sub-units connected in series [12]. The two-stage model is one of the so-called network DEA models (e.g. [1] [10]). In this study, we first assume for both stages an output-oriented model with the constant returns to scale (CRS). For Stage 1 and sub-unit 1 of unit q we solve:

$$\begin{aligned} &\text{maximize} && \phi_{1q} - \varepsilon\,(\mathbf{e}^T\mathbf{s}_1^+ + \mathbf{e}^T\mathbf{s}_1^-), \\ &\text{subject to} && \mathbf{X}\boldsymbol{\lambda}_1 + \mathbf{s}_1^- = \mathbf{x}_q, \\ & && \mathbf{Y}\boldsymbol{\lambda}_1 - \mathbf{s}_1^+ = \phi_{1q}\,\mathbf{y}_q, \\ & && \boldsymbol{\lambda}_1 \geq \mathbf{0},\ \mathbf{s}_1^+ \geq \mathbf{0},\ \mathbf{s}_1^- \geq \mathbf{0}, \end{aligned} \tag{1}$$

where ϕ_{1q} is the efficiency score, $\boldsymbol{\lambda}_1$ is the vector of variables, $\mathbf{s}_1^+$, $\mathbf{s}_1^-$ are the vectors of slack variables, $\mathbf{x_q}$ and $\mathbf{y_q}$ are the vectors of inputs and outputs of the evaluated unit q, $\mathbf{X}$ is the matrix of inputs and $\mathbf{Y}$ is the matrix of outputs (intermediate products), $\mathbf{e}^T = (1, 1, \ldots, 1)$, and ε is the infinitesimal constant. The production unit q is technically efficient if the optimal value of variable ϕ_{1q} is one and all slack variables equal zero. For Stage 2 and sub-unit 2 of production unit q we solve:

$$\begin{aligned} &\text{maximize} && \phi_{2q} - \varepsilon\,(\mathbf{e}^T\mathbf{s}_2^+ + \mathbf{e}^T\mathbf{s}_2^-), \\ &\text{subject to} && \mathbf{Y}\boldsymbol{\lambda}_2 + \mathbf{s}_2^- = \mathbf{y}_q, \\ & && \mathbf{Z}\boldsymbol{\lambda}_2 - \mathbf{s}_2^+ = \phi_{2q}\,\mathbf{z}_q, \\ & && \boldsymbol{\lambda}_2 \geq \mathbf{0},\ \mathbf{s}_2^+ \geq \mathbf{0},\ \mathbf{s}_2^- \geq \mathbf{0}, \end{aligned} \tag{2}$$

where ϕ_{2q} is the efficiency score, $\boldsymbol{\lambda}_2$ is the vector of variables, $\mathbf{s}_2^+$, $\mathbf{s}_2^-$ are the vectors of slack variables, $\mathbf{y_q}$ and $\mathbf{z_q}$ are the vectors of inputs and outputs, $\mathbf{X}$ is the matrix of inputs and $\mathbf{Z}$ is the matrix of outputs, $\mathbf{e^T} = (1, 1, \ldots, 1)$, and ε is the infinitesimal constant. Finally, the efficiency of the whole unit q is computed from:

$$\begin{aligned} &\text{maximize} && \phi_q - \varepsilon\,(\mathbf{e}^T\mathbf{s}^+ + \mathbf{e}^T\mathbf{s}^-), \\ &\text{subject to} && \mathbf{Y}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{y}_q^*, \\ & && \mathbf{Z}\boldsymbol{\lambda} - \mathbf{s}^+ = \phi_q\,\mathbf{z}_q, \\ & && \boldsymbol{\lambda} \geq \mathbf{0},\ \mathbf{s}^+ \geq \mathbf{0},\ \mathbf{s}^- \geq \mathbf{0}, \\ & && \mathbf{y_q}^* = \mathbf{Y}\boldsymbol{\lambda}_1, \end{aligned} \tag{3}$$

where ϕ_q is the efficiency score, $\boldsymbol{\lambda}$ is the vector of variables, $\boldsymbol{\lambda}_1$ is the vector obtained from Stage 1 program (1), $\mathbf{s}^+$, $\mathbf{s}^-$ are the vectors of slack variables $\mathbf{y_q}$ and $\mathbf{z_q}$ are the vectors of inputs and outputs, $\mathbf{y_q}^*$ is the expected value of inputs (intermediate products) in case of efficiency in Stage 1, $\mathbf{Y}$ is the matrix of inputs and $\mathbf{Z}$ is the matrix of outputs, $\mathbf{e^T} = (1, 1, \ldots, 1)$, and ε is the infinitesimal constant. The models (1-3) are solved for each unit q =1, 2, ..., n.

For a comparison, we also solve the one-stage DEA model that does not divide production units to sub-units. The one-stage model treats each unit as "black box" by considering only inputs $\mathbf{X}$ and final outputs $\mathbf{Z}$. The model has the form:

$$\begin{aligned} &\text{maximize} && \theta_q - \varepsilon\,(\mathbf{e}^T\mathbf{s}^+ + \mathbf{e}^T\mathbf{s}^-), \\ &\text{subject to} && \mathbf{X}\boldsymbol{\mu} + \mathbf{s}^- = \mathbf{x}_q, \\ & && \mathbf{Z}\boldsymbol{\mu} - \mathbf{s}^+ = \theta_q\,\mathbf{z}_q, \\ & && \boldsymbol{\mu} \geq \mathbf{0},\ \mathbf{s}^+ \geq \mathbf{0},\ \mathbf{s}^- \geq \mathbf{0}, \end{aligned} \tag{4}$$

where θ_q is the efficiency score, $\boldsymbol{\mu}$ is the vector of variables. The two-stage and the one-stage DEA models with variable returns to scale (VRS) can be defined analogically.

3 Results

All computations presented here were made by the optimization software LINGO (developed by LINDO Systems). First, we used the CRS model for efficiency evaluation of health systems. The CRS model gives the most pessimistic estimates of real efficiency of health systems because one can hardly expect that the relation between health services and health is linear and that an assumption of constant returns to scale holds. However, this worst-case scenario represents a useful benchmark. The efficiency scores of the CRS model are summarized in Table 2 that shows scores calculated for each stage, total efficiency calculated by the two-stage DEA model, and for a comparison of models, efficiency scores calculated by the traditional one-stage DEA model. Second, we calculated the efficiency score by the VRS model that is probably more appropriate to model the relations between the inputs and outputs in a health system, especially in Stage 2 (Table 3).

Country Year	**Stage 1 Resources/Services**			**Stage 2 Services/Health**			**Two-Stage Model Total Efficiency**			**One-Stage Model Resources/Health**		
	1990	**2000**	**2010**	**1990**	**2000**	**2010**	**1990**	**2000**	**2010**	**1990**	**2000**	**2010**
Albania		0.65			1.00			0.65			1.00	
Armenia	0.36	0.17		0.38	1.00		0.14	0.17		0.31	0.38	
Austria	0.72	0.73	0.67	0.32	0.28	0.27	0.23	0.20	0.18	0.44	0.34	0.28
Belgium		0.66	0.63		0.30	0.31		0.20	0.20		0.46	0.45
Bulgaria	0.61			0.32			0.19			0.41		
Czech Rep.	0.82	0.76	0.68	0.24	0.22	0.22	0.20	0.17	0.15	0.48	0.38	0.36
Denmark		0.66			0.45			0.29			0.52	
Estonia		0.68	0.65		0.38	0.42		0.26	0.28		0.49	0.59
Finland		0.84			0.39			0.32			0.54	
France		0.71	0.70		0.29	0.31		0.21	0.21		0.39	0.39
Germany		0.71	0.70		0.27	0.22		0.19	0.16		0.39	0.34
Greece	0.44	0.59		0.48	0.44		0.21	0.26		0.41	0.40	
Hungary		0.97	0.86		0.21	0.22		0.20	0.19		0.48	0.45
Iceland	0.69			1.00			0.69			1.00		
Israel		0.81			0.31			0.25			0.53	
Italy		0.60			0.35			0.21			0.34	
Japan		1.00	0.85		0.39	0.35		0.39	0.29		0.65	0.57
Korea			1.00			0.29			0.29			0.64
Latvia	0.64	0.75	0.66	0.27	0.37	0.38	0.17	0.28	0.25	0.39	0.49	0.52
Lithuania		0.68	0.69		0.30	0.29		0.20	0.20		0.38	0.38
Luxembourg			0.61			0.86			0.52			1.00
Mexico		0.49	0.63		1.00	0.86		0.49	0.54		1.00	1.00
Netherlands	0.47	0.52		0.44	0.45		0.21	0.23		0.51	0.53	
Norway			0.79			0.41			0.33			0.47
Poland			0.79			0.34			0.26			0.58
Portugal	0.44	0.51	0.57	0.66	0.58	0.52	0.29	0.30	0.30	0.51	0.52	0.52
Romania		1.00	1.00		0.35	0.35		0.35	0.35		0.66	0.54
Slovakia		0.91			0.24			0.22			0.41	
Slovenia			0.82			0.38			0.31			0.64
Sweden	0.69	0.79	1.00	0.57	0.57	0.52	0.39	0.45	0.52	0.50	0.54	0.65
Turkey		0.70	1.00		0.77	0.33		0.53	0.33		1.00	0.78
UK		0.75			0.41			0.31			0.65	
Average	0.59	0.71	0.76	0.47	0.45	0.39	0.27	0.29	0.29	0.50	0.54	0.56

Table 2 Efficiency Scores 1990 – 2010 (Constant Returns to Scale)

Country Year	Stage 1 Resources/Services			Stage 2 Services/Health			Two-Stage Model Total Efficiency			One-Stage Model Resources/Health		
	1990	2000	2010	1990	2000	2010	1990	2000	2010	1990	2000	2010
Albania		0.69			1.00			0.97			1.00	
Armenia	0.37	0.18		0.89	1.00		0.89	0.90		0.89	0.90	
Austria	0.73	0.74	0.68	0.93	0.96	0.99	0.93	0.96	0.99	0.93	0.96	0.99
Belgium		0.66	0.64		0.95	0.98		0.95	0.98		0.96	0.99
Bulgaria	0.61			0.88			0.88			0.88		
Czech Rep.	0.83	0.77	0.69	0.88	0.92	0.95	0.88	0.92	0.95	0.88	0.92	0.95
Denmark		0.69			0.95			0.94			0.95	
Estonia		0.73	0.75		0.88	0.94		0.88	0.94		0.88	0.94
Finland		0.85			0.96			0.95			0.96	
France		0.93	0.90		0.97	1.00		0.97	1.00		0.97	1.00
Germany		1.00	1.00		0.97	0.98		0.97	0.98		0.96	0.99
Greece	0.45	0.59		0.95	0.96		0.94	0.96		0.95	0.96	
Hungary		0.98	0.87		0.88	0.91		0.88	0.91		0.89	0.92
Iceland	1.00			1.00			1.00			1.00		
Israel		0.84			0.97			0.97			0.97	
Italy		0.77			0.99			0.98			0.99	
Japan		1.00	1.00		1.00	1.00		1.00	1.00		1.00	1.00
Korea			1.00			0.98			0.98			1.00
Latvia	0.65	0.78	0.69	0.86	0.87	0.91	0.86	0.87	0.91	0.86	0.87	0.91
Lithuania		0.70	0.70		0.89	0.91		0.89	0.90		0.89	0.91
Luxembourg			1.00			1.00			1.00			1.00
Mexico		0.49	0.63		1.00	1.00		0.95	0.98		1.00	1.00
Netherlands	0.47	0.52		0.95	0.97		0.94	0.96		0.95	0.97	
Norway			0.82			1.00			1.00			1.00
Poland			0.81			0.94			0.93			0.95
Portugal	0.45	0.52	0.58	0.92	0.95	0.99	0.91	0.94	0.98	0.92	0.95	0.98
Romania		1.00	1.00		0.88	0.91		0.88	0.91		0.90	0.92
Slovakia		0.94			0.90			0.90			0.91	
Slovenia			0.87			0.99			0.99			0.99
Sweden	0.69	0.81	1.00	0.97	0.99	1.00	0.95	0.98	1.00	0.96	0.98	1.00
Turkey		0.70	1.00		0.93	0.93		0.93	0.93		1.00	0.98
UK		0.80			0.97			0.96			0.99	
Average	0.63	0.75	0.83	0.92	0.95	0.97	0.92	0.94	0.96	0.92	0.95	0.97

Table 3 Efficiency Scores 1990 – 2010 (Variable Returns to Scale)

In both CRS and VRS models, we can observe that the two-stage DEA model gives lower efficiency scores than the one-stage DEA model, although in the case of the VRS model the differences in average efficiency are very low. The CRS model (the worst-case efficiency scenario) gives for some countries extremely low efficiency scores that can hardly be acceptable by any health analysts; therefore it seems that the VRS model describes the reality better.

Regarding the long-term trends in efficiency of health systems, the CRS and the VRS models suggest some improvements in the cases of both one- and two-stage DEA models. The CRS and VRS models also show a similar positive trend in the efficiency of health resources to produce health services (Stage 1), but the models

differ in the efficiency trend of health services to produce health of population (Stage 2). The limit of this argument is that the sets of countries with the calculated efficiency scores are not the same for the three observed years.

4 Conclusion

Efficiency evaluation in health care represents a very challenging application field. We showed that the total efficiency of national health systems can be divided into efficiency of health resources and efficiency of health services. In such a case, the two-stage DEA model can be used to evaluate efficiency of national health systems. The observed trend in average efficiency scores suggests some improvements in efficiency of national health systems in the long term. This was observed in the cases of both one- and two-stage DEA models, for both constant- and variable-returns-to-scale models.

Acknowledgements

The research was funded by the Czech Science Foundation (project no. P403/12/1387) and was prepared with the institutional support for the long-term development of science and research at the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Castelli, L., Pesenti, R., and Ukovich, W.: A Classification of DEA Models When the Internal Structure of the Decision Making Units Is Considered. *Annals of Operations Research* **173** (2010), 207–235.
[2] Charnes, A., Cooper, W. W., and Rhodes, E.: Measuring the Inefficiency of Decision Making Units. *European Journal of Operational Research* **2** (1978), 429–444.
[3] Charnes, A., Cooper W. W., Lewin, A. Y, and Seiford, L. M. (eds.): *Data Envelopment Analysis: Theory, Methodology and Applications.* Kluwer Academic Publishers, Boston, 1994.
[4] Cooper, W. W., Seiford, L.W., and Tone, K.: *Data Envelopment Analysis*. Kluwer Academic Publishers, Boston, 2000.
[5] Cooper, W. W., Seiford, L.W., and Zhu, J. (eds.): *Handbook of Data Envelopment Analysis*. Kluwer Academic Publishers, Boston, 2004.
[6] Dlouhý, M.: Production Functions and Efficiency Evaluation in Health Care. In: *Operational Research for Health Policy: Making Better Decisions* (Brailsford, S., and Harper, P., eds.). Peter Lang, Bern, 2007.
[7] Dlouhý, M., Jablonský, J., and Novosádová, I., Využití analýzy obalu dat pro hodnocení efektivnosti českých nemocnic [Use of Data Envelopment Analysis for Efficiency Evaluation of Czech Hospitals]. *Politická ekonomie* **54** (2007), 60–71.
[8] Hollingsworth, B.: Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science* **6** (2003), 203–218.
[9] Jablonský, J., and Dlouhý, M.: *Modely hodnocení efektivnosti produkčních jednotek [Models of Efficiency Evaluation of Production Units]*, Professional Publishing, Prague, 2004. (ISBN 80-86419-49-5)
[10] Kao, C.: Network Data Envelopment Analysis: A review, *European Journal of Operational Research*, **239** (2014), 1–16.
[11] Novosádová, I., and Dlouhý, M.: Hodnotenie technickej efektívnosti nemocníc a odmeňovanie zdravotníkov [Evaluation of Technical Efficiency of Acute Hospitals and Its Relation To Wages of Personnel]. *Ekonomický časopis* **55** (2007), 783–792.
[12] Sexton, T. R., and Lewis, H. F.: Two-Stage DEA: An Application to Major League Baseball. *Journal of Productivity Analysis* **19** (2003), 227–249.
[13] Votápková, J., and Šťastná, L.: Efficiency of Hospitals in the Czech Republic. *Prague Economic Papers* **22** (2013), 524–541.
[14] World Health Organization: *The World Health Report 2000 - Health Systems: Improving Performance.* World Health Organization, Geneva, 2000.

Predicting Inflation by the Main Inflationary Factors: Performance of TVP-VAR and VAR-NN models

Anna Dobešová[1], David Hampel[2]

Abstract. A suitable way for forecasting inflation is to do it using main inflationary factors. Such factors can be sorted to domestic and foreign sets. One-way and two-way relations between them and inflation can be considered. Therefore, vector autoregressive model (VAR) seems to be a proper tool for modelling the reality. However, basic VAR model can suffer from insufficient forecasting performance caused by its linear nature. We employ two nonlinear vector autoregressive alternatives for predicting inflation: Time-Varying Parameter VAR model with stochastic volatility and VAR Neural Network model. In both cases we select the specification with the best combination of inflationary factors. Neural Networks are flexible tool which can be easily adjusted to an autoregressive form. Resulting VAR-NN models produce accurate inflation forecasting, but they take essential information mainly from the previous inflation observations and ignore the other series. Compared to that, TVP-VAR model is a statistical method applicable only for specific sort of problems. It is successful in incorporating information from all modelled series. TVP-VAR model leads to better forecasting performance than VAR-NN.

Keywords: inflation, TVP-VAR model, VAR-NN model.

JEL Classification: C45, C32, E31
AMS Classification: 62M45, 62H12, 62P20

1 Introduction

Inflation is one of the basic macroeconomic indicators. Currently, central bank policy is based on the inflation targeting regime, so it is crucial for them to perform accurate inflation predictions presented within inflation reports. For private entities, future inflation can serve as a scale of risk, return on investment and the size of the actual sales.

Compared to other macroeconomic variables, inflation has become less predictable in recent decades [14]. This process stepped so much that some authors state that, in the case of inflation, subjective forecasts achieve better results than the predictions made on the basis of mathematical and economic models [8]. At this point, it is necessary to point out several facts. Firstly, inflation is a complex unit consisting of a several thousand items. This determines the primary limitation of predicting accuracy. Secondly, key inflation forecasting horizon for central banks is a medium-term period. It means that the predictions are made for a period of 12–24 months. Forecasts designed for such a long period suffers from a high degree of uncertainty. More authors, who tested institutional forecasting accuracy, declared that naive predictions for the medium term may be more accurate than official institutional forecasts and that errors and variance of forecasts for higher horizons are substantially increasing, see for example [1], [13]. Therefore, we can assume that inflation can be predicted with acceptable errors within a one year period. For further horizons, general inflation mean is predicted instead of exact values [8]. It is necessary to access forecasting tools with realistic requirements.

The third fact, which can cause deficiencies in inflation predicting, covers the selected inflation models and their use. Generally, methods for inflation forecasting can be divided into theoretical models (i.e. economic-based tools) and empirical models (general mathematical and statistical techniques). Theoretical models are used in the institutional sphere mainly. Dynamic Stochastic General Equilibrium systems (DSGE), models based on Phillips curve and systems operating with the output gap are examples of such models. In the private or professional sphere so-called empirical approach dominates. It is prevailingly derived from the Box-Jenkins methodology: autoregressive (AR) and vector autoregressive (VAR) models, factor-augmented VAR models (FAVAR), time-varying parameter VAR models (TVP-VAR) etc. It is necessary to highlight the fact that the empirical techniques are often approached from too theoretical standpoint, which may markedly reduce their accuracy in the inflation predicting. For example, Elliot and Timmermann [8] mention as one of the possible empirical mod-

[1] 1 Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, anna.dobesova@mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: david.hampel.uso@mendelu.cz.

els for forecasting inflation TVP-VAR model of Primiceri (see [15]) with specification consisting interest rate, inflation and unemployment. Such specification is suitable for studying the transmission mechanism (the effectiveness of monetary policy) and includes also the Phillips curve relationship. However, the involved variables are not sufficient for predicting inflation. Inflation forecasting based on interest rate entails many difficulties, referred as "price puzzle" effect [16]. For the purpose of inflation forecasting the specification of the model should be modified.

Currently, the ability of private entities to predict inflation and other macroeconomic variables is comparable with the capacity of official institutions in publishing predictions; see [1], [9]. Inflation is largely a global phenomenon. More than two-thirds of inflation may depend on foreign factors, and changes in domestic inflation retroactively affect the global economy [4]. Under these conditions, it must be empirically suitable way to predict inflation using the main inflationary factors, both domestic and international ones. Forecasting inflation using the main economic factors is less common in the literature, but achieves a good performance if factors are selected correctly ([3]). Also central banks seek ways for improving their forecasts by adding other empirical techniques [1].

These reasons motivate us to employ the empirical procedure for the selection of suitable economic indicators for predicting inflation in this article. For predicting we use two modern nonlinear tools: TVP-VAR model with stochastic volatility and neural network VAR (VAR-NN) model. Both models are capable to estimate breaks and changes in economy. TVP-VAR model with stochastic volatility, proposed by Primiceri (see [15]), is a common tool for macroeconomic analysis and it is a representative of the mathematical and statistical approach. In contrast, neural networks fall into a branch of computational statistics; they are one main methods of artificial intelligence. Thanks to their flexibility, they can be deployed on a wide range of problems. For our purposes, we adapt this tool to the vector autoregression form. Neural network models are not often used approach for predicting inflation, because they require a specific debugging procedure that must be respected to achieve good performance [11].

Theoretically, there is a large amount of potential inflationary factors. We build on our previous experience [6], [7] and employ nine key factors: both domestic and global factors are included in the analysis. The aim is to establish VAR-NN and TVP-VAR models in the most appropriate specification for forecasting inflation and to evaluate the predictive power of the models. We focus on the one-step and two-step predictions. Subsequently, through knowledge of the resulting model behavior, we can extend our conclusions to multistep prediction. Although the models are built empirically, they are theoretically interpretable in terms of impact of estimated factors on inflation.

2 Data and Methods

Data of three European countries – Czech Republic, Slovakia and Germany – enter the analysis. The predicted variable is inflation in the form of HICP. As the main inflationary factors we include: GDP, three-month interbank interest rate, labor costs, unemployment, euro area inflation (the source of these items is the Eurostat database); the nominal dollar exchange rate (source database is Fxtop); M3 monetary aggregate (taken from OECD); crude oil Brent prices and the index of food prices (the source is the database of the World Bank). Generally, included are foreign (oil prices, food prices, inflation, foreign exchange rate) and domestic (GDP, interest rate, M3, labor costs, unemployment) economic indicators. We consider simultaneous relationships with delays among them, so it is advisable to model them by the vector autoregressive model.

All variables are converted to changes against to the same period of the previous year. By this manner we have reached their commensurability and adjusted seasonality. Neither TVP-VAR nor VAR-NN models strictly require stationary variables. However, we logarithm and subsequently differentiate all variables to reach stationary series, because by this manner we achieve better forecasting capabilities and easier estimation for the VAR-NN model.

We perform reverse operations for the interpretation purposes. Series are acquired in quarterly observations from the first quarter of 2001 to the fourth quarter of 2014. The latest observation is omitted from the model for creating a pseudo-prediction. German series have been taken from the first quarter of 2006 due to the missing values of unemployment for older period in Eurostat database. Gradually we compose TVP-VAR and VAR-NN models with different combinations of factors and choose the most appropriate ones based on the following methodology.

For $t = s + 1, \dots, n$, TVP-VAR model is given as

$$y_t = c_t + B_{1t} y_{t-1} + \cdots + B_{st} y_{t-s} + e_t,\ e_t \sim N(0, \Omega_t), \tag{1}$$

where y_t is $(k \times 1)$ vector of observed variables, $B_{1t}, \dots, B_{st}$ are $(k \times k)$ matrices of time-varying parameters, Ω_t is time-varying covariance matrix. Recursive relationship is given by decomposition $\Omega_t = A_t^{-1}\Sigma_t\Sigma_t A_t'^{-1}$, where A_t is lower triangular matrix with diagonal elements equal to one and $\Sigma_t = \mathrm{diag}(\sigma_{1t}, \dots, \sigma_{kt})$. Let's define β_t as a row vector $B_{1t}, \dots, B_{st}$, $a_t = (a_{1t}, \dots, a_{qt})'$ as a row vector A_t and $h_t = (h_{1t}, \dots, h_{kt})$, where $h_{it} = \log\sigma_{it}^2$. Time-varying parameters follow random walk process:

$$\begin{matrix} \beta_{t+1} = \beta_t + u_{\beta t}, \\ a_{t+1} = \alpha_t + u_{\alpha t}, \\ h_{t+1} = h_t + u_{ht}, \end{matrix} \quad \begin{pmatrix} \varepsilon_t \\ u_{\beta t} \\ u_{at} \\ u_{ht} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & \Sigma_\beta & 0 & 0 \\ 0 & 0 & \Sigma_\alpha & 0 \\ 0 & 0 & 0 & \Sigma_h \end{pmatrix}\right), \tag{2}$$

for $t = s+1, \dots, n$, with $e_t = A_t^{-1}\Sigma_t\varepsilon_t$, where Σ_a and Σ_h are diagonal matrices, $\beta_{s+1} \sim N(\mu_{\beta_0}, \Sigma_{\beta_0})$, $a_{s+1} \sim N(\mu_{a_0}, \Sigma_{a_0})$ and $h_{s+1} \sim N(\mu_{h_0}, \Sigma_{h_0})$. Further details can be found in [10] and [15]. Parameters of the model are estimated by the Markov Chain Monte Carlo (MCMC) method. We employ algorithm of Nakajima, for details see [10]. Our models have the greatest predictive performance with lag equal to 1 (and an error term has the character of white noise). Different combinations of inflationary factors are tested gradually in the models estimated with 1,000 iterations. We choose the model that achieves the best predictive accuracy on inflation measured by MSPE (Mean Square Prediction Error). Model with the selected combination of inflationary factors is estimated again with 10,000 iterations. However, there were no significant differences between the MCMC estimation with 1,000 and 10,000 iterations, so we can state that final models seems to be stable.

When defining VAR-NN model, it is necessary to modify feed-forward back propagation neural network to solve the problem of vector autoregression. We can write mathematically expressed model as follows:

$$\hat{y}_t = a^L = g(\Theta_0^L \mathrm{a}_0^{L-1} + \Theta_1^L \mathrm{a}_1^{L-1} + \cdots + \Theta_{s_{L-1}}^L \mathrm{a}_{s_{L-1}}^{L-1}), \tag{3}$$

where $\hat{y}_t$ is vector of estimated variables, $t = 1,2, \dots, n$, which is equal to neurons of the last layer a^L, L means number of layers, s_l is number of neurones (units) in the layer l. Further, Θ^l is a weight matrix, which drives transmission of intermediate inputs from the layer l to the layer $l+1$. Finally, a_i^l means activation of the i-th unit in the layer l and g is the identity transition function of the target layer.

Analogously, the units of the hidden layers are functions of all units of the previous layer, creating a complex estimated hypothesis

$$\mathrm{a}_{s_l}^l = f(\Theta_0^l \mathrm{a}_0^{l-1} + \cdots + \Theta_{s_{l-1}}^l \mathrm{a}_{s_{l-1}}^{l-1}), \tag{4}$$

where f is transitive function of the hidden layers (we use hyperbolic tangent). The input layer only repeats the individual inputs $y_{t-1}, \dots, y_{t-s}$. We use back-propagation algorithm for estimation of parameters. Price for the flexibility of neural networks is the impossibility of testing statistical assumptions and the necessity of its gradual debugging. Lag equal to one was selected as the best. Various combinations of inflationary factors were subsequently tested on network topology with 1–3 hidden layers and 1–10 neurons in each layer. For tuning was selected such a combination of factors, which achieved the best average performance on the test set for all 30 networks.

Quality of inflation prediction was not taken into account to prevent overfitting of the network and maintain sufficient performance on the testing data. For networks with selected factors, there were 60 times randomly initialized weights for topologies with one to three hidden layers and 1–10 neurons in each layer. The best network was selected according to the network performance on the test set. Such approach provides similar performance also for real dataset. Gradual debugging of possible models with continuous restriction of the test models set is typical in the field of "statistical learning".

Models with a combination of four variables (inflation plus 3 factors) seem to be the most suitable for both TVP-VAR and VAR-NN models. Adding another variable had no positive effect and cause overfitting of the model. The calculations are conducted in the computational system MATLAB R2015a.

3 Results and discussion

The best combination of variables for predicting inflation by TVP-VAR model in the Czech Republic is exchange rate, labor costs and foreign inflation. One-step predictions are plotted in *Figure 1*, where we can see that inflation is predicted very accurately. The last value for the fourth quarter of 2014 was forecasted by the two-step pseudo-prediction with error equal to -0.1 percentage points (*Table 1*). Variables labor costs, unemployment and foreign inflation were selected for the VAR-NN model. It is slightly different selection in comparison to the TVP-VAR model. Strong links to the labor market (mainly through labor cost) are economically justifiable. Also

interconnection with inflation in the euro area is clearly visible. Forecasting accuracy of the VAR-NN is good (*Figure 1)*, but estimated inflation mainly reproduces information from past inflation and ignores the additive information from the other series (its character is thus close to the one-dimensional autoregressive estimate). This fact is not simply explainable by the nature of model, because a topology of the VAR-NN comprises 2 layers, each with a single hidden neuron: four estimated values are distributed by transmission from one neuron of last hidden layer. Topology with a few hidden layers as well as a small number of neurons is expectable: the problem is relatively simple for neural networks, which are generally applicable to more complex data. Autoregressive character of the estimated VAR-NN model is also reflected in worse inflation one-step forecast MSPE compared to TVP-VAR (*Table 1*). Nevertheless, it is a fairly good forecast and the last value pseudo-prediction error was -0.1, so just as for TVP-VAR.

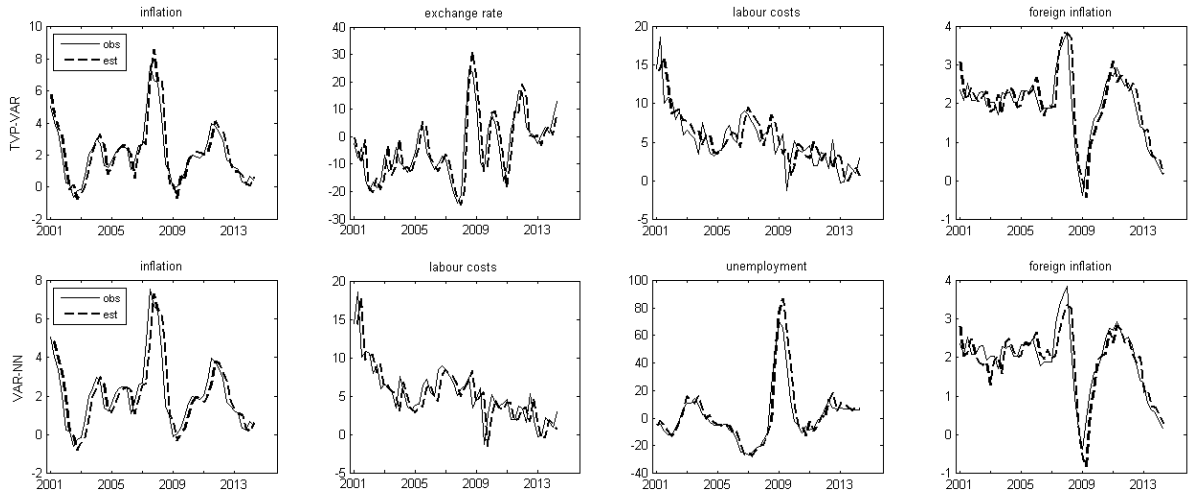


Figure 1 One-step (two-step for the last value) predictions for the Czech Republic, the first row TVP-VAR, the second row VAR-NN

When predicting inflation using TVP-VAR model in Germany, the most suitable variables are M3 aggregate, labor costs and a global index of food prices (*Figure 2*). Connection with the labor market through labor costs is reflected similarly to the Czech Republic. Foreign inflation is not taken into account, since the Eurozone inflation is largely identical to the German inflation. In contrast, the model of world food prices came as a representative of the foreign factors. In the case of Germany, it is strongly reflected a major force of TVP-VAR models: include information from the past development of the other variables. As visible in *Figure 2*, the curve of predicted inflation is more "bumpy" than the observed inflation. There is a bit worse prediction of the labor costs, which points to the rather one-sided causality from labor costs to inflation, which may be caused by fundamental reforms of the labor market in Germany in the last decade, see for example [2].

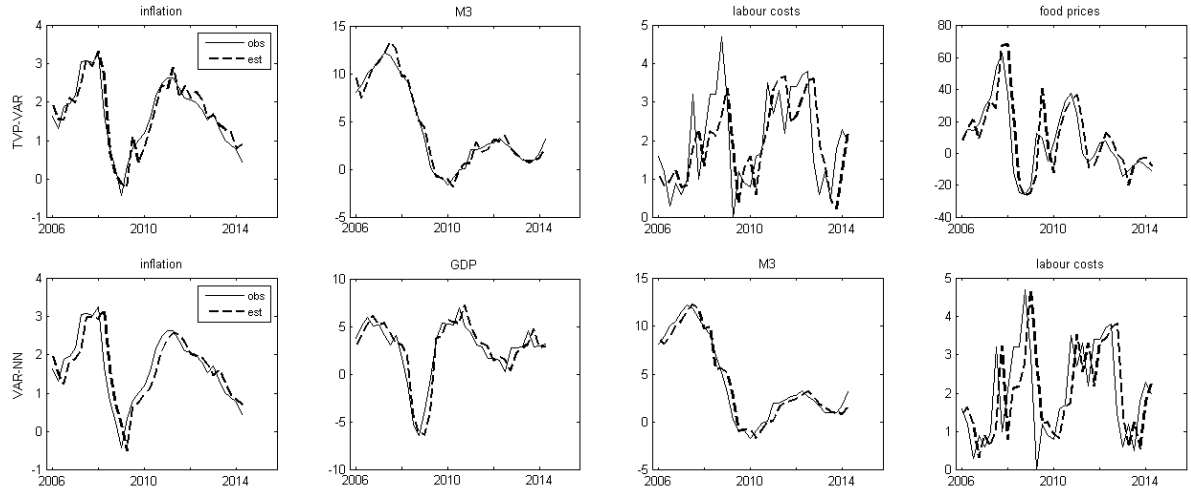


Figure 2 One-step (two-step for the last value) predictions for Germany, the first row TVP-VAR, the second row VAR-NN

NN-VAR model predicted inflation and other series quite well (*Figure 2*). Precision of prediction is given in *Table 1*. Forecasts for the last quarter of 2014 seem less accurate in comparison to the Czech Republic, but

Deutsche Bundesbank in the official forecasts of June 2014 is worse [5]: it expected the annual average inflation at value 1.3 (actually, the average was 0.8; our VAR-NN prediction was 1.1). Autoregressive character of estimated VAR-NN emerged as an advantage in this case.

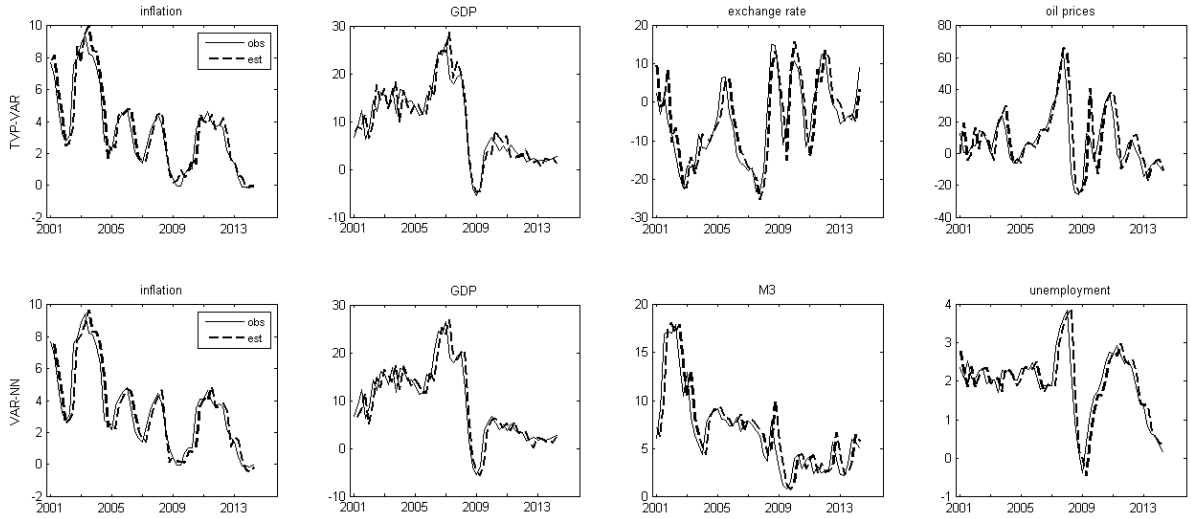


Figure 3 One-step (two-step for the last value) predictions for Slovakia, the first row TVP-VAR, the second row VAR-NN

For the inflation forecast at Slovakia using TVP-VAR we employed GDP, exchange rate and oil prices (*Figure 3*). Foreign factors have relatively large representation. Again, estimation is very accurate; for the VAR-NN model it holds too (see *Table 1*). Employed models forecast inflation more accurate than National Bank of Slovakia: the two-step forecast of the second quarter of 2014 (comparable with the horizon of our forecast) has error -0.7 percentage point [12].

	Czech republic		**Germany**		**Slovakia**	
	One-step predictions		**One-step predictions**		**One-step predictions**	
	TVP-VAR	**VAR-NN**	**TVP-VAR**	**VAR-NN**	**TVP-VAR**	**VAR-NN**
Variables	**Inflation** **Exchange rate** **Labour costs** **Foreign p.**	**Inflation** **Labour Costs** **Unemployment** **Foreign p.**	**Inflation** **M3** **Labour costs** **Food prices**	**Inflation** **GDP** **M3** **Labour costs**	**Inflation** **GDP** **Exchange rate** **Oil prices**	**Inflation** **GDP** **M3** **Unemployment**
MSPE	0.00033	0.00054384	2.73E-07	0.00011455	0.0021	0.00065937
	Last value two-step prediction		**Last value two-step prediction**		**Last value two-step prediction**	
	TVP-VAR	**VAR-NN**	**TVP-VAR**	**VAR-NN**	**TVP-VAR**	**VAR-NN**
Error	-0.1	-0.1	-0.5	-0.3	0	-0.1

Table 1 Accuracy of inflation prediction: TVP-VAR and VAR-NN with the best specification,

Table with calculated accuracy of predictions of basic VAR model and TVP-VAR model in the specification of the corresponding VAR-NN system is added for the comparison of systems with the same variables (*Table 2*).

	Czech republic		**Germany**		**Slovakia**	
	One-step predictions		**One-step predictions**		**One-step predictions**	
	TVP-VAR	**VAR(2)**	**TVP-VAR**	**VAR(3)**	**TVP-VAR**	**VAR(1)**
MSPE	0.00041	0.0032	6.00072E-06	6.5184E-04	0.0025	0.0052

Table 2 Accuracy of inflation prediction: TVP-VAR and simple VAR estimated on the same dataset as the VAR-NN model in *Table 1*

Compared to standard techniques, TVP-VAR and VAR-NN provide very good predictions especially with regard to the complexity of the predicted indicator. Inflation forecasting using autoregressive neural network models with finding the main inflationary factors is not common in practice (they are applied as one-dimensional models usually, see [11]), but it carried satisfactory results. Neural networks are flexible method that we customized to solve presented type of problem. In comparison to TVP-VAR models they showed slightly lower performance and worse consideration of external information. Generally, when compared our predictions to the values predicted by official institutions, our models gave better results. For a more accurate assessment a broader comparison should be done, but our results support the idea of other authors (for example [1], [9]) that private subjects are able to make predictions qualitatively comparable with predictions of official institutions. Moreover, errors in institutional predictions can be deliberate because of institutional discretionary freedom and their effort to influence inflation expectations. Note that comparison with the Czech National Bank prediction could not be performed, because they deal with inflation measured as CPI, while we model the HICP.

4 Conclusions

Existing applications in forecasting inflation may give poor results. We verified a suitable empirical approach which generates high-quality inflation forecasts. Selection of relevant economic indicators instead of focusing on rigid economic assumptions can significantly improve the inflation forecasts. TVP-VAR models with stochastic volatility and VAR-NN models differ from the classical econometric modelling. For both models, we cannot test the significance of variables; suitable specification and parameter settings are founded by the gradual iterative seeking through the set of all possible models. Therefore, this approach is closer to "computational statistics" than traditional statistical estimate.

Acknowledgements

Supported by the grant No. PEF_DP_2015_025 of the Internal Grant Agency PEF MENDELU.

References

[1] Abreu, I.: *International organisations' vs. private analysts' forecasts: an evaluation*. Banco de Portugal, Lisboa, 2011.
[2] Akyol, M., Neugart, M., Pichler, S.: Were the Hartz Reforms Responsible for the Improved Performance of the German Labour Market? *Economic Affairs* **33** (2013), 34-47.
[3] Bruneau C., and Bandt, O. de, Flageollet, A., Michaux, E.: Forecasting inflation using economic indicators: the case of France. *Journal of Forecasting* **26**, 1 (2007), 1-22.
[4] Ciccarelli, M., and Mojon, B.: Global Inflation. *The Review of Economics and Statistics* **92** (2010), 524-532.
[5] Deutsche Bundesbank: *Outlook for the German economy – macroeconomic projections for 2014 and 2015 and an outlook for 2016*. Deutsche Bundesbank, Berlin, 2014. Available at http://www.bundesbank.de/Redaktion/EN/Downloads/Publications/Monthly_Report_Articles/2014/2014_06_outlook.pdf?__blob=publicationFile.
[6] Dobešová, A. and Hampel, D.: Asymmetries in Effects of Monetary Policy and their Link to Labour Market - Application of TVP-VARs to Three European Countries. In: *Proceedings of the 32nd International Conference Mathematical Methods in Economics 2014*. Palacký University, Olomouc, 2014, 156-161.
[7] Dobešová, A., and Kraváčková, L.: Main inflationary factors in the Czech Republic and their change caused by economic crisis. *Forum Statisticum Slovacum* **2014**, 2 (2014), 25-30.
[8] Elliot, G., and Timmermann, A.: *Handbook of Economic Forecasting*, **2**. North Holland, Amsterdam, 2013.
[9] Gamber, E., and Smith, J.: Are the Fed's inflation forecasts still superior to the private sector's? *Journal of Macroeconomics* **31**, 2 (2009), 240-251.
[10] Nakajima, J.: Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications. *Monetary and Economic Studies* **30**, 9 (2011), 107–142.
[11] Nakamura, E.: *Inflation forecasting using a neural network. Economics Letters* **86**, 3 (2005), 373-378.
[12] Národná banka Slovenska: *Strednodobá predikcia: 2.Q 2014. Národná banka Slovenska*, Bratislava, 2014. Available at http://www.nbs.sk/_img/Documents/_Publikacie/PREDIK/2014/P2Q-2014.pdf.
[13] Öller, L.E., Barot, B.: *The accuracy of European growth and inflation forecasts. International Journal of Forecasting* **16**, 3 (2000), 293-315.
[14] Stock, J., and Watson, M: Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking* **39**, (2007), 3-33.
[15] Primiceri, G.: Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* **72** (2005), 821–852.
[16] Rabanal, P.: Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model. *Journal of Economic Dynamics and Control* **31**, 3 (2007), 906-937.

$M/E_n/1/m$ queueing system subject to two types of failures

Michal Dorda[1], Dušan Teichmann[2]

Abstract. The paper deals with a finite single-server queueing system with a server subject to two classes of breakdowns – the first class is represented by non-preemptive failures, the second class by so-called catastrophes. The non-preemptive failures do not interrupt the service process; we consider that it is possible to finish the service of the customer before we begin to repair the server. The catastrophes interrupt the service of the customer and all the customers found in the system are flushed out of the system. Moreover, the system rejects all the customers when the server is under repair after the catastrophe (that means the system is always empty during the repair of the catastrophic failure). We consider that the customers enter the system according to the Poisson process and they can wait in a queue that has a limited capacity equal to m-1. The customers are served by the single server according to the FCFS service discipline, service times are Erlang distributed. Times between the non-preemptive failures and the catastrophes and times to repair are exponentially distributed. The queueing system is modelled as a multi-dimensional Markov chain for which we present a linear equation system to obtain stationary probabilities of the individual states of the system; the equation system is solved numerically via software Matlab. On the basis of the probabilities we are able to compute some performance measures. At the end of the paper we present some results of experiments carried out with the presented mathematical model in order to present its functionality.

Keywords: $M/E_n/1/m$, queueing, failures, catastrophes, Matlab

JEL Classification: C44
AMS Classification: 60K25

1 Introduction – Our Inspiration to Create the Model

Let us start the article with a brief description of a practical example from which we drew inspiration to create the mathematical model that is described in the paper. We are specialized in modelling transport processes and therefore we develop mathematical and simulation models in this area. One of our fields of interest is modelling processes at railway stations, especially at so-called marshalling yards.

Marshalling yards are specialized railway yards that carry out a lot of tasks – see for example [1]. However, the one of the most important tasks is classification of inbound freight trains that enter the marshalling yard to be classified according to final destinations of individual wagons forming the freight train. The inbound freight trains stop on receiving tracks, where the inbound freight trains are prepared for their classification. The classification process is carried out over a hump (it is a small hill over which a hump track leads); the set of wagons is pulled by a shunting locomotive towards the hump and decoupled wagons (or groups of wagons) run by gravity onto selected classification tracks. Let us call the classification process of the inbound freight trains primary shunting.

In practice it is usual that some industrial tracks often lead to the marshalling yards. That means, apart from the inbound freight trains also trains coming from the industrial tracks enter the marshalling yard; such trains must be classified at the marshalling yard as well – let us call the process secondary shunting. Due to the fact that only single hump is found at the marshalling yard (in the conditions of the Czech Republic), when such trains are being classified, the inbound freight trains must wait; that means the primary function of the hump cannot be carried out. That is the reason why we consider secondary shunting to be non-preemptive failures of the hump – primary shunting is interrupted by secondary shunting.

Moreover, we must take into account the fact that the hump (including the follow-up infrastructure) consists of several technical devices and facilities and each of them may break down from time to time. When an element of the hump (including the follow-up infrastructure) is broken down then it is not possible to classify any trains

[1] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 325 754, e-mail: michal.dorda@vsb.cz.
[2] VŠB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, phone: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.

over the hump (that means primary and secondary shunting cannot be carried out). In such cases it is necessary to classify the trains in a different way (if it is possible). For example, the railway station "Ostrava hlavní nádraží" consists of two marshalling yards that are named "Ostrava levé nádraží" and "Ostrava pravé nádraží". If the hump at "Ostrava pravé nádraží" cannot be used due to some technical problems, all the freight trains waiting at "Ostrava pravé nádraží" are shunted to "Ostrava levé nádraží" where the trains can be classified.

On the basis of the short description of the process it is possible to conclude that the classification process can be considered (from our point of view) a queueing system subject to two classes of failures. As mentioned earlier in the text customers entering the queueing system are represented by individual requests for primary shunting whilst requests for secondary shunting correspond to the first class of failures. The first class failures are considered to be non-preemptive; that means we must finish the classification of the inbound freight train before we can start to carry out secondary shunting. The second class failures, which we assume in the model, correspond to so-called catastrophes [6] – the classification process (primary or secondary shunting) must be interrupted and no train can be classified when the catastrophe is under repair. All the trains found at the yard leave it (the yard empties) because we must classify them in a different way. And finally, no trains enter the yard when the hump is under repair.

In our previous work we published the mathematical models in which the first [2] or second class [4] of failures was assumed. In this article we decided to incorporate both classes of failures into a general mathematical model. However, the concept of several classes of the server failures is not novel. In the past, some authors studied such queueing systems. We can mention for example paper [5], the authors assumed so-called disruptive and non-disruptive server interruptions.

The paper is organized as follows. In Section 2 all the necessary assumptions are made and the general mathematical model is presented. In Section 3 some results of experiments are shown in order to demonstrate the solvability of the presented model.

2 Mathematical model

Let us start the section with some basic assumptions we must make to create the mathematical model. We model the queueing system that consists of a server and a queue with a finite capacity equal to m-1; that means the system has the capacity for m customers. Please note that the value of m is greater than or equal to 2. The customers are served according to the ordinary FCFS service discipline. Let us assume that the costumers come to the system in a Poisson process that is defined by the parameter $\lambda > 0$; that means costumer inter-arrival times are exponentially distributed with the same parameter λ. Costumer service times are Erlang distributed with the shape parameter $n \geq 2$ and the scale parameter $n\mu > 0$. As mentioned earlier in the text, the customers represent the inbound freight trains that are prepared to be classified over the hump.

Times between failures and times to repair are assumed to be distributed exponentially for both classes of the failures; see the parameters corresponding to the individual random variables in Table 1. As discussed in Section 1, the first class failures, that are non-preemptive, do not interrupt the service of the customer that is currently being served. The concept of the non-preemptive failures is close to so-called server vacations. The difference lies in the fact, that usual queueing models with server vacations consider that the server goes on a vacation only when is empty. In our case the repair of the non-preemptive failure can start even if there are some costumers in the system. The second class failures, that are called the catastrophes, immediately interrupt the service of the customer or the repair of the first class failure and empty the system. Moreover, all the costumers and first class failures entering the system when the catastrophe is under repair are rejected.

Random variable	Meaning in practice	Probability distribution	Parameters
Costumer inter-arrival times	Primary shunting	Exponentially distributed	$\lambda > 0$
Customer service times		Erlang distributed	$n \geq 2$, $n\mu > 0$
Times between failures – first class failures	Secondary shunting	Exponentially distributed	$\eta_1 > 0$
Times to repair – first class failures		Exponentially distributed	$\zeta_1 > 0$
Times between failures – second class failures	Failures of the necessary infrastructure	Exponentially distributed	$\eta_2 > 0$
Times to repair – second class failures		Exponentially distributed	$\zeta_2 > 0$

Table 1 The summary of all the random variables used in the model

Let us define three discrete random variables denoted as K, P and F. The variable K describes the number of the customers found in the system; the variable takes its values from the set $\{0,1,...,m\}$. The variable P expresses

how many phases of the service the server has already finished; the variable takes the values from the set {0,1,...,*n*-1}. The last variable F takes 4 values from the set {0,1,2,3}, where the meaning of the individual values is as follows:

- If the variable F is equal to 0, there is no failure in the system.
- If the variable F is equal to 1, the first class failure (the non-preemptive failure) is waiting to be repaired because we have to finish the service of the customer that is currently being served.
- If the variable F is equal to 2, the first class failure is under repair.
- If the variable F is equal to 3, the second class failure (the catastrophe) is under repair.

We defined the random variables in order to describe all the possible states of the modelled queueing system; the states are described by triplets (*k*,*p*,*f*). The state space of the system can be expressed as the union of four subsets:

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4 .$$

The meaning of the individual subsets is as follows:

- The subset $\Omega_1 = \{(k, p, f): k \in \{1,\ldots,m\}, p \in \{0,\ldots,n-1\}, f = 0\} \cup \{(0,0,0)\}$ includes the states for which it holds: *k* customers are in the system, *p* phases of the service of the customer have already been finished and there is no failure in the system. To these states we have to add the state (0,0,0) which corresponds to the state in which the system is idle (and empty).
- The subset $\Omega_2 = \{(k, p, f), k \in \{1,\ldots,m\}, p \in \{0,\ldots,n-1\}, f = 1\}$ includes the states for which it holds: *k* customers are in the system, *p* phases of the service of the customer have already been finished and there is the first class failure (non-preemptive) of the server waiting to be repaired.
- The subset $\Omega_3 = \{(k, p, f), k \in \{0,\ldots,m-1\}, p = 0, f = 2\}$ covers the states for which it holds: *k* customers are waiting in the system and the first class failure of the server is being repaired.
- The subset $\Omega_4 = \{(0,0,3)\}$ includes only the state for which it holds that the system is empty and the second class failure of the server is under repair.

The modelled system forms a multidimensional Markov chain. The structure of the chain can be depicted by the state transition diagram – see Figure 1.

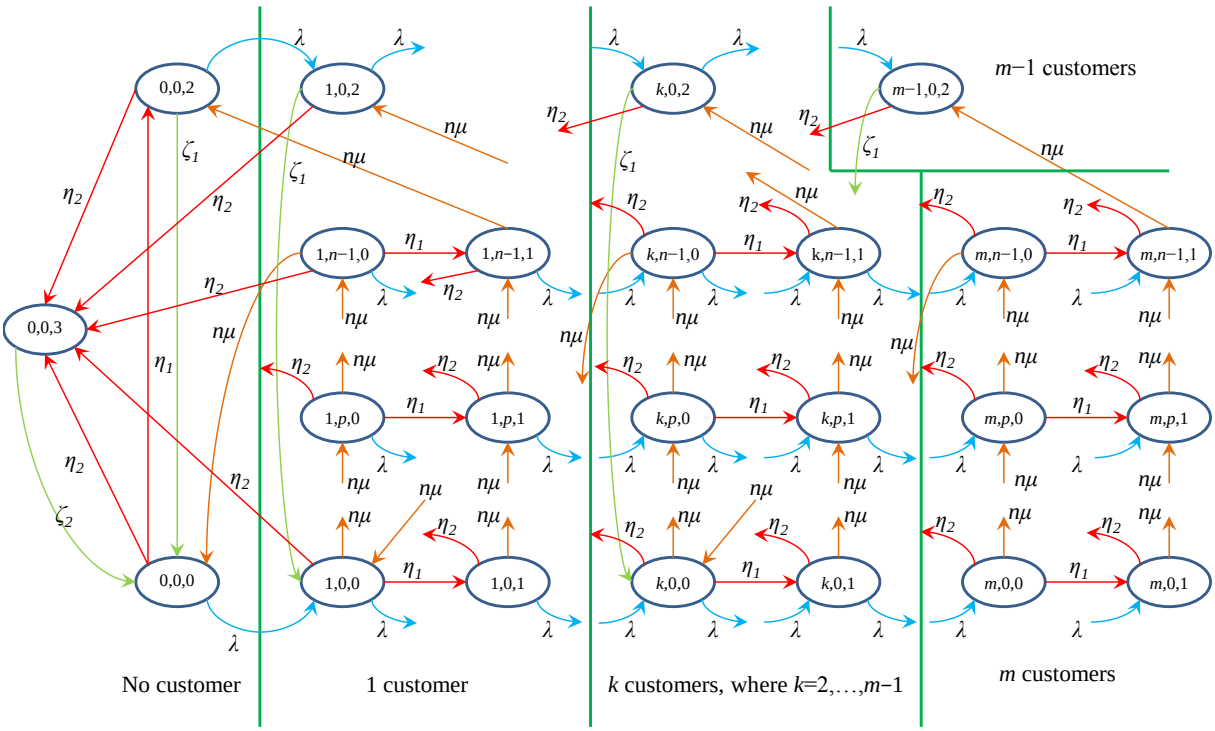


Figure 1 The state transition diagram of the modelled queueing system

On the basis of the state transition diagram we can write down linear equations for the individual stationary probabilities. The linear equation system has the following form:

$$(\lambda+\eta_1+\eta_2)\cdot P_{(0,0,0)} = n\mu\cdot P_{(1,n-1,0)} + \zeta_1\cdot P_{(0,0,2)} + \zeta_2\cdot P_{(0,0,3)}, \quad (1)$$

$$(\lambda+n\mu+\eta_1+\eta_2)\cdot P_{(k,0,0)} = \lambda\cdot P_{(k-1,0,0)} + n\mu\cdot P_{(k+1,n-1,0)} + \zeta_1\cdot P_{(k,0,2)} \text{ for } k=1,\ldots,m-1, \quad (2)$$

$$(n\mu+\eta_1+\eta_2)\cdot P_{(m,0,0)} = \lambda\cdot P_{(m-1,0,0)}, \quad (3)$$

$$(\lambda+n\mu+\eta_2)\cdot P_{(1,0,1)} = \eta_1\cdot P_{(1,0,0)}, \quad (4)$$

$$(\lambda+n\mu+\eta_2)\cdot P_{(k,0,1)} = \eta_1\cdot P_{(k,0,0)} + \lambda\cdot P_{(k-1,0,1)} \text{ for } k=2,\ldots,m-1, \quad (5)$$

$$(n\mu+\eta_2)\cdot P_{(m,0,1)} = \eta_1\cdot P_{(m,0,0)} + \lambda\cdot P_{(m-1,0,1)}, \quad (6)$$

$$(\lambda+n\mu+\eta_1+\eta_2)\cdot P_{(1,p,0)} = n\mu\cdot P_{(1,p-1,0)} \text{ for } p=1,\ldots,n-1, \quad (7)$$

$$(\lambda+n\mu+\eta_1+\eta_2)\cdot P_{(k,p,0)} = \lambda\cdot P_{(k-1,p,0)} + n\mu\cdot P_{(k,p-1,0)} \text{ for } k=2,\ldots,m-1 \text{ and } p=1,\ldots,n-1, \quad (8)$$

$$(n\mu+\eta_1+\eta_2)\cdot P_{(m,p,0)} = \lambda\cdot P_{(m-1,p,0)} + n\mu\cdot P_{(m,p-1,0)} \text{ for } p=1,\ldots,n-1, \quad (9)$$

$$(\lambda+n\mu+\eta_2)\cdot P_{(1,p,1)} = \eta_1\cdot P_{(1,p,0)} + n\mu\cdot P_{(1,p-1,1)} \text{ for } p=1,\ldots,n-1, \quad (10)$$

$$(\lambda+n\mu+\eta_2)\cdot P_{(k,p,1)} = \eta_1\cdot P_{(k,p,0)} + \lambda\cdot P_{(k-1,p,1)} + n\mu\cdot P_{(k,p-1,1)} \text{ for } k=2,\ldots,m-1 \text{ and } p=1,\ldots,n-1, \quad (11)$$

$$(n\mu+\eta_2)\cdot P_{(m,p,1)} = \eta_1\cdot P_{(m,p,0)} + \lambda\cdot P_{(m-1,p,1)} + n\mu\cdot P_{(m,p-1,1)} \text{ for } p=1,\ldots,n-1, \quad (12)$$

$$(\lambda+\zeta_1+\eta_2)\cdot P_{(0,0,2)} = \eta_1\cdot P_{(0,0,0)} + n\mu\cdot P_{(1,n-1,1)}, \quad (13)$$

$$(\lambda+\zeta_1+\eta_2)\cdot P_{(k,0,2)} = n\mu\cdot P_{(k+1,n-1,1)} + \lambda\cdot P_{(k-1,0,2)} \text{ for } k=1,\ldots,m-2, \quad (14)$$

$$(\zeta_1+\eta_2)\cdot P_{(m-1,0,2)} = n\mu\cdot P_{(m,n-1,1)} + \lambda\cdot P_{(m-2,0,2)} \quad (15)$$

$$\zeta_2\cdot P_{(0,0,3)} = \eta_2\cdot P_{(0,0,0)} + \eta_2\cdot\sum_{k=1}^{m}\sum_{p=0}^{n-1}\sum_{f=0}^{1} P_{(k,p,f)} + \eta_2\cdot\sum_{k=0}^{m-1} P_{k,0,2}\,. \quad (16)$$

Because an equation is redundant, to get a unique and admissible solution of the linear equation system we must omit one equation – for example equation (16) – and replace it by normalization condition (17) that states the sum of all the stationary probabilities must be equal to 1:

$$P_{(0,0,0)} + \sum_{k=1}^{m}\sum_{p=0}^{n-1}\sum_{f=0}^{1} P_{(k,p,f)} + \sum_{k=0}^{m-1} P_{k,0,2} + P_{(0,0,3)} = 1. \quad (17)$$

After omitting equation (16) we get the equation system of $2\cdot m\cdot n+m+2$ linear equations that is formed by equations (1) up to (15) plus equation (17) with $2\cdot m\cdot n+m+2$ unknown stationary probabilities. To solve it using Matlab it is necessary re-label the individual states because we must work with matrices in Matlab. Therefore we applied an alternative one-dimensional state description in the following form:

- The states (k,p,f) for k=1,...,m, p=0,...,n-1, and f=0,1 can be denoted by the expression $f\cdot m\cdot n+p\cdot m+k$,
- The states (k,0,2) for k=0,...m-1 can be denoted by the expression $2\cdot m\cdot n+k+1$.
- The state (0,0,0) is labelled as $2\cdot m\cdot n+m+1$.
- The state (0,0,3) is labelled as $2\cdot m\cdot n+m+2$

After calculating the stationary probabilities we are able to compute some performance measures of the modelled queueing system.

For the mean number of the costumers in the service *ES* we can write:

$$ES = \sum_{k=1}^{m}\sum_{p=0}^{n-1}\sum_{f=0}^{1} P_{(k,p,f)}\,. \quad (17)$$

The mean number of the customers waiting in the queue *EL* is given by formula (18):

$$EL = \sum_{k=2}^{m}(k-1)\cdot\sum_{p=0}^{n-1}\sum_{f=0}^{1} P_{(k,p,f)} + \sum_{k=1}^{m-1} k\cdot\sum_{p=0}^{n-1} P_{(k,p,2)}\,. \quad (18)$$

The sum of the values of *ES* and *EL* corresponds to the mean number of the customers found in the system *EK*. Regarding our example the performance measures *ES*, *EL* and *EK* relate to primary shunting.

For the mean number of the broken servers *EF* it holds formula (19):

$$EF = \sum_{k=0}^{m-1} P_{(k,0,2)} + P_{(0,0,3)}\,. \quad (19)$$

In the case of our example the value of *EF* can be interpreted as the probability the hump cannot carry out primary shunting because the hump is occupied by secondary shunting or is out of service.

The customer is rejected when the system if full or when the catastrophe is under repair. Therefore we can write for the probability of rejection P_{rej}:

$$P_{rej} = \sum_{p=0}^{n-1} \sum_{f=0}^{1} P_{(m,p,f)} + P_{(m-1,0,2)} + P_{(0,0,3)}. \tag{20}$$

3 Experimental results

In the section we present some results of the numerical experiments we carried out with the model to test its solvability. We used data from paper [3]; from the paper we took over the values of the parameters m, λ, n, $n\mu$, η_1 and ζ_1. The parameter η_2 took values from 0.0001 up to 0.001 min^{-1} with the step 0.0001 min^{-1} and the parameter ζ_2 from 0.001 min^{-1} up to 0.01 min^{-1} with the step 0.001 min^{-1}. Please note that the values of the parameters we took over are real and were acquired by surveys done at the marshalling yard "Ostrava pravé nádraží" whilst the values of the parameters η_2 and ζ_2 are fictional and are not based on reality. The summary of all the parameters used for the experiments is given in Table 2. The results of the experiments are depicted graphically – see Figures 2, 3 and 4. The figures show the dependences of the individual performance measures *ES*, *EL*, *EK*, *EF* and P_{rej} on the parameters η_2 and ζ_2.

Random variable	Applied value of the parameter
Costumer inter-arrival times	λ = 0.01520 min^{-1}
Customer service times	n = 10, $n\mu$ = 0.63622 min^{-1}
Times between failures - first class failures	η_1 = 0.01037 min^{-1}
Times to repair - first class failures	ζ_1 = 0.02464 min^{-1}
Times between failures - second class failures	η_2 = 0.0001 – 0.001 min^{-1} (the step 0.0001 min^{-1})
Times to repair - second class failures	ζ_2 = 0.001 – 0.01 min^{-1} (the step 0.001 min^{-1})

Table 2 The parameters for the numerical experiments

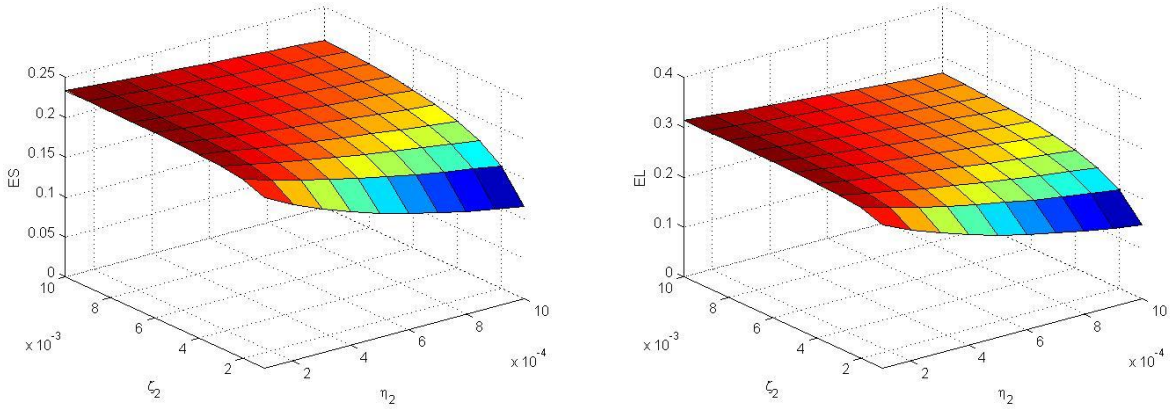


Figure 2 The dependence of *ES* and *EL* on η_2 and ζ_2

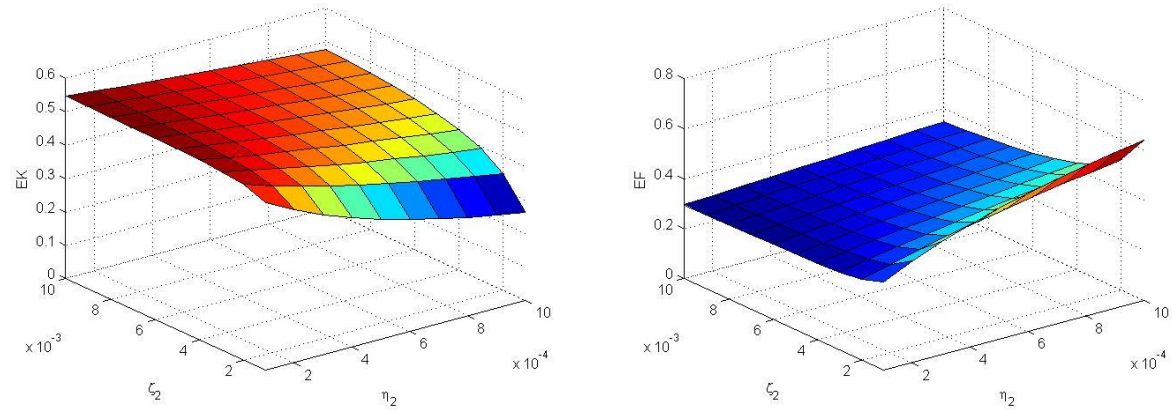


Figure 3 The dependence of *EK* and *EF* on η_2 and ζ_2

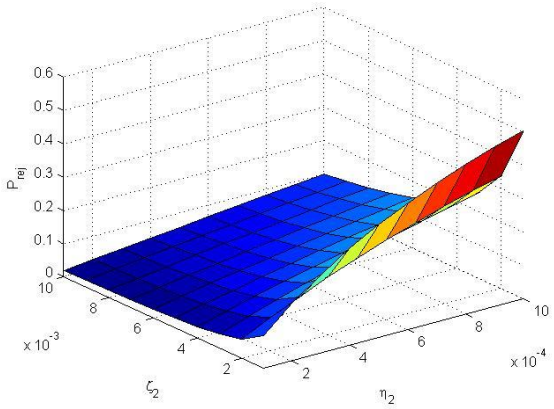


Figure 4 The dependence of P_{rej} on η_2 and ζ_2

4 Conclusions

In the article we demonstrated the single-server queueing model subject to two classes of the failures (the non-preemptive failures and the catastrophes), with the input buffer with the finite capacity, and with the Erlang distributed service times. Due to the complicated structure of the multi-dimensional Markov chain that corresponds to the modelled system, the presented mathematical model is solved numerically via Matlab software. We demonstrated the solvability of the proposed mathematical model on some numerical experiments. The numerical experiments with the mathematical model confirmed the following dependences:

- The values of *ES*, *EL* and *EK* decrease with the increasing value of η_2 and with the decreasing value of ζ_2.
- The values of *EF* and P_{rej} increase with the increasing value of η_2 and with the decreasing value of ζ_2.

Input data we used for the experiments are partially based on reality, only the values of the parameters η_2 and ζ_2 are fictional. Therefore, in the frame of our future research we would like to do a survey in order to find out estimations of the parameters.

References

[1] Boysen, N., Fliedner, M., Jaehn, F., and Pesch, E.: Shunting yard operations: Theoretical aspects and applications. *European Journal of Operational Research* **220.1** (2012), 1-14.

[2] Dorda, M., and Teichmann, D.: On finite single-server queue subject to non-preemptive breakdowns. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013, Part I.* (Vojáčková, H. ed.). College of Polytechnics Jihlava, Jihlava, 2013, 141-146.

[3] Dorda, M., and Teichmann, D. Modelling of freight trains classification using queueing system subject to sreakdowns. *Mathematical Problems in Engineering*, vol. 2013, Article ID 307652, 11 pages, 2013. doi:10.1155/2013/307652.

[4] Dorda, M., and Teichmann, D.: M/E_n/1/*m* queuing system subject to catastrophes. In: *Proceedings of the 32nd International Conference MME2014 – Mathematical Methods in Economics* (Talašová, J., Stoklasa, J. and Talášek, T. eds.). Palacký University, Faculty of Science, Olomouc, 2014, 168-173.

[5] Fiems, D., Maertens, T., and Bruneel, H.: Queueing systems with different types of server interruptions. *European Journal of Operational Research* **188.3** (2008), 838-845.

[6] Kumar B. K., Krishnamoorthy A., Madheswari S. P., and Basha S. S.: Transient analysis of a single server queue with catastrophes, failures and repairs. *Queuing Systems* **56** (2007), 133-141.

On the Crossing Number of Cartesian Products

Emília Draženská [12]

Abstract.

There are known the exact values of the crossing numbers of Cartesian products of cycles with all graphs of order at most four and with several graphs on five, six or seven vertices. In this paper, we extend these results by determining the exact values of the crossing numbers of Cartesian products of a few graphs with seven vertices and seven edges with cycles.

Keywords: graphs, drawings, crossing numbers.

JEL classification: C02
AMS classification: 05C10; 05C38

1 Introduction

Let us consider a simple *graph* G whose vertex set is V and edge set E. A *drawing* D of the graph G is a representation of a graph G in a plane in such a way, that each vertex is represented by a point in $\mathbb{R}^2$ and each edge by a curve between its two endpoints. A *crossing* of two edges is the intersection of the interiors of the corresponding curves. The *crossing number* $cr(G)$ of a graph G is the minimum number of edge crossings in any drawing of G in the plane. The drawing with a minimum number of crossings, must be a *good drawing*, that means, that in the drawing: (a) no edge crosses itself, (b) adjacent edges do not cross, and (c) no two edges cross more than once.

The problem of determining the crossing number of a given graph has been studied in graph theory and also in computer science, VLSI-layout. Many researchers have focused on optimizing the VSLI circuit layout. VLSI theory has expanded in many directions. One of them is layout theory which studied the efficiency of embedding the graphs in the plane according to VLSI rules. One of the main problem is minimizing the number of wire crossings in a circuit. C. Thompson introduced a graph-theoretic model for very large integration circuitry [21].

Garey and Johnson [9] have proved that compute the crossing number for a given graph is a very difficult problem, it is NP–complete problem, in general.

Cartesian product of two graphs is one of a several classes of graphs for which the crossing number is studied. The *Cartesian product* $G_1 \Box G_2$ of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is a graph with vertex set $V = V_1 \Box V_2$ and edge set $E = \{\{(x_1, y_1), (x_2, y_2)\}; x_1 = x_2$ and $\{y_1, y_2\} \in E_2$ or $y_1 = y_2$ and $\{x_1, x_2\} \in E_1\}$. That means, two vertices $(x_1, y_1), (x_2, y_2)$ are adjacent in $G_1 \Box G_2$ if their first coordinates are the same and their second coordinates are adjacent in G_2, or if their second coordinates are the same and their first coordinates are adjacent in G_1.

Let C_n be the cycle of length n, P_n be the cycle of length n and S_n be the star isomorphic to complete bipartite graph $K_{1,n}$. The researchers try to stay exact value of the crossing numbers for the Cartesian products $G \Box C_n$, $G \Box P_n$ and $G \Box S_n$ for some specific graphs G. Beineke and Ringeisen in [4], Jendrol' and Ščerbová in [12], Klešč in [14], [15], [16] determined the crossing numbers of the Cartesian products of all graphs on four vertices with cycles, paths and stars. Klešč, Richter and Stobert in [18], and Klešč and Kocúrová in [17] gave the crossing numbers of $G \Box C_n$ for several graphs of order five. Harary et al. [11] conjectured that the crossing number of $C_m \Box C_n$ is $(m-2)n$ for all m, n satisfying $3 \leq m \leq n$. This has been proved only for m, n satisfying $n \geq m,\ m \leq 7$, see [1], [2], [3], [4], [5], [18], [19], [20]. It was recently proved by Glebsky and Salazar [10] that the crossing number of $C_m \Box C_n$ equals its long-conjectured value at least for $n \geq m(m+1)$.

[1]The research was supported by the Slovak VEGA grant No. 1/0389/15.

[2]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, e-mail: emilia.drazenska@tuke.sk

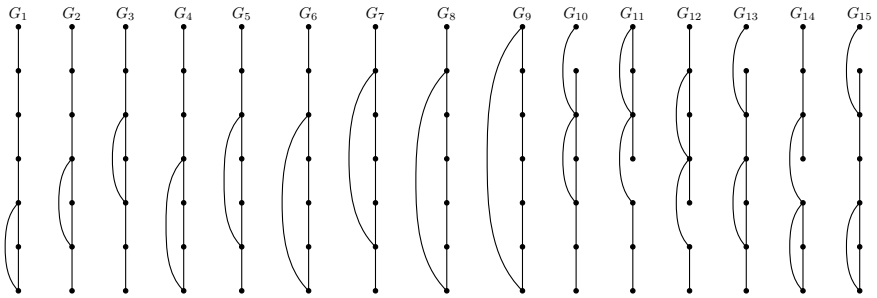


Figure 1: Several seven vertex graphs

In [6], [7] and [8] there are given the crossing numbers of $G \,\square\, C_n$ for several graph G on six vertices. It seems natural to investigate the crossing numbers of Cartesian products with cycles and graphs on seven vertices. We give the exact values of the crossing numbers for Cartesian products $G \,\square\, C_n$ for several graphs G on seven vertices and with seven edges. There are 33 such a graphs, we are interested in 15 of them (see Figure 1).

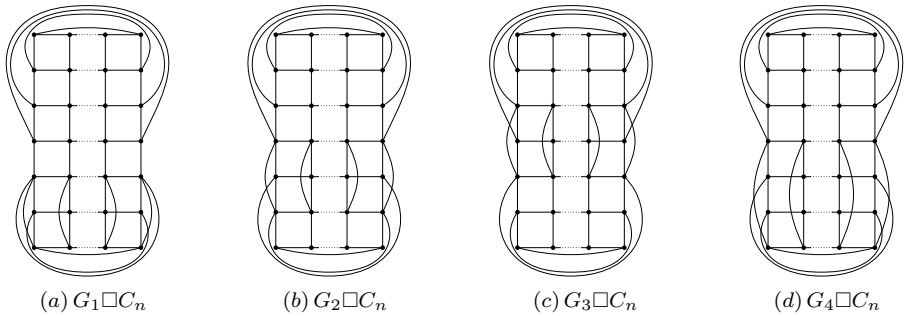
2 The crossing number of the Cartesian products of several seven vertex graphs with cycles

Theorem 1. *For $n \geq 3$, $cr(G_1 \square C_n) = cr(G_2 \square C_n) = cr(G_3 \square C_n) = n$. For $n \geq 4$, $cr(G_4 \square C_n) = cr(G_5 \square C_n) = 2n$. For $n \geq 5$, $cr(G_6 \square C_n) = cr(G_7 \square C_n) = 3n$. For $n \geq 6$, $cr(G_8 \square C_n) = 4n$. For $n \geq 7$, $cr(G_9 \square C_n) = 5n$.*

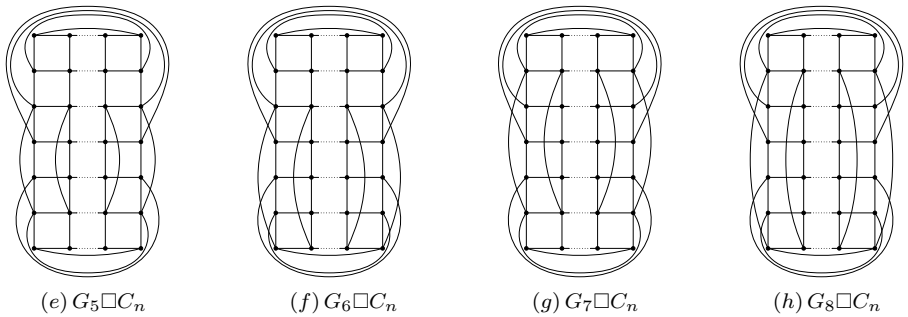
Proof: Every graph $G_j \square C_n$ contains n copies of the graph G_j. We sketch a drawing of $G_j \square C_n$ and a number of crossings in this drawing is an uppper bound for the crossing number of the graph $G_j \square C_n$. On the other hand, if a graph $G_j \square C_n$ contains some graph as its subgraph for which crossing number is published, it is a lower bound for the crossing number of the graph $G_j \square C_n$. Figures 2(a), 2(b) and 2(c) show the drawings of the graphs $G_1 \square C_n$, $G_2 \square C_n$ and $G_3 \square C_n$. The edges of every copy the graphs G_1, G_2 and G_3 are crossed once, so the crossing number of $G_1 \square C_n$, $G_2 \square C_n$ and $G_3 \square C_n$ is at most n. Figures 2(d) and 2(e) show the drawings of the graphs $G_4 \square C_n$ and $G_5 \square C_n$ in which the edges of n copies of subgraphs isomorphic to G_4 and G_5 are crossed twice, so $cr(G_4 \square C_n) \leq 2n$ and $cr(G_5 \square C_n) \leq 2n$. The drawings in Figures 2(f) and 2(g) show the graphs $G_6 \square C_n$ and $G_7 \square C_n$ with $3n$ crossings. Thus, $cr(G_6 \square C_n) \leq 3n$ and $cr(G_7 \square C_n) \leq 3n$. There is a drawing of the graph $G_8 \square C_n$ with $4n$ crossings (see Figure 2(h)), we have $cr(G_8 \square C_n) \leq 4n$. In second step, we find the lower bounds of crossing numbers $cr(G_j \square C_n)$ for $j = 1 \ldots 15$. The graph $C_3 \square C_n$ is a subgraph of the graphs $G_1 \square C_n$, $G_2 \square C_n$ and of the graph $G_3 \square C_n$ and in [20] it is proved that $cr(C_3 \square C_n) = n$, so the crossing number of $G_1 \square C_n$, $G_2 \square C_n$ and $G_3 \square C_n$ is at most n. The graphs $G_4 \square C_n$ and $G_5 \square C_n$ contain the graph $C_4 \square C_n$ as a subgraph and it is proved in [4], [5] that $cr(C_4 \square C_n) = 2n$, thus $cr(G_4 \square C_n) \geq 2n$ and also $cr(G_5 \square C_n) \geq 2n$. The graph $C_5 \square C_n$ is a subgraph of the graph $G_6 \square C_n$ and of the graph $G_7 \square C_n$. In [18] it is published that $cr(C_5 \square C_n) = 3n$. We have the lower bound $3n$ for the crossing numbers of $G_6 \square C_n$ and $G_7 \square C_n$. The graph $G_8 \square C_n$ contains the graph $C_6 \square C_n$ as a subgraph and in [19], [2] is proved that $cr(C_6 \square C_n) = 4n$, $cr(G_8 \square C_n) \geq 4n$ holds. The upper and lower bounds of crossing numbers for graphs $G_j \square C_n$ for $j = 1, \ldots 8$ are the same. Thus, we have exact values of crossing numbers of corresponding graphs. The graph G_9 is isomorphic to C_7 and using results in [1], [3], $cr(C_7 \square C_n) = 5n$. □

Theorem 2. *For $n \geq 6$, $cr(G_{10} \square C_n) = cr(G_{11} \square C_n) = cr(G_{12} \square C_n) = 2n$, $cr(G_{10} \square C_3) = cr(G_{11} \square C_3) = cr(G_{12} \square C_3) = 4$, $cr(G_{10} \square C_4) = cr(G_{11} \square C_4) = cr(G_{12} \square C_4) = 6$, $cr(G_{10} \square C_5) = cr(G_{11} \square C_5) = cr(G_{12} \square C_5) = 9$.*

Proof: In Figures 3(a), 3(b) and 3(c) there are the drawings of the graphs $G_{10} \square C_n$, $G_{11} \square C_n$ and $G_{12} \square C_n$ with $2n$ crossings. Thus, $cr(G_j \square C_n)$ is at most $2n$, for $j = 10, 11, 12$. In Figures 3(d), 3(e) and 3(f) are the drawings of the graphs $G_{10} \square C_5$, $G_{11} \square C_5$ and $G_{12} \square C_5$ with nine crossings. Thus, $cr(G_{10} \square C_5) \leq 9$, $cr(G_{11} \square C_5) \leq 9$ and $cr(G_{12} \square C_5) \leq 9$. Deleting the edges of one copy of the graph

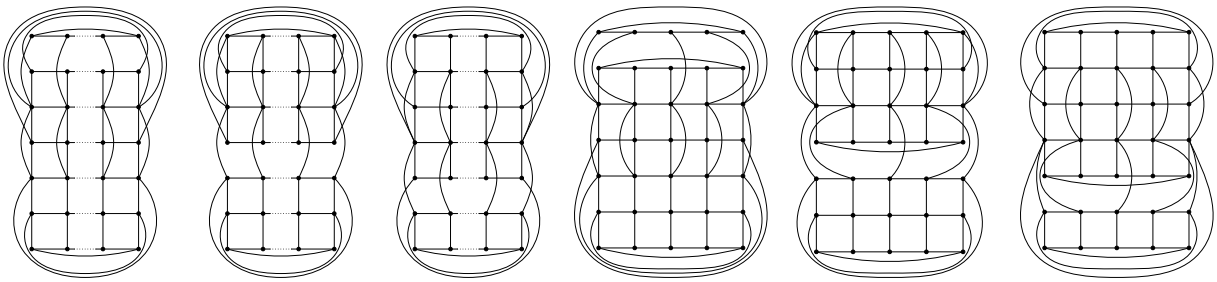


(a) $G_1\square C_n$ (b) $G_2\square C_n$ (c) $G_3\square C_n$ (d) $G_4\square C_n$



(e) $G_5\square C_n$ (f) $G_6\square C_n$ (g) $G_7\square C_n$ (h) $G_8\square C_n$

Figure 2: Cartesian products with graphs G_i for $i = 1, 2, \ldots 8$



(a) $G_{10}\square C_n$ (b) $G_{11}\square C_n$ (c) $G_{12}\square C_n$ (d) $G_{10}\square C_5$ (e) $G_{11}\square C_5$ (f) $G_{12}\square C_5$

Figure 3: Cartesian products with graphs G_i for $i = 10, 11, 12$

G_{10}, G_{11} or G_{12} with three crossings from corresponding drawing, we obtain a subdivision of the graph $G_{10}\square C_4$, $G_{11}\square C_4$ or $G_{12}\square C_4$ with six crossings. So, $cr(G_{10}\square C_4) \leq 6$, $cr(G_{11}\square C_4) \leq 6$, $cr(G_{12}\square C_4) \leq 6$. Deleting the edges of one copy of the graph G_{10}, G_{11} or G_{12} with three crossings and of one copy with two crossings from a drawing, there is a subdivision of the graph $G_{10}\square C_3$, $G_{11}\square C_3$ or $G_{12}\square C_3$ with four crossings. Hence, $cr(G_{10}\square C_3) \leq 4$, $cr(G_{11}\square C_3) \leq 4$, $cr(G_{12}\square C_3) \leq 4$. The graphs G_{10}, G_{11} and also the graph G_{12} contain the graph H (see Figure 4) as a subgraph. Thus, the graph $H\square C_n$ is a subgraph of the graphs $G_{10}\square C_n$, $G_{11}\square C_n$ and $G_{12}\square C_n$. In [16] is proved that $cr(H\square C_3) = 4$, $cr(H\square C_4) = 6$, $cr(H\square C_5) = 9$ and $cr(H\square C_n) = 2n$ for $n \geq 6$. Hence, the crossing numbers of $G_{10}\square C_n$, $G_{11}\square C_n$ and $G_{12}\square C_n$ are at least $2n$ for $n \geq 6$. The crossing numbers of $G_{10}\square C_3$, $G_{11}\square C_3$ and $G_{12}\square C_3$ are at least 4, the crossing numbers of $G_{10}\square C_4$, $G_{11}\square C_4$ and $G_{12}\square C_4$ are at least 6 and of $G_{10}\square C_5$, $G_{11}\square C_5$ and $G_{12}\square C_5$ are at least 9. This fact completes the proof. □

Figure 4: The graph H

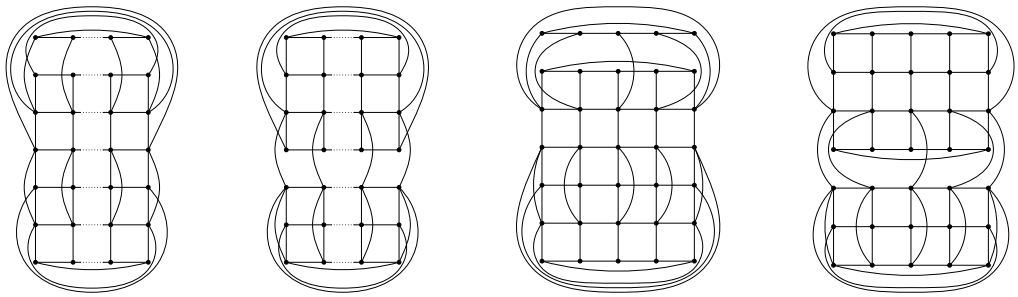
Theorem 3. *For* $n \geq 6$, $cr(G_{13}\square C_n) = cr(G_{14}\square C_n) = 2n$, $cr(G_{13}\square C_3) = cr(G_{14}\square C_3) = 2$, $cr(G_{13}\square C_4) = cr(G_{14}\square C_4) = 4$, $cr(G_{13}\square C_5) = cr(G_{14}\square C_5) = 8$.

Proof: Both graphs G_{13}, G_{14} contain the graph K (see Figure 5) as a subgraph. The facts $cr(H\square C_3) = 4$, $cr(H\square C_4) = 6$, $cr(H\square C_5) = 9$, $cr(H\square C_n) = 2n$ for $n \geq 6$ (see [7]), confirm that the crossing numbers

for the graphs $G_{13}\square C_n$ and $G_{14}\square C_n$ are at least 4, 6, 9 and $2n$ for $n = 3, 4, 5$ and $n \geq 6$. Figures

Figure 5: The graph K

6(c) and 6(d) show the drawings of the graphs $G_{13}\square C_5$ and $G_{14}\square C_5$ with nine crossings. Deleting the edges of one copy of the graph G_{13} (G_{14}) with three crossings from the drawing of $G_{13}\square C_5$ ($G_{14}\square C_5$), result is the drawing of the subdivision of $G_{13}\square C_4$ ($G_{14}\square C_4$) with six crossings. Hence $cr(G_{13}\square C_4) \leq 6$ ($cr(G_{14}\square C_4) \leq 6$). Next, we delete the edges of one more copy of the graph G_{13} (G_{14}). We have a subdivision of $G_{13}\square C_3$ ($G_{14}\square C_3$) with four crossings, Thus $cr(G_{13}\square C_3) \leq 4$ ($cr(G_{14}\square C_3) \leq 4$). In Figures 6(a) and 6(b) are the drawings of the graph $G_{13}\square C_n$ and $G_{14}\square C_n$ with $2n$ crossings. It implies, that $cr(G_{13}\square C_n) \leq 2n$ and $cr(G_{14}\square C_n) \leq 2n$. These lower and upper bounds confirms that the theorem holds. □

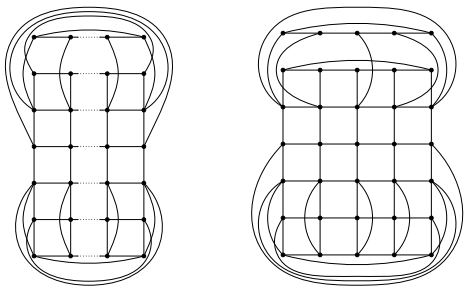


(*a*) $G_{13}\square C_n$ (*b*) $G_{14}\square C_n$ (*c*) $G_{13}\square C_5$ (*d*) $G_{14}\square C_5$

Figure 6: Cartesian products with graphs G_i for $i = 13, 14$

Theorem 4. *For* $n \geq 6$, $cr(G_{15}\square C_n) = 2n$, $cr(G_{15}\square C_3) = 4$, $cr(G_{15}\square C_4) = 6$, $cr(G_{15}\square C_5) = 9$.

Proof: The graph G_{15} contains two vertex disjoint subgraphs S_3 and C_3. So $cr(G_{15}\square C_n) \geq cr(S_3\square C_n) + cr(C_3\square C_n)$ for every $n \geq 3$. It means that $cr(G_{15}\square C_3) \geq 1 + 3 = 4$, $cr(G_{15}\square C_4) \geq 2 + 4 = 6$, $cr(G_{15}\square C_5) \geq 4 + 5 = 9$ and $cr(G_{15}\square C_n) \geq n + n = 2n$ for $n \geq 6$. Using Figures 7(a) and 7(b) (including deleting the edges of copy of subgraphs isomorphic to G_{15} in the middle and one more), one can easy see that $cr(G_{15}\square C_3) \leq 4$, $cr(G_{15}\square C_4) \leq 6$, $cr(G_{15}\square C_5) \leq 9$ and $cr(G_{15}\square C_n) \leq 2n$ for $n \geq 6$. Thus the proof is done. □



(*a*) $G_{15}\square C_n$ (*b*) $G_{15}\square C_5$

Figure 7: Cartesian products with graphs G_{15}

References

[1] Adamson, J., Richter, R. B.: Arrangements, circular arrangements and the crossing number of $C_7 \times C_n$, *J. Combinatorial Theory-B* **90** (2004), 21–39.

[2] Anderson, M., Richter, R. B., Rodnay, P.: The crossing number of $C_6 \times C_6$, *Congr. Numerantium* **118** (1996), 97–107.

[3] Anderson, M., Richter, R. B., Rodnay, P.: The crossing number of $C_7 \times C_7$, *Congr. Numerantium* **125** (1997), 97–117.

[4] Beineke, L. W., Ringeisen, R. D.: On the crossing numbers of products of cycles and graphs of order four, *J. Graph Theory* **4** (1980), 145–155.

[5] Dean, A. M., Richter, R. B.: The crossing number of $C_4 \times C_4$, *J. Graph Theory* **19** (1995), 125–129.

[6] Draženská, E.: The crossing number of $G \square C_n$ for the graph G on six vertices, *Mathematica Slovaca*, **61** (2011), 675–686.

[7] Draženská, E., Klešč, M.: On the crossing numbers of $G \square C_n$ for graphs G on six vertices, *Discussiones Mathematicae Graph Theory*,**36** (2007), 109–119.

[8] Draženská, E., Klešč, M.: The crossing numbers of products of cycles with 6–vertex trees, *Tatra Mountains Math. Publ.*, **36** (2007), 109–119.

[9] Garey, M. R., Johnson, D. S.: Crossing number is NP-complete, *SIAM J. Algebraic and Discrete Methods* **4**, (1983) 312–316.

[10] Glebsky, L. Y., Salazar, G.: The crossing number of $C_m \times C_n$ is as conjectured for $n \geq m(m+1)$, *J. Graph Theory* **47** (2004), 53–72.

[11] Harary, F., Kainen, P. C., Schwenk, A. J.: Toroidal graphs with arbitrarily high crossing numbers, *Nanta Math.* **6** (1973), 58–67.

[12] Jendroľ, S., Ščerbová,M.: On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$, *Časopis pro pěstování matematiky* **107** (1982), 225–230.

[13] Kleitman, D. J.: The crossing number of $K_{5,n}$ *J. Comb. Theory* **9**, (1970) 315–323.

[14] Klešč, M.: On the crossing numbers of Cartesian products of stars and paths or cycles, *Mathematica Slovaca* **41** (1991), 113–120.

[15] Klešč, M.: The crossing number of $K_{2,3} \times C_3$, *Discrete Mathematics* **251** (2002), 109–117.

[16] Klešč, M.: Some crossing numbers of products of cycles, *Discussiones Mathematicae - Graph theory* **25** (2005), 197–210.

[17] Klešč, M., Kocúrová, A.:The crossing number of products of 5-vertex graphs with cycles, *Discrete Mathematics* **307** (2007), 1395–1403.

[18] Klešč, M., Richter, R. B., Stolbert, I.: The crossing number of $C_5 \times C_n$, *J. Graph Theory* **22** (1996), 239–243.

[19] Richter, R. B., Salazar, G: The crossing number of $C_6 \times C_n$, *Australasian Journal of Combinatorics* **23** (2001), 135–144.

[20] Ringeisen, R. D., Beineke, L. W.: The crossing number of $C_3 \times C_n$, *J. Combinatorial Theory* **24** (1978), 134–136.

[21] Thompson, C. D.: "Area-time complexity for VLSI", *Proceedings of the eleventh annual ACM symposium on Theory of computing, ser. STOC 79*, (1979) 81–88.

Stock valuation based on prospect theory for continuous distribution

Renata Dudzińska-Baryła[1]

Abstract. Originally prospect theory was intended to evaluate risky alternatives composed of at most two non-zero outcomes, but a lot of risky decisions have more outcomes or even have continuous distribution of outcomes. The lack of consistency with stochastic dominance is also considered as "disadvantage" of prospect theory. Therefore the cumulative prospect theory was proposed. However researchers [1, 2] have reported that decision-makers' choices frequently violate the stochastic dominance. In such situations the prospect theory better describes choices. The paper exploits the proposal of Rieger and Wang [10] which extends the application of prospect theory to the decision alternatives with continuous distributions. In our paper we show the application of this approach based on real market data. Piasecki and Tomasik [9] stated that logarithmic rates of return of majority of stocks have normal-inverse Gaussian (NIG) distribution. We show the practical application of prospect theory for continuous distribution (assuming NIG and normal distributions) to create rankings of stocks quoted on Warsaw Stock Exchange.

Keywords: prospect theory, continuous distribution, stocks, normal-inverse Gaussian distribution.

JEL Classification: D81
AMS Classification: 91B06

1 Introduction

Prospect theory is a descriptive theory, which describes the way decision-makers make their decisions. The formal model is similar to the model based on the expected utility theory, but here the authors, Kahneman and Tversky [7], have also taken into account certain psychological factors based on the observation of actual decision-making process: a subjective valuation of the relative outcomes of the decision as well as a subjective valuation of probabilities.

The prospect theory is often criticized for being incompatible with the stochastic dominance, widely considered to be the basis for rational choices. This inconsistency was "corrected" in the cumulative prospect theory [11], but in the opinion of many researchers in situations where decision-makers preferences are incompatible with stochastic dominance, their choices are better described by the prospect theory than the cumulative prospect theory.

The undoubted disadvantage of prospect theory approach is the ability to evaluate decision alternatives with at most two non-zero outcomes. This defect prevents the use of it on a larger scale, because most of the real decision alternatives have more than two possible outcomes. The extension of the prospect theory for decision alternatives with a larger number of outcomes was proposed by different researchers, e.g. Karmarkar [8], Fennema and Wakker [6], Camerer and Ho [3]. On the other hand Rieger and Wang [10] proposed extension of prospect theory for non-discrete distributions. This approach offers new opportunities for applications of prospect theory, as many economic variables, such as rates of return have continuous distributions.

Rieger and Wang in [10] announced an article by Hens, Mayer and Rieger, in which the authors intended to practically apply the prospect theory for continuous distributions to historical stock market data, but such article, according to our knowledge, has not yet been published. Therefore, the aim of this paper is to show a practical example of the use of prospect theory for continuous distributions to the evaluation of selected Polish stocks.

In section 2 main principles of the prospect theory are described. Section 3 presents the proposal of Rieger and Wang, which extends the prospect theory to continuous distributions. In section 4 the prospect theory for continuous distributions is applied for creating rankings of companies belonging to the WIG20 index. In section 5 we summarize our results.

[1] University of Economics in Katowice, Department of Operations Research, 1 Maja 50, 40-287 Katowice, Poland, renata.dudzinska@ue.katowice.pl.

2 Prospect theory

The aim of the prospect theory is to describe real choices made by decision-makers. Kahneman and Tversky extended the concept of the expected utility theory by additionally taking into account the psychological aspects of the decision-making process. They use gains and losses relative to some reference point instead of the absolute outcomes of decision. Moreover, the utility function was replaced by the value function $v(x)$, which is convex for relative losses and concave for relative gains. The structure of the function reflects risk aversion for gains and simultaneous risk-prone behavior when faced with losses. In addition, the two-part form of the value function $v(x)$ allows the formal modeling of loss aversion. Another important feature of the prospect theory is the consideration of subjective probabilities which are probabilities weighted by means of the S-shaped function $w(p)$. This function reflects the behavior of decision-makers who overestimate small probabilities and underestimate the large ones.

The valuation of the decision alternative with the probability distribution $p = (x_1, p_1; x_2, p_2)$, where $p_1 + p_2 \leq 1$ and $x_1 x_2 < 0$ is calculated according to the formula

$$PT(p) = w(p_1)v(x_1) + w(p_2)v(x_2) \tag{1}$$

If both outcomes are gains or losses ($x_1 > x_2 > 0$ or $x_1 < x_2 < 0$) then the formula is as follows

$$PT(p) = v(x_2) + w(p_1)\left[v(x_1) - v(x_2)\right] \tag{2}$$

The analytical forms of functions $v(x)$ and $w(p)$ have been proposed only in 1992 [11], when the cumulative prospect theory was presented. Since then functions, as proposed by Tversky and Kahneman, are often used in empirical research. Also in this study, we adopt this approach and assume the following form of functions

$$v(x) = \begin{cases} x^{0,88}, & \text{for } x \geq 0 \\ -2{,}25(-x)^{0,88}, & \text{for } x < 0 \end{cases} \tag{3}$$

$$w(p) = \frac{p^{\gamma}}{\left(p^{\gamma} + (1-p)^{\gamma}\right)^{1/\gamma}} \tag{4}$$

where $\gamma = 0{,}69$ if probability concerns loss and $\gamma = 0{,}61$ in the case of gain.

3 Prospect theory for continuous distributions

The disadvantage of the approach proposed by Kahneman and Tversky is that it can be applied only for the decision alternatives with at most two non-zero outcomes. As the authors themselves have noted [7] the extension of formulas (1)-(2) to the decision alternatives with more outcomes is simple and in the case of alternatives with a large number of outcomes it may require extra simplification in the editing phase. For the decision alternative with multiple outcomes, both positive and negative, the following evaluation formula is proposed [3, 6]:

$$\mathrm{PT}(p) = \sum_{i=1}^{n} w(p_i)v(x_i) \tag{5}$$

The nonlinear transformation of probabilities in (4) causes that the sum of weights is less than unity. This is not a desirable property, especially in the case of decision alternatives with large or infinite number of outcomes, because it can cause $\mathrm{PT}(p)$ to diverge to $+\infty$. Karmarkar [8] has observed similar property of weights in SWU and has suggested introducing a standardization of weights. Then the weights add up to one and a weight of each outcome depends on the probabilities of other outcomes. The same procedure introduced in the formula (4) results in the following formula:

$$\mathrm{PT}(p) = \frac{\sum_{i=1}^{n} w(p_i)v(x_i)}{\sum_{i=1}^{n} w(p_i)} \tag{6}$$

Rieger and Wang [10] have proposed the method for the evaluation of decision alternative with continuous distribution. They divided the outcome space into n intervals (each of $1/n$ width) and approximated continuous probability measure p by a sequence of Dirac measures. They proved that if $p(x)$ is a probability function with exponential decay in infinity, $v(x)$ is a function of C^1 class, and for a weighting function $w(p)$ there exists some $\kappa \in (0,1)$ and some finite number $C > 0$ such that $\lim_{\varepsilon \to 0} \frac{w(\varepsilon)}{\varepsilon^{\kappa}} = C$ then the value of a decision alternative is calculated as

$$\mathrm{PT}(p) = \frac{\int_{-\infty}^{\infty} v(x)\big(p(x)\big)^{\kappa}\, dx}{\int_{-\infty}^{\infty} \big(p(x)\big)^{\kappa}\, dx} \tag{7}$$

Properties of measure (7) are the same as properties of the measure originally proposed by Kahneman and Tversky, including the lack of compatibility with the stochastic dominance for the decision alternatives with more than two outcomes.

4 Rankings of stocks – empirical example

Rieger and Wang in [10] announced an article by Hens, Mayer and Rieger, in which the authors intended to practically apply the prospect theory for continuous distributions on historical stock market data, but such article, according to our knowledge, has not yet been published.

The aim of this work is to evaluate the group of polish stocks. We assume various probability distributions of their rates of return, i.e. normal distribution (NORM) and normal-inverse Gaussian distribution (NIG). The motivation for choosing such distributions is that the normal distribution has the desired properties, but the NIG distribution is the most often empirically justified in the stocks analysis. Piasecki and Tomasik analyzed various distributions of logarithmic returns of all polish stocks quoted in 1992-2010 and in 93% of cases they failed to reject the hypothesis that the logarithmic returns have the NIG distributions.

The data for this research are quotations of twenty stocks (group of the biggest stocks which comprise WIG20 index) from the period January-December 2014. For each stock we calculated the daily rates of return ($R1 = \frac{Q_t - Q_{t-1}}{Q_{t-1}}$), logarithmic daily returns ($R2 = \ln \frac{Q_t}{Q_{t-1}}$) and a hundredfold of logarithmic daily returns ($R3 = 100 \ln \frac{Q_t}{Q_{t-1}}$), where Q_t is a quotation in time t.

For a **normal distribution** $N(\mu, \sigma^2)$ the estimator for μ is the sample mean and the estimator for σ^2 is the sample variance. Skewness and excess kurtosis equal 0.

The **normal inverse Gaussian distribution** is a subclass of the generalized hyperbolic distribution. Its probability density function has the following form:

$$p_{\mathrm{NIG}}(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\eta)^2}\right)}{\pi\sqrt{\delta^2 + (x-\eta)^2}} \exp\left[\left(\delta\sqrt{\alpha^2 - \beta^2}\right) + \beta(x-\eta)\right] \tag{8}$$

where η is a location parameter, α – a tail heaviness parameter, β – an asymmetry parameter, δ – a scale parameter and $K_1(\cdot)$ is a modified Bessel function of the second kind. The NIG distribution can model fat-tailed and skewed distributions.

In the prospect theory the outcomes are understood as gains and losses relative to some reference point. In our research we assume a "neutral" reference point of value 0. At the same time we are aware that the value of the reference point can influence the ranking of stocks (see [5]). For each stock we estimated the parameters of the normal distribution and we tested the null hypothesis that empirical distribution is the same as estimated normal distribution (Lilliefors test in the *nortest* Package of R software). Tables 1 contains means, standard deviations and p-values for $R1$, $R2$ and $R3$. For each stock means and standard deviations for $R1$ and $R2$ are almost the same (the biggest difference equals 0.0005) and every parameter of R3 is a hundredfold of the adequate parameter of R2. Only for seven stocks we fail to reject the null hypothesis about the normal distribution with significance level of 0.05 (p-values marked in bold).

	*R*1			*R*2			*R*3		
Stock	**mean**	**sd**	**p-value**	**mean**	**sd**	**p-value**	**mean**	**sd**	**p-value**
ALIOR	0.0000	0.0160	0.0460	-0.0002	0.0160	0.0278	-0.0176	1.5958	0.0278
ASSECOPOL	0.0005	0.0137	0.0013	0.0004	0.0137	0.0010	0.0417	1.3742	0.0010
BOGDANKA	-0.0009	0.0156	0.0007	-0.0011	0.0157	0.0004	-0.1067	1.5659	0.0004
BZWBK	0.0000	0.0156	**0.4316**	-0.0001	0.0156	**0.5614**	-0.0133	1.5604	**0.5614**
EUROCASH	-0.0007	0.0207	0.0006	-0.0009	0.0207	0.0014	-0.0912	2.0718	0.0014
JSW	-0.0043	0.0245	0.0258	-0.0046	0.0247	0.0208	-0.4636	2.4682	0.0208
KERNEL	-0.0007	0.0320	0.0007	-0.0012	0.0324	0.0009	-0.1168	3.2389	0.0009
KGHM	-0.0002	0.0173	0.0096	-0.0003	0.0173	0.0058	-0.0324	1.7313	0.0058
LOTOS	-0.0012	0.0175	0.0178	-0.0013	0.0179	0.0078	-0.1323	1.7864	0.0078
LPP	-0.0006	0.0224	0.0003	-0.0009	0.0224	0.0006	-0.0877	2.2421	0.0006
MBANK	0.0001	0.0164	**0.0887**	0.0000	0.0164	**0.0636**	-0.0016	1.6374	**0.0636**
ORANGEPL	-0.0005	0.0164	0.0103	-0.0007	0.0165	0.0081	-0.0658	1.6460	0.0081
PEKAO	0.0001	0.0158	**0.5211**	0.0000	0.0158	**0.4075**	-0.0018	1.5837	**0.4075**
PGE	0.0007	0.0163	**0.6868**	0.0006	0.0163	**0.5814**	0.0597	1.6324	**0.5814**
PGNIG	-0.0004	0.0186	**0.4728**	-0.0006	0.0187	**0.3305**	-0.0587	1.8734	**0.3305**
PKNORLEN	0.0009	0.0176	**0.0502**	0.0007	0.0177	**0.0565**	0.0709	1.7685	**0.0565**
PKOBP	-0.0003	0.0125	**0.8317**	-0.0004	0.0125	**0.7363**	-0.0391	1.2495	**0.7363**
PZU	0.0004	0.0132	0.0035	0.0003	0.0132	0.0021	0.0318	1.3231	0.0021
SYNTHOS	-0.0010	0.0160	0.0324	-0.0011	0.0160	0.0192	-0.1148	1.6044	0.0192
TAURONPE	0.0007	0.0161	0.0280	0.0006	0.0161	0.0168	0.0581	1.6115	0.0168

Table 1 Means, standard deviations and p-values for normal distributions of rates of return ($R1$), logarithmic returns ($R2$) and $R3$

The parameters of NIG distribution for each stock were estimated using *GeneralizedHyperbolic* Package. Table 2 contains additional parameters of rates of return: e.g. skewness and excess kurtosis as well as p-values for the Kolmogorov-Smirnov test of goodness-of-fit which was used to test the null hypothesis that the empirical distribution is the same as the estimated NIG distribution. For all stocks we failed to reject the null hypothesis, thus we assume that rates of return of all analyzed stocks can have NIG distribution.

	*R*1			*R*2			*R*3		
Stock	**skew**	**ex.kurt**	**p-value**	**skew**	**ex.kurt**	**p-value**	**skew**	**ex.kurt**	**p-value**
ALIOR	0.1094	0.6393	0.4014	0.0474	0.6105	0.4074	0.0474	0.6105	0.4067
ASSECOPOL	-0.2363	2.8618	0.9363	-0.3356	3.0477	0.9319	-0.3356	3.0477	0.9314
BOGDANKA	-0.0411	1.2093	0.8957	-0.1154	1.2039	0.8937	-0.1154	1.2039	0.8916
BZWBK	0.0303	0.1639	0.9131	-0.0200	0.1785	0.9122	-0.0200	0.1785	0.8295
EUROCASH	0.1806	1.2562	0.7841	0.0819	1.2146	0.7850	0.0819	1.2146	0.7875
JSW	-0.0545	1.0198	0.7858	-0.1654	1.0690	0.7981	-0.1654	1.0690	0.7968
KERNEL	-0.4972	7.2336	0.8164	-0.9841	10.1127	0.8159	-0.9841	10.1127	0.8162
KGHM	-0.1335	0.8230	0.8551	-0.2055	0.8422	0.8581	-0.2055	0.8422	0.8579
LOTOS	-1.7123	13.7712	0.7489	-2.0808	17.1923	0.7515	-2.0808	17.1923	0.7499
LPP	-0.0881	2.1916	0.6457	-0.2306	2.4273	0.6487	-0.2306	2.4273	0.6503
MBANK	-0.0633	0.8016	0.9551	-0.1314	0.8242	0.9523	-0.1314	0.8242	0.9522
ORANGEPL	-0.2031	1.8884	0.5363	-0.2978	1.9960	0.5288	-0.2978	1.9960	0.5259
PEKAO	-0.2892	0.6949	0.8660	-0.3520	0.8344	0.8599	-0.3520	0.8344	0.8448
PGE	-0.1763	0.6499	0.8852	-0.2410	0.7670	0.8840	-0.2410	0.7670	0.8843
PGNIG	-0.5217	2.3349	0.8811	-0.6399	2.8207	0.8738	-0.6399	2.8207	0.8732
PKNORLEN	-0.3152	2.3644	0.9964	-0.4308	2.7120	0.9961	-0.4308	2.7120	0.9961
PKOBP	-0.1006	0.4019	0.9938	-0.1455	0.4585	0.9917	-0.1455	0.4585	0.9941
PZU	-0.4402	2.3450	0.9049	-0.5228	2.5386	0.9050	-0.5228	2.5386	0.9048
SYNTHOS	-0.2752	0.6540	0.8246	-0.3367	0.6987	0.8333	-0.3367	0.6987	0.8326
TAURONPE	-0.0262	0.1932	0.3364	-0.0786	0.1830	0.3548	-0.0786	0.1830	0.3549

Table 2 Skewness, excess kurtosis and p-values for NIG distributions of rates of return ($R1$), logarithmic returns ($R2$) and $R3$

For every stock represented by the distribution function, we calculated the prospect theory value using the formula (7) with the value function of the form (3):

$$\mathrm{PT}(p) = \frac{\int_{-\infty}^{0} -2{,}25(-x)^{0{,}88}\left(p(x)\right)^{0{,}69}dx + \int_{0}^{+\infty} x^{0{,}88}\left(p(x)\right)^{0{,}61}dx}{\int_{-\infty}^{0}\left(p(x)\right)^{0{,}69}dx + \int_{0}^{+\infty}\left(p(x)\right)^{0{,}61}dx} \tag{9}$$

In formula (9) as the probability function $p(x)$ we assumed the probability density function of the normal distribution as well as the NIG distribution (formula (8)) with estimated parameters. Integrals used in (9) were calculated in R software. For continuous distributions the integrals are complicated and the amount of required computations is quite substantial, eg. to create a ranking of 20 stocks we had to compute 80 integrals. Based on the prospect theory values (not presented here) we created the rankings of stocks (table 3). Number 1 is the stock with the highest prospect theory value. Both rankings for $R1$ do not look similar. Five stocks out of all twenty differ by more than 10 positions. The same situation is for both rankings for $R2$ – four stocks differ by more than 9 positions. However both rankings for $R3$ ($R2$ multiplied by 100) are much more similar. No significant difference in rankings for hundredfold of logarithmic returns in various periods of bull and bear market were also obtained in [4].

The position of stock in the ranking depends on the assumed probability distribution function as well as the way the rate of return is defined. For example assuming the NIG distribution, PKNORLEN is the third best stock if $R1$ or $R3$ is considered and the worst of all stock if the logarithmic return ($R2$) is taken into account. Assuming the normal distribution PKNORLEN is the sixth best stock regardless of the way a rate of return is defined.

Table 4 shows the Spearman correlation coefficients of rankings for different ways of calculating rates of return and for assumed distributions (NORM and NIG). If we assume normally distributed rates of return each pair of rankings is very strong positively correlated. In other words, for the normal distribution, the position of stock do not depend so much on the way the rate of return is calculated. Assuming the NIG distribution the rankings differ much more.

Stock	***R*1**		***R*2**		***R*3**	
	NORM	**NIG**	**NORM**	**NIG**	**NORM**	**NIG**
ALIOR	10	8	10	4	10	7
ASSECOPOL	2	10	2	3	2	1
BOGDANKA	**13**	**2**	13	13	13	12
BZWBK	8	9	8	2	8	8
EUROCASH	**17**	**6**	**17**	**7**	17	14
JSW	20	20	20	19	20	20
KERNEL	**19**	**4**	**19**	**8**	19	19
KGHM	11	18	11	17	11	13
LOTOS	16	17	16	16	16	16
LPP	**18**	**5**	**18**	**6**	18	15
MBANK	9	15	9	14	9	9
ORANGEPL	12	7	12	5	12	10
PEKAO	7	1	7	1	7	11
PGE	5	14	5	12	4	6
PGNIG	15	19	15	18	14	17
PKNORLEN	6	3	**6**	**20**	6	3
PKOBP	3	11	3	9	5	4
PZU	**1**	**12**	1	10	1	2
SYNTHOS	14	16	14	15	15	18
TAURONPE	4	13	4	11	3	5

Table 3 Rankings of stocks based on the prospect theory values

NORM	*R*2	*R*3
*R*1	1	0.993985
*R*2		0.993985

NIG	*R*2	*R*3
*R*1	0.572932	0.227068
*R*2		0.302256

Table 4 Spearman correlations for pairs of rankings for the normal and NIG distributions

Although the idea of the prospect theory for continuous distribution is very interesting, the literature has so far lacked the papers presenting the practical use of the prospect theory principles in the evaluation of random rates of return. The reasons for such a situation can be seen in some difficulties that are encountered in the practical use of the this concept. The decision-maker has to know (or assume) the probability distribution functions and estimates their parameters. It is well known that the rates of return often do not have a normal distribution. The estimation of other probability functions and testing its goodness-of-fit requires a good knowledge of statistical issues that goes beyond the basic knowledge of an investor.

5 Conclusions

The idea of the prospect theory for continuous distribution has two important features: it allows to describe the choices which are inconsistent with stochastic dominance as well as it can be applied in the real economic and financial problems. In our research the rankings of stocks not only depends on the assumed probability distribution but also on the way the rates of return are calculated. In our research rates of return of all stocks can be modeled by the NIG distribution, and the normal distribution of rates of return can be assumed only for several stocks. Assuming the NIG distribution the position of stock in the ranking largely depends on the way the rate of return is defined. On the other hand the position of stock also depends on the probability distribution if ordinal or logarithmic rates of return are calculated. It is worth noting that for the same methodology of creating the ranking (prospect theory) the mathematical definition of a rate of return as well as its distribution function can largely influence the ranking. It is mainly visible especially if we assume the NIG distribution which is often considered in the analysis of stocks recently.

References

[1] Birnbaum, M., Martin, T.: Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In: *Emerging perspectives on judgment and decision research. Cambridge series on judgment and decision making* (Schneider, S., and Shanteau, J. eds.). Cambridge University Press, Cambridge, 2003, 84–107.

[2] Birnbaum, M., Navarrete, J.B.: Testing descriptive utility theories: violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty* **17** (1998), 49–78.

[3] Camerer, C., Ho, T.-H.: Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty* **8** (1994), 167–196.

[4] Dudzińska-Baryła, R.: Ocena wariantów decyzyjnych o rozkładach ciągłych na gruncie teorii perspektywy (Valuation of decision alternatives with continuous distributions based on prospect theory). *Studia Ekonomiczne. Zeszyty Naukowe Wydziałowe Uniwersytetu Ekonomicznego w Katowicach* (in print).

[5] Dudzińska-Baryła, R.: Wpływ zmiany wartości punktu referencyjnego na ocenę wariantu decyzyjnego na gruncie kumulacyjnej teorii perspektywy (Influence of the change of reference point on the valuation of decision alternative based on the prospect theory). In: *Metody i zastosowania badań operacyjnych'13*. Uniwersytet Łódzki, Łódź, (in print).

[6] Fennema, H., Wakker, P.: Original and cumulative prospect theory: a discussion and empirical differences. *Journal of Behavioral Decision Making* **10** (1997), 53–64.

[7] Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. *Econometrica* **47** (1979), 263–291.

[8] Karmarkar, U.S.: Subjectively weighted utility: a descriptive extension of the expected utility model. *Organizational Behavior and Human Performance* **21** (1978), 61–72.

[9] Piasecki, K., Tomasik, E.: *Rozkład stóp zwrotu z instrumentów polskiego rynku kapitałowego (Probability Distributions of Rates of Return of the Polish Capital Market instruments)*. edu-Libri, Kraków-Warszawa, 2013.

[10] Rieger, M.O., Wang, M.: Prospect theory for continuous distributions. *Journal of Risk and Uncertainty* **36** (2008), 83–102.

[11] Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** (1992), 297–323.

A note on effect of errors in input parameters on mean-variance efficient portfolios

Jitka Dupačová and Tomáš Petráš[1]

Abstract. Mean-variance efficient portfolios are influenced by errors due to approximation, estimation and incomplete information. Therefore, the obtained results – recommendations for the risk and portfolio manager, should be carefully analyzed. This note presents results of a simulation study devoted to the output analysis with respect to perturbed input data – expected yields and elements of their covariance matrix. The motivation comes from results of the simulation study [2] whose conclusions about the prevailing importance of expectations turn out to be substantially influenced by the chosen value of the model parameter that quantifies the level of the risk aversion of the investor. Our simulation study complements these results comparing the influence of perturbed values of expectations, variances and covariances of yields for the whole range of the risk aversion parameter.

Keywords: Markowitz model, mean-variance efficient portfolios, perturbed parameters

JEL classification: D81, G11, C61
AMS classification: 91G10, 90C31

1 The Markowitz mean-variance model

The Markowitz model is a static, single period model which assumes a frictionless market. It applies to small rational investors whose investments cannot influence the market prices and who prefer higher yields to lower ones and smaller risks to larger ones. Let us recall the basic formulation: The composition of portfolio of I assets is given by weights of the considered assets, $x_i, i = 1, \dots, I$, $\sum_i x_i = 1$. The unit investment in the i-th asset provides the random return ρ_i over the considered fixed period. The assumed probability distribution of the vector ρ of returns of all assets is characterized by a vector of expected returns $E\rho = \mu$ and by a fixed covariance matrix $\Sigma = [\mathrm{cov}(\rho_i, \rho_j),\ i, j = 1, \dots, I]$ whose main diagonal consists of variances of individual returns. This allows to quantify the "yield from the investment" x as the expectation $\mu(x) = \sum_i x_i \mu_i = \mu^\top x$ of its total return and the "risk of the investment" x as the variance of its total return, $\sigma^2(x) = \sum_{i,j} \mathrm{cov}(\rho_i, \rho_j) x_i x_j = x^\top \Sigma x$. According to the assumptions, the investors aim at maximal possible yields and, at the same time, at minimal possible risks – hence, a typical decision problem with two criteria, "max" $\{\mu(x), -\sigma^2(x)\}$ or "min" $\{-\mu(x), \sigma^2(x)\}$. The mean-variance efficiency introduced by Markowitz is fully in line with general concepts of multiobjective optimization. Accordingly, mean-variance efficient portfolios can be obtained by solving various optimization problems.

In accordance with [2] we use the scalarization technique and we shall present the results for the parametric quadratic program

$$\max_{x \in \mathcal{X}} \{\mu^\top x - \lambda x^\top \Sigma x\} \tag{1}$$

where the value of parameter $\lambda \geq 0$ reflects the level of investor's risk aversion and $\mathcal{X} = \{x \in \mathbb{R}^I : x_i \geq 0 \, \forall i,\ \sum_i x_i = 1\}$; in general, the approach is valid for an arbitrary nonempty bounded convex polyhedron.

Problem (1) is a convex quadratic program and there exist various solution techniques and theoretical results concerning its stability in dependence on elements of μ, Σ, and parameter λ. See e.g. Chapter 5.3 of [1] for the general theory and [3, 4, 5, 8] for applications to the Markowitz model. In the sequel, we shall assume that Σ is a positive definite matrix, briefly $\Sigma \in S^+$.

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, Prague 8, 18675, Czech Republic, dupacova@karlin.mff.cuni.cz

For a *fixed* value λ, the objective function $f(x;\mu,\Sigma;\lambda)$ is linear in elements of μ,Σ, hence, for each $\lambda \geq 0$ the optimal value

$$\varphi(\mu,\Sigma;\lambda) = \max_{x\in\mathcal{X}}\{\mu^\top x - \lambda x^\top \Sigma x\}$$

is a convex function in μ, Σ, hence continuous on $\mathbb{R}^I \times S^+$. It means, inter alia, that for consistent estimates μ^ν, Σ^ν, of μ,Σ, the optimal value $\varphi(\mu^\nu,\Sigma^\nu;\lambda)$ is a consistent estimate of the true optimal value. This assertion can be complemented by the rates of convergence. Using the "delta" method, cf. [10], section 7.2.2 for a succinct explanation, for asymptotically normal estimates μ^ν, Σ^ν, asymptotic normality can be proved, and asymptotic confidence region derived.

Moreover, for each $\lambda > 0$ and for all $(\mu,\Sigma) \in \mu \times S^+$, there exists unique optimal solution $x^*(\mu,\Sigma;\lambda)$, a continuous vector function of μ,Σ; cf. Theorem 5.3.2 of [1]. However, in general, its asymptotic distribution is a mixture of normal distributions.

In this note we shall assume that problem (1) was solved for certain reference or nominal values of elements of μ,Σ and we shall study influence of perturbations in these values on the output by simulation, regardless of their origin. Of course, the results depend on the value of the model parameter λ; notice the evident differences for $\lambda = 0$, i.e. no influence of perturbations in Σ and possibly multiple optimal solutions, and $\lambda > 0$.

Notice that for multinormal $\mathcal{N}(\mu,\Sigma)$ distribution of returns the value of λ may come from an underlying problem, e.g. from maximization of expected concave utility of the total return $\sum_i \rho_i x_i$, or from minimization of its VaR or CVaR, cf. [9].

2 Simulation study

In conclusions of [11], there are *The Top 10 Points to Remeber* in applications of stochastic programming models to asset, liability and wealth management, including the Markowitz mean-variance model. Let us quote:

> The point# 1: "Means are by far the most important part of the distribution of returns, especially the direction"...

Indeed, this was also the conclusion of [2] for the Markowitz model (1):

> "...errors in means are over ten times as damaging as errors in variances, and over twenty times as damaging as errors in covariances."...

The simulation study of [2] was based on monthly returns in 1.1.1980 – 1.12.1989 of 10 randomly selected stocks from the Dow Jones Industrial Average Index for fixed $\lambda = 0.02$. The influence of changes in the nominal parameters θ_0 containing selected elements of μ_0, Σ_0 obtained from historical data on the optimal, mean-variance portfolio was quantified using values of the Cash Equivalent Loss (CEL). For Markowitz model (1) CEL is equal to the relative error ratio

$$CEL_{\bar\theta} := \Big|\frac{\varphi(\theta_0;\lambda) - \varphi(\bar\theta;\lambda)}{\varphi(\theta_0;\lambda)}\Big|,$$

where $\bar\theta$ denotes the perturbed values of selected parameters; cf. [2]. Perturbations of components θ_{0i} of θ_0 were generated randomly according to

$$\bar\theta_i = \theta_{0i}(1 + k\varepsilon_i) \tag{2}$$

where ε_i are iid $N(0,1)$ random variables and the error magnitude is fixed by $k = 0.05, 0.10, 0.15, 0.20$. The average values of CEL obtained by [2] for 100 data perturbations according to (2), separately for perturbed means $\bar\mu$, variances $\bar\Sigma_{var}$ and covariances $\bar\Sigma_{cov}$, are contained in the first column of Table 1. The next columns of the table contain our extensions of these results for higher values of parameter λ.

Figure 1 plots the average CEL values (in %) as a function of λ for $k = 0.1$ Perturbations of means are the prevailing factor for for λ close to 0, but there is an evident change in ranking the importance of perturbations which appears approximately starting with $\lambda = 0.1$. For $\lambda \geq 0.5$ the average values of CEL remain approximately constant.

	λ=0.02	λ=0.2	λ=1	λ=2
k=0.05				
$CEL_{\bar{\mu}}$	$5.41 \cdot 10^{-3}$	$8.50 \cdot 10^{-4}$	$2.21 \cdot 10^{-5}$	$5.43 \cdot 10^{-6}$
$CEL_{\bar{\Sigma}_{var}}$	$6.75 \cdot 10^{-4}$	$3.75 \cdot 10^{-3}$	$2.22 \cdot 10^{-3}$	$2.11 \cdot 10^{-3}$
$CEL_{\bar{\Sigma}_{cov}}$	$2.02 \cdot 10^{-4}$	$1.73 \cdot 10^{-3}$	$9.74 \cdot 10^{-4}$	$9.26 \cdot 10^{-4}$
k=0.10				
CEL_{μ}	$2.37 \cdot 10^{-2}$	$3.35 \cdot 10^{-3}$	$8.53 \cdot 10^{-5}$	$2.10 \cdot 10^{-5}$
$CEL_{\bar{\Sigma}_{var}}$	$2.57 \cdot 10^{-3}$	$1.86 \cdot 10^{-2}$	$1.13 \cdot 10^{-2}$	$1.07 \cdot 10^{-2}$
$CEL_{\bar{\Sigma}_{cov}}$	$9.70 \cdot 10^{-4}$	$8.53 \cdot 10^{-3}$	$4.92 \cdot 10^{-3}$	$4.64 \cdot 10^{-3}$
k=0.15				
CEL_{μ}	$5.25 \cdot 10^{-2}$	$7.54 \cdot 10^{-3}$	$1.89 \cdot 10^{-4}$	$4.62 \cdot 10^{-4}$
$CEL_{\bar{\Sigma}_{var}}$	$5.55 \cdot 10^{-3}$	$4.70 \cdot 10^{-2}$	$2.84 \cdot 10^{-2}$	$2.70 \cdot 10^{-2}$
$CEL_{\bar{\Sigma}_{cov}}$	$2.70 \cdot 10^{-3}$	$2.16 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$
k=0.20				
$CEL_{\bar{\mu}}$	$8.46 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$3.34 \cdot 10^{-4}$	$8.09 \cdot 10^{-5}$
$CEL_{\bar{\Sigma}_{var}}$	$9.71 \cdot 10^{-3}$	$8.74 \cdot 10^{-2}$	$5.72 \cdot 10^{-2}$	$5.51 \cdot 10^{-2}$
$CEL_{\bar{\Sigma}_{cov}}$	$4.21 \cdot 10^{-3}$	$2.71 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	$1.60 \cdot 10^{-2}$

Table 1 Average values of CEL for error model (2) using data from [2].

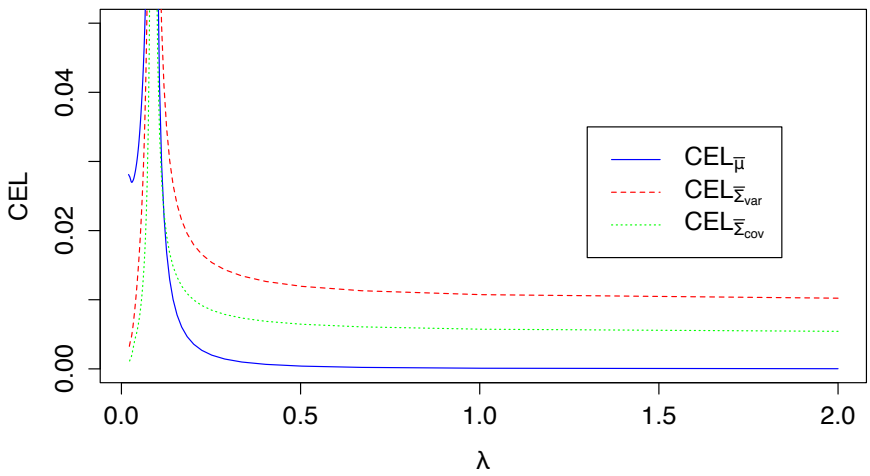


Figure 1 Dependence of CEL on risk aversion for error model (2) with $k = 0.1$.

Motivated by [6], we repeated the simulation experiment for a different construction of perturbations:

$$\bar{\theta}_i = \theta_{0i} + k\varepsilon_i\Delta_i \tag{3}$$

where ε_i are again iid random variables with $N(0,1)$ distribution and Δ_i reflects the spread of the corresponding parameter values in the sample. The results are similar, see Table 2 and Figure 2. For details see [7].

	λ=0,02	λ=0,2	λ=1	λ=2
k=0,05				
$CEL_{\bar{\mu}}$	$5.13 \cdot 10^{-3}$	$6.55 \cdot 10^{-4}$	$1.52 \cdot 10^{-5}$	$3.28 \cdot 10^{-6}$
$CEL_{\bar{\Sigma}_{var}}$	$6.56 \cdot 10^{-5}$	$5.74 \cdot 10^{-4}$	$3.99 \cdot 10^{-4}$	$3.86 \cdot 10^{-4}$
$CEL_{\bar{\Sigma}_{cov}}$	$2.42 \cdot 10^{-4}$	$2.23 \cdot 10^{-3}$	$1.21 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$
k=0,10				
$CEL_{\bar{\mu}}$	$1.91 \cdot 10^{-2}$	$2.60 \cdot 10^{-3}$	$6.40 \cdot 10^{-5}$	$1.41 \cdot 10^{-5}$
$CEL_{\bar{\Sigma}_{var}}$	$2.64 \cdot 10^{-4}$	$2.32 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$
$CEL_{\bar{\Sigma}_{cov}}$	$1.08 \cdot 10^{-3}$	$1.02 \cdot 10^{-2}$	$5.29 \cdot 10^{-3}$	$4.94 \cdot 10^{-3}$
k=0,15				
$CEL_{\bar{\mu}}$	$4.03 \cdot 10^{-2}$	$5.84 \cdot 10^{-3}$	$1.46 \cdot 10^{-4}$	$3.32 \cdot 10^{-5}$
$CEL_{\bar{\Sigma}_{var}}$	$5.96 \cdot 10^{-4}$	$5.59 \cdot 10^{-3}$	$3.96 \cdot 10^{-3}$	$3.84 \cdot 10^{-3}$
$CEL_{\bar{\Sigma}_{cov}}$	$2.75 \cdot 10^{-3}$	$2.02 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$
k=0,20				
$CEL_{\bar{\mu}}$	$6.73 \cdot 10^{-2}$	$1.03 \cdot 10^{-2}$	$2.61 \cdot 10^{-4}$	$6.04 \cdot 10^{-5}$
$CEL_{\bar{\Sigma}_{var}}$	$1.08 \cdot 10^{-3}$	$1.12 \cdot 10^{-2}$	$8.03 \cdot 10^{-3}$	$7.80 \cdot 10^{-3}$
$CEL_{\bar{\Sigma}_{cov}}$	$4.79 \cdot 10^{-3}$	$2.99 \cdot 10^{-2}$	$1.84 \cdot 10^{-2}$	$1.74 \cdot 10^{-2}$

Table 2 Average values of CEL for error model (3).

3 Conclusions

Our main conclusions can be summarized in the following two points:

- Small perturbations may cause visible relative errors in the optimal variance adjusted expected return of the portfolio.
- The main source of errors need not be the expected return, the performance depends on the model parameter λ which quantifies the level of the risk aversion of the investor.

Similar results based on simulation studies can be obtained also for maximization or minimization of other mean-risk objectives, e.g. for the mean-CVaR objective, cf. [6].

Acknowledgements

Supported by the grant No. P402/12/G097‘DYME/Dynamic Models in Economics‘ of the Czech Science Foundation.

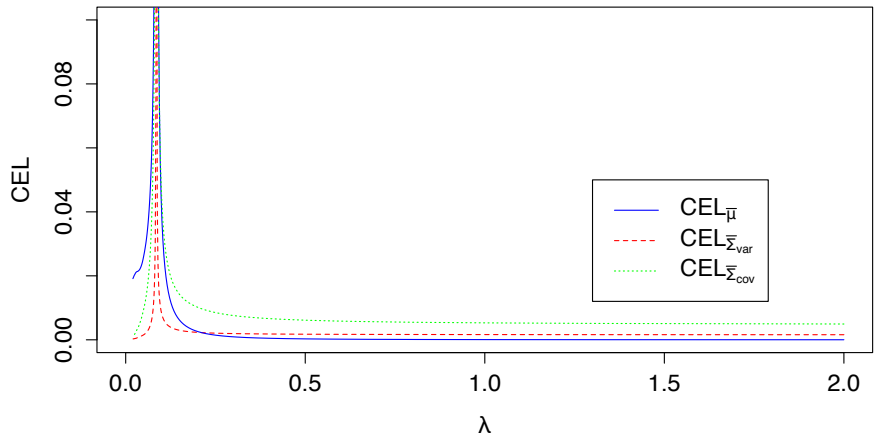


Figure 2 Dependence of CEL on risk aversion for error model (3) with $k = 0.1$.

References

[1] Bank, B., Guddat, J., Klatte, D., Kummer, B., and Tammer, K.: *Non-Linear Parametric Optimization.* Akademie-Verlag, Berlin, 1982.

[2] Chopra, W. K., and Ziemba, W. T.: The effect of errors in means, variances and covariances on optimal portfolio choice, *J. of Portfolio Mgt.* **19** (1993), 6–11.

[3] Dupačová, J.: Portfolio Optimization via Stochastic Programming: Methods of Output Analysis. *MMOR* **50** (1999), 245–270.

[4] Dupačová, J.: Output analysis and stress testing for mean-variance efficient portfolios, in: *Proceedings of 30th International Conference Mathematical Methods in Economics*,Ramík, J., and Stavárek, D. (eds.). Silesian University, School of Business Administration, 2012, pp. 123–128.

[5] Dupačová, J., and Kopa, M.: Output analysis and stress testing for risk constrained portfolios, Chapter 5 in: *Quantitative Risk Management: Theory and Practice*, C. Zopounidis, and E. Galariotis, eds. Wiley 2015.

[6] Kaut, M., Vladimirou, H. Wallace, S.W., and Zenios, S. A.: Stability analysis of portfolio management with conditional value-at-risk. *Quantitative Finance* **7** (2007), 397–409.

[7] Petráš, T.: *Robustness of the Markowitz portfolios* [In Slovak]. Master thesis, Faculty of Mathematics and Physics, Charles University, Prague (2014). Supervisor J. Dupačová.

[8] Ruszczyński, A., and Vanderbei, R. J.: Frontiers of stochastically nondominated portfolios, *Econometrica* **71** (2003), 1287–1297.

[9] Rockafellar, R. T., and Uryasev, S.: Optimization of conditional value-at-risk. *Journal of Risk* **2** (2000), 21–41.

[10] Shapiro, A., Dentcheva, D., and Ruszczyński, A.: *Lectures on Stochastic Programming. Modeling and Theory.* MPS-SIAM Series on Optimization, SIAM-MPS, Philadelphia 2009.

[11] Ziemba, W. T.: *The Stochastic Programming Approach to Asset, Liability and Wealth Management.* Research Foundation of CFA Institute (2003).

Switching Value of Public Projects

Stanislava Dvořáková[1], Petr Jiříček[2]

Abstract. The goal of the paper is analysing the switching value which is part of sensitivity analysis for pubic project socio-economic efficiency evaluation by means of cost-benefit analysis pursuant the EU methodology. Determining the switching value of a variable means finding the value of a critical variable at which the economic net present value (ENPV) criterion of a project does not reach the accepted minimum. The critical variable of a project will be represented by the resulting amount of socio-economic cash-flow of the project in the examined year when a crisis situation in its implementation occurs.
The switching value analysis will be carried out by means of a mathematical model using function, which corresponds to the mathematical representation of the criterion for evaluating the socio-economic efficiency of a public project. The model will be designed for switching value analysis of both conventionally designed projects and for nonconventional projects. The created model will be applied in a case study for calculating the switching value of variables of a particular supposed public project under the conditions of the amount of social discount rate determined by the EU.

Keywords: switching value, nonconventional public projects, cost-benefit analysis, rational function, sensitivity analysis, economical net present value.

JEL Classification: C20, H43
AMS Classification: 65H04

1 Introduction

Public project social efficiency evaluation is executed by means of cost-benefit analysis, which, based on theoretical principles of the optimising Kaldor–Hicks criterion, quantifies benefits from a project (definition in [1]). In the area of public project evaluation, cost-benefit analysis uses the modern apparatus of investment analysis used in commercial sector and working on the bases of discounting cash-flows from the investment [2]. The purpose of a public project implementation is reaching the state when its positive benefits overweigh the losses. Public project impact evaluation is used in the first degree on the project financial analysis level when accounting revenues and costs of a project (including investment costs) are defined and quantified on the basis of so called financial cash-flows. On the level of project economic analysis, these financial cash-flows are aggregated with quantified positive and negative externalities caused by the project into the resulting socio-economic cash-flow. To judge social efficiency of public project, the EU methodology for public project social efficiency evaluation uses economic net present value ($ENPV$), which is the discounted value of socio-economic cash-flows. If $ENPV$ is positive, the project is eligible for implementation and will receive support from the European funds [3]. During public project implementation, however, there are crisis situations in their construction, operational or liquidation stages which can change the resulting benefit of the project. The EU methodology therefore uses sensitivity analysis to identify critical input variables both in the financial analysis stage and in the resulting economic project analysis and one of the components of sensitivity analysis is also identifying so called *switching value* of the project (see [3]).

2 Goals and methods

The goal of the paper is to model the switching value of a critical variable as part of sensitivity analysis on the level of socio-economic cash-flows of a public project. Sensitivity analysis focuses on identifying critical variables of a project when individual input/independent variables of the project (cash-flow) gradually change by certain percentage and the consequent changes of output/dependent variable, here $ENPV$, are monitored. Changes are monitored both on the basis of project financial analysis, i.e. on the level of commercial parameters – accounting revenues and expenses – and on the resulting basis of project economic analysis, i.e. on the level of quantifying positive externalities (benefits) and negative externalities (costs, losses). Based on the *ceteris paribus* rule, only one input variable is always changed and the other parameters stay the same. According to the

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz.
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, jiricek@vspj.cz.

EU methodology, critical variables are the ones whose positive or negative change by 1% causes the resulting change of the output variable ($ENPV$) by 5% [3].

Such small selected percentage changes, however, may not necessarily correspond to variable changes in real life projects, where more significant crisis situations can occur in both their investment, operational and liquidation stages. For this reason, the EU methodology also deals with the calculation of the threshold values of variables by means of which critical value of the criterion of public project social efficiency evaluation can be identified. The equation $ENPV = 0$ is considered the critical value of the criterion at which the project is rejected. According to [3] it can alternatively be required that the resulting economic net present value does not exceed the minimum level given by the project donor.

The examined input variable of the project in the model will thus be the resulting value of socio-economic cash-flow in the selected examined year of a public project life span when the model crisis situation is expected. *The threshold value* of this critical cash-flow will cause that the resulting benefit from project implementation will equal zero and the situation when a public project brings no social benefit occurs. *The switching value* can be defined in the form of a percentage deviation from the given variable (more in [4]).

In the basic version, switching value analysis will be carried out by means of a mathematical model using rational function of the sixth degree, which corresponds to mathematical representation of the criterion for evaluation of socio-economic efficiency of a 6-year-long public project financed from the European funds, i.e. of economic net present value ($ENPV$).

The model for switching value analysis will be designed both for a conventionally proposed project and for the variant of a nonconventional project where multiple alternations of project caused positive and negative cash-flows occur. The created model will consequently be applied in a case study for calculating the switching value of variables of specifically proposed public project under the conditions of social discount rate determined by the EU for the programming period of 2007–2013.

3 Results and discussion

In case of public projects of development character, the net present value method ($ENPV$) says which cash-flow caused by the investment will be left after deducting investment costs in the pre-projected life span

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}, \tag{1}$$

where CF_t is the cash-flow of the public development project (i.e. the benefits) minus costs, losses of the project,

k is the common social discount rate (the required return on public investment from the perspective of the investor or donor of the project)

n is the period of economic lifespan of the project (in our model case $n = 6$),

$ENPV$ is economic net present value of the public project reflecting the benefit from the project from a wider social economic view of the given investment.

The criteria for public development project efficiency evaluation, which essentially correspond to the criteria used for business sector (see [6]):

$ENPV > 0$ – Investment will increase net economic wealth of the investor (project is accepted).

$ENPV < 0$ – Investment will decrease net economic wealth of the investor (project is rejected).

$ENPV = 0$ – Investment will not change net economic wealth of the investor (indifference between accepting and rejecting the project).

For identifying switching value, we use a model public project where the project cash-flow equals $CF = [CF_0, CF_1, CF_2, CF_3, CF_4, CF_5, CF_6] = [-3, 1, 1, 1, 1, 1, 1]$.

In case of conventional projects there is only one sign change in the CF number sequence; in case of nonconventional project, there are more than one sign changes in the sequence of cash-flow stream.

To determine the switching value we define the following model scenarios:

A. Classical conventional project with the cash-flow sequence $CF = [-, + + + + ++]$.

B. Conventional project with critical introduction of operational stage with the sequence $CF = [-, - + + + ++]$.

C. Nonconventional project with crisis situation during the course of operational stage with the sequence $CF = [-, + + - + ++]$.

D. Nonconventional project with crisis situation in the final liquidation stage with the sequence $CF = [-, + + + + +-]$.

In Figure 1 we can see construction of individual curves (1) of the project, (definition in [7]) for simulating the above stated scenarios of the public project model. In all the scenarios, the investment curves have only one positive real root, namely $k_0 = 0.243$ for scenario A, $k_0 = 0.062$ for B, $k_0 = 0.084$ for C and $k_0 = 0.136$ for scenario D.

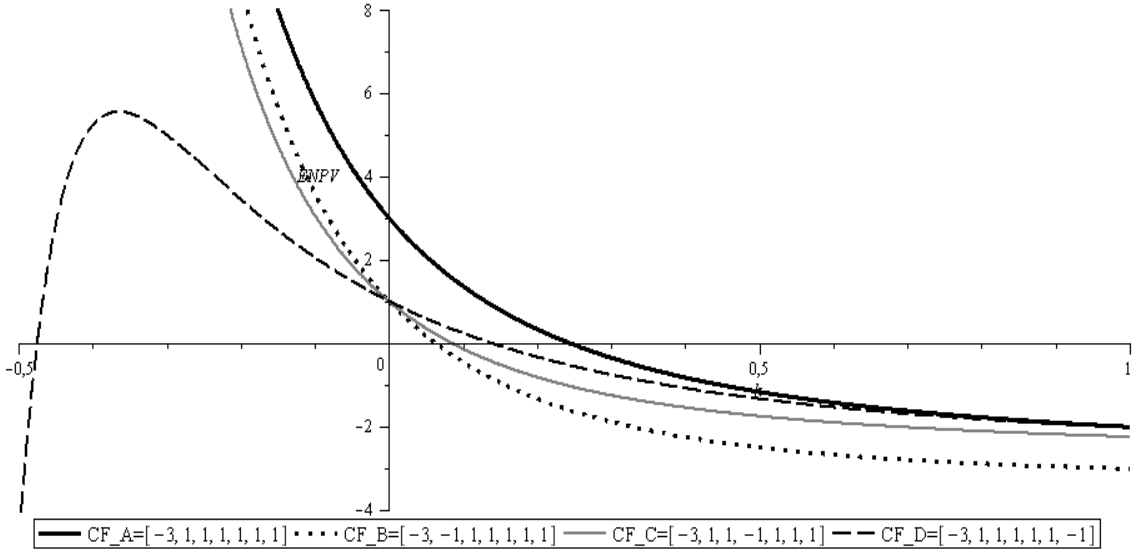


Figure 1 Investment curves of the project $ENPV = f(k)$ for individual scenarios

Now we will identify the threshold values of input variables, i.e. the project cash-flows for a selected discount rate. It will be social discount rate determined by the donor of projects, i.e. by the European Union for financial support from European funds in the programming period of 2007–2013. The rate of 5.5 % p.a. is valid for financing public project of the EU member states from the Structural Funds and Cohesion Fund.

From formula (1) a general formula for calculating the variable CF_i (for $i = 0, \dots, n$) can be derived:

$$CF_i = ENPV_{sw}(1 + k_{sw})^i - \sum_{t=0}^{i-1} \frac{CF_t}{(1 + k_{sw})^{t-i}} - \sum_{t=i+1}^{n} \frac{CF_t}{(1 + k_{sw})^{t-i}} = -\sum_{t=0}^{i-1} \frac{CF_t}{1.055^{t-i}} - \sum_{t=i+1}^{n} \frac{CF_t}{1.055^{t-i}}, \qquad (2)$$

while the condition $ENPV_{sw} = 0$ (see above) and discount rate $k_{sw} = 0.055$ was considered in our case.

In Table 1 we analyse the threshold values of critical variables in conventional projects, i.e. such values of cash-flows in individual years of model project scenarios for which the economic net present value of the project is equal to zero and the project does not bring social benefit. The switching value height subsequently indicates the percentage deviation from the initial estimated cash-flow value. The deviation is necessarily negative as only a drop of project cash-flow value causes a drop of benefit judged on the basis of the $ENPV$ criterion (for significant criteria see above). Pursuant the results in Table 1 it can be stated that in the examined model scenarios (conventional project and conventional project with critical introduction of the initial stage), the switching value in case of unit cash-flows rises proportionally to the length of ongoing project implementation. Yet the model scenario B is clearly more sensitive and even a slight change of switching value indicates a loss of benefit from the project measured by $ENPV$.

scenario	A			B		
Year	Proposed CF	Threshold CF value	Switching value	Proposed CF	Threshold CF value	Switching value
0	-3	-5.00	-67%	-3	-3.10	-3%
1	1	-1.11	-211%	-1	-1.11	-11%
2	1	-1.22	-222%	1	0.89	-11%
3	1	-1.34	-234%	1	0.88	-12%
4	1	-1.47	-247%	1	0.88	-12%
5	1	-1.61	-261%	1	0.87	-13%
6	1	-1.75	-275%	1	0.86	-14%

Table 1 Switching values and cash-flow threshold values of individual conventional project scenarios

In Table 2 we analyse the threshold value of variables in nonconventional projects with multiple alternation of positive and negative cash-flows, namely for the scenarios with a crisis in the operational stage of the project and with a crisis in the final liquidation stage of the project. We can see that due to negative cash-flows, the sensitivity of the project benefit measured by means of $ENPV$ to the change of the variables decreased. It is obvious that the more distant a negative cash-flow (meaning a crisis in the project) is from the beginning of the project implementation, the larger negative changes of cash-flows are necessary to reject the project.

scenario	C			D		
year	Proposed CF	Threshold CF value	Switching value	Proposed CF	Threshold CF value	Switching value
0	-3	-3.292	-10%	-3	-3.545	-18%
1	1	0.692	-31%	1	0.425	-58%
2	1	0.675	-33%	1	0.393	-61%
3	-1	-1.343	-34%	1	0.360	-64%
4	1	0.638	-36%	1	0.325	-68%
5	1	0.618	-38%	1	0.288	-71%
6	1	0.597	-40%	-1	-1.752	-75%

Table 2 Switching values and cash-flow threshold values of individual nonconventional project scenarios

In the following Figures 2 and 3, the initial cash-flows in individual simulated scenarios of the model project and threshold values of cash-flows are graphically presented, while there is a loss of benefit at the critical value of $ENPV = 0$. The percentual deviation between both the variables in individual years of the project life span is the demanded switching value.

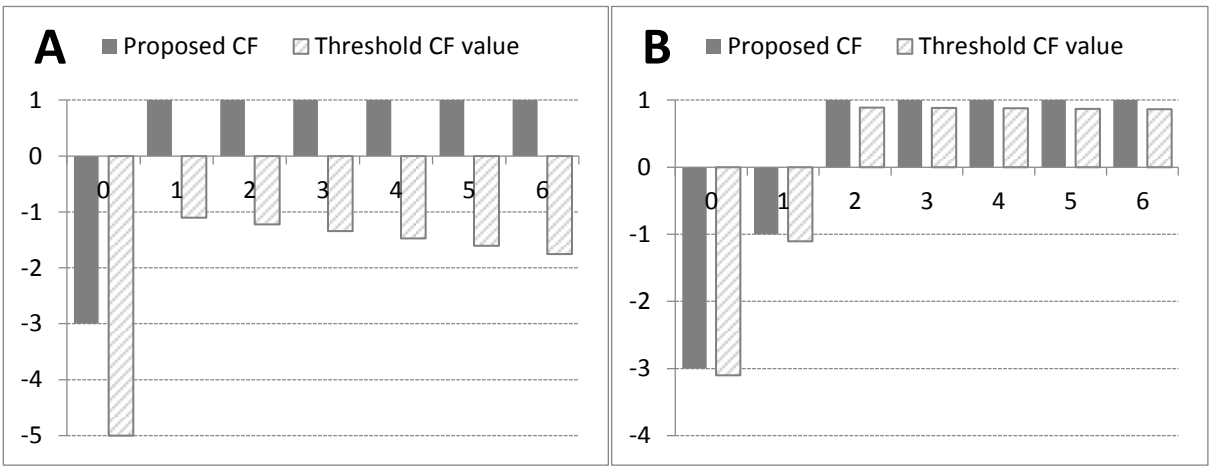


Figure 2 Cash-flow threshold values of individual scenarios of conventional projects

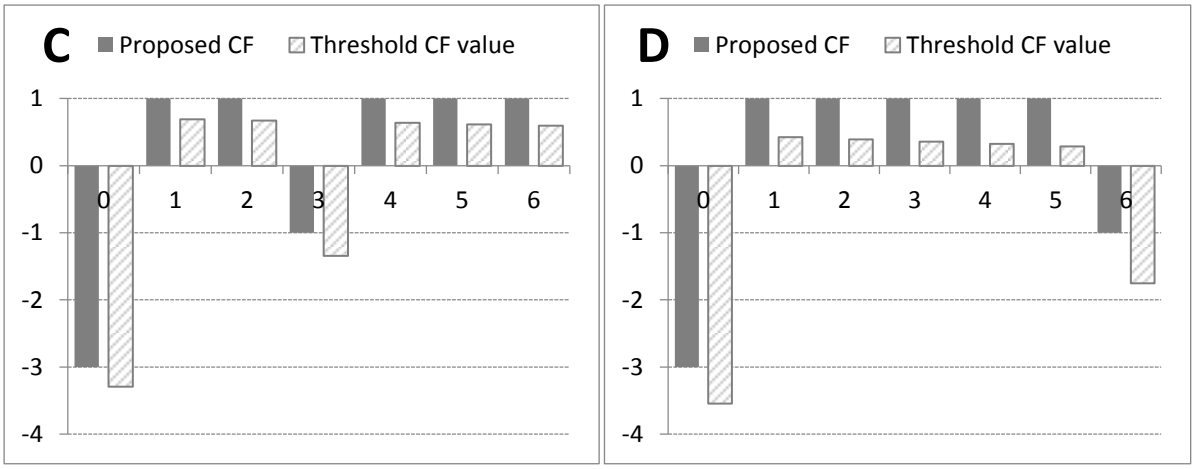


Figure 3 Cash-flow threshold values of individual scenarios of nonconventional projects

The model for switching value calculation will now be tested on a real-life project financed from the European funds. It is a project of the technical centre and electronic filing service of the Olomouc Region, which was proposed for implementation in 2013–2022. In the model, the project switching values are determined for the period of 13 years based on cash-flows quantification on the level of project economic analysis. From the analysis of Table 3 it can be stated that the project is proposed with very low sensitivity to cash-flow changes in individual project years. Unlike the basic model, this model shows considerable fluctuations in switching values.

t	Year	Project CF in mil. CZK	Threshold value in mil. CZK	Switching value
0	2010	-0.58	-159.67	-27613%
1	2011	-57.27	-225.12	-293%
2	2012	-0.30	-177.38	-59025%
3	2013	35.58	-151.23	-525%
4	2014	35.58	-161.51	-554%
5	2015	35.58	-172.35	-584%
6	2016	35.58	-183.78	-617%
7	2017	35.58	-195.85	-650%
8	2018	15.58	-228.58	-1567%
9	2019	29.58	-228.01	-871%
10	2020	29.58	-242.17	-919%
11	2021	29.58	-257.12	-969%
12	2022	29.58	-272.89	-1023%

Table 3 Switching values and cash-flow threshold values of the case study

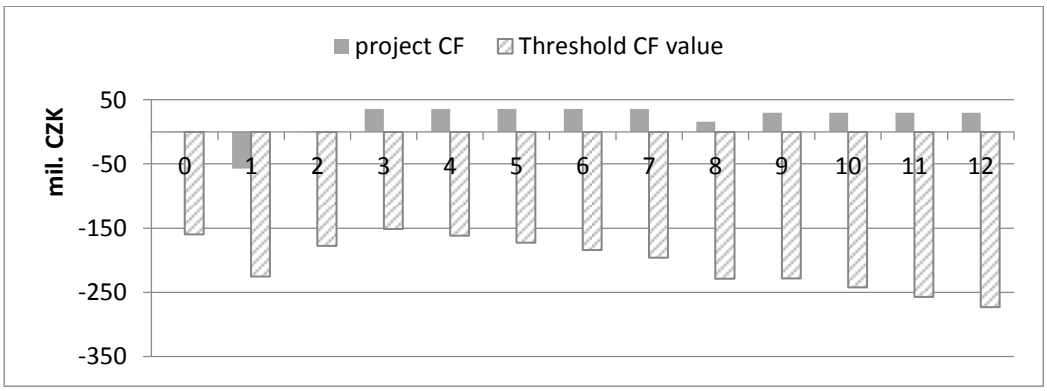


Figure 4 Cash-flow threshold values of individual scenarios of nonconventional projects

4 Conclusion

Determining the switching value of a project can be defined as finding such a value of the percentual deviation from the project's cash-flows that will cause its zero benefit measured by means of economic net present value. That $ENPV$ is the basic method for evaluating social efficiency of public projects financed from the European funds. The switching value indicates the sensitivity of the output variable of the project to the change of the input variable while following the ceteris paribus rule. Low switching value means high sensitivity and low stability of the proposed project efficiency and high values, on the contrary, mean low sensitivity of the project benefit to the cash-flow change, while on the level of economic project analysis it cannot be distinguished whether these are cash-flows or flows from externalities valued via shadow prices. However, from the case study analysis it can be deduced that extremely high switching values may not necessarily mean high stability of the project social efficiency but rather an oversized quantification of the cash-flow from externalities. This indication could consequently be confirmed or disproved by a more detailed analysis of cash-flow quantification, especially in the area of positive externalities valuation. Further development of the proposed model can be carried out in the direction of a more detailed decomposition of cash-flows or in the area of multi-parametrical determining of switching value (e.g. related to discount rate).

References

[1] Mishan, E. and Quah, E.: *Cost-Benefit Analysis.* 5th edition. London: Routledge, 2007. 316 p. ISBN 978-0-415-34991-8

[2] Brealey, R. and Myers, S. and Allen, F.: *Principles of Corporate Finance*. 2th. edition. New York: McGraw-Hill Irwin, 2011, 524 s. ISBN 978-0-07-353074-1

[3] *Guide to Cost-Benefit Analysis of Investment Projects* [online]. Brussels: European Commission. DG Regional Policy. 2008 [cit. 2013-11-17]. Available at http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf

[4] Belli, P. and Anderson, J.R. and Barnum, H.N. and Dixon, J.A. and Tan, J.P.: *Economic Analysis of Investment Operations*. 1st edition. World Bank Institute. Washington: 2001. 263 p. ISBN 0-8213-4850-7

[5] *Studie proveditelnosti projektu Technologické centrum a elektronické spisové služby* [online]. Olomoucký kraj. 2011.[cit. 2015-01-08]. Available at http://www.kr-olomoucky.cz/clanky/dokumenty/202/priloha-c1-studie-proveditelnosti-technologicke-centrum-a-elektronicka-spisova-sluzba.pdf

[6] Dvořáková, S. and Jiříček, P.: *Modelling Financial Flows of Development Projects Subsidized from European Funds*. In: 31st International Conference on Mathematical Methods in Economics. Jihlava: VŠPJ. 2013. p. 119-128. ISBN 978-8097035-79-4

[7] Dvořáková, S. and Jiříček, P.: Závislost průběhu investiční křivky na změně vstupních parametrů pro potřeby hodnocení projektů metodou cost-benefit analýzy. In: Matematika v ekonomické praxi. Jihlava: VŠPJ. 2012. s. 119-128. ISBN 978-80-87035-62-7

Monopoly profit maximization model in the conditions of perfect price differentiation at a non-linear price setting

Prof., Dr. Michal Fendek [1], Prof., Dr. Eleonora Fendeková [2]

Abstract. We can speak of a price differentiation of a monopoly production when a monopoly uses its market position or economic power to set and enforce a market price in accordance with its interests and thereby ensures maximum profit on a relevant market. We can say that a monopoly uses its position to reach a monopoly profit which exceeds profit on a perfect competition market.
A tool to reach this goal is a market price and a monopoly is at relatively wide liberty to set this price. The very practice of defining different prices to the same products is widely implemented in a market environment and is not at all a typical attribute of monopolies only. Most companies use price differentiation of some kind whether it's a price differentiation of regional markets in a country of origin, a price differentiation between domestic and foreign markets, between wholesale and retail or simply between different customers. Bloom [2] shows, that only small percentage of companies does not use any kind of price differentiation towards its customers at all.
In this article we discuss the general aspects of quantitative analysis of monopoly profit differentiation models while analyzing an optimization monopoly profit maximization model in the conditions of perfect price differentiation at a non-linear price setting more closely. This type of price differentiation is also known as second degree price differentiation. In this case a monopoly sets a differentiated price by setting different prices for different purchase volumes.

Keywords: Perfect price rice differentiation, profit maximization model, non-linear price setting, optimality conditions.

JEL Classification: D11, D43, L1144
AMS Classification: 49M05

1 Introduction

Price differentiation at a non-linear price setting is also known as a second degree price differentiation. Essentially it is a price differentiation when a monopoly sets different prices for different volumes of purchases which is generally called volume rabat. With this price differentiation a price-supply function $p = s(x)$ defines price *p*, for which a monopoly is willing to sell *x* units of a product. Monopoly revenue $e(x)$ is then also a non-linear function of a sale volume, that is a product of a price and respective supply in a form:

$$e(x) = s(x)x \tag{1}$$

Let's now assume that an *i*-th consumer S_i for i=1, 2,..., m is willing to purchase x_i units of a price differentiated good at a price $p_i = s(x_i)$ and is therefore willing to pay an amount of

$$r_i = s(x_i)x_i \tag{2}$$

monetary units. Ultimately, for a consumer as well as for a monopoly it is only important that an i-th consumer is willing to pay for a purchase of x_i units of a product the amount r_i monetary units. Therefore a monopoly de facto doesn't analyze the differentiated price $p = s(x)$, but instead it is interested in determining an optimal combination (r_i, x_i).

[1] University of Economics Bratislava, Department of Operations Research and Econometrics, Dolnozemska cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: fendek@dec.euba.sk.
[2] University of Economics Bratislava, Department of Business Economics, Dolnozemska cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: nfendek@dec.euba.sk.

.

2 Monopoly profit maximization model in the conditions of perfect price differentiation

Following we will address the analysis of monopoly profit maximization models while applying the different schemes of price differentiation.

While formulating the model, we will assume a somewhat simplified assumption that on a relevant market there is only one "average" or "aggregated" consumer who experiences utility expressed in monetary units by a real utility function $u(x)$while purchasing x units of a product. A monopoly uses this situation in a way that it offers such a combination of price and supply(p^*, x^*), which maximizes monopoly's profit.

We have to realize, however, that a monopoly offers "all or nothing" choice, meaning that price p^* is valid only in case of purchasing exactly x^* units of a product with differentiated price and monopoly revenue will be

$$r^* = p^* x^*.$$

In other words, a consumer either purchases x^* units of a product at a price p^*, or has no possibility to purchase the product.

Let's now have a look at a cost function of a monopoly:

$$n(x) = nv(x) + n_F \tag{3}$$

while
$n(x): R \longrightarrow R$ – continuous and differentiable total cost function,
$nv(x): R \longrightarrow R$ - continuous and differentiable variable cost function,
n_F – fixed costs.

Assuming that total costs only account for monopoly's variable costs and we abstract from fixed costs, cost function of a monopoly is:

$$n(x) = nv(x)$$

Profit function of a monopoly as a difference between its revenue and costs would then be:

$$\pi(x) = px - n(x) \tag{4}$$

or

$$\pi(r, x) = r - n(x) \tag{5}$$

Waldman, D. E. – Jensen, E. J. [9] show, that a monopoly identifies an optimal, profit maximizing, combination of total revenues and supply (r^*, x^*) based on a solution of an optimization problem in a form:

$$\pi(r, x) = r - n(x) \rightarrow max \tag{6}$$

subject to

$$u(x) \geq r \tag{7}$$

Constraint (7) guarantees rationality of consumer's behavior, who is willing to spend r financial funds on x units of a product only in a case that his feeling of satisfaction from the purchase expressed in monetary units using utility function $u(x)$ will at least be as high as his expenses. Let us remind that r expresses not only monopoly revenue from selling its production in volume x for monopoly price p but has an important alternative intuitive interpretation. At the same time it represents consumer's willingness to spend r financial funds on purchase of supplied volume of products at a given market price.

Since a monopoly obviously expects its revenues at a level of customer's feeling of maximum satisfaction, the constraint (19) is actualized as equality and further we will analyze a monopoly profit maximization problem in a form

$$\pi(r,x) = r - n(x) \rightarrow max \tag{8}$$

subject to

$$u(x) = r \tag{9}$$

In Bazaraa, M. - C. M. Shetty, C.M. [1] we can see, that this optimization problem of mathematical programming represents maximization problem on bound extremum. Let us express the problem in a standard form – as a minimization problem, while we add the conditions of non-negativity of variables:

$$-\pi(r,x) = -r + n(x) \rightarrow min \tag{10}$$

subject to

$$u(x) = r \tag{11}$$
$$x, r \geq 0 \tag{12}$$

For optimization problem (10) ... (12 we can formulate generalized Lagrange function:

$$\mathcal{L}(r,x,\lambda) = -r + n(x) - \lambda(u(x) - r) \tag{13}$$

Kuhn – Tucker optimality conditions for Lagrange function (13) are:

$$\frac{\partial \mathcal{L}(r,x,\lambda)}{\partial x} \geq 0 \qquad \frac{\partial \mathcal{L}(r,x,\lambda)}{\partial r} \geq 0 \qquad \frac{\partial \mathcal{L}(r,x,\lambda)}{\partial \lambda} = 0$$

$$x\frac{\partial \mathcal{L}(r,x,\lambda)}{\partial x} = 0 \qquad r\frac{\partial \mathcal{L}(r,x,\lambda)}{\partial r} = 0 \tag{14}$$

$$x \geq 0 \qquad r \geq 0$$

Jarre, F. - Stoer, J [6] show, that after substituting analytical form of Lagrange function (14) we can restate Kuhn – Tucker optimality conditions of a consumer utility function maximization problem as follows:

$n'(x) - \lambda u'(x) \geq 0$ (14.1) $\quad -1 + \lambda \geq 0$ (14.4) $\quad u(x) - r = 0$ (14.7)

$x\big(n'(x) - \lambda u'(x)\big) = 0$ (14.2) $\quad r(-1+\lambda) = 0$ (14.5)

$x \geq 0$ (14.3) $\quad r \geq 0$ (14.6)

We can see that if a monopoly has an interest to enforce such an optimal combination of price and supply of its production, represented by revenue and supply vector$(r^\star, x^\star)$, which would maximize its profit $\pi(r,x) = r - n(x)$, there must exist a Lagrange multiplier λ^*, for which a Kuhn – Tucker optimality conditions (9) are met – meaning that a variables vector $(x^\star, r^\star, \lambda^*)$ is a solution to the system of equations and inequalities (14.1),..., (14.7).

Let us now have a closer look at Kuhn –Tucker optimality conditions (14) for a monopoly profit maximization problem in the conditions of an optimal combination of price-differentiated product supply and price. Assuming that a monopoly supplies positive optimal volume $x^\star > 0$ of product with a differentiated price, its optimum revenues will also be positive and $r^\star = px^* > 0$. From positive optimal revenue and validity of optimality condition (14.5) results an optimal value of Lagrange multiplier $\lambda^* = 1$. But if $\lambda^* = 1$, then for optimal positive value of monopoly production $x^* > 0$ and concurrent validity of optimality condition (14.2) following relations are valid

$$x^*\big(n'(x) - \lambda^* u'(x^*)\big) = 0 \ \wedge \ \ x^* > 1 \ \ \Rightarrow \ \ n'(x) - \lambda^* u'(x^*) = 0 \tag{15}$$
$$n'(x) - \lambda^* u'(x^*) = 0 \ \wedge \ \lambda^* = 1 \Longrightarrow$$

$$u'(x^*) = \ n'(x) \tag{16}$$

Relation (16) confirms a very important fact, namely that a monopoly supplies such an optimal production volume x^*, for which marginal utility equals marginal production costs.

Let us now analyze one specific situation of monopoly cost structure. Let's assume that fixed costs continue to be zero while variable costs are linear, so the cost function is $n(x) = cx$, resulting in marginal costs to be constant and equal unitary variable costs *c* of the monopoly production, while

$$\frac{dn(x)}{dx} = \frac{d(cx)}{dx} = c$$

Relation (16) can be then expressed as:

$$u'(x^*) = c \tag{17}$$

So the relation (16) in the end represents an inverse function of monopoly supply and supply function can be expressed in a form:

$$x = (u')^{-1}(c) = s(c) \tag{18}$$

So supply is a function of monopoly costs, which is of course logical and it confirms a connection between producer's technological level and his supply. Pepall, L. – Richards, D. - Norman , D. [7] show, that deciding an optimal combination of price and supply a monopoly chooses pare to-optimal solution – marginal willingness of consumers to purchase supplied volume of products corresponds with monopoly marginal costs. In other words a monopoly accepts a state that it can only better its market position if a consumer's situation gets worse. Nevertheless, a producer is in a situation when he takes and advantage of all the pareto-optimal volume of supply so he reaches maximum profit

$$max\ \ \pi(r^*, x^*) = r^* - n(x^*) \geq 0$$

while a consumer seemingly doesn't consume at all – his feeling of utility is fully "eliminated" by his expenses

$$u(x^*) - \ \ r^* = 0$$

Let us now examine a hypothetical situation when a monopoly supplies the same volume on this market which he could have supplied on a market of perfect competition. As a matter of fact, on the market of perfect competition a supplier produces exactly the amount that makes price and marginal costs as well as supply and demand consistent. The concurrence of these two conditions results in

$$p(x) = \ \ c \tag{19}$$

which exactly matches the condition (17), from which a relation (18) for supply inverse function was derived. Naturally, in this hypothetical case the benefits of sale in a competitive equilibrium are divided totally differently. A consumer reaches non-negative utility

$$u(x^*) - \ \ cx^* \geq 0$$

and on the other hand a company reaches no profit, as

$$\pi(r^*, x^*) = r^* - n(x^*) = p(x^*)x^* - cx^* = cx^* - cx^* \ = 0.$$

3 Conclusion

Based on the formalized analytical tools we showed that if a producer has enough market power to not only accept the market price but to be able to significantly influence and create it, he can quite effectively use his knowledge of consumer behavior to optimize a combination of supply and price of his product. As a matter of fact, it is a rational use of the information complex about the behavior of a consumer with specifically structured market basket, where they separately analyze consumer's utility regarding purchase of optimal volume of a price-differentiated product and this utility is represented in monetary units. Other goods in the market basket are being studied without any further specification of their volumes or range as one "aggregated good" and utility regarding purchase of these other goods is represented in monetary units as a simple sum of expenses spent on the purchase.

Regarding this fact, let us just remind that a marginal utility of a last unit purchase expressed in monetary units at the same time represents willingness of a consumer to pay for a product a price corresponding with this marginal utility. In other words, at the end it is not relevant how a monopoly discriminates the consumers. It doesn't matter

whether it applies an “all or nothing” rule or whether an individual units of production are sold at a differentiated prices corresponding to marginal willingness of a consumer to buy this unit of a product.

Acknowledgements

Supported by the grant No. 1/0697/15 “Optimization models and methods as tools of effective regulation in the modern theory of network industries organization” of the Grant Agency VEGA, Slovakia. (Scientific Grant Agency of the Ministry of Education, science, research and sport of the Slovak Republic and the Slovak Academy of Sciences)

References

[1] Bazaraa, M. - C. M. Shetty, C.M.: Nonlinear Programming: Theory and Algorithms. Wiley-Interscience. New York ,2006
[2] Blum, U. - Müller, S. - Weiske von Gabler, A.: Angewandte Industrieökonomik. Theorien - Modelle – Anwendungen. Berlin: Verlag Gabler, 2006..
[3] Fendekova, E. - Fendek, M.: Podmienky optimálnosti Kuhna-Tuckera v modeloch rovnováhy trhu sieťových odvetví. Kuhn – Tucker optimality conditions in equilibrium models of network industries market In Politická ekonomie : teorie, modelování, aplikace. - Praha : Vysoká škola ekonomická, 2012. ISSN 0032-3233, 2012, roč. 60, č. 6, s. 801-821..
[4] Fendekova, E. - Fendek, M.: Modely cenovej regulácie sieťových odvetví. Models of Price Regulation of Network Industries, In Ekonomický časopis: journal for economic theory, economic policy, social and economic forecasting. - Bratislava: Ekonomický ústav SAV : Prognostický ústav SAV, 2010. ISSN 0013-3035, 2010, roč. 58, č. 10, s. 1039-1054
[5] Fendekova, E. - Fendek, M.: Models of regulation of network industries in Slovakia. In International journal of economics and business research. - UK : Inderscience Enterprises Ltd, 2009. ISSN 1756-9850, 2009, vol. 1, no. 4, s. 479-495.
[6] Fendek, M, - Fendekova, E.: Optimization Model of a Monopoly Production Price Differentiation. The 26th European Conference on Operational Research, Rome, Italy, 2013.
[7] Jarre, F. - Stoer, J.: Optimierung. Berlin: Springer Verlag. ISBN3-540-43575-1, 2004.
[8] Pepall, L. – Richards, D. - Norman , D.: Industrial Organization: Contemporary Theory and Empirical Applications.New York: Wiley-Blackwell, 2008.
[9] Thomemann, U.: Operations Management. Pearson Studium, ISBN-10: 3827371201, 2005.
[10] Waldman, D. E. – Jensen, E. J.: Industrial Organization: Theory and Practice. New York: Addison Wesley 2006.

Testing Gibrat's law for small and medium-sized manufacturing firms: empirical evidence for the Czech Republic

Roman Fiala[1]

Abstract. The paper focuses on the investigation of the relationship between firm size and firm growth for the small and medium-sized manufacturing enterprises (SMEs) from the Czech Republic during 2007–2012. The study has two aims. Firstly to investigate the validity of Gibrat's law over the entire period 2007 to 2012 and secondly to examine if there is any convergence toward the Gibrat's law through time. The validity of Gibrat's law was tested with the help of linear regression model with first-order autoregressive process. Over the entire period 2007–2012, it was found statistically significant positive relationship between firm size and firm growth, so Gibrat's law was rejected. However there is a convergence (with the exception of the period 2010–2011) toward the validity of Gibrat's law through time. While for the period 2007–2008 there is statistically significant relationship between firm size and firm growth and Gibrat's law is rejected, if was found no statistically significant relationship between firm size and firm growth for period 2011–2012 and Gibrat's law for this period holds.

Keywords: Gibrat's law, firm size, firm growth, small and medium-sized enterprises.

JEL Classification: L11, L26
AMS Classification: 62M10

1 Introduction

Many scholars have focused on the relationship between firm growth and firm size. These researchers have been influenced by Robert Gibrat [15]. Gibrat's [15] law states that firm growth is a random walk, independent of company size. Gibrat's law is known the Law of Proportionate Effect (LPE) as well.

The findings of the studies, which are focused on the Gibrat's law, are not unified. A lot of studies have rejected the Law of Proportionate Effect, for instance AmirKhalkhali and Mukhopadhyay [2], Almus and Nerlinger [1], Oliveira and Fortunato [23], Dunne and Hughes [10]; Feizpour, Mahmoudi and Soltani [14], Goddard, Wilson and Blandon [16], Evans [12] or Evans [13]. Other studies tended to confirm the Gibrat's law (for example Del Monte and Papagni [9], Simon and Bonini [24] or Hymer and Pashigian [19]). Some authors rejected Gibrat's law only in some industries and in others confirm the validity (for example Chen and Lu [5] or Aslan [3]). Studies differ in many aspects: (1) data (length of time series, sample of firms, country, industry), (2) chosen empirical model and (3) firm size measurement.

Some authors (for instance Lottti, Santarelli and Vivarelli [21] or Tang [26]) deal with the difference between the verification of Gibrat's law in the short run and long run. Both studies describe two models of passive and active learning, which suggest, that Gibrat's law tends to reject in the short run and tends to hold in the long run. First argument for this hypothesis is a Bayesian model of noisy selection – according this model of passive learning, efficient companies grow and survive and inefficient enterprises tends to decline and fail (Jovanovic [20], Lotti, Santarelli and Vivarelli [21]). Lotti, Santarelli and Vivarelli [21, p. 33] simply describe this model: „*…firms are initially endowed with unknown, time-invariant characteristics, i.e., ex-ante efficiency parameters, while ex-post the prior distribution is updated as evidence comes in which leads some firms to discover that they are more efficient than others. Thus, each firm has to decide on its strategy: whether to exit, to continue at the same size, to expand, or to reduce its productive capacity.*" After the noisy selection process has been completed, the law of proportionate effect tends to confirm (Tang [26]).

The active learning model was designed by Ericson and Pakes [11]. Teruel-Carrizosa [27, p. 360] aptly and shortly describes this model: "*...firms could modify their own level of efficiency by increasing their investments. However, these firms have to consider investment by other firms and external shocks. This means that, while a firm makes a great effort to invest, it should also take into account investment by its competitors.*"

[1] College of Polytechnics Jihlava, Department of economic studies, Tolstého 16, 586 01 Jihlava, roman.fiala@vspj.cz.

This paper has two aims. Firstly to verify the validity of Gibrat's law over the entire period 2007 to 2012 and secondly to investigate if there is any convergence toward the Gibrat's law through time. Both goals will be verified for the small and medium-sized manufacturing enterprises (SMEs) from the Czech Republic with the help of database Albertina CZ Gold Edition.

2 Data and methodology

The data for this study come from the database Albertina CZ Gold Edition. It was used data about the small and medium-sized manufacturing enterprises (SMEs) from the Czech Republic for the period 2007 to 2012. For the measurement of firm size are used revenues. Manufacturing sector is section C according to CZ-NACE classification.

For definition of small and medium-sized manufacturing enterprises was used Commission Recommendation 2003/361/EC of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. The small and medium-sized enterprises employ fewer than 250 employees and have an annual turnover fewer than 50 million EUR, and/or annual balance sheet total fewer than 43 million EUR [6]. Firms were divided on the basis of data from the year 2007.

According to Daunfeldt and Elert [8] we use only data such companies, which were in 2007 at least 5 years in the industry and survived throughout the entire period 2007 to 2012. In case of inclusion of all firms, the results could be biased because smaller companies have a higher expected probability of exit than their larger counterparts. There is not included new entrants during 2007 to 2012 in the data, because these new firms should have specific development. The descriptive statistics are shown in table 1.

year	N	Mean	Std. Dev.
2007	6343	10.171	1.735
2008	6343	10.100	1.749
2009	6343	9.959	1.758
2010	6343	9.873	1.815
2011	6343	9.894	1.865
2012	6343	9.814	1.930

Table 1 Natural Logs of revenues in thousand CZK

To verify the validity of Gibrat's law we use the approach of Daunfeldt and Elert [8]. They estimate the validity of Gibrat's law using this model

$$lnS_{jt}^{i} = \alpha_{j0} + \alpha_{j1}.\ln S_{j(t-1)}^{i} + \theta_{jt}.T_t + u_{jt.} \quad (1)$$

Where $S_{jt}^{i}$is the size of i-th firm of j-th industry in time t, $\theta_{jt}.T_t$is a vector of time specific fixed effects. To estimate the Gibrat's law validity, we modify the original model (equation 1) and use this form

$$lnS_{it} = \alpha_0 + \alpha_1.\ln S_{i(t-1)} + \alpha_2.NACE_i + \alpha_{3k}.T_t.NACE_i + u_{t.} \quad (2)$$

Where $S_{it}$is the size of i-th firm in time t, $NACE_i$is the dummy variable for industry using 5-digit NACE classification of i-th firm, $\alpha_2.NACE_i$ is the vector of industry specific fixed effects, $\alpha_{3k}.T_t.NACE_i$ is a vector of time and industry specific fixed effects. The values of parameter $\alpha_1$indicate if the the Gibrat's law is valid or not. The Gibrat's law holds if $\widehat{\alpha_1}$equal to one. The value smaller than one implies, that small firm grow faster than large and the value higher than one, that large firm grow faster than small.

We use the revenue as an indicator of firm size. The revenue represents the real revenues which are calculated using consumer price index published by the Czech Statistical Office [7]. The indicator "revenue" includes revenues from sales of goods and services. Like Daunfeldt and Elert [8], to estimate the parameters of model we use OLS estimator. Because of heteroskedasticity and serial correlation problem, we use OLS estimator with cluster-robust standard errors. To confirm or reject Gibrat's law, we test null hypothesis $H0: (\widehat{\alpha_1}) = 1$ versus $H1: (\widehat{\alpha_1}) \neq 1$ using F-test.

3 Results and discussion

We estimate validity of Gibrat's law using linear regression model with first-order autoregressive process. Firstly, it was investigated Gibrat's law for the entire period 2007–2012. It was used two versions of equation (2). Model (1) contains only the time specific fixed effect, which captures time-variant heterogeneity in growth rates. Model

(2) includes also industry specific fixed effect and industry and time specific fixed effect capturing industry-variant heterogeneity in growth rates.

As you can see in the table 2, over the entire period 2007–2012, it was found statistically positive relationship between firm size and firm growth, so Gibrat's law was rejected for both models (1) and (2). For this whole period 2007 to 2012 smaller manufacturing enterprises grew faster than their larger counterparts in the Czech Republic.

These results are consistent for example with Almus and Nerlinger [1], Oliveira and Fortunato [23] or Goddard, Wilson and Blandon [16]. Almus and Nerlinger [1] focused on the situation in West Germany in the period 1990 to 1996 and found, that smaller companies have larger growth potential than larger firms. The same findings found in their study Oliveira and Fortunato [23], which investigated data of Portuguese manufacturing industry for the period 1990 to 2001. Goddard, Wilson and Blandon [16] verified Gibrat´s law in manufacturing sector in Japan in the period from 1981 to 1996. According this study, Gibrat´s law should be rejected.

	Model (1)	**Model (2)**
$\ln.S_{t-1}$ (α_1)	0.990501***	0.984754***
	(0.0022663)	(0.0027642)
T_t fixed effects	Yes	-
$NACE_j$ fixed effects	-	Yes
$T_t.NACE_j$ fixed effects	-	Yes
Constant	0.0254399	1.301418
	(0.024631)	(1.133732)
R^2	0.9391	0.9429
N	31715	31715
F-test[a]	17.57	30.42
p-value	0.0000	0.0000

Table 2 Estimation of Gibrat´s law validity - period 2007-2012

Notes: ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level, robust standard errors in brackets, a. F- test of H0: $\alpha_1=1$.

According to Lotti, Santarelli and Vivarelli [21], Gibrat's law was investigated year-by-year (five separate estimates) using equation (2) (see table 3) with the aim to reveal, if there is convergence toward Gibrat's law through time (ex post). The results presented in table 3 indicate that consistently with Lotti, Santarelli and Vivarelli [21], convergence toward Gibrat-like behavior through time is found ex post. While in the first two periods (2007-2008 and 2008-2009) is Gibrat's law rejected at the 1 percent level (p-value of F-test is 0.0000), for the following periods (2009-2010, 2010-2011, 2011-2012) is Gibrat's law confirmed at the 1 percent level. In the period 2010-2011 is Gibrat's law rejected at the 5 percent level.

These findings are consistent with the studies of Tang [26] and Lotti, Santarellia and Vivarelli [21], in which was Gibrat's law rejected for the entire period, but convergence toward this Law occurs through time. Reason for this finding should be probably models of active and passive learning mentioned above.

Years	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012
ln.St-1 (α1)	0.9673502	0.9689738	0.9901258	0.9869681	1.006721
	(0.0071224)	(0.0067236)	(0.0067608)	(0.0065164)	(0.0056736)
NACE	Yes	Yes	Yes	Yes	Yes
Constant	1.431736	0.9804877	-0.4345214	-0.464511	0.114717
	(1.100845)	(0.5752426)	(0.3717485)	(0.3672849)	(0.1139327)
R2	0.9455	0.9361	0.9361	0.9432	0.9432
N	6343	6343	6343	6343	6343
F-testa	21.01	21.29	2.13	4.00	1.40
p-value	0.0000	0.0000	0.1442	0.0456	0.2362

Notes: ***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level, robust standard errors in brackets, a. F- test of H0: $\alpha_1=1$.

Table 3 Estimation of Gibrat´s law validity (model 2) – year-by-year estimation

4 Conclusions and suggestions for further research

Overall, this study analysed the relationship between firm growth and firm size for the small and medium-sized manufacturing enterprises from manufacturing sector for the period 2007-2012.

This paper had two goals - to investigate the validity of Gibrat‘s law over the period 2007 to 2012 and to verify if there is any convergence toward the Gibrat’s law through time.

Over the entire period 2007–2012, Gibrat’s law was rejected, smaller companies grew in this period faster than larger firms. However, year-by-year estimation revealed a convergence toward the validity of Gibrat’s law, which is the main finding of this paper.

One of the theme of the future research can be gender structure of top level management and impact of this factor on the Gibrat’s law. For instance, according to Slabá [25], management belongs to the most important stakeholders. Women’s behaviour can be specific (in comparison to men’s behaviour) and this fact can influence the validity of the Gibrat’s law.

The results of Watson [28] indicate that female-controlled SMEs have relatively lower levels of external funding in comparison with male counterparts. The cause of this fact could be risk averse behavior of women rather than bank discrimination. Hedija [17] and Hedija [18] concludes that female managers in comparison to their male counterparts have a lower tendency to apply wage discrimination against women. This fact can be motivated for women and positively influenced their performance. Empirical studies show that firms owned by women are mostly smaller, often have lower profits and revenues (for example Aspray and Cohoon [4].

Acknowledgements

Supported by the internal grant of the College of Polytechnics Jihlava „The relationship among firm size, firm growth and gender characteristics“.

References

[1] Almus, M., and Nerlinger, E. A.: Testing “Gibrat’s Law” for Young Firms – Empirical Results for West Germany. *Small Business Economics* **15** (2000): 1–12.

[2] AmirKhalkhali, S., and Mukhopadhyay, A. K.: On the relationship between firm size and firm growth: some Canadian results. In: *Proceedings of the Pennsylvania economic association 2008 conference. Slippery Rock University of Pennsylvania*. (Baumgardner, G., Baier, M. B., eds.). Slippery Rock, 2005, 24–30.

[3] Aslan, A.: Testing Gibrat’s law: empirical evidence from panel unit root tests of Turkish firms. *MPRA: Munich Personal RePEc Archive*. 2008. [Online]. Available at: http://mpra.ub.uni-muenchen.de/10594/1/mpra_paper_10594.pdf. [Accessed: 2015, March 18].

[4] Aspray, W., and Cohoon, J. M.: *Gender Differences in Firm Size, Growth, and Persistence: A Review of Research Literature on Women’s Entrepreneurship in the Information Technology Field*. (2007). [Online]. Available at: https://www.ncwit.org/sites/default/files/resources/1_entrepreneurialgrowth.pdf. [Accessed: 2015, April 21]

[5] Chen, J. R., and Lu, W. C.: Panel unit root tests of firm size and its growth. *Applied Economics Letters* **10** (2003): 343–345.

[6] Commission Recommendation 2003/361/EC of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. [Online]. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:124:0036:0041:EN:PDF. [Accessed: 2015, April 17].
[7] Czech statistical office. *Indexy spotřebitelských cen - inflace - časové řady*. [Online]. Available at: http://www.czso.cz/csu/redakce.nsf/i/isc_cr. [Accessed: 2015, January 12].
[8] Daunfeldt, S-O., and Elert, N. 2013. When is Gibrat's law a law? *Small Business Economics* **41** (2013): 133–147.
[9] Del Monte, A., and Papagni, E.: R&D and the growth of firms: empirical analysis of a panel of Italian firms. *Research Policy* **32** (2003): 1003–1014.
[10] Dunne, P., and Hughes, A.: Age, size, growth and survival: UK companies in the lates 1980s. *The Journal of Industrial Economics* **42** (1994): 115–140.
[11] Ericson, R., and Pakes, A.: Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies* **62** (1995): 53–82.
[12] Evans, D. S.: Tests of alternative theories of firm growth. *Journal of Political Economy* **95** (1987): 657-674.
[13] Evans, D. S.: The Relationship Between Firm Growth, Size and Age: Estimates for 100 Manufacuring Industries. *The Journal of Industrial Economics* **35** (1987): 567–581.
[14] Feizpour, M. A., Mahmoudi, V. and Soltani, E.: The Validity of Gibrat´s Law: Evidence From Manufacturing Industry In Iran: 1995-1998. *The International Business & Economics Research Journal* **9** (2010): 33–36.
[15] Gibrat, R. *Les inégalités économiques*. Librairie du Receuil Sirey, Paris, 1931.
[16] Goddard, J., Wilson, J., and Blandon, P.: Panel tests of Gibrat's Law for Japanese manufacturing. *International Journal of Industrial Organization* **20** (2002): 415–433.
[17] Hedija, V.: The Effect of Female Managers on Gender Wage Differences. *Prague Economic Papers* **24** (2015), 38–59.
[18] Hedija, V.: Gender Pay Gap in Different Sectors of Czech Economy. In: *32nd International Conference Mathematical Methods in Economics 2014* (Talašová, J., Stoklasa, J., and Talášek, T., eds.). Palacký University, Olomouc, 2014, 275–280.
[19] Hymer, S., and Pashigian, P.: Firm size and rate of growth. *Journal of Political Economy*, **70** (1962): 556–569.
[20] Jovanovic, B.: Selection and evolution of industry. *Econometrica* **50**, 649–670.
[21] Lotti, F., Santarelli, E., and Vivarelli, M.: Defending Gibrat's Law as a long-run regularity. *Small Business Economics*, **32** (2009): 31–44.
[22] Nassar, I. A., Almsafir, M. K., and Al-Mahrouq, M. H.: The Validity of Gibrat's Law in Developed and Developing Countries (2008-2013): Comparison Based Assessment. *Procedia - Social and Behavioral Sciences* **129** (2014): 266–273.
[23] Oliveira, B., and Fortunato, A. Testing Gibrat´s Law: Empirical Evidence from a Panel of Portuguese Manufacturing Firms. *International Journal of the Economics of Business* **13** (2006): 65-81.
[24] Simon, H. A., and Bonini, C. P.: The size distribution of business firms. *American Economic Review* **48** (1958): 607–617.
[25] Slabá, M.: Stakeholder analysis in the bank sector. In: *9th International Scientific Conference Financial Management of Firms and Financial Institutions*. VŠB-TU Ostrava, Ostrava, 2013, 817–826.
[26] Tang, A.: Does Gibrat's law hold for Swedish energy firms? [Online]. Available at: http://www.oru.se/PageFiles/13614/Version%20April%20%202013.pdf. [Accessed: 2015, March 20].
[27] Teruel-Carrizosa, M.: Gibrat's law and the learning process. *Small Business Economics*, **34** (2010): 355–373.
[28] Watson, J.: External Funding and Firm Growth: Comparing Female- and Male-Controlled SMEs. *Venture Capital* **8** (2006), 33–49.

Biform games in supply chains

Petr Fiala[1]

Abstract. Supply chain is defined as a system of suppliers, manufacturers, distributors, retailers and customers where material, financial and information flows connect participants in both directions. The ongoing actions in the supply chain are a mix of cooperative and non-cooperative behavior of the participants. The paper proposes to use biform games for the analysis of supply chains. A biform game is a combination of non-cooperative and cooperative games, introduced by Brandenburger and Stuart (2007). It is a two-stage game: in the first stage, players choose their strategies in a non-cooperative way, thus forming the second stage of the game, in which the players cooperate. The biform game approach can be used for modeling general buyer-supplier relationships in supply chains. First, suppliers make initial proposals and take decisions. This stage is analyzed using a non-cooperative game theory approach. Then, suppliers negotiate with buyers. In this stage, a cooperative game theory is applied to characterize the outcome of negotiation among the players over how to distribute the total surplus. Each supplier's share of the total surplus is the product of its added value and its relative negotiation power. A specific model of this type for analyzing supply chains is presented.

Keywords: Supply chain, game theory, biform games, non-cooperation, cooperation.

JEL Classification: C70
AMS Classification: 91A

1 Introduction

Supply chain is defined as a decentralized system with layers of suppliers, manufacturers, distributors, retailers and customers where material, financial and information flows connect participants in both directions. A supply chain is a complex and dynamic supply and demand network of agents, activities, resources, technology and information involved in moving a product or service from supplier to customer. Most supply chains are composed of independent units with individual preferences. Each unit will attempt to optimize his own preference. Behavior that is locally efficient can be inefficient from a global point of view. Supply chain management has generated a substantial amount of interest both by managers and by researchers. There are numerous opportunities to create hybrid models that combine competitive and cooperative behavior.

There are many concepts and strategies applied in designing and managing supply chains (see Simchi-Levi et al., 1999). The expanding importance of supply chain integration presents a challenge to research to focus more attention on supply chain modeling (see Tayur et al., 1999). In supply chain behavior is much inefficiency. The so-called bullwhip effect (see Lee, 1997, Tayur et al., 1999), describing growing variation upstream in a supply chain, is probably the most famous demonstration of system dynamics in supply chains. There are some known causes (see Lee 1997, Tayur et al 1999), of the bullwhip effect: information asymmetry, demand forecasting, lead-times, batch ordering, supply shortages and price variations. Information sharing of customer demand has an impact on the bullwhip effect and other inefficiencies in supply chains (Fiala 2005). Researchers in supply chain management now use tools from game theory to help managers to make strategic operational decisions in complex multi-agent supply chain systems.

The evolution of supply chain management recognized that a business process consists of several decentralized firms and that operational decisions of these different entities impact each other's profit, and thus the profit of the whole supply chain. To effectively model and analyze decision making in such multi-agent situation where the outcome depends on the choice made by every agent, game theory is a natural choice. Game theory has become a useful instrument in the analysis of supply chains with multiple agents, often with conflicting objectives. The paper analyzes allocation decisions in supply chains. Equilibrium search in supply chains is a very important problem. Allocation games are used for behavior modeling of supply chains and focus on allocation of resources, capacities, costs, revenues and profits. A profit allocation two-stage procedure for equilibrium in supply chains is proposed, based on combination of non-cooperative and cooperative game approaches.

The rest of the paper is organized as follows. Section 2 summarizes the basics of the game theory applicable in the allocation of profit in supply chains. In Section 3, the problem formulation and an outline of the procedure

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, pfiala@vse.cz

are provided. Non-cooperative part of the problem is analyzed in the Section 4. A cooperative approach for profit allocation is presented in Section 5. Section 6 presents conclusions.

2 Game theory background

This section summarizes some of the basic non-cooperative and cooperative concepts of the game theory that are applied in the proposed approach for profit allocation in supply chains. John von Neumann and Oskar Morgenstern (1944) is the classic work upon which modern game theory is based. Since then, an extensive literature on game theory was published. For example, Myerson's book (1997) provides a clear and thorough examination of the models, solution concepts, results, and methodological principles of non-cooperative and cooperative game theory. Game theory models situations where players make decisions to maximize their own utility, while taking into account that other players are doing the same, and that decisions made by players, impact others utilities. There is a broad division of game theory into two approaches: the cooperative and the non-cooperative approach. These approaches, though different in their theoretical content and the methodology used in their analysis, are really just two different ways of looking at the same problem.

The non-cooperative theory of games is strategy oriented; it studies what one may expect the players to do. The non-cooperative theory is a "micro" approach in that it focuses on precise descriptions of what happens. The field of supply chain management has seen, in recent years, a wide variety of research papers that employ game theory to model interaction between players. Cachon and Netessine (2004) provide an excellent survey and state of art especially non-cooperative game techniques. The concept of using non-cooperative agents to formulate allocation mechanisms in a game theoretical setting is closer to the classical market concept than solutions employing cooperative strategies. Most non-cooperative allocation strategies in distributed systems consist of following steps:

- The formulation of utility functions for the system participants.
- The formulation of best response strategies.
- The existence of Nash equilibrium is proved in the system of multiple agents.
- Efficiency is measured compared to achievable welfare.

An *n*-player non-cooperative game in the normal form is a collection

$$\{N = \{1,2,\dots,n\};\ X_1, X_2, \dots, X_n;\ \pi_1(x_1, x_2, \dots, x_n), \pi_2(x_1, x_2, \dots, x_n), \dots, \pi_n(x_1, x_2, \dots, x_n)\ \}, \quad (1)$$

where N is a set of n players; X_i, $i = 1, 2, \dots, n$, is a set of strategies for player i; $\pi_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, n$, is a pay-off function for player i, defined on a Cartesian product of n sets X_i, $i = 1, 2, \dots, n$.

Decisions of other players than player i are summarized by a vector

$$\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, x_{i+1} \dots, x_n). \quad (2)$$

A vector of decisions $(x_1^0, x_2^0, \dots, x_n^0)$ is a Nash equilibrium of the game if

$$x_i^0(\mathbf{x}_{-i}^0) = \text{argmax}_{x_i}\, \pi_i(x_i, \mathbf{x}_{-i}) \forall i = 1,2, \dots, n. \quad (3)$$

A Nash equilibrium is a set of decisions from which no player can improve the value of his pay-off function by unilaterally deviating from it.

Stackelberg games are strategic games with 2 players. They are also called leader-follower games. The leader plays first, anticipating the decision of the follower, and the follower has no other choice than to act optimally as anticipated by the leader. Such games generally reach a compromise situation, called the Stackelberg equilibrium.

The leader's optimal decision, denoted x_1^0, is computed recursively from the knowledge of the follower's optimal response function $x_2^0(x_1)$:

$$x_1^0 = \text{argmax}_{x_1}\, \pi_1(x_1, x_2^0(x_1)) \text{ and } x_2^0 = x_2^0(x_1^0)\,. \quad (4)$$

When the demand is stochastic than the newsvendor model can be applied. The newsvendor model is not complex, but it is sufficiently rich to study important questions in supply chain coordination. In a standard newsvendor problem the price is assumed to be fixed but the problem is to analyze contracts for supply chain coordination with price-dependent stochastic demand.

Cooperative game theory looks at the set of possible outcomes, studies what the players can achieve, what coalitions will form, how the coalitions that do form divide the outcome, and whether the outcomes are stable and robust. Nagarajan and Sošić (2008) review the existing literature on applications of cooperative games to

supply chain management. They also deal with certain methodological issues when modeling supply chain problems. The paper focuses on applications in supply chains with two central questions of cooperative games:

- What are feasible outcomes and how the players in a coalition allocate the outcomes?
- What are stable coalitions?

Allocation mechanisms are based on different approaches such as negotiations, auctions, Shapley values, etc. When modeling cooperative games is advantageous to switch from the game in normal form to the game in the characteristic function form. The characteristic function of the game with a set of n players N is such function $v(S)$ that is defined for all subsets $S \subseteq N$ (i.e. for all coalition) and assigns a value $v(S)$ with following characteristics:

$$v(\varnothing) = 0\,, \tag{5}$$

$$v(S_1 \cup S_2) \geq v(S_1) + v(S_2)\,, \tag{6}$$

where S_1, S_2 are disjoint subsets of the set N. The pair (N, v) is called a cooperative game of n players in the characteristic function form.

A particular allocation policy, introduced by Shapley (1953) has been shown to possess the best properties in terms of balance and fairness. So called Shapley vector is defined as

$$\mathbf{h} = (h_1, h_2, ..., h_n), \tag{7}$$

where the individual components (Shapley values) indicate the mean marginal contribution of i-th player to all coalitions, which may be a member. Player contribution to the coalition S is calculated by the formula:

$$v(S) - v(S - \{i\}). \tag{8}$$

A complicating factor is that with the increasing number of n players is rapidly increasing number of coalitions and complexity of their production. Shapley value for the i-th player is calculated as a weighted sum of marginal contributions according to the formula:

$$h_i = \sum_S \left\{ \frac{(|S|-1)!\,(n-|S|)!}{n!} . \left[v(S) - v(S - \{i\}) \right] \right\}, \tag{9}$$

where the number of coalition members is marked by symbol $|S|$ and the summation runs over all coalition $i \in S$.

A biform game is a combination of non-cooperative and cooperative games, introduced by Brandenburger and Stuart (2007). It is a two-stage game: in the first stage, players choose their strategies in a non-cooperative way, thus forming the second stage of the game, in which the players cooperate. The biform game approach can be used for modeling general buyer-supplier relationships in supply chains. First, suppliers make initial proposals and take decisions. This stage is analyzed using a non-cooperative game theory approach. Then, suppliers negotiate with buyers. In this stage, a cooperative game theory is applied to characterize the outcome of negotiation among the players over how to distribute the total surplus. Each supplier's share of the total surplus is the product of its added value and its relative negotiation power.

3 Problem and solving formulation

The problem is formulated as a supply chain with layers of suppliers, producers, retailers and customers. Suppliers form a layer with m agents and provide m types of resources to producers. The layer of producers is represented by n agents. These agents produce one type of product. The production is characterized by consumption of m resources to produce one unit of the final product. Each production agent is characterized by its available production resources. The resource capacity constraints compare the total availability of resources in the production layer with total consumption of resources to produce total number of q units of products. Producers send the products to retailers. Retailers meet price-dependent stochastic demand of customers. This problem is solved by two-stage procedure based on combination of no-cooperative and cooperative game approaches.

The first stage solves problems by price-dependent stochastic demand of customers:

- How to get maximal profit from customers.
- How to allocate the maximal profit between retailers and producers.

The problems are solved by non-cooperative manner. A Stackelberg game is formulated between the layer of producers and the layer of retailers as a newsvendor problem with pricing. Retailers seek to maximize total profit from the sale and try to align goals with producers on a contract basis and share the total profit with them. The

maximization of the profit is by the resource capacity constraints. The equilibrium point (p^0, q^0) is given by values of total number of q production units and optimal price p.

A large number of papers have been published that proposed analyze mechanisms for supply chain coordination. Mechanisms based on non-cooperative game theory usually propose establishment of coordinating contracts. A retailer can usually collect demand information easier than a producer and he has a better motivation for optimally determining sales quantities and prices. There are many types of contracts. The basic type is a wholesale price contract. With a wholesale price contract (Larievier, 1999) the supplier charges the retailer w per unit purchased. The producer knows exactly what retailer will order at every wholesale price and bears no responsibility for the product. All uncertainty regarding the producer profit is foisted onto the retailer. The wholesale price contract coordinates the chain only if the producer earns a non-positive profit. So the producer clearly prefers a higher wholesale price. As a result, the wholesale price contract is generally not considered a coordinating contract. The richer contracts differ from wholesale price contracts by allowing the producer to assume some of the risk arising from stochastic demand. With a buy back contract (Pasternack, 1985) the producer charges the retailer w per unit purchased, but pays the retailer b per unit remaining at the end of the season. The retailer should not profit from left over inventory, so assume $b \leq w$. There is assumed that a returns policy on the decentralized chain introduces no additional cost beyond that incurred by the centralized system.

A specific buyback contract is used for coordination. The layer of producers as leader proposes the wholesale price w and the buyback price b. The layer of retailers as follower accepts the prices to coordinate the system. The allocation of the total profit between retailers and producers is given by splitting parameter λ ($0 \leq \lambda \leq 1$). The value of the parameter λ is negotiated by retailers and producers.

In the second stage, producers address the following issues:

- How the determine the optimal coalition structure.
- How to allocate the profit among the members of the optimal coalition.

The problems are solved by cooperative manner. These agents compete to be members of a coalition and are willing to cooperate to produce products and sell them to customers through retailers. The optimal coalitions are determined according to the maximal profit with respect the resource capacity constraints for the coalition.

The maximal profit is allocated among the members of the coalitions by Shapley values. Shapley value has been shown to possess the best properties in terms of balance and fairness.

4 First stage: non-cooperative problem

We consider a supply chain in one-period setting in which the layer of producers sells a product to the layer of retailers facing stochastic demand from consumers. We assume that stochastic demand u has a continuous distribution $F(u)$ with density $f(u)$. The demand distribution and cost information are common knowledge. Define the failure rate function of the u distribution as

$$g(u) = \frac{f(u)}{1 - F(u)} \tag{10}$$

and the generalized failure rate function as

$$h(u) = ug(u). \tag{11}$$

Assume the demand distribution has strictly increasing generalized failure rate property (IGFR). Many distributions have the IGFR property, including the uniform, the normal, the exponential, the gamma, and the Weibull. We define the following quantities: q retailer's total order quantity; c producer's unit production cost; p retail price. The setting can be characterized as a newsvendor problem.

Centralized solution

Centralized solution is a benchmark for the decentralized supply chain. The centralized chain is considered as an integrated firm that controls production and sales to customers. The profit of an integrated firm for stocking level q is

$$\pi(q) = (p - c)q - p\int_0^q F(\mathrm{u})du\,. \tag{12}$$

The problem is concave in q and the optimal solution is given by

$$q^0 = F^{-1}\left(\frac{p - c}{p}\right). \tag{13}$$

The maximum system profit $\pi(q^0)$ is completely determined by the production level q^0.

Decentralized solution can be improved by contracting. The contract coordinates the chain if it induces the choice of the centralized system's optimal stocking level q^0. The approach is based on a specific buy back contract for the price-dependent stochastic demand.

Price-dependent stochastic demand

Chen and Cheng (2012) presented price-dependent revenue sharing mechanism. Little work has been done on the combined problem of supply chain coordination with price-dependent stochastic demand (Yao et al., 2006). The contracts proposed for coordination with price-independent stochastic demand are not applicable for coordination of supply chains with price-dependent stochastic demand.

We will analyze the multiplicative form of price-dependent stochastic demand

$$D(p, u) = y(p)u\,, \tag{14}$$

a function of p and u, where u is a random variable independent of p and $y(p)$ is continuous, nonnegative, twice differentiable function. The expectation of D is specified by a function $y(p)$ for any given price p:

$$E[D(p, u)] = y(p). \tag{15}$$

The expected profit for centralized solution for any output level q and price p is:

$$\pi(p, q) = E\{p[\min(q, D(p, u)] - cq\} = E\{(p - c)q - p\max(0; q - D(p, u))\} = (p - c)q - py(p)\int_0^{\frac{q}{y(p)}} F(u)du \tag{16}$$

The objective is to choose (p^0, q^0) to maximize the expected profit $\pi(p, q)$.

By fixing price p the problem reduces to standard newsvendor problem without pricing and the optimal level of production

$$q^0 = y(p)F^{-1}\left(\frac{p-c}{p}\right). \tag{17}$$

By substituting it into the expected profit

$$\pi(p) = y(p)[(p - c)F^{-1}\left(\frac{p-c}{p}\right) - p\int_0^{F^{-1}(\frac{p-c}{p})} F(u)du]. \tag{18}$$

The problem is now with only one decision variable p and the optimal price p^0 can be obtained by solving

$$\frac{d\pi(p)}{dp} = 0\,. \tag{19}$$

The assumptions of the existence and uniqueness of the optimal solution (p^0, q^0) are concavity of deterministic part of demand function $y(p)$ and IGFR property of stochastic part of demand function u.

The proposed contract for coordination of the decentralized supply chain is a specific buy-buck contract. The wholesale price w and the buy-buck price b are specified:

$$w = \lambda\,(p - c) + c, \tag{20}$$

$$b = \lambda\, p, \tag{21}$$

$$\text{where } 0 \le \lambda \le 1. \tag{22}$$

By the setting of the prices w and b the retailer's profit and the supplier's profit for any chosen output level q and price p are

$$\pi_R = E\,\{p[\min(q, D(p, u)] - wq + b\max(0; q - D(p, u))\} = E\,\{(p - w - c)q - (p - b)\max(0; q - D(p, u))\} =$$

$$= (1 - \lambda)\,E\{(p - c)q - p\max(0; q - D(p, u))\} = (1 - \lambda)\,\pi\,, \tag{23}$$

$$\pi_P = E\{(w - c)q - b\max(0; q - D(p, u))\} = E\{\lambda\,(p - c)q - \lambda\, p\max(0; q - D(p, u))\} = \lambda\,\pi. \tag{24}$$

From previous expressions of the retailer's profit and the producer's profit, it is clear that the retailer and the producer solve the same problem as the centralized supply chain and the sum of the retailer's profit and the supplier's profit is equal to the profit of the centralized supply chain. The parameter λ characterizes a splitting of the total profit between the retailer and the supplier.

5 Second stage: cooperative problem

In the considered problem the layer of producers is represented by n agents. The set of production agents is denoted $N = \{1, 2, \ldots, n\}$. These agents compete to be members of a coalition $S \subseteq N$ and are willing to cooperate to produce products and sell them to customers through retailers. A coalition S is defined as a subset of the set N of n producers with characteristic vector $\mathbf{e}(S) \in \{0, 1\}^n$ such that

$e_j(S) = 1$, if $j \in S$, and
$e_j(S) = 0$, otherwise.

The production is characterized by consumption of m resources. The resource vector $\mathbf{r} = (r_1, r_2, \ldots, r_m)$ represents consumption of m resources to produce one unit of the final product. Each agent is characterized by its available production resources. The availability is defined by an availability matrix $\mathbf{A} = [a_{ij}]$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, where a_{ij} is the amount of resource i available at agent j. The resource capacity constraints for coalition S are given

$$q\mathbf{r} \leq \mathbf{A}\mathbf{e}(S). \tag{25}$$

The cooperative problem is formulated as to maximize profit of producers by a production quantity q and a coalition structure S subject to resource capacity constraints

$$\begin{aligned} &\pi_P = E\{(w-c)q - b\max(0; q - D(p,u))\} \rightarrow \max \\ &\text{subject to} \quad q\mathbf{r} \leq \mathbf{A}\mathbf{e}(S), \\ &\qquad\qquad q \in \mathrm{R},\ \mathbf{e}(S) \in \{0,1\}^n. \end{aligned} \tag{26}$$

Problem (26) can be solved for given vectors $\mathbf{e}(S)$. The total profit maximization is achieved for the grand coalition, i.e. $\mathbf{e}(S) = \mathbf{1}$. But some smaller coalitions can give maximal profit also. The maximal profit for producers is denoted by Π_P. For the profit allocation, it is necessary to identify all the coalitions that achieve this maximal profit by testing problem (26) with given maximal profit.

The coalitions with lower potential profit than the maximal get 0. The individual profit Π_{Pi} for the members from the coalitions with the maximal profit is allocated by Shapley values

$$\Pi_{Pi} = \Pi_P \sum_S \left\{ \frac{(|S|-1)!\,(n-|S|)!}{n!} \right\}. \tag{27}$$

Computation of the Shapley value allocation requires computing the solution of the problems (26) for all coalitions $S \subseteq N$. This can be time consuming for large sets of producers.

The procedure of the profit allocation algorithm can be summarized in following steps:

Step 1: Solve the problem (26) with resource capacity constraints for the grand coalition to obtain the price p^0 and the optimal level of production q^0. Compute the maximal total expected profit.

Step 2: Negotiations between retailers and producers how to allocate the total profit is given by splitting parameter λ ($0 \leq \lambda \leq 1$).

Step 3: Set the wholesale price vector w computed by (20) and buyback price b computed by (21).

Step 4: Identify all the coalitions that achieve the maximal profit by testing problem (26) with given maximal profit for producers Π_P.

Step 5: Compute the Shapley value allocation (27) to allocate the expected profit among the producers.

6 Conclusions

The aim of this paper is to propose mechanism for profit allocation in supply chains. The proposed procedure is based on a biform game and comes from the fact that the ongoing actions in the supply chain are a mix of cooperative and non-cooperative behavior of the participants. A combination of non-cooperative and cooperative game approaches is used. In the non-cooperative part, a coordination mechanism based on a specific buy-back contract is applied between producers and customers with price-dependent stochastic demand. The contract has desirable features: full coordination of the supply chain, flexibility to allow any division of the supply chain's profit, and easy to use. The cooperative part is merely focused on two concepts, coalition formations by resource capacity constraints and profit sharing. Profit sharing is carried out on the recognized concept of Shaply value. The analysis of the simple cases for the approach gives recommendations for more complex real problem. The approach seems to be useful and promising for next research. There are some possible extensions of the approach and some areas for further research, for example other types of games can be analyzed.

Acknowledgements

Supported by the grant No. 13-07350S of the Grant Agency of the Czech Republic and by Grant No. IGA F4/54/2015, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Brandenburger, A., Stuart, H.: Biform games. *Management Science*, **53 (**2007), 537-549.

[2] Cachon, G.: Supply Chain Coordination with Contracts. In: *Handbooks in Operations Research and Management Science: Supply Chain Management*. (Graves, S., de Kok, T., eds.). Elsevier, Amsterdam, 2003, 227-339.

[3] Cachon, G., Netessine, S.: Game Theory in Supply Chain Analysis. In: *) Handbook of Quantitative Supply Chain Analysis: Modeling in the e-Business Era* (Simchi-Levi, D., Wu, S. D., Shen, M., eds.). Kluwer, Boston, 2004, 13-65.

[4] Chen, J. M., Cheng, H. L.: Effect of the price-dependent revenue-sharing mechanism in a decentralized supply chain. *Central European Journal of Operations Research*, **20** (2012), 299-317.

[5] Fiala, P.: Information sharing in supply chains. *Omega*, **33** (2005), 419 - 423.

[6] Hennet, J. C., Mahjoub, S.: Supply network formation as a biform game. In: *Management and Control of Production Logistics*. University of Coimbra, Coimbra, 2010, 108-113.

[7] Lariviere, M. A.: Supply Chain Contracting and Coordination with Stochastic Demand. In: *Quantitative models for supply chain management* (Tayur, S., Ganeshan, R., Magazine, M., eds.). Kluwer, Boston, 1999, 233-268.

[8] Myerson, R. B.: *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, 1997.

[9] Nagarajan, M., Sošić, G.: Game-Theoretic Analysis of Cooperation among Supply Chain Agents: Review and Extensions. *European Journal of Operational Research*, **187** (2008), 719-745.

[10] Pasternack, B.: Optimal pricing and returns policies for perishable commodities. *Marketing Science*, **4** (1985), 166-176.

[11] Shapley, L. S.: A value for n-person games. In: *Contributions to the Theory of Games II*. (Tucker, A. W., Luce, R. D., eds.). Princeton University Press, Princeton, 1953, 307-317.

[12] Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E.: *Designing and Managing the Supply Chain: Concepts, Strategies and Case studies*. Irwin/ Mc Graw-Hill, Boston, 1999.

[13] Simchi-Levi, D., Wu, S. D., Shen, Z. J.: *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. Kluwer, Boston, 2004.

[14] Tayur, S., Ganeshan, R., Magazine, M. (eds): *Quantitative models for supply chain management*. Kluwer, Boston, 1999.

[15] von Neumann, J., Morgenstern, O.: *Theory of games and economic behavior*. Princeton University Press, Princeton, 1944.

[16] Yao, L., Chen, Y., Yan, H.: The newsvendor problem with pricing: Extensions. *International Journal of Management Science and Engineering Management*, **1** (2006), 3-16.

Burgernomics Revisited: Regional Purchasing Power

Jakub Fischer[1], Hana Lipovska[2]

Abstract. The aim of this paper is to estimate inter-area price levels in the Czech Republic. For this purpose, we developed the Regional Big Mac Index as a variation of the standard Big Mac Index which was calculated by *The Economist*. A unique dataset was used of Big Mac sales and prices in 89 McDonald's restaurants for June 2013. The analysis was performed by employing OLS model, cluster analysis and standard statistical methods. We found that Regional Big Mac Index is not suitable for estimating inter-area price levels in the Czech Republic, as it is only moderately correlated with the regional purchase power parity.

Keywords: Big Mac Index, cluster analysis, inter-area price levels, Regional purchase power parity,

JEL Classification: C43
AMS Classification: 91B82

1 Introduction

Regional purchasing power (RPP) is one of the most important, yet least explored, economic indicators. Regional policy in the European Union has replaced the Common Agricultural Policy in terms of importance as well as the budget expenditure. Intra-area price levels are needed particularly for comparing regional GDP and cohesion policies. However, Cadil et al. [1] claims that true regional price levels are not reported. Estimates of the regional purchasing power were not available for the Czech Republic till the recent studies of Cadil et al. [1] and Musil [7]. Their unique analyses are based on data from 2007, moreover, their approach is demanding in terms of methodology as well as data availability.

For regional research and policymakers both at the European Union and national level, it would be great help if a simple, cheap, ready and easy-to-use indicator could replace the RPP indicator at least as an initial, rough estimate. The aim of this paper is to estimate an alternative regional purchasing power index called **Regional Big Mac Index** (*RBMI*) and to compare and contrast this index with the results presented in Cadil et al. [1] (hereafter referred to as *RPPP CMMK* according to the initials of authors' last names).

RBMI was developed with the aim of this study according to the classical Big Mac Index introduced by *The Economist* in 1986. The Big Mac Index serves as simple way to calculate the overvaluation and undervaluation of currencies against the US dollar based on the theory of purchasing-power parity (The Economist [10]). The idea behind it is that a basket of identical goods costs the same everywhere. In this case the only good in the basket is a Big Mac. However, a Big Mac itself consists of numerous inputs and the Big Mac's price reflects those inputs prices (e.g. wages, rents, the price of beef or of lettuce (Parsley and Wei [8]). Nevertheless, as far as the authors are aware, this approach has never been applied on inter-area price levels.

The rest of the paper is organized as follows. First, we shall introduce the data and methodology we used. We then present the results of our analysis. The *RBMI* for the Czech NUTS 3 regions is calculated first and secondly results are compared with the findings of Cadil et al. [1]. Finally, the strengths and weaknesses of the *RMBI* approach are discussed.

2 Methodology and data

For the aim of our analysis a unique dataset, provided by McDonald's Czech Republic, was used. This dataset consists of prices and sales volumes of Big Macs in 89 McDonald's restaurants in June 2013. The reported prices used by *The Economist* are based on the *menu board* prices, which do not vary significantly between restaurants. In 93% of restaurants the menu board price was 70 CZK, in four restaurants 72 CZK and in two restaurants (situated at Vaclav Havel Airport and in Rozvadov, border crossing with Germany) 79 CZK. The average price of Big Mac was therefore 70.3 CZK, which correspondents to the price reported by *The Economist* for the first half of 2013 (The Economist dataset [10], data for January 2013). Only 16% of all Big Macs were however sold for

[1] University of Economics in Prague, Department of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, fischerj@vse.cz
[2] Masaryk University, Faculty of Economics and Administration, Lipova 41a, Brno, Czech Republic, lipovska@mail.muni.cz

the menu board price. The majority of burgers were sold in special price offers. In total there were 18 different price offers for all restaurants, however, each restaurant offered from 5 to 9 distinct price offers for a Big Mac (nearly two thirds of them offered 6 different price offers in total).

The weighted Regional Big Mac Index (*RBMI*) was calculated as follows for each of the fourteen Czech regions on the level of CZ-NUTS 3 (which correspondents to the districts or *kraj* in the Czech Republic)[3]. For rth region ($r = 1 \dots 14$), the matrix Q^r of quantities (2.1) and matrix P^r of prices (2.2) was created, n being the number of restaurants in the regional dataset. Each row correspondents to the appropriate restaurant in the $r-th$ region.

$$Q^r = \begin{bmatrix} q_{1,1} & \cdots & q_{1,18} \\ \vdots & \ddots & \vdots \\ q_{n,1} & \cdots & q_{n,18} \end{bmatrix} \tag{2.1}$$

$$P^r = \begin{bmatrix} p_{1,1} & \cdots & p_{1,18} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,18} \end{bmatrix} \tag{2.2}$$

For $r-th$ region, the total turnover V_r was calculated as the sum of the sale volumes for each restaurant:

$$V_r = \sum_{j=1}^{n} \sum_{i=1}^{n} Q_{ij}^r P_{ij}^r \tag{2.3}$$

Similarly, the total quantity T_r sold in $r-th$ region was calculated:

$$T_r = \sum_{j=1}^{18} \sum_{i=1}^{n} Q_{ij}^r \tag{2.4}$$

The average weighted regional price $\widehat{P_r}$ for $r-th$ region was then calculated as the ratio of total regional turnover and total regional quantity.

$$\widehat{P_r} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} Q_{ij}^r P_{ij}^r}{\sum_{j=1}^{18} \sum_{i=1}^{n} Q_{ij}^r} = \frac{V_r}{T_r} \tag{2.5}$$

The average price $\bar{P}_{CZ}$ for the Czech Republic was calculated as the share of the total turnover for all Czech McDonald's restaurants $\left(\sum_{r=1}^{14} V_r\right)$ and the total quantity sold in the reference period for all rants$\left(\sum_{r=1}^{14} T_r\right)$.

$$\bar{P}_{CZ} = \frac{\sum_{r=1}^{14} V_r}{\sum_{r=1}^{14} T_r} \tag{2.6}$$

Finally, individual regional purchasing parity $RBMI_r$ was computed for each region as the regional-Czech price ratio (2.7).

$$RBMI_r = \frac{\widehat{P_r}}{\bar{P}_{CZ}} \times 100\% \tag{2.7}$$

Roughly one fourth of McDonald's restaurants are situated close to international transport hubs (e.g. Vaclav Havel Airport) or along their highways. Prices in these restaurants do not reflect purchasing power in the relevant region, because such restaurants are situated away from local customer's normal shopping areas. Besides the *RBMI*, the modified index *RBMI**, which excludes those restaurants, was calculated.

Furthermore, we used data on the monthly *regional unemployment rate* published by the Czech Ministry of Labour and Social Affairs (for June 2013, CZSOa [2]). The quarterly *regional wages* w_r published by the Czech Statistical Office (CZSOb [3], data for second quarter of 2013) were used in the form of an index (2.8), which takes the average wage w_{CZ} in the Czech Republic as basement levels:

[3] NUTS stays for *Nomenclature of Units for Territorial Statistics*.

$$W_r = \frac{w_r}{w_{CZ}} \times 100\% \tag{2.8}$$

Similarly the index (2.9) GDP_r was counted for each region using data on the *Gross Domestic Product per capita* (*GDP p.c.*, published by CZSOc [4] for 2013). For *r*th region the gdp_r was used as a nominator and average GDP in the Czech Republic gdp_{CZ} as denominator:

$$GDP_r = \frac{gdp_r}{gdp_{CZ}} \times 100\% \tag{2.9}$$

Finally, the matrix of *regional food prices* published by CZSO was used (CZSOd [5]). This matrix covers the prices for such items as beef, pork, milk, eggs, potatoes and tomatoes. For each, the commodity respondent index (2.10) was calculated:

$$COMMODITY_r^i = \frac{price_r}{price_{CZ}} \times 100\% \tag{2.10}$$

For the further analysis, the standard linear regression model (OLS) was estimated. For this OLS model, the dummy variables $D_i, i = 1, \dots, 3$ were created on the basis of a cluster analysis (see Řezanková et al. [9]). The centroid model based on the k-means algorithm was used. This cluster analysis divided our data into four groups (clusters) according to the modified Regional Big Mac Index (*RBMI**) and Regional purchasing power (*RPPP CMMK*) estimated by Cadil et al. in [1] (see figure 1).

In spite of the small size of the data set ($n = 14$), the Pearson correlation coefficient ρ was used to compare our results with RPPP CMMK.

Pearson correlation coefficient	strength
$\langle \pm \mathbf{0,7}; \pm \mathbf{1} \rangle$	strong
$\langle \pm \mathbf{0,3}; \pm \mathbf{0,69} \rangle$	medium
$\langle \pm \mathbf{0,1}; \pm \mathbf{0,29} \rangle$	weak
$\langle \mathbf{0,0}; \pm \mathbf{0,09} \rangle$	none

Table 1 Strength of the correlation. Source: Jackson [6].

For the estimation of a linear regression model, the econometric software Gretl (version 1.9.14) was used. The statistical analysis was computed with the statistical software STATISTICA (version 12).

3 Results

According to (2.7) fourteen Regional Big Mac Indices *RBMI* and modified Regional Big Mac Indices *RBMI** were calculated. There is a medium-strength relationship between *RPPP CMMK* and *RBMI* ($\rho = 0.31$) and stronger, but still only a medium relationship between *RPPP CMMK* and *RBMI** ($\rho = 0.42$). It is apparent that the modified *RBMI**, after excluding restaurants close to international transport hubs, represents regional purchasing power better than the general *RBMI*. Table 2 shows the calculated *RBMI* and *RBMI** and compares them with *RPPP CMMK* based on the one common basket of goods and services[4].

NUTS 3			RPPP CMMK	RBMI	RBMI*	direction
10	PHA	Praha	119.7	103.3	104.3	+
20	STC	Stredocesky	101.9	102.5	101.7	+
31	JHC	Jihocesky	97.9	98.8	100.1	+
32	PLK	Plzensky	97.1	104.9	102.1	-
41	KVK	Karlovarsky	101.4	103.0	104.4	+
42	ULK	Ustecky	94.9	95.8	94.8	+
51	LBK	Liberecky	101.4	97.3	98.6	-
52	HKK	Kralovehradecky	96.4	102.3	102.5	-

[4] Cadil et al. [1] estimated RPPP even for regional baskets, which take into account different weights of individual items in the consumer's basket. RPPP based on the one basket does not differ significantly from the RPPP based on regional baskets.

53	PAK	Pardubicky	98.2	94.7	95.9	+
63	VYS	Vysocina	95.6	103.1	102.8	-
64	JHM	Jihomoravsky	103.4	98.6	98.8	-
71	OLK	Olomoucky	96.9	93.9	95.1	+
72	ZLK	Zlinsky	100.8	94.7	96.0	-
80	MSK	Moravskoslezsky	96.7	89.7	90.9	+

Table 2 Regional Big Mac Index (RBMI) modified Regional Big Mac Index (RBMI*) and Regional Purchasing Power, source: own computation and Cadil et al. [1].

Firstly we analyze the *direction* of both indices. If both the *RPPP CMMK* and $RBMI^{(*)}$ are above 100 or below 100, the direction is positive. If one of indices is above 100 while the other is below 100, direction is negative. There are positive directions in 8 regions: *Praha*, *Stredocesky*, *Jihocesky*, *Karlovarsky*, *Ustecky*, *Pardubicky*, *Olomoucky* and *Moravskoslezsky*. Three regions (*Plzensky*, *Kralovehradecky* and *Vysocina*) are above 100 according to *RBMI*, not the *RPPP CMMK*, and finally three regions (*Liberecky*, *Jihomoravsky* and *Zlinsky*) are below 100 according to *RBMI*, not the *RPPP CMMK*.

Based on the average price $\bar{P}_{CZ}$ for the Czech Republic and on the regional purchasing power *RPPP CMMK*, the theoretical prices $\widetilde{P}_r$ of regional Big Mac were calculated and compared with the real weighted regional prices$\widehat{P}_r$. Difference between theoretical and real prices was than calculated according to (2.8):

$$\delta_r = \widetilde{P}_r - \widehat{P}_r \tag{3.1}$$

As Figure 1 clearly depicts, a Big Mac is mostly overvalued in the historical land of Bohemia (except the region *Kralovehradecky*, *Pardubicky* and *Praha*), where the price should be according to the *RPPP CMMK* higher. The greatest exception is capital city Prague (*Praha*), where Big Mac price should rise by 7 CZK. On the other hand, a Big Mac is undervalued in all Moravian regions.

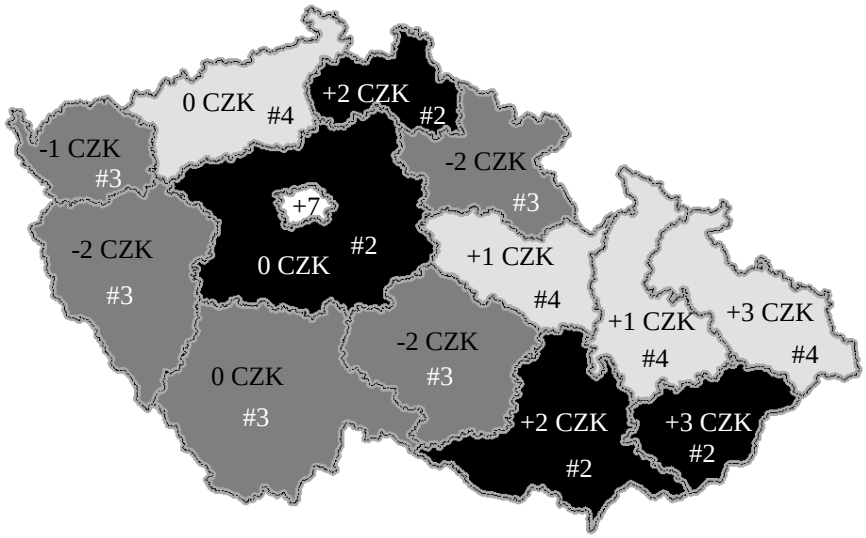


Figure 1 Map to show the division into four clusters based on the RPPP CMMK and RBMI. Differences δ_r between the theoretical and real price. Own calculation.

As the correlation matrix (table 3) clearly shows, Real Purchasing Power Parity as calculated by Cadil et al. [1] is more correlated with the modified *RBMI**, which excludes restaurants close to motorways ($\rho = 0.42$) in comparison to the general *RBMI* ($\rho = 0.31$). While *RPPP CMMK* is correlated more strongly with regional average wages ($\rho = 0.88$) and regional *Gross Domestic Product per capita* ($\rho = 0.91$), *RBMI** is correlated with regional unemployment ($\rho = -0.64$). While *RPPP CMMK* is more correlated with pork than with beef, in case of *RBMI** the opposite is true, which makes sense in light of Big Mac composition.

	RBMI	**RBMI***	**RPPP CMMK**
RBMI	1.00	**0.96**	0.31
RBMI*	**0.96**	1.00	0.42
RPPP	0.31	0.42	1.00

U	**-0.60**	**-0.64**	-0.48
W	0.37	0.39	**0.88**
GDP p. c.	0.33	0.38	**0.91**
BEEF	**0.55**	**0.57**	0.37
PORK	0.50	0.51	**0.63**

Table 3 Correlation matrix. Own calculation.

Based on the cluster analysis, 14 Czech NUTS–3 region were divided into four groups as shown in table 3 (those clusters are depicted also in figure 1). These regions serve for three dummy variables D_1, D_2, D_3 (with cluster 4 being the basement) which were used in the linear regression model.

	regions	**RPPP**	**RBMI***
cluster 1	PHA	119.7	104.3
cluster 2	STC, LBK, JHM, ZLK	101.9	98.8
cluster 3	KVK, PLK, JHC, VYS, HKK	97.7	102.4
cluster 4	ULK, PAK, OLK, MSK	96.7	94.2

Table 4 Results of cluster analysis. Own calculation.

Several specifications of linear regression model were tested, with model (3.2) proving most useful:

$$RBMI^* = \beta_0 + \beta_1 WAGE + \beta_2 BEEF + \beta_3 D_3 + \varepsilon \tag{3.2}$$

White HCE estimator was used for estimation. The model was satisfactorily tested for multicollinearity and the normality of residuals. The general misspecification of functional form was tested by Ramsey RESET test. The estimated liner regression model (table 4) explains 84% of data variability, which means that it fits the data very well. The modified Regional Big Mac Index *RBMI** can be explained by three regressors according to this model: (i.) regional wage index *WAGE*, which varies between 85.3 for *Karlovarsky* region and 130.7 for *Praha*, (ii.) dummy variable *D3_BOHEMIA*, which equals 1 if the region belongs to cluster 3 and 0 in other cases, and finally (iii.) the regional price index of beef *BEEF*, which varies between 93.8 for *Moravskoslezsky* region and 105.4 for *Stredocesky* region.

If the wage index *WAGE* grows by ten percentages points, the RBMI* grows by 2 percentage points. If beef-price index *BEEF* grows by ten percentages points, the *RBMI** grows by six percentages points. Finally, if region belongs to the "Bohemia" cluster 3, *RBMI** grows by 6.8 percentage points.

Variable	**Parameter estimate**	**Standard error**	**p-value**	
const.	21.7	19.9	0.3	
WAGE	0.2	0.03	0.00	***
D3_BOHEMIA	6.8	1.3	0.00	***
BEEF	0.6	0.2	0.02	**
R^2	0.84			

Table 4 The estimated OLS model of RBMI. Own calculation.

4 Conclusion

In this contribution, the suitability of the Regional Big Mac Index instead of traditional Regional purchase power parity was tested. Inter-area price levels, as estimated by Eurostat or Cadil et al. [1], are difficult to calculate and data-demanding as well as time consuming. If the Regional Big Mac Index proved to be strongly correlated with *RPPP* and reflected the inter-area level for each or at least most of the Czech regions, it would be useful to compute *RBMI* regularly, because this methodology is much easier.

However, it was found in this paper that Regional Big Mac Index does not reflect inter-area price levels in the Czech Republic well enough. The strength of correlation between the general and modified Regional Big

Mac Index, and the Regional Purchase Power Parity is merely moderate. This was shown in our results where it overvalued the regional price level in three regions and undervalued regional price level in other three regions. Unlike Regional Purchase Power Parity as calculated by Cadil et al. [1], Regional Big Mac indices are not strongly correlated with Gross Domestic Product per capita, nor the index of regional wages. On the other hand, the Regional Big Mac index reflects well the unemployment rate in Czech regions. With every additional one percentage point of unemployment, the Regional Big Mac Index falls by 1.7%. By comparison of inter-area price levels as estimated by Cadil et al. and of Regional Big Mac Index, Big Mac is theoretically undervalued in Moravia, Prague and the region *Liberecky*. This finding points to one of the greatest virtues and vices of the Big Mac Index; as it is based only on commodities, it is relatively easy to calculate; on the other hand it weakly reflects the reality on the regional level.

It is clear from our analysis that the Big Mac Index is not suitable for estimations of inter-area price levels. Several reasons can be detected: (1) The majority of McDonald's restaurants are situated in the district cities, which differ from the rest of the relevant regions. (2) Big Mac sales might be influenced by tourism in every region; they do not necessarily reflect the preferences and purchasing power of local citizens. (3) Data available represents sales in just one month (June 2013), so it is impossible to distinguish if the price-and-sales mix was not the result of local price-war between McDonald's and other restaurants.

Based on our results, it is also necessary to ask whether the Big Mac Index failing so fatally on the regional level is really suitable for purchasing power parity on the national level. At least, the results published by *The Economist* should next time be taken even more as a curiosity than as a serious indicator quoted by the national press.

Acknowledgements

This work was written thanks to the Institutional Support to Long-Term Conceptual Development of the Research Organisation of the Faculty of Informatics and Statistics of the University of Economics, Prague and was supported by the project of No. MUNI/A/1223/2014 funded by the Masaryk University. We would like to express our sincere than and appreciation both to Ivana Prosecká, McDonald's Finance Director for the Czech and Slovak Republic, and Michaela Polydorová, Sr. Accounting Manager in McDonald's Cooperation, for providing us with the unique dataset as well as for the fruitful discussion on the pricing.

References

[1] Cadil, J., Mazouch, P., Musil, P. and Kramulova J.: True regional purchasing power: evidence from the Czech Republic. *Post-Communist Economies* **26** (2014), 241–256.

[2] CZSOa: Nezaměstnanost v Jihomoravském kraji k 30. červnu 2013 [Unemployment in South Moravia District to 30th June 2013]. Available from: http://www.czso.cz/xb/redakce.nsf/i/nezamestnanost_v_jihomoravskem_kraji_k_30_cervnu_2013

[3] CZSOb: Průměrná hrubá měsíční mzda ve 2. čtvrtletí 2013 [Average gross monetary wage in 2nd quarter 2013]. Available from: http://www.czso.cz/xs/redakce.nsf/i/prumerna_hruba_mesicni_mzda_ve_2_ctvrtleti_2013

[4] CZSOc: Regionální účty v krajích ČR [Regional accounts in Czech regions]. Available from:

[5] CZSOd: Průměrné spotřebitelské ceny vybraných potravinářských výrobků [Average consumer prices of selected food products] CEN0150UP_KR, 24th week 2013. Available from: http://vdb.czso.cz/vdbvo/tabparam.jsp?cislotab=CEN0150UP_KR&voa=tabulka&go_zobraz=1&aktualizuj=Aktualizovat&cas_1_46=201324

[6] Jackson, S. L.: *Research methods and statistics: a critical thinking approach. 3rd ed.* Heinle Cengage Learning., Australia, 2008.

[7] Musil, P.: Cenové přepočty v národních účtech [Price calculation in national accounts], *dissertation thesis*, University of Economics, Prague (2014).

[8] Parsley, D. C. and Wei, S.-J.: A Prism into the PPP Puzzles: The Micro-foundations of Big Mac Real Exchange Rates. *Economic Journal, Royal Economic Society* **117** (2007), 1336–1356.

[9] Řezanková, H., Húsek, D. and V. Snášel: Shluková analýza dat [Cluster Analysis]. Praha: Professional Publishing, 2009

[10] The Economist: The Big Mac index – Interactive currency-comparison tool (2015). Available from http://www.economist.com/content/big-mac-index.

The Czech Republic and its neighbors: Analysis of spatial macroeconomic dynamics

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Regional macroeconomic dynamics in the Czech Republic and its neighbors may not be efficiently analyzed without properly accounting for the spatial structure of interactions and interdependencies. Spatial econometric models constitute a useful toolbox for a broad range of quantitative analyses and evaluation techniques. In this contribution, we assess spatial dynamics of unemployment in the Czech Republic and its neighbors. This analysis is performed at the NUTS2 level.
Within the scope of regional competitiveness indicators, a regression model for unemployment is estimated. We simulate the choices of alternative spatial structure specifications in order to approach and evaluate the robustness of our research and its conclusions. We conclude that spatial approach to econometric analysis of regional macroeconomic dynamics provides additional useful insight and robustness to a potentially broad range of spatially defined economic analyses that are carried out while relying on regional (spatial) data.

Keywords: Spatial econometrics, regional dynamics, unemployment rate.

JEL Classification: C23, C31, C52, E66
AMS Classification: 91B72

1 Introduction

Spatial econometric models account for the presence of spatial effects (such as economic spill-overs) when analyzing the relationships between variables through regression models and other related estimation methods. Spatial quantitative models play an ever more important role in regional macroeconomic and social analyses, real estate studies, agricultural & ecological applications, epidemiology and in many other non-economic fields of research. For this type of analysis, data need to be geo-coded using the latitude/longitude geographic coordinates system, as distances and common borders are used to estimate spatial dependencies.

Moran [10] and Geary [7] are often cited as the founding fathers of spatial econometrics, yet the actual framework for contemporary applied spatial econometrics was provided by Cliff and Ord in [3], [4] and their other publications, by introducing a relatively flexible spatial weights specification. Spatial weights are usually calculated in a two-step approach: First, spatial matrix is used to define neighbors (spatially close variable observations) using a dummy variable technique, where each element of the square spatial matrix equals 1 if the two spatial units are neighbors and 0 otherwise. Then, **spatial weights matrix** ($\boldsymbol{W}$) is constructed by row-standardizing the spatial matrix, so that the row weights sum up to 1. A simple 4-unit (4x4) example is provided next:

$$\text{Spatial matrix} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}; \qquad \boldsymbol{W} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

From the first row (and column) of the symmetric Spatial matrix we may observe that the first unit (say, region or city) is a neighbor of units 2, 3 and 4; the second row shows that unit 2 is a neighbor of units 1 and 3 (not a neighbor of unit 4), etc. Diagonal elements are set to zero by definition (units are not neighbors to themselves).

Spatial matrix construction is, perhaps surprisingly, the most ambiguous part of the otherwise well rooted methodology of spatial model specification and estimation. In many cases, it requires extensive geographical (polygon-based) mapping datasets and specialized software. **Contiguity** approach is a theoretically simple yet computationally complex rule, defining two units as neighbors if they share a common border. **Distance-based** approach usually constructs the spatial matrix by defining two units as neighbors if their distance does not exceed some ad-hoc predefined threshold. Alternatively, we may denote a preset number of *k* nearest units as neighbors (KNN method). This method solves for differences in areal densities (*k* neighbors are ensured for each unit), yet

[1] VŠE Praha, nam. W. Churchilla 4 Praha 3 Prague 130 67 Czech Republic. e-mail: formanek@vse.cz
[2] VŠE Praha, nam. W. Churchilla 4 Praha 3 Prague 130 67 Czech Republic. e-mail: husek@vse.cz

it leads to potentially flawed asymmetric spatial matrices. Distances between units are measured using **centroids**, their conveniently chosen representative positions. Depending on model focus, data availability and researcher's individual preferences, centroids may be pure geographical center points, locations of main cities, population-base weighted positions, transportation network based, etc. The variety of available neighbor definition approaches, the choice of centroids and selection of maximum neighbor distance thresholds imply that researchers usually need to consider several different choices (spatial structure settings) in order to verify model stability and robustness. As far as spatial matrices are concerned, there usually isn't a single right solution and researches often look for the most useful or interpretable model setup.

Before estimating spatial econometric models, we apply preliminary tests for spatial autocorrelation in the observed (cross-sectional) variables. Many spatial autocorrelation test statistics are available in [1] and from other sources, yet **Moran's I** seems to be the most widely used (spatial matrix symmetry is required, yet relatively straight-forward transformation algorithms for non-symmetric matrices are available):

$$I(k)_t = \left(\frac{n}{S}\right) \boldsymbol{z}_t' \boldsymbol{W} \boldsymbol{z}_t \, (\boldsymbol{z}_t' \boldsymbol{z}_t)^{-1} , \tag{1}$$

where $\boldsymbol{z}_t$ is the vector of n spatial observations (n units) of the variable k under scrutiny at time t. $S = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ is the standardization factor corresponding to the sum of all elements of the spatial weights matrix $\boldsymbol{W}$. The expected value of Moran's I under the null hypothesis of no spatial autocorrelation is: $-1/(n-1)$. As in [12], we use $var(I(k)_t)$ to calculate a z-score and test for statistical significance: whether neighbor units are more similar to one another than they would be under spatial randomness. The sign of Moran's I discriminates between positive and negative spatial autocorrelation.

Once significant spatial dependence in observed data is verified, spatial regression may be used to account for such situation. Again, various estimation methods are available. We use a **spatial lag model** when focusing on the analysis of spatial interactions in the dependent variable (within this paradigm, the dependent variable is the one with spatial structure). A general formula for the spatial lag model and its reduced form may be written as

$$\boldsymbol{y}_t = \rho \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{u}_t , \tag{2}$$

$$(\boldsymbol{I} - \rho \boldsymbol{W}) \boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{u}_t , \tag{3}$$

where $\boldsymbol{y}_t$ is the vector of all y_{it} spatial units' observations at time t, $\boldsymbol{I}$ is the $n \times n$ identity matrix, $\boldsymbol{X}_t$ is a matrix of regressors (includes the intercept element, may include lagged variables). Maximum likelihood (ML) approach is used to estimate both the spatial dependence parameter ρ and the regression coefficients $\boldsymbol{\beta}$ which are used to explain the variability of individual y_{it} observations that is not explained spatially. As usual, $\boldsymbol{u}_t$ and its elements u_{it} describe the random portion of the regression model. Individual observations of the dependent variable y_{it} in (2) and (3), may be expressed in terms of weighted averages of their neighbors' values: $\boldsymbol{W}\boldsymbol{y}_t$ is the spatial lag of $\boldsymbol{y}_t$ and we may write $SpatialLag(y_{it}) = \sum_j w_{ij} y_{jt}$, where y_{it}, y_{jt} both refer to spatially defined dependent variable observations at time t and w_{ij} are the row-standardized spatial weights – for each y_{it} , i-th row of $\boldsymbol{W}$ is used (zeros on the diagonal). Additional detailed discussion is provided in [1].

If the research does not focus on spatial dependence structure and its analysis, we may still take advantage of the **spatial error model** in order to account for (correct for) the spatial nature of the data observed. In this case, we deal with spatial autocorrelation by introducing spatially correlated errors $\boldsymbol{u}_t = \lambda \boldsymbol{W} \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t$ that are justified by the presumed existence of unobservable features associated with location or with spatially defined omitted variables. By analogy to equations (2) and (3), the spatial error model and its reduced form may be expressed as

$$\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta} + \lambda \boldsymbol{W} \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t , \tag{4}$$

$$(\boldsymbol{I} - \lambda \boldsymbol{W}) \boldsymbol{y}_t = (\boldsymbol{I} - \lambda \boldsymbol{W}) \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t , \tag{5}$$

where λ is the spatial dependence parameter and $\boldsymbol{\beta}$ coefficients are used to explain the variability in y_{it} observations that is not explained by the spatial nature of the observed data. While $\boldsymbol{u}_t$ suffers from spatial autocorrelation, $\boldsymbol{\varepsilon}_t$ is a true random element. Estimated models (3) and (5) may be directly compared upon their maximized log-likelihoods. A Durbin-like specification test for equation (5) is based on the *spatial common factor hypothesis* that exploits the fact that model (5) may be expressed in spatial lag form (3) if spatially lagged regressors (spatial lag of the k-th regressor is defined as: $SpatialLag(x_{kit}) = \sum_j w_{ij} x_{kjt}$) are explicitly included in $\boldsymbol{X}_t$ and if specific *common factor constraints* on $\boldsymbol{\beta}$ coefficients hold. A simple likelihood ratio test may be used to verify the validity of the null common factor hypothesis: each β coefficient of a spatially lagged regressor equals the negative of the product of ρ and the β coefficient of the corresponding non-lagged regressor. Additional information on spatial

models such as (2) to (5), further spatial correlation tests and model selection and validation procedures are provided in [1], as well as complementary literature references.

As the $n \times n$ dimension of the $\boldsymbol{W}$ matrix is determined by the number of spatial units, computational limits for the methods described here exist. However, econometric models comprised of, say, 500 and more units may be easily handled using ordinary PC configurations and the freely available R software (http://www.r-project.org), especially when spatial matrices are sparse (each unit has a relatively limited amount of neighbors). Last but not least, spatial regression models have been extended to encompass panel data methods. This approach is well described in [2] and, [5], where many useful references are provided.

The remainder of this paper is structured as follows: Section two focuses on regional competitiveness and unemployment dynamics to be quantified using spatial models applied to the Czech Republic and its neighbors at the regional level NUTS2 (as defined by the Nomenclature of Territorial Units for Statistics). Section three discusses data issues and regression model specification, section four summarizes the empirical results, their significance, relevance and robustness. Section five and the list of references conclude our contribution.

2 Regional competitiveness and unemployment dynamics

At the microeconomic (firm) level, the notion of competitiveness is fairly clear and based on the individual firm's capacity to compete, grow and make profit through products or services that meet market requirements in terms of price and quality (other relevant market factors may play a role). In contrast, at the macro level, there is a less clear concept of country-based or **regional competitiveness**, despite of the fact that competitiveness is frequently presented as the main goal of economic policy actions. Usually, competitiveness is perceived as some combination of productivity, employment (unemployment) rate, living standards, foreign direct investment (FDI) attractiveness, etc. However, individual and collective political preferences, country-specific short term fluctuations in key macroeconomic indicators, general public inattention and other rather amorphous influences play a potentially significant role in competitiveness definition, leading to diverse interpretations and possible misunderstandings.

Some authors (such as Krugman, [8]) consider the concept of national and regional competitiveness to be mostly meaningless and potentially dangerous, arguing that the analogy between nations (regions) and firms is inherently flawed: unsuccessful firms ultimately go out of business whereas no equivalent situation for a nation or region may occur as far as developed world is concerned. We would refer the readers to [9] for detailed and structured discussion of macroeconomic competitiveness aspects as we proceed with this contribution by focusing on spatially defined macroeconomic data that are consistent with the mainstream '**consensus view**' (as referred to in [9] and other texts) of regional competitiveness indicators.

In this contribution, our motivation is two-fold: it is our belief that successful regional macroeconomic performance (generally speaking, competitiveness) may be reasonably assessed in terms of living standards dynamics (i.e. GDP per capita in absolute or relative terms) and in terms of unemployment and its dynamics. Also, both variables (GDP per capita and unemployment) are often cited as constituent parts of regional competitiveness (please refer to [9], [6] and [11] for in-depth discussion). Using the econometric analysis paradigm outlined in previous section, selected regional competitiveness indicators may be assembled into a relatively simple, yet useful and theoretically well-defined model: we aim to explain unemployment dynamics in terms of relative per capita productivity (its short term dynamics) and using a convenient technological advantage indicator (high-tech sector labor force proportion). Once spatial dependencies in the data are properly addressed, such model (exact specification is provided in the next section) may serve as a tool for discerning the influence of geographically defined conditions from factors that may be - at least potentially - influenced by macroeconomic policies (regionally focused fiscal stimuli, establishing of technological parks, applied research funding, etc.).

3 Data used and model framework

Over the last ten years, Eurostat (http://ec.europa.eu/eurostat) has made a great progress in harmonization and availability of spatial macroeconomic and socio-economic data at 'lower' levels of NUTS aggregation (NUTS2, NUTS3). Still, in most data series relevant for this study, we face serious data availability issues. At the NUTS2 level, there is a great likelihood of missing observations across regions during each year included in our dataset (we use annual data, as relevant quarterly observations are completely unavailable). Moreover, data observations are not missing at random but country-wise (systematically), which may lead to serious biases in the estimated regression models, if not addressed properly.

For our analysis, we use R-software (eurostat package) to extract GDP per capita, unemployment and high-tech employment data for a total of 76 NUTS2 regions of the Czech Republic and it neighbors. The complete set of 76 spatial (cross sectional) observations at the NUTS2 level is comprised of 8 regions in the Czech Republic, 4 in Slovakia, 9 in Austria, 16 in Poland and 39 in Germany (of those, 8 are former East-German regions; 9 including

the unified Berlin). Specifically, the following Eurostat databases are used (identification by table code): nama_r_e2gdp, lfst_r_lfu3rt and htec_emp_reg2. Although annual observations in some spatial data series may cover the period from 1999 to 2014, observation availability (incompleteness) issues and temporal relevance considerations led us to focus on data from the time period 2010-2012 only (unless stated otherwise).

Model framework and specification

As mentioned in the previous section, the application part of our contribution is based upon regressing regional unemployment on relative GDP dynamics (first differences) and a convenient technological advantage measure (relative employment in high-tech industries). We have determined that the OLS estimator is not appropriate for such task, as Moran's I statistics (1) indicates strong and positive spatial autocorrelation for all observed data series considered. All Moran's I results are significant at the 5% significance level (we reject the null of spatial randomness) and were found to be relatively stable across different annual observations (2000 – 2012) and across varying spatial structure definitions, obtained by using a wide range of maximum neighbor distances for $\boldsymbol{W}$ calculations. Due to space limitations, individual Moran's I results are omitted here. All skipped statistics, figures and tables mentioned here and in the following paragraphs are available from the authors upon request, along with the datasets and R scripts used.

As a next step in our analysis and following the general approach outlined in [1], we use specialized Lagrange multiplier test statistics for spatial autocorrelation, in order to decide whether spatial lag or spatial error model should be used. At the 5% significance level, only the spatial lag model is supported by the observed data. Hence, using the general specification (2), equation (6) is the chosen model used for estimation and subsequent analysis of regional unemployment dynamics of the selected countries:

$$Unemp_{it} = \rho\, \boldsymbol{w}_i^T \boldsymbol{Unemp}_t + \beta_0 + \beta_1\left(GDP_{i,t-1} - GDP_{i,t-2}\right) + \beta_2 TechEmp_{it} + u_{it}\,, \qquad (6)$$

where $Unemp_{it}$ is the general rate of unemployment for NUTS2 region i at time t (2012), $\boldsymbol{Unemp}_t$ is a vector of all *Unemp* observations for the year 2012, $GDP_{i,t-1}$ is region's GDP per capita (current EUR prices of 2011) expressed as percentage of EU average (also in current 2011 prices). The expression $\left(GDP_{i,t-1} - GDP_{i,t-2}\right)$ therefore describes the y-o-y percentage point change in GDP-based "relative strength" of region i. This variable serves as a proxy for region's y-o-y change (from 2010 to 2011) in macroeconomic competitiveness. In contrast to using GDP_{it} in our model (6), the temporal lag and dynamic nature of $\left(GDP_{i,t-1} - GDP_{i,t-2}\right)$ is consistent with neoclassical macroeconomics, allowing time for the transmission mechanisms to take place. $TechEmp_{it}$ describes the percentage of employees working in the "high-tech industry" (NACE r.2 code HTC) and $\boldsymbol{w}_i^T$ is the i-th row of the spatial weight matrix $\boldsymbol{W}$ used to define the spatial structure of the spatial lag model.

4 Empirical results and interpretation

Spatial model paradigm may suffer from the generally perceived ambiguity in spatial matrix definition and construction. Therefore, to provide supporting evidence for our results, we test **specification robustness** of the model (6) against changes in definition of its spatial structure. Multiple estimations of model (6) are performed with the dataset available, while varying the distance-based spatial weights matrix $\boldsymbol{W}$. We start with a relatively sparse $\boldsymbol{W}$ matrix constructed using maximum neighbor distance threshold set to 160 km. Thresholds lower than 160 km lead to the existence of at least one island unit (a region with zero neighbors), which breaks down the ML estimation of the spatial model. From the initial 160 km value, we increase the threshold distances by 10-kilometer iterations, up to a maximum neighbor distance of 1.000 km (results obtained using neighbor distance thresholds beyond 1.000 km are rather non-informative). A total of 85 $\boldsymbol{W}$-based alternative spatial structure-specifications of model (6) are produced, estimated and summarized in figure 1. For each estimated spatial structure definition, figure 1 shows the corresponding Akaike information criteria (AIC), maximized log-likelihood (LL) statistics and the estimated coefficients $\boldsymbol{\beta}$ and ρ along with their asymptotic (+/-) one standard error bands.

Maximum neighbor distance thresholds between 210 and 270 km result in fairly similar AIC values that are close to the best (lowest) AIC statistic, observed at the 240 km threshold (shown by dotted vertical lines in figure 1). The LL statistics mirrors our AIC results and the described model selection process leads us to the choice of model (6) with the $\boldsymbol{W}$ matrix calculated using the 240 km threshold. As may be observed from table 1, all estimated coefficients are significant at the 5% level and the Wald statistic indicates high overall model significance. Moreover, our results are reasonably consistent and stable across the whole 210 - 270 km maximum neighbor distance interval. At the same time, the results in figure 1 show that coefficients β_0, β_2 and ρ (Intercept, TechEmp and spatial dependence) remain stable well beyond a 350 km maximum neighbor distance threshold. The stability of β_1 over this extended neighborhood threshold interval is less prominent. Model instability at the lower end of the distance threshold interval should not be viewed as puzzling: 160 km is the minimum distance that avoids isolated units (islands) for the spatially defined dataset, but such spatial structure is not realistic (providing very few neighbors

to the observed regions). Similarly, high distance thresholds are not realistic either, as they are not consistent with plausible regional interactions and spillover effects for unemployment. This may be observed from the AIC and LL statistics, as well as from the behavior of individual regression coefficients as plotted in figure 1.

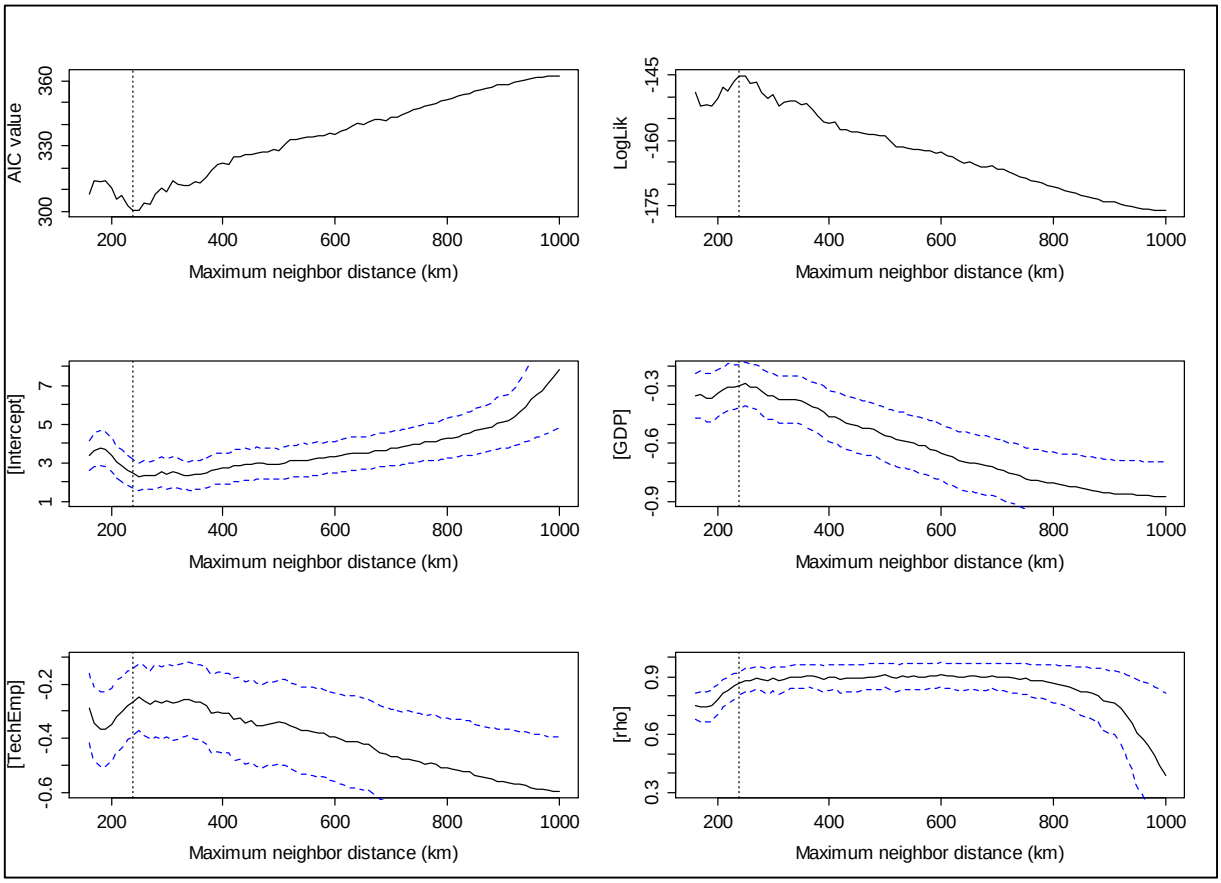


Figure 1 Stability: AIC, LL and coefficient values as functions of maximum neighbor distance thresholds

Technicalities of model estimation may be summarized by stating that the spatially defined equation (6) was processed using R-software (spdep package). Also, as apparent from table 1, four spatial-unit observations were missing in our dataset. Hence, our model of choice is estimated using only 72 out of the total 76 NUTS2 regions described in section 3. In this particular situation, we are able to treat the NAs as 'missing at random' and estimate the model using a simple complete-cases-only approach (all missing data belong to German NUTS2 regions where only 4 out of 39 observations are missing and the NA values come from two separate NUTS1-level regions).

```
Dependent variable: Y20.64_2012 (maximum distance threshold = 240 km)

Type: lag
                                    Estimate Std. Error z value  Pr(>|z|)
(Intercept)                          2.46708    0.74520  3.3106 0.0009308
I(EUR_HAB_EU_2011 - EUR_HAB_EU_2010) -0.30115    0.11337 -2.6563 0.0078997
HTC_2012                             -0.26669    0.12473 -2.1381 0.0325051

Rho: 0.86344, LR test value: 62.442, p-value: 2.7756e-15
Asymptotic standard error: 0.062937 z-value: 13.719, p-value: < 2.22e-16

Wald statistic: 188.22, p-value: < 2.22e-16
Log likelihood: -145.1291 for lag model
ML residual variance (sigma squared): 2.9052, (sigma: 1.7045)
Number of observations: 72, Number of parameters estimated: 5
AIC: 300.26, (AIC for lm: 360.7)
LM test for residual autocorrelation: test value: 0.032973, p-value: 0.85591
```

Table 1 Model (6) – R estimation output

Overall, we may summarize the estimation output as follows: we find strong evidence for spatial autocorrelation in the data, as the spatial dependence coefficient is relatively large ($\rho = 0.86$) and highly significant. The other regression coefficients are significant at the 5% level (or lower) and their signs and magnitudes are consistent with

standard macroeconomic approaches to the subject matter. **Predictive properties** of the model at its spatial structure described above are satisfactory and fairly comparable across all the NUTS2 regions included in this study. Moreover, we do not observe any significant country level (NUTS0) prediction power differences. There is no bias towards, say, accurately predicting unemployment in German regions (35 of the total observed 72 regions) at the expense of other spatial units. Some proportion of this rather surprising behavior may be reconciled, among other factors, to the 8 (9 including Berlin) NUTS2 regions belonging to former East Germany; see [6] for detailed FDI-focused regional/spatial analysis of former East Germany, the Czech Republic and Poland. Structured econometric evaluation of model's predictive properties is omitted due to space limitations.

5 Conclusions

It may be argued that the model framework as described in this contribution is relatively simple and therefore much of the spatial effect is attributable to some omitted variable factor. However, spatial dependence may be interpreted as a proxy for a number of real, yet practically unobservable spatial effects. Spatial dependencies, interactions and their dynamic features are very difficult to conveniently define and structure in a way that would facilitate informative and harmonized quantification. In practical terms - tasks such as consistently measuring cross-border work commuting preferences at the NUTS2 level, accounting for administrative/qualification employment barriers between countries (NUTS0 level), quantifying the impact of language differences, aerial distances vs. topology, etc. - would inherently introduce many subjective decisions and disputable features to any quantitative model. Hence, spatial lag models provide a useful, interpretable and functional approach towards regional (macroeconomic) data analysis.

This contribution provides supporting evidence for the use of regionally focused macroeconometric models. Although spatial econometric models have been neglected and perhaps underappreciated by the general public, they are able to provide interpretable and robust evidence and supporting material for decision making processes at different central authority levels. The general specification of spatial models, as outlined by equations (2) to (5), may be easily customized as for their functional form and variable selection, depending on the particular tasks and research goals assigned.

Acknowledgements

This research project was supported by Grant No. IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Anselin, L., and Rey, S. J. (eds.): *Perspectives on Spatial Data Analysis*. Springer Verlag, Berlin, 2010.
[2] Baltagi, B. H., Bresson, G. and Pirotte, A.: Forecasting with spatial panel data. *Computational Statistics and Data Analysis* **56** (2012), 3381-3397.
[3] Cliff, A. D., and Ord, J. K.: The problem of Spatial autocorrelation. In*: London Papers in Regional Science 1, Studies in Regional Science*, (Scott, A. J., ed.), Pion, London, 1969, 25-55.
[4] Cliff, A. D., and Ord, J. K.: *Spatial Processess: Models and Applications*. Pion, London. 1981.
[5] Fingleton, B.: Forecasting with dynamic spatial panel data: practical implementation methods. *Economics and Business Letters* **3**(4) (2014), 194-207.
[6] Gauselmann, A., Marek, P., and Angenendt, J. P.: MNE's Regional Location Choice - A Comparative Perspective on East Germany, the Czech Republic and Poland. *IWH Discussion Papers*, **8** (2011), 1-32.
[7] Geary, R. C.: The contiguity ratio and statistical mapping. *The Incorporated Statistician* **5** (1954), 115-145.
[8] Krugman, P.: Competitiveness: A Dangerous Obsession. *Foreign Affairs* **2** (1994), 28-44.
[9] Martin, R.: *A study on the factors of regional competitiveness – a draft final report for the European Commission Directorate-General Regional Policy*. Cambridge University Press, Cambridge, 2004
[10] Moran, P. A. P.: Notes on continuous stochastic phenomena. *Biometrica* **37** (1950), 17-23.
[11] Piras, G., and Arbia, G.: Convergence in per-capita GDP across EU-NUTS2 regions using panel data models extended to spatial autocorrelation effects. *STATISTICA* **2** (2007), 157-172.
[12] Ullah, A., and Giles, D.E.A. (eds.): *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 1998.

Stochastic analysis of profitability of the pig breeding process

Ludvík Friebel[1], Naděžda Kernerová[2], Jana Friebelová[1]

Abstract. The main goal of this contribution is to find and apply an efficient simulation model of pig breeding at a particular company in order to improve its economic results. The main inputs for the model were statistical distributions for number of inseminations and number of weaned piglets. In addition we have used conditional probability distributions for particular animal among each litter. Costs at particular parts of the breeding process were used as the additional simulation inputs, together with number of parities, length of farrowing interval, number of weaned piglets depending on a parity number, lactation length, and the gestation length of sows. Next inputs were price of produced piglets, price of insemination doses etc. The empirical curve of sow profitability depending on the number of parities, which can help to decide about the culling of sows from the herd, was also determined in the research. The obtained results and recommendations are discussed at the end of the paper.

Keywords: simulation; probability distribution; conditional probability, pig breeding; weaning of piglets; culling of sows

JEL Classification: C13, C15, P4
AMS Classification: 60H30

1. Introduction

Many authors have dealt with the economy of piglet production farms, e.g. Kleinhanss and Werner [4], Pelletier et al. [5]. The key factors of economy of piglet production system can be divided into two groups – exogenous and endogenous. Exogenous factors, such as the price of feeding, the price of energy and the price of pork meat, can be influenced negatively. In contrast, the endogenous factors can be managed with success.

Piglet production is intimately related to the reproduction process of sows and many different factors other than feeding may affect the final results. Hence, the increasing number of new variables and constraints affecting piglet production make difficult to explore all possible management alternatives to find the best one. Therefore, sow herd management models can play an important role to optimise management alternatives or to explore new ones (Rodriguez-Sanchez et al. [10]).

Herd management is the process by which certain goals of the farm manager, expressed as amount of product, are achieved by consuming a corresponding amount of production factors. In order to be able to combine these factors in an optimal way it is necessary to know the main interrelations among them and their influence on the final productivity of the system.

The optimal replacement policy of sows is described by Rodriguez et al. [9]. Plà et al. [8] have developed a production model using Markow chains and Semi-Markow chains (Plà et al., [7]). A review of mathematical models for sow herd management is described by Plà [6]. Simulation model, which includes comparison of economic performance of two breeds of sows, was introduced by the authors in the article Friebel et al., [1]. The relation between culling rate and production efficiency in nucleus herd is assessed in Houška [2].

In this study, we focused on the distribution of time spent by farrowing batches at different parts of the production process during farrowing interval. Because of the stochastic character of animal reproduction, we have developed a stochastic simulation model covering costs and incomes arising during the whole reproductive cycle.

2. Material and methods

The pig breeding process is depicted in Figure 1. At the top of Figure, we can see the time axis. Total capacity of the model farm (620 sows) is divided into five particular sections - facilities. Capacity of particular facilities is given in Table 1.

Facility	Capacity

1 Faculty of Economics, University of South Bohemia in České Budějovice, Studentská 13, 370 05 České Budějovice, Czech Republic, friebel@ef.jcu.cz
2 Faculty of Agriculture, University of South Bohemia in České Budějovice, Studentská 13, 370 05 České Budějovice, Czech Republic, kernerova@zf.jcu.cz

Breeding	170
Gestation	280
Farrowing	130
Gilts acclimatization	40
Piglet rearing	1700

Table 1: Capacity of sections

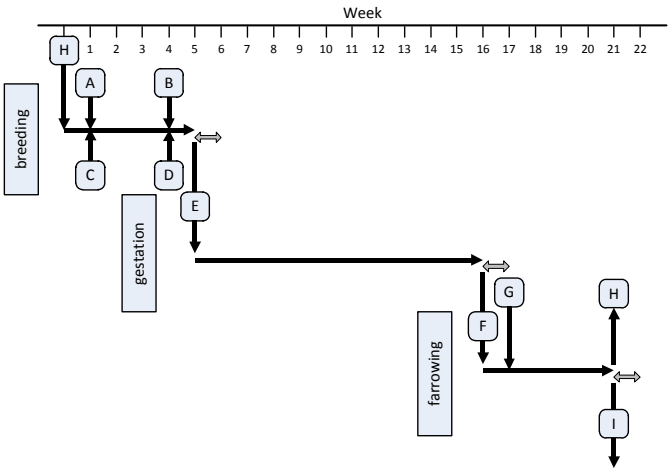


Figure 1: Reproduction diagram, A – oestrus, insemination, B – detection of pregnancy, C – herd completion, D – detection of heat, E – movement do gestation facility, F – movement to farrowing facility, G –farrowing, H – weaning of piglets, I – movement to breeding facility, ⇔ disinfection of particular facility

Because the duration of reproduction cycle is biologically set, we can only change the lactation length. However, the possible change of the above mentioned stage has to be ethologically admissible. For the purpose of simulation model we considered minimal weaning interval of 21 days, which is also minimal lactation interval.

The second parameter of the pig breeding process which can be successfully changed and which has a great economic impact is the culling policy. Keeping the inefficient sows from the point of view fertility in reproduction process has an undesirable impact on the economy of the whole farm. Inefficiency of the sows can be assessed from the point of view of pregnancy, fertility and veterinary costs.

Pregnancy is evaluated with a number of insemination and oestrus which are needed for becoming pregnant in the particular reproduction cycle. During the first and second reproduction cycle, it is possible to tolerate a higher number of inseminations; typical are three or four attempts. In the third and the next reproduction cycle, the number of possible insemination falls to two. It means unless the sow become pregnant after the second oestrus, it is culled from herd. In our study, we defined a critical value of oestrus for each reproduction cycle as an element of vector of pregnancy – **a**. This vector has got seven elements, because we consider seven reproduction cycles in maximum for the reasons stated below.

Fertility is the next important culling reason. It is represented with the number of weaned piglets. At the first farrowing, there should be at least seven weaned piglets, otherwise the sow is culled. For the next farrowing, the required number of the weaned piglets increases to nine. It is possible to concede a lower number of piglets after the fourth or fifth reproductive cycle. In our study, we defined a critical (minimal) value of weaned piglets for each reproduction cycle as a component of fertility vector – **b**.

The change of both the weaning time and the culling rules has a great impact on distributions of sows at particular facilities. To change internal dispositions of a real farm may be difficult, thus possible changes, especially of the pregnancy rules, are limited.

1.1 Data

The data file included 5771 records of Czech Landrase (CL) sows. Each record included the birth date, date of culling, reason for culling, date of farrowing, date of weaning, number of weaned piglets in each litter, number of insemination during farrowing interval, etc. The data had to be processed in order to gain necessary simulation inputs see below.

1.2 Costs analysis

In order to achieve as an accurate model as possible, the costs in particular phases of the production cycle are described by cost functions, whose argument is the time spent on particular stands. The cost functions

include veterinary costs (vaccination, deworming, and disinfection of facilities), energy costs (lighting, operation of automatic feeding systems, ventilation, heating etc.), water costs and wages. This part of the cost functions are for the given phase of the reproduction cycle constant, but they differ significantly from each other, see Table 1 at http://home.ef.jcu.cz/~friebel/simulations/costs2.pdf. This is caused primarily by a different arrangement of particular farming parts given by different needs of sows during the reproduction cycle.

The cost functions include also feeding costs which differ not only from the point of view of particular phases of the reproduction cycle, but also during the stay of a sow and its piglets in a given stand as a feeding dose changes.

The first operation cost of the commercial herd is purchasing a sow. In the model we calculate the price to be CZK 6,000. These new sows need to be acclimatised before they are integrated into the herd, this phase is considered to be 40 days. The construction of the cost function for the acclimatisation stay is depicted at http://home.ef.jcu.cz/~friebel/simulations/costs2.pdf.

After integrated into the herd, the sow is placed in the breeding facility. The cost function for this phase is constructed using presented in Table 1 and 2 at http://home.ef.jcu.cz/~friebel/simulations/costs2.pdf.

The time spent in this phase changes for particular sows and it is connected with their ability to get pregnant. This ability is also connected with insemination dose cost and is calculated to CZK 139.

After getting pregnant, sows are removed to gestation facilities, where they are for a constant period of time. The costs calculated for this phase are based on the data presented in Table 1 and 3 at http://home.ef.jcu.cz/~friebel/simulations/costs2.pdf.

The farrowing facility is, considering the costs, the most important part of the reproduction farming. The highest costs of the commercial herd cover energy, as piglets require higher temperature, and also wages, due to the necessary assistance of keepers when the sows give birth. It is similar with the water consumption, see Table 3. The most significant, considering the total costs in the farrowing facility, are feeding costs of lactating sows and of complementary feeding of their piglets.

Due to the variability, caused by the different weaning time of particular simulations and the different number of piglets in particular iterations, it is neither possible to generalise the costs in the farrowing facility or to use a cost function. In our model, we decided to separate feeding doses for sows and feeding doses for piglets.

To calculate the total costs of one sow in the farrowing facility, we used the data from Table 1 and 4. It is necessary to take into account different weaning times. We consider the weaning time to be 21–25 day. During the last five days in the farrowing facility, the feeding doses decrease, see Table 4, in order to stop lactation. It is necessary to add to the feeding costs also the daily costs presented in Table 1. The feeding costs of piglets in the farrowing facility are the product of the average costs per feeding day of a piglet in the farrowing facility (CZK 1.14, which is complementary feeding) and their number generated during every iteration, see below.

Costs of piglet rearing are calculated as the product of the number of weaners and the costs of a feeding day in the piglet rearing facility see table 1 at http://home.ef.jcu.cz/~friebel/simulations/costs2.pdf.

1.3 Incomes

The main income of a pig breeding farm is from sold piglets. Piglets are shipped in a large scale and the total weight and the total number is taken into account. For this reason, the average weight and average price per kg is considered for the simulation purposes. The income per sold piglet for purpose of simulation model was computed as the multiple of average price and average weight and is set to CZK 1,568, in our model p_p.

The additional income consists of revenues for culled sows, which are sold for meat. We calculate in the model as the multiple of average weight of culled animals and the average price of meat and we set it for CZK 6,250. This income can partially balance costs for purchasing and acclimatisation of gilts. The difference between the purchasing price of gilts and the revenue, when they are sold after the culling, is represented by so called replacement cost, in our model c_r.

1.4 Simulated inputs

Number of inseminations

The number of inseminations influences the time spent on breeding facility. As well the price of insemination dose is one of the important costs. The number of inseminations was monitored only up to the 7th parity for the same reasons as for the litter size, see below. In order to create this simulation input was used the Distribution Fitting function in the program @RISK. The most suitable were the Truncated Binomial distribution.

Because the probability of conception differs from the increasing number of litters, we had to suggest a particular probability distribution function for each litter. The number of inseminations necessary for farrowing is generated according to this distribution during every iteration.

Litter size - number of piglets in a particular litter

At first we investigated dependency of number of weaned piglets in consecutive parities for a particular animal. We hypothesized that fertility is linked with particular animal.

Correlation indices in table 2 verified this hypothesis. Pdf for the second and next parity is considered as conditional on the previous parity. Hence litter size in the each fallowing parity depends on the previous one. The litter size was monitored only up to 7th parity thereafter the efficiency of the sow slopes down mainly because of piglet mortality.

***i*-1**	***i***	***r***	***p* value**
1	2	0.3439	0.000
2	3	0.3381	0.001
3	4	0.2356	0.018
4	5	0.2471	0.021
5	6	0.2802	0.005
6	7	0.2947	0.003
7	8	0.4798	0.000

Table 2: Corelation indices for consecutive parities

We dealt with the problem of inhomogeneous Markov chains in this, because general fertility expressed in number of weaned piglets in consecutive parities can be characterized with multidimensional probability distributions, where number of dimensions is number of parities.

In our study we had to fit empirical distribution with theoretical function separately for sows with the same number of piglets in previous parity in order to obtain pdf for particular litter. The most suitable was the truncated Binomial distribution. The problem occurred in the case of less frequent combinations, where data were insufficient for fitting with theoretical pdf. In this case we had to use empirical pdf.

1.5 Simulation model

Simulation inputs

- Costs function in particular facility f_a (breeding), f_b (gestation), f_c (farrowing), f_d (rearing);
- Price of sold piglets p_p;
- Sow's replacement cost c_r;
- Price of insemination dose c_o;
- Costs per piglets and day spent on farrowing facility c_{pc};
- Number of inseminations before becoming pregnant for i-th litter $r_i \sim R_{1i}$;
- Number of weaned piglets for *i*-th litter $p_i \sim R_{2i\text{-}1i}$;
- Culling rules see Table 3, vectors $\boldsymbol{a}$ and $\boldsymbol{b}$;
- Time spent in breeding facility T_b (87 days);
- Time securing gestation T_a;
- Maximal number of farrowings l;
- Time from farrowing to weaning T_w;
- Time from birth to shipping of piglets T_s (70 days).

Number of litters i	**Maximal number of inseminations** a_i	**Minimal number of piglets** b_i
1	4	7
2	3	9
3	2	9
4	2	10
5	2	10
6	2	10
7	2	9

Table 3: Culling rules

The simulation model is described in Figure 2, where *i* denotes the order of the litter. The stands are denoted with *a, b, c, d* (*d* is the rearing facility).

Simulations outputs

- Income per sow per cycle p_r and per year p_y;
- Time spent by a sow in particular stands during the *i*-th litter t_{ai}, t_{bi}, t_{ci};

- Time spent by piglets of the *i*-th litter in stand *d* (rearing facility) $t_{di}=p_i(T_s-T_w)$;
- Costs of particular stands *a* to *d* and particular litter: c_{ai}, c_{bi}, c_{ci}, c_{di}.

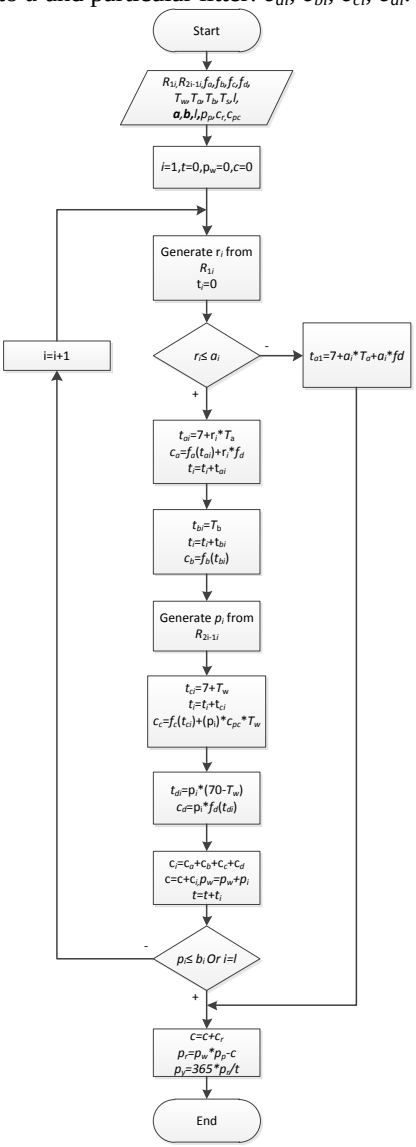


Figure 2: Simulation model

3. Results and conclusions

In our model, we were working with an option, after which litter the sows should be culled and the weaning time. In fact, we were looking for the extreme of a multi-variable function, in this case of two variables. In calculus, this would be solved using partial derivatives. However, it is not possible in our case. Therefore, we had to test all allowable combinations of the weaning time and the maximum number of litters and determine the maximum average profit per sow.

At first, we performed 100,000 iterations for different weaning periods and different values of the maximal number of farrowing. We tested five possible periods of weaning for 21–25 days and seven maximal number of farrowing for 3–7. The total profitability per year and sow is shown in Table 4.

Max no. of farrowings	Weaning interval [day]				
	21	**22**	**23**	**24**	**25**
3	9884	9872	9861	9850	9839
4	10436	10423	10411	10399	10387
5	10828	10815	10802	10790	10777
6	10864	10851	10838	10826	10813
7	10824	10810	10799	10786	10774

Table 4: Profitability in CZK (Profit per sow and year) 100,000 iterations

It seems as the most efficient to cull sows after the sixth litter. The difference between the fifth, sixth and seventh litter is minimal. The reason is the minimum number of sows which get to a higher than the sixth reproduction cycle. For these reasons, it is common in production farming to leave sows in the cycle only to the sixth litter.

Considering the weaning process, the model proved that shorter weaning time brings higher profits. Unlike the number of litters (see the previous paragraph), the differences are most evident. To be more objective, it is necessary to note that the shorter weaning time may cause problems in the following inseminations. The reason is the unfinished uterine involution of some sows. Opinions on this topic differ, see Kiracofe [3]. In modern commercial herds, the weaning time of 21 days is commonly used.

The average length of the meantime in our model is 150 days and the turnover rate is 2.43. When considering the theoretical value of 141 days (21 day weaning time), it represents an increase by nine days due to repeated inseminations.

Another output of the simulation model is the capacity of particular sections during the year. The values presented in Table 5, show how the capacity of these sections should be set up in the chosen operation mode. Considering the total capacity of 580 sows in the farm and the mode arising from the simulation, the original dispositions of the farm should be changed in accordance in respect to the following table.

Facility	Breeding	Gestation	Farrowing	Total	Piglet rearing
Time (mean value)	110.81	191.20	62.99	365	
Places new	176	304	100	580	1874

Table 5: Distribution of spent time between sections [day]

The analysis also shows that the original capacity of the piglet rearing facility would be, considering the optimal profitability, too low, as the model comes with the mean required of 1,874 places. The average number of litter per sow till its culling is 3.421.

As was mentioned in the methodology part, one of the most important questions of the pig breeding process with a great economic impact is the culling policy. This policy is closely connected with purchasing gilts and herd replenishing. Incomes for culled sows can balance the cost for purchasing gilts only partially. The most important culling reason in our model was farrowing – 57.2 %. The second reason for culling was pregnancy. – 34.8 %. The age represented with number of parities was from that point of view culling negligible – only 8% of sows were culled after sixth parity.

References

[1] Friebel, L., Friebelová, J., Marinov, L.: Stochastic analysis of costs during the pig breeding process. In *Proceedings of 31st International Conference on Mathematical Methods in Economics 2013*, College of Polytechnics Jihlava, Jihlava, 2013, 189–194.

[2] Houška, L.: The relationship between culling rate, herd structure and production efficiency in a pig nucleus herd. *Czech J. Anim. Sci.*, 54(8), 2009, 365–375.

[3] Kiracofe, G.H.: Uterine involution: its role in regulating postpartum intervals. (Review). *J. Anim. Sci.*, 51(2), 1980, 16–28.

[4] Kleinhanss, W.: Development of the incomes, liquidity and stability of pig holdings. *Berichte uber Landwirtschaft*, 87(3), 2009, 428–443.

[5] Pelletier, N., Lammers, P., Stender, D. et al.: Life cycle assessment of high- and low-profitability commodity and deep-bedded niche swine production systems in the Upper Midwestern United States. *Agr. Sys.*, 103 (9), 2010, 599–608.

[6] Plà, L.M.: Review of mathematical models for sow herd management. *Livest. Sci.*, 106(2-3), 2007, 107–119.

[7] Plà, L.M., Faulin, F.J, Rodríguez, S.V.: A linear programming formulation of a semi-Markov model to design pig facilities. *J. Oper. Res. Soc.*, 60(5), 2009, 619–625.

[8] Plà, L.M., Pomar, C., Pomar, J.: A Markov decision sow model representing the productive lifespan of herd sows. *Agr. Sys.*, 76(1), 2003, 253–272.

[9] Rodriguez, S., Jensen, T.B., Plà, L.M., Kristensen, A.R.: Optimal replacement policies and economic value of clinical observations in sow herds. *Liv. Sci.*, 138(1-3), 2011, 207–219.

[10] Rodriguez-Sanchez, S.V., Pla-Aragones, L.M., Albornoz, V.M.: Modelling tactical planning decisions through a linear optimization model in sow farms. *Liv. Sci.*, 143(2-3), 2012, 162–171.

Combined System of Continuous Double Auction and Automatic Market Maker in Prediction Market

Mikuláš Gangur [1]

Abstract. The contribution deals with the proposal of the prediction market financial structure with respect to increasing the volume of trading as well as liquidity that are the key factors of the prediction market quality. The described approach modifies the trading mechanism of the market and it integrates the system of continuous double auction and the principles of automatic market maker into this mechanism. This way the combined system of continuous double auction and automatic market maker is created. In the contribution the process of price determination by means of the logarithmic market scoring rule principles is explained. The fruitfulness of the newly proposed system, i.e. the usage of orders generated by tradesmen, is compared with the previously used system of the constant price in case of Dutch auction.

Keywords: Prediction market, automatic market maker, continuous double auction, market scoring rule, automatic order maker.

JEL classification: C65
AMS classification: 65C50

1 Introduction

Prediction markets (PM) belong among the alternative tools for collecting information spread among numerous experts. These markets are used for the evaluation of the success rate of the assigned forecasts - predictions and as speculative markets simulate the activity of a stock exchange at which such titles are traded that are related to forecasting a particular event or those related to a value of an estimated parameter. The value of titles is given by the extent of confidence of the sellers and buyers in a given event or a value of a parameter. The current market price can be interpreted as an estimate (forecast) of the probability of an event or an estimated value of parameters. Prediction markets use mechanisms of classic stock markets on the basis of the information shared by the individual participants of the market by means of the price of shares.

In prediction market implementation we follow the structure of market. The setting of this structure designates the quality and fruitfulness of market. According to [7] the construction of PM can be divided up to three areas: Choice of Forecasting Goal, Incentives for participation and information revelation and Financial market design.

This contribution focuses on financial market design that establishes the principles and mechanisms of trading on the market i.e. the selection of market and trading rules. The financial markets mostly use the following mechanisms: Continuous Double Auction (CDA), Automatic Market Maker (AMM), Call Auctions (CA) and Dynamic Pari-Mutuel Market (DPMM). PM mostly use two main trading mechanisms - CDA and AMM.

In the CDA system the orders of participants are matched and if matching is successful the transaction is executed. The AMM system uses the Market Scoring Rule (MSR), and in this system the share prices are determined continuously by administrator (manually or by means of automata). The market participants don't trade with each other, but they trade through AMM that continuously determines sales and buying prices. Most PMs use only one of above mentioned mechanisms. In this contribution the proposal of combined trading mechanism is introduced. The share prices are set by the market participants and they are simultaneously complemented by automatically generated orders with prices

[1]University of West Bohemia/Faculty of Economics, Department of Economics and Quantitative Methods, Tylova 11, Plzeň, Czech Republic, gangur@kem.zcu.cz

that are calculated by means of MSR principles. It is the price determination that seems to be the key problem of the proposed combined trading mechanism.

2 The share price determination

If there are no orders to sell or buy inserted in the orders book (OB) by participants in CDA system, the system automatically generates its own orders and it inserts the orders into OB under administrator id. The generator plays the role of a participant through the administrator id and this way it takes part in trading in CDA system. The key problem is the new order price and the determination of the amount of shares.

The basic idea of determination of new prices results from R. Hanson works (i.e. [3]) in the area of automatic market maker (AMM) in prediction markets. R. Hanson uses the scoring rule (SR) principle, that determines the payoff for the determination of the correct probability distribution $\mathbf{p}$ among particular states i of the considered events. The shares tied with the given states of event predict the event occurrence or nonoccurence. The logarithmic market scoring rule $s_i(\mathbf{p}) = b \cdot log(p_i)$ (LMSR) is one of the most employed rules in AMM. The construction of the applied cost function for the share price determination is derived from this rule.

According to the LMSR the participant that changes the probability distribution of the state with low probability (a small amount of contributed information) is priced higher than that who changes the distribution of the state with higher probability (more contributed information). The change of low probability corresponds to an intuitive concept of change in the situation, when we don't have much information about the considered event state (little money have been spent for the shares of the given state and, therefore, the share price is low). That is why any change is more influental and also more priced than in the situation when the share price is high (traders put more money into shares i.e. the shares of the given state aggregate more information). The change, i.e. inserting more information, influences the share price less, i.e. it changes the probability distribution less.

Example 1

As an example of event with different states the 'Percentual profit of political party ABC in election' is shown. The number of states is determined by the issuer (market administrator). The event can be represented by mutually disjointed states - shares, which are displayed in the table 1, together with the current price and the number of shares held by the market participants.

Profit in % of votes	$\langle 0-20)$ %	$\langle 20-40)$ %	$\langle 40-60)$ %	$\langle 60-80)$ %	$\langle 80-100\rangle$ %
Current price/Number of held shares	89/30	90/40	69/60	25/50	10/10

Table 1 Distribution of held shares for all event states (Source: own)

□

R. Hanson describes the behavior of traders as a permanent change of price instead of a process of selling and buying shares. If we take a look at this process as selling and buying, we are interested just in the share price, that we sell or buy it for. If we assume that the final evaluation of share is 1 virtual currency unit (VCU) in the occurrence of the event state and 0 VCU in the nonoccurrence of the event state, we can replace AMM that uses MSR by the market maker using the cost function (cost-function based market maker). The value of this cost function depends on the number of the sold or bought shares [1] i.e. it depends on distribution of marketed shares among event states. If we denote q_i as the number of all sold (held) shares of i-th event state and we want to determine the price of any number of shares, we can use the following cost function corresponding to LMSR:

$$C(\mathbf{q}) = b \log \left(\sum_j \exp(q_j/b) \right) \tag{1}$$

On the basis of the introduced prescription the current price of i-th state share can be determined as a partial derivative of the cost function (1) with respect to the q_i. The price, determined in this way, is considered only for a few traded shares. For the price determination of any number of shares we use the change of cost function (1) with respect to the old price [6].

In the next part we can interpret cost function C (1) for shares with final values 1/0 as a total amount of money, that all traders spend for all shares (shares of all event states) up to now. The parameter b has to be set to avoid the loss of AMM and, at the same time, to determine the amount of price change depending on the number of shares in the transaction. In automatic price determination the parameter b is set with respect to the range of shares price (0-100) as $100 \times 100 = 10000$.

By means of the cost function C we determine the price c_i, that traders pay or obtain if they demand or supply quantity q^i shares of the selected event state i. The calculated price is the difference between the value of all the held shares after the transaction (the value of cost function C after transaction) and the cost function value before the transaction. This price is then the price of transaction and it doesn't define the current share price, because such price works only for a small number of shares in the transaction.

If traders demand q^i shares, we determine the price with respect to the following prescription:

$$c_i = C(q_1, \cdots, q_i + q^i, \cdots, q_n) - C(q_1, \cdots, q_i, \cdots, q_n) \tag{2}$$

n is the number of the event states

The price can be determined similarly in case of the supply of shares from traders. In this case we subtract the amount q^i of supplied shares.

If the price c_i is positive, the trader pays it (administrator cashes it). If it is negative, the trader cashes this sum of money (administrator pays). The sign determines the direction of transaction from the view of administrator (AMM).

3 The price determination in the combined model of CDA and AMM

The creator of AMM R. Hanson described the possible integration of orders book and automatic market maker (AMM) in [2]. In this case the AMM trades all the supplied or demanded shares directly on the basis of the calculated price, or it may trade only part of them according to the calculated share price. In the price calculation AMM uses MSR instead the cost function and it doesn't take the current share price into account, and the computation itself by means of numerical methods is time consuming.

The CDA and AMM systems were used also on the Zocalo prediction market [5]. Its author introduces the used principle of CDA and AMM integration in [4]. In the described system the price determination by means of LMSR is isolated from the resulting CDA prices and the market participants select the final price by comparing prices from both the systems. This combined model is limited only to the event states with two-values result (0/1). Moreover, the applied LMSR principle, proposed by R. Hanson, considers only the amount of the supplied or demanded shares without the inclusion of the volume of the demanded or supplied financial means i.e. the total volume of transaction.

In the calculation of the supplied or demanded price for the generated order the system doesn't work only with the amount of the traded shares, but it also includes the total volume of transaction. It, at the same time, takes the current share price into account. It employs the principle of the price determination with the cost function for the determination of the value (price) of the distributed portfolio change. The change is consequence of executed transaction. The value of change is the basis for the determination of the change of the share price which corresponds to the change of the distributed portfolio with regard to the demand or supply.

With respect to the difference between the AMM model, described in the previous part, and the combined CDA and AMM model, we will, at first, substitute the quantities q_i of the held shares in the description (2) by the total value of the held shares for the considered event. We determine the total value along to the following prescription:

$$v_i = \sum_j q_{i,j} \times c_i^{last} \tag{3}$$

where v_i is the total value of the held shares for i-th state, $q_{i,j}$ is the quantity (number) of shares for i-th state, held by j-th trader and c_i^{last} is the current (last) price of shares for i-th state

Another question is what can be the share quantities q^i in (2) substituted by ? The quantity q^i states how many pieces of shares are demanded or supplied by trader. While only one order to buy or sell is avalaible in AMM for the price, determined by AMM, and the trader just sets the number of shares to buy or sell, in the CDA system more orders with different prices and different numbers of shares are inserted in OB all at once. What method do we use to determine the volume of means v^i, meant for the transaction, that substitute q^i ?

In the case of buying or selling such modification of principle of maximizing sales volume was choosen that is used for quotation of shares. According to this method the system selects such a transaction order that demands or supplies most financial means and most contributes to the change of probability distribution for the considered event.

$$v^i = \max_j(q_{i,j} \times c_{i,j}) \tag{4}$$

where v^i is the total sales of the transaction for the shares of the selected state i with respect to all shares of the states tied with the event, $q_{i,j}$ is the number of shares for i-th state, demanded (supplied) by j-th trader and $c_{i,j}$ is the share price for i-th state demanded (supplied) by j-th trader

On the basis of the above process it is possible to modify the prescription (2) as follows:

$$\triangle c_i = C(v_1, \cdots, v_i + v^i, \cdots, v_n) - C(v_1, \cdots, v_i, \cdots, v_n) \tag{5}$$

where n is the number of event states and $\triangle c_i$ is the price of transaction

and the situation is similar in cases where shares are supplied by traders.

Except for substituting the share quantity in (2) by the total value of the held shares and by the sales of transaction, another difference can be found in (5). While in AMM system traders pay (obtain) the price of transaction c_i for the implemented change in shares distribution and the system doesn't work with the current share price, in the combined AMM and CDA system the calculated value is the change of the current price, because the prices are determined by traders regardless of AMM directly in the range of the share values. At the begining the prices are set by administrator with respect to his/her preferences[1]. That is why the final price of share c_i^{supply} in this model is determined as the current price c_i^{last} that is modified by the calculated price of the whole transaction $\triangle c_i$ per one share of transaction. If trader buys the share, ($\triangle c_i$ is positive), the final price increases, if trader sells ($\triangle c_i$ is negative), the final price decreases.

Another price correction monitors exceeding the maximum share price c_i^{max} for the given event state or depreciation of the share price bellow determined minimal price c_i^{min}. The final share price for i-th state (new current price) is computed with respect to the following prescription:

$$c_i^{supply/demand} = \max(c_i^{min}, \min(c_i^{last} + round(\frac{\triangle c_i}{q_i} + 0.5), c_i^{max})) \tag{6}$$

The administrator (AMM) is also 'trader' and therefore tries to sell the most profitably or to buy the most cheaply. That is why it is possible to correct the share price with respect to the prices demanded or supplied by traders. In case of shares sold by administrator the system selects the maximal price from calculated price (6), and the referential price c_i^{ref}, that is demanded by traders, and at the same time from the current price c_i^{last}. In case of purchase the administrator offers minimal price of the calculated price or referential price that is supplied by traders, and, at the same time, from the current price.

Referential price can be determined by different methods e.g. the maximal price from the demanded prices in case of shares sold by administrator or the minimal price from supplied prices in case of shares purchased by administrator. Another referential price is the price from order with maximal turnover (see prescription (4)). This approach takes into account the main priority of the system administrator, which is to ensure as much volume of trades and market liquidity as possible.

[1] AMM determines the start price of share with help of C cost function (1) on the basis of initial distribution $q_1, \cdots, q_n$, that is set by system administrator.

Example 2

In the example 1 for prediction of the percentual election profit of ABC party the shares distribution before the transaction for each of the five states is displayed in table 1. The trader inserted the demand order for 20 shares with price 70 VCU of event state *Party ABC receives 40-60 % votes in election.* The current price of the share tied with the event state is 69 VCU.

According to (3) the distribution of the held share values is **v**=(2670,3600,4140,1250,100), and according to (5) the change of price is determined by the following prescription:

$$\triangle c_3 = C(2670, 3600, 4140 + 20 \times 70, 1250, 100) - C(2670, 3600, 4140, 1250, 100) = 349.30$$

according to (6) the supply price is calculated as

$$c_3^{supply} = \max(1, \min(69 + round(\frac{349.30}{20} + 0.5), 100)) = 87$$

This price is compared with the maximal demanded referential price 70 VCU, with current price 69 VCU, and the maximal price $c_3^{final} = max(87, 70, 69) = 87$ is selected. The shares are supplied by administrator (AMM) for price 87 VCU in number of demanded shares, corresponding to referential price.

Similarly, the share price demanded by administrator (AMM) can be determined in case of supply from traders. □

4 The comparison of automatic order generators

In this part the results of the use of two generators are compared. The first one is the previously used automatic generator that issues shares with constant price (GCP) and second one is the proposed automatic combined generator with price that is calculated according to the method in part 2 (CG).

The generator with constant price was activated in the tested group under the similar conditions as the combined generator from the previous part in the second tested group. In the experiment just administrator's supplies were analyzed. The constant price was determined as maximal of 95 VCU and of the price offered by trader as the demanded price. For every unsold shares the price depreciated 1 VCU every day in the system of Dutch auction as well as in CG. The number of shares was set with respect to the planed investment and the constant price in contrast to the combined generator. That's why this number of supplied shares is in the most cases smaller than the demanded number by traders.

Amount	Administrator supply - Number	Administrator supply - Volume	Average price per share
Generated CG	1052	78941	75.04
Accepted CG	854	61502	72.02
Fruitfulness CG	81.18 %	77.91 %	-
Generated GCP	600	57230	95.11
Accepted GCP	270	24272	89.90
Fruitfulness GCP	45 %	42.41 %	-

Table 2 The trades of generated sell orders with CG and GCP (Source: own)

The consequent values CG (GCP) are presented in table 2. The smaller number of issued (and also traded) shares in the GCP (600/270) as compared with CG (1052/854) is result of different method of determining the number of the newly issued shares.

Following the displayed values the fruitfulness of both generators can be compared according to the number of the issued shares and the volume of trades. The fruitfulness 81.18 % as opposed to 45 % traded shares of the newly issued shares number is markedly higher with CG. Similarly, the fruitfulness measured by the volume of potential and executed trades is higher in favour of CG (77.91 %) as compared with GCP (42.41 %).

It can be stated from presented comparisons that prices generated by CG are more accepted by traders, more trades are executed, the price corresponds to the current distribution of particular event shares and this way the price values the information contributed by traders.

For the sake of the completeness it is needed to compare the activities of both groups in the experiment period so that we exclude a potential effect of greater activity towards greater accepting of automatically generated orders. The activity is represented by a list of volume of trades for each participant of the given group in the given period.

Both groups are compared by means of average and median. The congruity of both the groups was tested both by means of the Levene's test and the nonparametric Mann-Whitney test with respect to data distribution in both groups which normality was not proved. Neither of tests confirmed the validity of the alternative hypothesis about the differences of averages (medians) (p-value=0.1325, ev. p-value=0.1467 on significance level $\alpha = 0.05$). The activities of both groups, given by the total volume of trades, are comparable and therefore it can be stated, that the detected difference between fruitfulness of CG and GCP in the newly issueed shares is not affected by the different activity of the participants in both groups, but it can be described as the result of the different methods of the compared generators.

5 Conclusion

The proposed combined automatic generator integrates the principles of the automatic market maker, used for price determination in case of automatic generation of the demand or supply orders, into the market mechanism of the continuous double auction. The generator was compared in a practical experiment with simple, previously used generator that determines the price as maximally possible and then it depreciates the price in periodical time intervals in the Dutch auction system. The main emphasis was put on the comparison of both generators' fruitfulness, i.e. the application of the generated orders according to the volume of the executed trades or. as the case may be, the frequency of transactions. The results of the comparison of the above generators show the newly proposed combined generator as being more successful.

In the contribution the method of price determination in the combined generator system is presented for binary predictions, i.e. the prediction of complementary events. These predictions answer the question whether the event occurs or does not occur. Another method must be used in case of estimating the predicted parameter values. An ongoing research study focuses on the price determination in case of these type of predictions. Generally, the evaluation of the estimated parameter value can be transformed into the presented algorithm for binary predictions. A special construction of complementary events is employed in this case. This approach allows a wider application of the combined generator for different types of predictions.

References

[1] Chen, Y.: *Mechanisms for prediction markets*, chapter 4. Wiesbaden, Verlag Gabler, 2011, ISBN 978-0-203-81552-6, 40–51.

[2] Hanson, R.: Book Orders for Market Scoring Rules. Working paper, George Mason University, 2003.

[3] Hanson, R.: Logarithmic market scoring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets* **1** (2007), 3–15.

[4] Hibbert, C.: Integrating book orders and market makers. [online], Retrieved May 2007 from `http://pancrit.blogspot.com/2007/01/integrating-book-orders-and-market.html`

[5] Hibbert, C.: Zocalo. [online], Retrieved July 2014 from `http://zocalo.sourceforge.net`

[6] Pennock, D.: Implementing Hanson's Market Maker. [online], Retrieved October 2006 from `http://blog.oddhead.com/2006/10/30/implementing-hansons-market-maker/`

[7] Spann, M., and Skiera, B.: Internet-based virtual stock markets for business forecasting. *Management Science* **49** (2003), 63–69.

Scheduling Based on the Critical Path Analysis

Pavel Gežík[1]

Abstract. The Critical Path Method (CPM) solves the shortest time of project and the earliest start and the latest end for each activity without making the project longer. But no consideration is given to the fact how many workers/machines would be needed to undertake these activities, or indeed to which worker or machine would be needed for each activity. In the Critical Path Analysis (CPA), scheduling for workers is related to bin-filling problems or similar problems called the knapsack problems which are known as the NP-hard problems.
This paper describes a simple heuristic method based on the Critical Path Analysis which provides a good result regarding to time of solving but does not necessarily give the optimal solution every time. This method uses earliest starts and latest ends of every activity of a project to determine the order of activities. There is also the calculation for the minimum number of workers required to complete the project and to match workers with activities.

Keywords: Scheduling, Critical Path Analysis, Heuristic Method, Project.

JEL Classification: C61
AMS Classification: 90B35

1 Introduction

This paper describes how the heuristic scheduling method can be applied to the Critical Path Analysis (CPA) problems but, importantly, it will be shown that such methods do not necessarily give the optimal solution every time. This scheduling problem is related to the bin-filling problems, and to a similar problem called the knapsack problem in which the items carried have not only particular weight, but also have an appropriate value. These problems are known as the **NP-hard problems** ([4], [5]) and that are often solved by heuristic methods which provides a good result regarding to time of solving.

Most of these problems are solved quite casually or automatically without explicit recognition of a problem. Sometimes an ordering method is determined essentially by chance; more often tasks are performed in the order in which they arise (for example the "first-come, first-served" solution). [3]

These types of problem are solved by using a technique called *the branch and bound method* or *method of shortest path*. Both methods aim to the optimal solution but they are difficult and they need inadequately solution time for a complex project with a large number of activities. The heuristics methods do not provide optimal solution every time but they take acceptable solution time. In [6], [7], known heuristics methods are *Assembly-Line Scheduling Method, Method of Moodie Young, Tonge´s Method, Arcus´s Method, Method of Killibridge and Western and Hoffmann´s Method*. All that methods are based on diversification of tasks to groups or phases.

2 Critical Path Method (CPM)

The Critical Path Method (CPM) is a method of the Critical Path Analysis (CPA) and solves the shortest time of project and the earliest start and the latest end for each activity without making the project longer. The CPM is a step-by-step project management technique for process planning that defines critical and non-critical activities with the goal to find shortest time possible to complete the project. Critical activities have to follow each other and by this way they create the critical path.

There [1], [2] are few ways to how to use the CPM but all of them find the *Early Start* (*ES*), *Early Finish* (*EF*), *Late Start* (*LS*), *Late Finish* (*LF*) and *Slack* for each activity. The calculation of the CPM is separated into two phases. In the first phase the early starts and finishes are calculated by seeking the maximal values in previous activities. In the second phase the late starts and finishes are calculated by seeking the minimal values in next activities. The calculation can be carried out by MS Excel with the functions MAX, MIN, and VLOOKUP from table array of predecessor and successor activities like in the example below.

[1] University of Economics in Bratislava, Faculty of Business Informatics, Department of Operations Research and Econometrics, Dolnozemská 1/b, 852 35 Bratislava, Slovak Republic e-mail: pavel.gezik@euba.sk.

In the following *table 1* there is an example which will be used to illustrate the problem. There are *activities* (a_i) for $i = 1, 2, \ldots, n$ where n is a set of project activities, the time t_i of each activity i (days in the example) as well as predecessor and successor $\boldsymbol{a_i}$. ES_i, EF_i, LS_i, LF_i and $slack_i$ of each activity i are calculated in the next columns. The project duration can be seen in the table like a maximal number in a column of finishes and critical path is composed of activities where $Slack_i = 0$.

In the example, project duration is 65 days and critical activities are a_2, a_7, a_9 and a_{10}.

$\boldsymbol{a_i}$	$\boldsymbol{t_i}$	***predecessor*** $\boldsymbol{a_i}$	***successor*** $\boldsymbol{a_i}$	$\boldsymbol{ES_i}$	$\boldsymbol{EF_i}$	$\boldsymbol{LS_i}$	$\boldsymbol{LF_i}$	$\boldsymbol{Slack_i}$
a_1	7	-	a_4, a_6, a_7	0	7	9	16	9
a_2	16	-	a_4, a_6, a_7, a_8	0	16	0	16	**0**
a_3	4	-	a_5	0	4	27	31	27
a_4	7	a_1, a_2	a_5	16	23	24	31	8
a_5	16	a_3, a_4	a_{10}	23	39	31	47	8
a_6	8	a_1, a_2	a_{10}	16	24	39	47	23
a_7	20	a_1, a_2	a_9, a_{11}	16	36	16	36	**0**
a_8	13	a_2	a_9, a_{11}	16	29	23	36	7
a_9	11	a_8, a_7	a_{10}	36	47	36	47	**0**
a_{10}	18	a_5, a_6, a_9	-	47	**65**	47	**65**	**0**
a_{11}	8	a_8, a_7	-	36	44	57	**65**	21

Table 1 CPM example

3 Scheduling

In the previous part, the CPM solves the shortest time of project and the earliest start and the latest end for each activity without making the project longer. But no consideration is given to the fact how many workers/ machines[2] would be needed to undertake these activities, or indeed to which worker or machine would be needed for each activity. These problems can be solved by the method of scheduling based on the CPA directly after the end of CPA.

This method of scheduling is based on the CPA because it uses the information from the CPA about the whole project and also about every activity. This method has the two parts. The first part needs early star and late finish for each activity to set order of activities and also it needs value of critical path (project duration) to calculate the number of workers. The second part is about scheduling and matching workers with the activities.

3.1 Order of activities and number of workers

This part is simple and there is no algorithm. The order of activities is given by sorting of activities according to early starts (ES_i) at the first level (ascending) and according to late finishes (LF_i) at the second level (ascending). The logic of sorting is depended on principles of the CPA and in the beginning activities with smaller early start have to be done. If two activities have same early starts then the first one which has to be done is the activity with smaller late finish. When two activities have same early starts and also late finish then they had to be sorted according to $Slack_i$ of activity (ascending).

In the example, there is the order of activities: a_2, a_1, a_3, a_4, a_7, a_8, a_6, a_5, a_9, a_{11} and a_{10} and it is sort in below *table 2*.

$\boldsymbol{a_i}$	$\boldsymbol{a_2}$	$\boldsymbol{a_1}$	$\boldsymbol{a_3}$	$\boldsymbol{a_4}$	$\boldsymbol{a_7}$	$\boldsymbol{a_8}$	$\boldsymbol{a_6}$	$\boldsymbol{a_5}$	$\boldsymbol{a_9}$	$\boldsymbol{a_{11}}$	$\boldsymbol{a_{10}}$
$\boldsymbol{ES_i}$	0	0	0	16	16	16	16	23	36	36	47
$\boldsymbol{LF_i}$	16	16	31	31	36	36	47	47	47	65	65
$\boldsymbol{slack_i}$	0	9	27	8	0	7	23	8	0	21	0

Table 2 Order of activities

[2] Workers are used in the whole paper but it can be exchanged for machines, groups of workers or employees.

The *workers number* (*W*) calculation uses the sum of activities and the critical path from the CPA. There are the simple assumptions about needs of workers for all activities of and as many workers will be needed as the sum of activities divided by the *project duration* (*T*). The project duration means the length of the critical path from the CPA.

In the calculation, the following rules are assumed:
1. each activity requires only one worker;
2. if there is an activity that can be started, no worker may be idle;
3. once a worker starts an activity, it must be continued by that worker until it is finished;
4. each worker can carry out any activity in the project.

The equation of workers number calculation will be $W = \sum_{i=1}^{n} a_i / T$.

In the example, the equation is $W = (16 + 7 + 4 + 7 + 20 + 13 + 8 + 16 + 11 + 8 + 18)/65 = 128/65 = 1{,}9692$; that means there will be two workers needed in the project.

3.2 Scheduling and matching workers with activities

This second part of method is about matching the workers with the activities through the algorithm. The algorithm calculates the *beginnings* (b_j) of activities and *idle times* (IT_j) for every worker j (j = 1, 2, … w). Algorithm uses the set of activities (N) from the first part of the method. The order of activities in the N is very important and it has to be strictly followed up.

Algorithm

These are the steps of the algorithm[3]:
1. set $b_j = 0$ for j = 1, 2, … w;
2. choose the first a_i from N (then a_i is removed from N), find *min* b_j and calculate new $b_j = max(ES_i; b_j) + t_i$ for j = 1, 2, … w; if $ES_i > b_j$ then $IT_j = ES_i - b_j$;
3. if N is empty, then it ends and if not repeat the step 3.

In the example:
- $b_1 = 0$ and $b_2 = 0$ and $N = (a_2, a_1, a_3, a_4, a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_2: $min(0; 0) = 0 \Rightarrow b_1 = max(0; 0) + 16 = 16$ and $N = (a_1, a_3, a_4, a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_1: $min(16; 0) = 0 \Rightarrow b_2 = max(0; 0) + 7 = 7$ and $N = (a_3, a_4, a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_3: $min(16; 7) = 0 \Rightarrow b_2 = max(0; 7) + 4 = 11$ and $N = (a_4, a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_4: $min(16; 11) = 0 \Rightarrow b_2 = max(16; 11) + 7 = 23$ and $N = (a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10})$, $IT_2 = 5$;
- for a_7: $min(16; 23) = 0 \Rightarrow b_1 = max(16; 16) + 20 = 36$ and $N = (a_8, a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_8: $min(36; 23) = 0 \Rightarrow b_2 = max(16; 23) + 13 = 36$ and $N = (a_6, a_5, a_9, a_{11}, a_{10})$;
- for a_6: $min(36; 36) = 0 \Rightarrow b_1 = max(16; 36) + 8 = 44$ and $N = (a_5, a_9, a_{11}, a_{10})$;
- for a_5: $min(44; 36) = 0 \Rightarrow b_2 = max(23; 36) + 16 = 52$ and $N = (a_9, a_{11}, a_{10})$;
- for a_9: $min(44; 52) = 0 \Rightarrow b_1 = max(36; 44) + 11 = 55$ and $N = (a_{11}, a_{10})$;
- for a_{11}: $min(55; 52) = 0 \Rightarrow b_2 = max(36; 52) + 8 = 60$ and $N = (a_{10})$;
- for a_{10}: $min(55; 60) = 0 \Rightarrow b_1 = max(47; 55) + 18 = 73$ and N is empty;

The calculation can be realized also in a simple way where every matching of activities for workers has the following form $b_j \mid a_i(t_i)$. Subsequently the calculated duration of project is maximal value of b_j and idle time is like a note inside of the calculation. This way of calculation for this example looks:

order of activities: $a_2, a_1, a_3, a_4, a_7, a_8, a_6, a_5, a_9, a_{11}, a_{10}$;

t_i (in order): $a_2 = 16$, $a_1 = 7$, $a_3 = 4$, $a_4 = 7$, $a_7 = 20$, $a_8 = 13$, $a_6 = 8$, $a_5 = 16$, $a_9 = 11$, $a_{11} = 8$, $a_{10} = 18$;

ES_i (in order): $a_2 = 0$, $a_1 = 0$, $a_3 = 0$, $a_4 = 16$, $a_7 = 16$, $a_8 = 16$, $a_6 = 16$, $a_5 = 23$, $a_9 = 36$, $a_{11} = 36$, $a_{10} = 47$;

worker 1: $0 \mid a_2(16)$; $16 \mid a_7(20)$; $36 \mid a_6(8)$; $44 \mid a_9(11)$; $55 \mid a_{10}(18)$; 73

worker 2: $0 \mid a_1(7)$; $7 \mid a_3(4)$; $IT_2 = 5$; $16 \mid a_4(7)$; $23 \mid a_8(13)$; $36 \mid a_5(16)$; $52 \mid a_{11}(8)$; 60

[3] The lexicographic decision rule is used in the case of same value.

The duration of project for two workers is 73 days. The sum of idle times of the worker 2 is 5 days and all the activities are done after 60 days (last $s_{wpi} + t_i$). The sum of work days is 133 for both workers.

This project can be carried out in 65 days if the worker 3 is taken into account and this worker will accomplish the activities a_5 and a_6 in 24 days. Other workers will accomplish remaining activities in the calculated order. In such a case the activities of worker 1 are critical and the sum of them is equal to 65 days. The worker 2 finishes activities after 44 days. The project duration will be shorter but the sum of work days is same.

On the left side of the *figure 1*, there is solution illustration of the example calculation with a timeline. The right side of the *figure 1* illustrates the optimal solution.

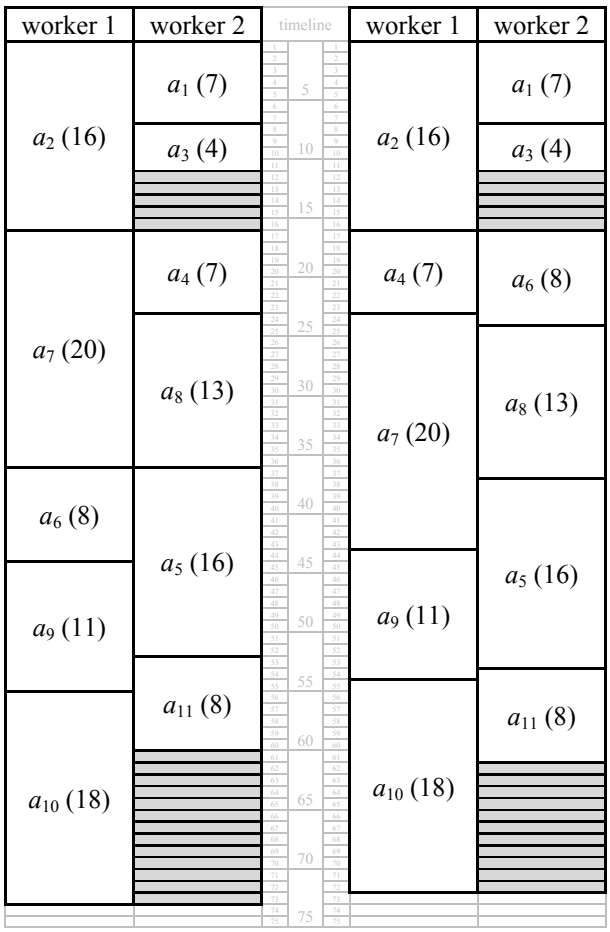


Figure 1 Solution

4 Conclusion

As it is shown in the comparison of solution obtained by this heuristic method and the optimal solution, there is a small difference. There were several examples in the testing of this method where the solution obtained by the heuristic method was same as the optimal solution and in the majority of examples they were close to the optimal solution. Base on this finding, the method can be used for basic scheduling.

The method has problems with some special characteristics in a project. There is the problem with the key activities which are known as the terms “bridges” or “articulation points" in graph theory. In case the project has this key activity (or more key activities) the graph of a project has to be divided into subprojects (subgraphs) and the method will be applied for each subgraphs.

Also the trivial subprojects in the project are problematic. The trivial subprojects are composed of activities which have only one predecessor and one successor. The activities with only one predecessor and successor activity can be merged into one activity and this “major” activity splits the project into two subprojects. In such a case the method can be used for both subprojects.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/0245/15 „Transportation planning focused on greenhouse gases emission reduction“.

References

[1] Brezina, I., Čičková, Z., Gežík, P.: *Sieťová analýza*. Ekonóm, Bratislava, 2012.
[2] Brezina, I., Pekár, J.: *Operačná analýza v podnikovej praxi*. Ekonóm, Bratislava, 2014.
[3] Conway, R. W., Maxwell, W. L, Miller, L. W.: Theroy of scheduling. Dover publications, New York, 2003.
[4] Čičková, Z., Brezina, I., Pekár, J.: Solving the real-life vehicle routing problem with time windows using self organizing migrating algorithm. In: Ekonomický časopis: časopis pre ekonomickú teóriu, hospodársku politiku, spoločensko-ekonomické prognózovanie = journal for economic theory, economic policy, social and economic forecasting. vol. 61. Ekonomický ústav SAV: Prognostický ústav SAV, Bratislava, 2013.
[5] Čičková, Z., Števo, S.: Flow shop scheduling using differential evolution. In: Management information systems: the International scientific journal, vol. 5, no. 2, p. 8-13. Subotica/Bratislava, 2010. <http://www.ef.uns.ac.rs/mis/archive.htm>.
[6] Líbal, V. et al.: *Organizace a řízení výroby*. Linde, Praha, 1999.
[7] Unčovský, L. et al.: *Modely sieťovej analýzy*. Alfa, Bratislava, 1991

Technical Equipment Scheduling Problem for Transport processes - Experimental Identification of Binary Linear Programming Benefits

Vojtěch Graf[1], Dušan Teichmann[2], Michal Dorda[3]

Abstract. Transport processes usually employ some technical equipment for its realizations; using the necessary technical equipment induces some costs. Let us consider that a transport process can be realized using different routes and the technical equipment can be transferred among destination and origin nodes of the defined transport processes. Let us consider that costs for every equipment transfer and costs produced by using the technical equipment are defined. The goal is to define the minimal number of technical facilities that are necessary to operate the defined transport processes. However, in practice we can often meet situations where the possible relocations of the facilities have zero benefit. Such situations usually occur if intervals between two subsequent beginnings of the process on the same route are short. The presented paper deals with an experimental definition of the situations for which the relocations of technical facilities do not yield any benefit. In our article we will restrict on the problems with two transport processes.

Keywords: Scheduling, linear programming, transport processes

JEL Classification: C61
AMS Classification: 90C05

1 Our Motivation

In industry or transport we often meet so-called transport processes that are realized using certain technical means or devices. Examples of such transport processes in industry are processes that secure transport of products among individual production stages. In transport we can mention for example processes that transport passengers or goods from their origins to destinations. For all the transport processes it is requested that utilization of the devices we need to realize the transport processes should be as high as possible.

One way how to fulfil the need for maximization of utilization is to minimize the number of the devices we use. In order to minimize the number of the necessary devices, we have to improve schedule of their operation. By better schedule we mean for example the possibility of using the same technical means on different transport routes or better balancing of requests for transport in time. It is more than obvious that basic requirements for such adaptation must be satisfied. Such basic requirements are as follows:

- Interoperability of all the technical devices (all the devices can be used for all the transport processes).
- It is possible to transfer the technical means among the individual transport routes.
- We can change times when the individual transport processes begin.

Such problems were successfully solved in the past using mathematical programming. The possibilities of technical device transfers among the transport routes were modelled using binary variables. However, approaches based on mathematical programming models that use binary variables may have an essential computational problem. Such problems usually occur in cases if the linear model consists of many binary variables. That is the reason why such problems are often solved by heuristic methods.

Practical pieces of experience with mathematical models reveal that results got by the models do not yield any positive effects in certain situations. That means the minimal number of the technical devices that serve the

[1] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, vojtech.graf@vsb.cz

[2] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, dusan.teichmann@vsb.cz

[3] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, michal.dorda@vsb.cz

transport routes can be found out using a known approach that is simpler than the approach based on mathematical programming. Therefore we think that it is necessary to try to estimate limits for which it holds that if the limits are exceeded, using mathematical models to minimize the number of the technical devices is useless.

The goal of the article is to compare experimentally results achieved by a mathematical model with results obtained by a simpler approach; the approach is discussed in Section 2. The experiment will be done for an example consisting of two transport routes.

2 State of the art

Scheduling of technical devices and equipment that secure transport processes by means of mathematical models was planned in the past. From many successful projects we can point out vehicle scheduling problem in bus public transport [1], [2] and [3]. The models were created either for a homogeneous vehicle fleet [2], [5] or for a heterogeneous vehicle fleet [6], [7]. The models offer results in the form of sequences of connections that should be served by individual vehicles of the given vehicle fleet. On the basis of these approaches the optimization system KASTOR was developed. The optimization program KASTOR, which was created in the 80's of the twentieth century, was employed to plan vehicle schedules in regional bus transport and its using yielded essential savings.

Scheduling based on the basis of above mentioned approaches was applied in other transport systems as well. For example paper [8] is focused on mathematical modelling of vehicle scheduling problem in the conditions of railway transport. Publications [9], [10], [11] apply mathematical modelling to solve scheduling problems in air transport.

To optimize planning of technical devices for transport processes heuristic methods [12] or dynamic programming [13] are often used.

The proposed mathematical models are general; that means they can be used in any conditions. However, the main assumption of their successful using is that the vehicles can transfer among the connections of the individual routes. We can also get a solution for which it holds that each vehicle serves only the connections of the single route. Everything is subordinated to the key criterion – to minimize the number of the vehicles we need to serve all the planned connections. The key criterion can be supported by a secondary criterion – to minimize the number of non-productive covered kilometers (the non-productive covered distances results from deadheading the vehicles among the routes).

The alternative approach to calculation how many vehicles we need is based on the assumption that each vehicle is assigned to the particular route in a given time period and its transfer to the connections of the different routes is not possible. In such cases we can calculate the number of the vehicles according to the following formula:

$$\sum_{j \in J} N_j = \sum_{j \in J} \left\lceil \frac{O_j}{L_j} \right\rceil, \tag{1}$$

where N_j is the number of the technical means we need to serve the transport route $j \in J$, O_j is the cycle time of the technical means that serves the transport process on the route $j \in J$, L_j is the headway between two consequent transport processes on the route $j \in J$.

Formula (1) results from the analogous formula given in publication [14] that was adapted for our problem. The formula is quite easy. Despite its simplicity the formula can be applied in practice, especially in cases when the headways between the subsequent connections are not too long. The square brackets on the left side of equation (1) express that the value of the fraction $\frac{O_j}{L_j}$ is rounded up to whole numbers.

3 Problem Formulation and Mathematical Model

Let us have a set of transport routes J; for the routes it holds that $|J| = n$. For each transport route $j \in J$ a set I_j of the transport processes that have to be realized in a regular interval L_j is given. For each transport process $i \in I_j$ its earliest possible beginning t_i, time duration T_i and maximal possible time shift b_i are defined. Our task is to create such sequence of the transport processes that are served by the same technical devices so that all the transport processes are served and the number of used technical devices is minimal. In our task we assume that all the technical devices are homogeneous from the point of view of their technical parameters and each technical device can be used for serving any transport process on any transport route. In addition we suppose that each transport process is served by single technical device.

Let us assume that it holds for our task that $\bigcup_{j\in J} I_j = I$, $\bigcap_{j\in J} I_j = \emptyset$ and $|I| = m$. In order to form the mathematical model, we must define an origin depot – it is a place where the technical devices are found before the beginning of the transport processes – and a destination depot – it is a place to where the transport devices return after serving all the planned transport processes. The origin depot has index 0 whilst the destination depot $m + 1$.

To simplify practical solving of the proposed models an incidence matrix A is used. Elements a_{ij} of the matrix take values 0 or 1, where $i \in I$ and $j \in J$. If $a_{ij} = 1$ then the transport process $i \in I$ is realized on the transport route $j \in J$. If $a_{ij} = 0$ then the transport process $i \in I$ is not realized on the transport route $j \in J$.

To solve the problem we must define two groups of variables. The first group of the variables is represented by the binary variables x_{ij}, where $i \in I \cup \{0\}$ and $j \in I \cup \{m + 1\}$. The binary variables model decisions about transfers of the technical devices to serve the transport process $j \in I$ after finishing the transport process $i \in I$. If $x_{ij} = 1$ after finishing the optimization experiment, then the transport device is transferred to serve the transport process $j \in I$ after serving the transport process $i \in I$. If $x_{ij} = 0$ after finishing the optimization process, then the transport device is not transferred to serve the transport process $j \in I$ after serving the transport process $i \in I$. If $x_{0j} = 1$, then the transport process $j \in I$ is served by the transport device that is found in the origin depot. If $x_{0j} = 0$, then the transport process $j \in I$ is not served by the transport device that is found in the origin depot. If $x_{im+1} = 1$ after finishing the optimization calculations, then the transport device goes to the destination depot after serving the transport process $i \in I$. If $x_{im+1} = 0$, then the transport device does not go to the destination depot after serving the transport process $i \in I$. The second group of the variables is formed by the non-negative variables z_i, where $i \in I$. The variables z_i model time shifts of the connections. Let us assume that the time shifts can be non-negative whole numbers from the interval $\langle 0; b_i)$.

The symbol *M* that is used in the mathematical model represents a so-called prohibitive constant; the symbol Z_0^+ expresses the set of non-negative whole numbers.

The mathematical model that we use to solve the problem has the following form:

$$\min f(x, z) = \sum_{i\in I} x_{0j} \tag{2}$$

subject to:

$$\sum_{i\in I\cup\{0\}} x_{ij} = 1 \qquad j \in I \tag{3}$$

$$\sum_{j\in I\cup\{m+1\}} x_{ij} = 1 \qquad i \in I \tag{4}$$

$$t_i + \sum_{k\in J} a_{ik}(z_k + T_k) \le t_j + \sum_{k\in J} a_{ik} z_k + M(1 - x_{ij}) \qquad i \in I \text{ and } j \in I \tag{5}$$

$$z_k \le L_k - 1 \qquad k \in J \tag{6}$$

$$x_{ij} \in \{0,1\} \qquad i \in I \cup \{0\},\ j \in I \cup \{m + 1\} \tag{7}$$

$$z_k \in Z_0^+ \qquad k \in J \tag{8}$$

Function (2) represents the optimization criterion – the number of the vehicles used in the task. The group of constraints (3) ensures that each transport process is served by the transport device. The group of constraints (4) ensures that each transport device which served the connection $i \in I$ is either transferred to serve the subsequent connection $j \in I$ or goes to the destination depot and its operation is terminated after serving the connection $i \in I$. The group of constraints (5) assures time admissibility of the sequence of the connections served by the same technical device. The left side of the constraint expresses when the transport process $i \in I$ is finished, whilst the right side of the constraint expresses when the transport process $j \in I$ begins. The group of constraints (6) must ensure for each transport process that its beginning falls into the given limits. The groups of constraints (7) and

(8) define domains of definition for both types of the variables used in the model. The model consists of $m^2 + 2m + n$ variables and $m^2 + 2m$ structural constraints.

4 Optimization Experiments

The experiments were realized in the conditions of two transport routes with the same origin place – see Figure 1.

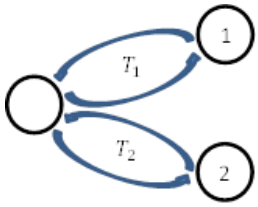


Figure 1 The routes used for the example

The experiments had two input parameters – the cycle time of serving the transport process on the transport route and the headway between two consecutive transport processes on the same transport route. It was given that 5 transport processes must be realized on each transport route. The initial value of the cycle time was equal to 40 minutes and the cycle time was gradually increased by 20 minutes up to 120 minutes. That means we used 5 different values of the cycle time – 40, 60, 80, 100 and 120 minutes. The initial headway was set to 10 minutes and the value was also gradually increased to 15, 20, 30, 40, 60 and 90 minutes. We assumed the same cycle time and headway for both transport routes.

In our experiment we compared the results got by two different approaches. The first approach presumes that the technical devices are permanently assigned to the transport routes. The minimal number of the devices is calculated using formula (1). The second approach is based on solving model (2) – (8). The mathematical model was solved using optimization software Xpress-IVE. The results we achieved by both approaches are presented in Tables 1 – 5.

In the left upper cell of each table the cycle time for both transport routes is shown. In the column below the cycle time the headways *L* are listed. Because we assumed the same cycle times and headways for both transport routes, the indexes of the transport routes are omitted. The calculated numbers of the technical means for the individual transport routes are listed in 4 columns. The columns entitled N_1 and N_2 show the number of the technical devices we need to serve all the planned connections on the transport routes 1 and 2; the values were computed using formula (1). One can notice that the corresponding values are equal in the columns; that is because we assumed the same cycle times and headway for both transport routes. The values in the next column are the sums of the values from the previous two columns. In the column $\sum_{j\in I} x_{0j}$ the numbers of the necessary technical devices obtained by solving model (2) – (8) are listed. The last two columns entitled z_1 and z_2 show the time shifts of the technical devices' departures on both transport routes. By shifting the departures it is possible to minimize the number of the technical devices.

$O = 40$ min	The number of the technical devices [-]				The time shift of the technical devices' departure [min]	
L (min)	N_1	N_2	$\sum_{j\in J} N_j$	$\sum_{j\in I} x_{0j}$	z_1	z_2
120	1	1	2	1	79	119
90	1	1	2	1	49	89
60	1	1	2	2	40	0
40	1	1	2	2	39	39
30	2	2	4	3	19	29
20	2	2	4	4	19	19
15	3	3	6	6	9	14
10	4	4	8	8	9	9

Table 1 Cycle time $O = 40$ minutes

$O = 60$ min	The number of the technical devices [-]	The time shift of the technical devices' departure [min]

L (min)	N_1	N_2	$\sum_{j\in J} N_j$	$\sum_{j\in I} x_{0j}$	z_1	z_2
120	1	1	2	1	59	119
90	1	1	2	2	60	0
60	1	1	2	2	59	59
40	2	2	4	3	19	39
30	2	2	4	4	29	29
20	3	3	6	6	19	19
15	4	4	8	8	14	14
10	6	6	12	10	0	9

Table 2 Cycle time $O = 60$ minutes

$O = 80$ min	The number of the technical devices [-]				The time shift of the technical devices' departure [min]	
L (min)	N_1	N_2	$\sum_{j\in J} N_j$	$\sum_{j\in I} x_{0j}$	z_1	z_2
120	1	1	2	2	80	0
90	1	1	2	2	80	0
60	2	2	4	3	39	59
40	2	2	4	4	39	39
30	3	3	6	6	19	29
20	4	4	8	8	19	19
15	6	6	12	10	0	14
10	8	8	16	10	0	9

Table 3 Cycle time $O = 80$ minutes

$O = 100$ min	The number of the technical devices [-]				The time shift of the technical devices' departure [min]	
L (min)	N_1	N_2	$\sum_{j\in J} N_j$	$\sum_{j\in I} x_{0j}$	z_1	z_2
120	1	1	2	2	100	0
90	2	2	4	3	79	89
60	2	2	4	4	19	56
40	3	3	6	5	19	39
30	4	4	8	7	19	29
20	5	5	10	10	0	19
15	7	7	14	10	0	14
10	10	10	20	10	0	9

Table 4 Cycle time $O = 100$ minutes

$O = 120$ min	The number of the technical devices [-]				The time shift of the technical devices' departure [min]	
L (min)	N_1	N_2	$\sum_{j\in J} N_j$	$\sum_{j\in I} x_{0j}$	z_1	z_2
120	1	1	2	2	119	119
90	2	2	4	3	59	89
60	2	2	4	4	59	59
40	3	3	6	6	39	39
30	4	4	8	8	29	29
20	6	6	12	10	0	19
15	8	8	16	10	0	14
10	12	12	24	10	0	5

Table 5 Cycle time $O = 120$ minutes

We can see in Tables 1 – 5 that the results obtained by model (2) – (8) are the same or better than the results got by using formula (1). Please note that we want to minimize the number of the necessary transport devices, therefore lower value means the better result. In the rows of the tables that are highlighted yellow the results achieved by the mathematical model are better than the results calculated using formula (1). We can also see that

with the increasing value of the cycle time O and the decreasing value of the headway L we got the inadmissible numbers of the technical devices when using formula (1). As written earlier in the text we assumed in the experiments that we must serve 10 transport processes. That means the maximal admissible number of the technical devices is equal to 10. However, in the rows that are highlighted red the results got by formula (1) correspond to more devices than the admissible maximum of 10 devices.

5 Conclusions

The article is focused on scheduling of the technical devices used for the transport processes. The goal of the schedule is to ensure the maximal utilization of the technical devices. We can fulfil the goal by minimization of the technical devices used to serve the planned connections. To determine the minimal number of the technical devices we can employ several approaches. In the paper we discussed and compared two of them. The first approach is based on the direct calculation of the number of the necessary technical devices for each transport route separately – using formula (1). The total number of the technical devices we need to serve all the connections operated on the transport routes we get as the sum of the numbers of the necessary technical devices for the individual transport routes. The second approach is based on solving mathematical model (2) – (8). The main goal of the article was to find experimentally the limits when the mathematical model does not yield better results. On the basis of the results achieved by both approaches we can observe some trends. If the fraction $\frac{O}{L}$ is not integer, we can observe savings in the necessary technical devices when the mathematical model is used. For example, the model gave better results for $\frac{O}{L} = 1{,}3\overline{3}$. On the other hand, if the fraction $\frac{O}{L}$ is integer, no savings were confirmed. For the value of the fraction $\frac{O}{L}$ for the particular route we can determine that the route can be left out of the optimization experiment. The technical devices that are assigned to the route cannot be used on other routes. We can formulate a hypothesis that if the fraction $\frac{O}{L}$ converges upwards to a whole number, the chance the mathematical model yields savings in the necessary technical devices decreases. In our future research we would like to examine the hypothesis more deeply.

References

[1] Černý, J., and Kluvánek, P.: *Základy matematickej teorie dopravy*. Alfa, Bratislava: 1990.

[2] Černý, J.: Optimalizační modely a metody pro oběhové rozvrhy v regionální dopravě. In: *Proceeding of International Conference Věda o dopravě*, , Faculty of Mechanical Engineering , ČVUT Praha, 2001, 54-59.

[3] Černá, A.: Optimalizace regionální autobusové dopravy. In: *Proceeding of International Conference Věda o dopravě*, Faculty of Mechanical Engineering , ČVUT Praha, 2001, 70-77.

[4] Černá, A., and Černý, J.: *Teorie řízení a rozhodování v dopravních systémech*. Jan Perner Transport Faculty, Pardubice, 2004.

[5] Palúch, S.: Bus scheduling as a graph coloring problem. *Communications* **5** (2003), 16-20.

[6] Palúch, S., Dubovec, J., and Hlačina, T.: Low-capacity busses in regional transport. In: *Proceeding of International Conference Vysoká škola jako facilitátor rozvoje společnosti a regionu*, European Polytechnic Institute, s.r.o., Kunovice, 2007, 224-228.

[7] Palúch, S., and Peško, Š.: Vehicle scheduling with heterogeneous bus fleet. In: *Proceeding of International Conference* of *Theoretical and Practical Issues in Transport*, University of Pardubice, Pardubice, 2010, 400-406.

[8] Brandalík, F.; Kluvánek, P.: *Operační analýza v železniční dopravě*. Alfa, Bratislava: 1984.

[9] Mercier, A., and Soumis, F.: An integrated aircraft routing, crew scheduling and flight retiming model. *Computers & Operations Research* **34 (**2007), 2251 – 2265.

[10] Bína, L., Bínová, H., Ploch, J., and Žihla, Z.: Provozování letecké dopravy a logistiky, CERN, Brno: 2014.

[11] Graf, V.: Optimalizace letového řádu charterového dopravce. VŠB-TU Ostrava, Ostrava: 2014.

[12] El Moudani, W., and Mora – Camino, F.: A dynamic approach for aircraft assignment and maintenance scheduling by airlines. *Journal of Air Transport Managament* **6** (2000), 233-237.

[13] Sriram, CH., and Haghani, A.: An optimization model for aircraft maintenance scheduling and reassignment. *Transportation Research Part A* **37 (**2003), 29-48

[14] Surovec, P.: Tvorba systému mestskej hromadnej dopravy, University of Žilina, Žilina: 1999.

Visualization of Bayesian managerial task solution by means of the tools of classical logic: classical versus adaptive approach

Simona Hašková[1]

The solutions of problems with conditional probability are based often on Bayes formula. In it the values of occurrences of examined phenomena in the form of prior and posterior probabilities are supplied. According to evolutionary biologists this approach is less psychologically acceptable for untrained individuals in mathematics, which results in many wrong solutions to problems of this type. This paper proposes alternative calculation procedures that are based on an adaptive approach. The classical probability formula and on it constructed model of Venn diagram (normative approach) will be transformed firstly into a frequency format, secondly into a probability decision tree (adaptive approach). Compared to the normative solution the usage of the adaptive approach facilitates comprehension by means of its visualization property and the possibility of checking the correctness of the solution. This is demonstrated in a Bayesian task of a managerial type. The adaptive procedure applied in it enables us to replace the original conditional probabilities by simpler Laplace probabilities that are easily understandable to managers without prejudice to the accuracy of the result.

Keywords: Bayesian inference, bounded rationality, probability decision tree, Venn diagram.

JEL Classification: C51, D81
AMS Classification: 62C10, 91B70, 91E10

1 Introduction

„Standard“ neoclassical economics examines the circumstances and consequences of the rational behaviour and decision making of individuals in terms of limited resources. The models are based on assumptions, which claim that individuals behave „economically“ in the sense of rational behaviour when evaluating the potential gains and losses resulting from the available options and that they are selfish and do not repeat their mistakes systematically. It also assumes that market forces operate that pull markets into a state of equilibrium, if not distorted e.g. by government regulations (Williamson, 1985).

In contrast, behavioural economists cannot see any obvious reason for the fact that an individual should be likened to a „homo economicus“ and why market forces should pull the economy into equilibrium. Their view is supported by empirical facts regarding the observed behaviour of individuals in real economy confirming that the assumption of the neoclassical economic theory concerning the ability of individuals to decide optimally is unrealistic. In their view decision-makers´ cognitive abilities are limited and described as „bounded rationality" (Simon, 1985).

The issue of bounded rationality has become a subject of the research of many scientists in the field of philosophy, economics and psychology. To the most important belong the research of Cohen (1992), Kahneman & Tversky (1996), Stein (1996), Gigerenzer (1996), Binmore (1999) and others, whose view of the decision-making process of individuals, or respectively their interpretation of empirical findings, has been the subject of many scientific discussions. The best known include the disputation between Tversky & Kahneman and Gigerenzer (Gigerenzer, 1996, Kahneman & Tversky, 1996) analysed in a number of scientific articles (e.g. Vranas, 1999; Polonioli, 2012), in which the authors present their arguments in favour of the opinion of one or other party. What is in principal going on? Generally there are two views on the rationality of individuals within their decision-making (Polonioli, 2012): a „*pessimistic*“ view, and an „*optimistic*“ view.

The pessimistic view is dated from 1969, when Kahneman and Tversky conducted a survey based on issues related to robust statistical estimates (Kahneman & Frederick, 2002). The results showed that most respondents believed that the law of large numbers is also applicable to small numbers, which had a negative impact on their final estimates. The misunderstanding of a small sample as the carrier of the entire population parameters on such a large scale led Kahneman & Tversky to the idea of heuristics of representativeness, which started

[1] VŠE Praha/Fakulta managementu, Jarošovská 1117/II, 377 01 Jindřichův Hradec, e-mail: haskovas@post.cz

research in the area of heuristics and distortion of judgments. Its results showed that individuals are prone to commit systematic errors in their judgments; the judgments based on the Bayesian inference being no exception.

The optimistic view is presented e.g. by G. Gigerenzer, who deems it unjustified to consider some forms of distortion described by Kahneman & Tversky as "errors" and "deception". Gigerenzer says that the reason for erroneous judgments may not be the „inability" of solvers to systematically not succumb to illusions, but to the fact that (Gigerenzer, 1998): "*If a human reasoning system enters an environment where statistical information is formatted differently from that encountered in the environment in which humans evolved, the reason algorithms may fail. But this does not imply that human minds are not built to reason the Bayesian way.*" And as he further says, the problem can be solved by the adaptation of mental algorithms to their environment.

In this article we follow and develop this Gigerenzer´s idea. The object of the paper is the issue of judgments based on conditional probabilities described by Bayes formula. In this context, we first analyse the normative approach presented by Kahneman & Tversky. We then interpret the abstract Bayesian formulas by means of the instrument of classical logic "Venn diagram", whose presentation performs one of the original contributions of this paper. The procedures of Bayesian inference and Venn diagram are followed by the adaptive approaches built on the adaptation of the Bayesian task towards the "evolutionary" capability of problem solvers. The first of them is based on Gigerenzer´s frequency approach derived from the problem presentation by Venn diagram, the second is based on the probabilistic decision tree. These diverse approaches are demonstrated in a managerial task of a probabilistic type.

2 Methodology

There are basically two approaches to a rational solution of specific tasks: one of them proceeds from "above" (the "normative" approach), the second from the "bottom" (the "case-based reasoning" approach). The process from "above" starts with the model of more general tasks (general formula), which is adapted to the particular problem by an appropriate choice of free parameters where necessary (Hašková & Zeman, 2014). The procedure from the "bottom", on the other hand, is built on the knowledge of the specifics of a particular task adequate to the chosen perspective on the problem. Contrary to the approach based on the general model, this approach can often find a *more acceptable* solution and / or a *more efficient* solution in terms of economy or user-friendliness.

The normative approach is presented in terms of Kahneman & Tversky idea of Bayesian inference (Griffin & Tversky, 1992) and by a visual model of Venn diagram, the construction of which is based on the known constructs described by formal logic by e.g. Shin (1994). The normative solution leans on the Venn diagram, which is then transformed firstly into a sequence of figures (frequency form (Meder & Gigerenzer, 2014)), secondly into a probabilistic decision tree in order to adapt the Bayesian task to economists and managers (the "case-based reasoning" approach).

3 The theoretical view: The normative approach to Bayesian problems

The estimates of the probability of phenomena are a normal part of economic life. Economic theory often relies on the so-called Bayesian statistics that view all uncertain phenomena as conditional probabilities. The Bayesian statistical analysis starts by assigning priori (initial) probabilities to the occurrences of the examined phenomena. Priori probabilities reflect what we think about these phenomena on the basis of our indigenous knowledge or subjective ideas, without knowing anything specific (e.g. Fleiss et al., 2003; Smith & Gelfand, 1992).

Let´s have some phenomenon (or respectively hypothesis) A with priori probability (or respectively verity) $P(A) > 0$. Then the subsequent awareness of the occurrence of some other phenomenon (or respectively information) B with the priori probability $P(B) > 0$, changes priori probability $P(A)$ to posterior (i.e. based on sensory experience) probability $P(A|B)$, which we call the conditional probability of event A if the event B occurred. In the same vein the $P(B|A)$ is the conditional probability of event B if phenomenon A occurred. If $A \wedge B$ means the simultaneous occurrence of events A and B, then this applies

$$P(A \wedge B) = P(B)P(A|B) = P(A)P(B|A) \tag{1}$$

From (1) it can be calculated (I) $P(A|B)$ by means of $P(A \wedge B)$ and $P(B)$, (II) $P(A \wedge B)$ by means of $P(B)$ and $P(A|B)$ and (III) $P(A|B)$ by means of $P(B|A)$, $P(A)$ and $P(B)$ – Bayesian formula in the form

$$P(A|B) = P(B|A)P(A)/P(B) \tag{2}$$

If the events A and B are independent then $P(A|B) = P(A)$ and $P(B|A) = P(B)$ – the conditional (posterior) and unconditional (priori) probabilities are equal. If event A is a partial case of event B, $P(A \wedge B) = P(A)$, and $P(A|B) = P(A)/P(B)$. If, however, A and B are mutually exclusive events, for which $P(A \wedge B) = 0$, then also $P(A|B) = P(B|A) = 0$.

Not to be overlooked, that the conditional probability represents the core of serious and often very heated debate between the so-called Bayesian statisticians, who are its proponents, and conventional (frequency) statisticians operating with classical instruments (to which, among other things p-values, permissible errors and significance levels belong), who are the opponents. Bayesian statisticians challenge the "frequencionists" due to "distortions" of their "logically inconsistent" ideas and frequency statisticians counterbalance with the fact that Bayesian inference requires determination of the priori probability expressing one´s faith before the experiment begins, while it is not clear at all how to actually choose these probabilities. More details of what Bayesian statistics are criticized for can be found in (Camerer & Loewenstein, 2004) and in (Saks & Uggerslev, 2010).

Among other things it firstly regards the claim that Bayesian updating requires separation of priori probabilistic inference from evaluation of new information, despite the fact that many of the cognitive procedures use priori information to interpret what is newly observed and thus violating a condition of separation; secondly, the claim that Bayesian updating does not predict the effects of the order of arrival of information that is common in memory because of the strength of new information (novelty effect) and due to the tendency to "test" the older information (primacy effect) – however, the way information is sorted shapes the resulting probabilistic judgment.

4 Visualization of the Bayesian formula by the Venn diagram: the normative approach

Inadequate consideration of the basic ratios in Bayesian tasks, according to Kahneman, reflects in the distorted judgments of problem-solvers, the consequence of which is a high rate of poor results. The Venn diagram, whose substantial benefit is the possibility of visual control when considering basic ratios, may help to reduce this deficiency. We explain this by means of a general approach to solving the problem in the field of management production.

4.1 The probabilistic managerial problem: Assignment

In the area of the management production, the manager needs to know the probabilities of possible causes of incurred deviations on the basis of which appropriate corrections can be made. The method of the continuous correction of deviations from the prescribed plan is being used. This closely relates to the following generally formulated task.

The event $\mathcal{U}$ occurred due to one of its possible causes. At the same time we know that:

- the imperfectly reliable indicia *I* is available (a sign or other relevant information), according to which *A* is the cause of event $\mathcal{U}$,
- the degree of indicia *I* reliability is a basic causal relationship *KP* (the probability of the occurrence of indicia *I* conditioned by the occurrence of cause *A*),
- basic statistical ratio *SP* is known, which is the priori share of cause *A* on the occurrence of event $\mathcal{U}$.

We ask: *What is the probability that the cause of event $\mathcal{U}$ is indeed the cause at which indicia I points?*

4.2 Solution: Bayesian´ rule application by means of the Venn diagram

In the case of the absence of the indicia we would consider *SP* as the probability of the cause *A*. On the other hand, $SP = 0.5$ has no informational value (the chance that *A* is either the cause or not, is equal and therefore cannot be qualifiedly decided only according to *SP*). Then conversely we would rely solely on the reliability of indicia (if it was available) and as the probability we would take *KP*. Both types of information can be connected with the Bayes rule as follows:

- $P(A) = SP$, or $P(\neg A) = 1 - SP$ the priori probability that *A* is, or is not the cause of the event $\mathcal{U}$,
- $P(I) \neq 0$ priori probability of the occurrence of indicia *I* within the occurrences of event $\mathcal{U}$,
- $P(I|A) = KP$, or $P(I|\neg A)$ posterior probability of the occurrence of indicia *I* conditioned by the occurrence of cause *A*, or by the occurrence of some cause other than *A*; $P(I|\neg A)$ is called a positive error of indicia *I* (*I* points at *A*, although *A* is not the cause of $\mathcal{U}$); generally $P(I|\neg A) = 1 - KP^*$, where KP^* is a degree of the reliability of indicia $\neg I$ that claims the opposite (*A* is not the cause of $\mathcal{U}$),
- $P(A|I)$ posterior probability of the occurrence of cause *A* conditioned by the occurrence of indicia *I*.

We ask about $P(A|I)$; from the relationship $P(A \wedge I) = P(A) \cdot P(I|A) = P(I) \cdot P(A|I)$ (where $P(A \wedge I)$ denotes the probability of the simultaneous occurrence of the cause A and indicia I within occurrences of event $\mathcal{U}$) then we get the Bayes formula $P(A|I) = P(I|A) \cdot P(A) / P(I)$. In it we only do not know the value of $P(I)$. To find it out the following Venn diagram shown in Fig. 1 will help us:

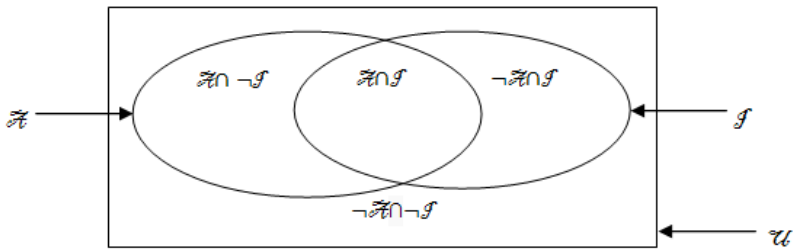


Figure 1 Venn diagram to the solution of the probabilistic managerial problem

In it, points inside the rectangle (the set $\mathcal{U}$ – universe of our consideration) represent occurrences of the referred event $\mathcal{U}$. The universe, by two ellipses drawn in it, is decomposed into the four designated mutually disjoint classes $\mathcal{A}\cap\neg\mathcal{I}$, $\mathcal{A}\cap\mathcal{I}$, $\neg\mathcal{A}\cap\mathcal{I}$ and $\neg\mathcal{A}\cap\neg\mathcal{I}$. Left ellipse (set $\mathcal{A}$) contains elements $\mathcal{U}$, whose cause is A (outside the ellipse, i.e. in the set $\neg\mathcal{A}$ are those elements of the universe $\mathcal{U}$, whose cause is not A). Right ellipse (set $\mathcal{I}$) comprises elements of $\mathcal{U}$, where indicia I is available (outside the ellipse, that is in the set $\neg\mathcal{I}$ are the elements of the universe $\mathcal{U}$ where I is not available). Symbols $\cap$ and $\neg$ occurring in the designation of decomposition classes are the operators of intersection of the respective sets and complement to the universe set. In the next the mentioned symbols will be supplemented by operator $\cup$ for unification of sets.

From Fig. 1 it is obvious that $\mathcal{I} = (\mathcal{A}\cap\mathcal{I})\cup(\neg\mathcal{A}\cap\mathcal{I})$ and therefore also $P(I) = P(A \wedge I) + P(\neg A \wedge I)$. By expressing the right side of the last expression by conditional probabilities, we get $P(I) = P(A) \cdot P(I/A) + P(\neg A) \cdot P(I|\neg A)$. Substituting the right side of the expression for $P(I)$ to the Bayes formula $P(A|I) = P(I|A) \cdot P(A) / P(I)$ we get:

$$P(A|I) = P(I|A) \cdot P(A) / (P(A) \cdot P(I|A) + P(\neg A) \cdot P(I|\neg A)) \tag{3}$$

and therefore it also applies

$$P(A|I) = SP \cdot KP / (SP \cdot KP + (1 - SP) \cdot (1 - KP^*)) \tag{4}$$

With help of Fig. 1 we can justify the validity of the assertion $P(I/\neg A) = 1 - KP^*$. From it is also apparent that $(\neg\mathcal{A}\cap\mathcal{I})\cup(\neg\mathcal{A}\cap\neg\mathcal{I}) = \neg\mathcal{A}$, therefore $P(\neg A) \cdot P(I|\neg A) + P(\neg A) \cdot P(\neg I|\neg A) = P(\neg A)$. Dividing both sides of the last equality by the expression $P(\neg A)$ we get $P(I|\neg A) + P(\neg I|\neg A) = 1$, from which it directly follows that $P(I|\neg A) = 1 - P(\neg I|\neg A) = 1 - KP^*$.

This general solution to the task has many specific applications in various fields (e.g. Mlodinow, 2009; Volf, 2014). The *benefit* of the Venn diagram is that it enables better understanding of the task, its control of correct solution and interpretation of the results.

5 Adaptive approaches to Bayesian-type problem solving

The Bayesian approach advocated e.g. by E. T. Jaynes in (2003) is, in terms of his theory, the extension of the classical Aristotelian logic to the case of statements of truth values that lie in the range between absolute truth and absolute untruth, which ranks it as fuzzy logic. His idea of the Bayesian approach is based on logical deduction. However, for the untrained individual in logic and mathematics this is less psychologically acceptable. This thesis is supported by evolutionary geneticists, who explain why it is so from their point of view (Ridley, 2003). This is because evolution equipped us with gathering and counting information obtained on the basis of case by case involving observation in the form of frequencies or gained from the analyses of scenarios assembled into a probabilistic decision tree rather than with formal statistical skills and deductive reasoning (Gigerenzer, 1998; Gigerenzer, 1996; Gigerenzer & Gaissmaier 2011).

5.1 The frequency approach to the Bayesian task

The adaptation of the Venn diagram into accessible vision to those who are not familiar with conditional probabilities is derived from Fig. 1 characterized by sets $\mathcal{A}\cap\neg\mathcal{I}$, $\mathcal{A}\cap\mathcal{I}$, $\neg\mathcal{A}\cap\mathcal{I}$ and $\neg\mathcal{A}\cap\neg\mathcal{I}$ with the number of

elements $|\mathcal{A}\cap\neg\mathcal{I}|$, $|\mathcal{A}\cap\mathcal{I}|$, $|\neg\mathcal{A}\cap\mathcal{I}|$ and $|\neg\mathcal{A}\cap\neg\mathcal{I}|$ (by the frequencies due to $|\mathcal{U}|$). This can be utilized for determination of the proper interpretation of the managerial task result obtained in the section 4.2.

Suppose that there is evidence *I* about which we know that its presence or absence tells us with 80 % reliability whether phenomenon *A* is or is not the cause of the considered event. Furthermore, we know that event *A* is the cause in 40 % of cases. According to task assignment in the section 4.1 we should estimate the probability *P(A|I)* that the phenomenon *A* is really the cause of the given event.

For the universe of 100 events a corresponding frequency schema in the form of the Venn diagram in Fig. 2 has the following form:

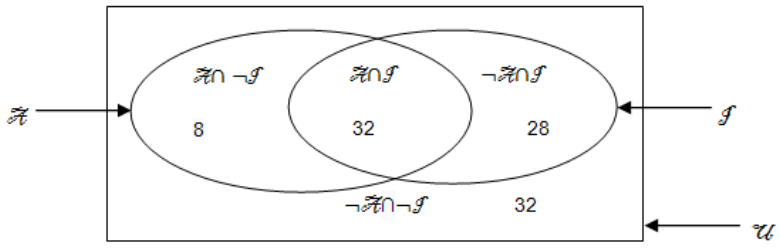


Figure 2 Frequency version of Venn diagram

From there it follows that the probability *P(A|I) = 32 / 60 = 0.53*.

5.2 The analysis of scenarios assembled into a probabilistic decision tree

The structure of the decision tree, whose purpose is (as in the case of the Venn diagram) to decompose the original set of objects into classes with different vectors of attribute values (e.g. the cause, the accuracy of indicia), corresponds in principle to the construction of the task in terms of different scenarios as shown in Fig. 3.

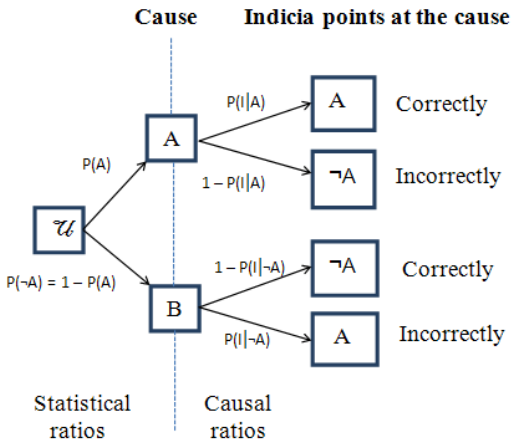


Figure 3 The probabilistic decision tree of adaptive approach

The leaves of the tree in Fig. 3 show all possibilities (correct and incorrect) of indicia references; each branch leading from the root to a leaf is some of the possible scenario. The probability of occurrence of each of them is given by the product of the edge evaluations of the branch. The leaves are mutually incompatible; their probabilities of occurrence can be, therefore, added. The answer to the question "*What is the probability that the cause of the event is indeed the cause at which indicia I points?*" is given as the ratio of probability of the correct reference of relevant indicia to the given cause to the probability of all references to it. Thus:

$$P(A|I) = p(A) \cdot p(I|A) / (p(A) \cdot p(I|A) + (1 - p(A)) \cdot P(I|\neg A)) \tag{5}$$

and also

$$P(\neg A|I) = p(\neg A) \cdot (1 - p(I|\neg A)) / (p(\neg A) \cdot (1 - p(I|\neg A)) + (1 - p(\neg A)) \cdot (1 - p(I|A))) \tag{6}$$

In this approach we avoid the pitfalls associated with complex reasoning about the priori and posterior probabilities of Bayesian formula and reach the solution in a more feasible way.

6 Discussion and conclusion

The conditional probability problems are the subject of many scientific analyses. The majority notice the descriptive side of things, which they interpret; the minority put forward proposals, which should lead to a reduction of errors that often occur in the solutions of these tasks. From the perspective of a normative approach to Bayesian task solving, the poor results of solvers are due to their inclination to various forms of distortion of judgment and the lack of effective comprehension of the Bayesian procedures. To avoid this requires an initial total understanding of a given problem. We propose the strategy for presenting the task in a more comprehensible way for economists and managers by means of two adaptive ways to the solution – by means of the Venn diagram and by the probability decision tree. The former corresponds to the task solution in the frequency format; the latter calculates the result from adequate scenarios. These tools facilitate the cognitive processes of understanding the Bayesian tasks to economists and managers and allow the feed-back control of the solution reached.

Acknowledgements

The creation of this contribution and its presentation at the conference MME 2015 was supported by the grant IGS, number FM/89/2015.

References

Camerer, C. & Loewenstein, G. (2004) *Behavioral economics: Past, present, future*. Princeton: Princeton University Press.

Cohen, J. (1992). A power primer. *Psychological bulletin*, Vol. 112(1), 155.

Fleiss, J.L., Levin, B., Paik, M.C. (2003) *Statistical methods for rates and proportions*. John Wiley & Son. Canada.

Gigerenzer, G. (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky. In: *Psychological Review*, Vol 103(3), 592-596.

Gigerenzer, G. (1998) Ecological intelligence: An adaptation for frequencies. In: *The evolution of mind*, 9-29.

Gigerenzer, G. & Gaissmaier, W. (2011) Heuristic Decision Making. In: *Annual Review of Psychology*, Vol. 62: 451-482.

Griffin, D. & Tversky, A. (1992) The weighing of evidence and the determinants of confidence. In: *Cognitive Psychology*, Vol. 24, Issue 3, 411–435.

Hašková, S. & Zeman, R. (2014) Analytical microeconomics as a search tool for the stock optimization. *Proceedings of the 40th International Conference on Mathematical Methods in Economics „MME 2014“*. Palacký University in Olomouc, 269-274.

Jaynes, E.T. (2003) *Probability Theory - The Logic of Science*. Cambridge University Press, Cambridge.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49-81.

Kahneman, D. & Tversky, A. (1996) On the reality of cognitive illusions. In: *Psychological Review*, Vol 103(3), 582-591.

Meder, B. & Gigerenzer, G. (2014) Statistical Thinking: No One Left Behind Probabilistic Thinking. In: *Advances in Mathematics Education*, 127-148.

Mlodinow, L. (2009). *Život je jen náhoda: jak náhoda ovlivňuje naše životy*. Slovart.

Polonioli, A. (2012). Gigerenzer's External Validity Argument against the Heuristics and Biases Project: An Assessment. In: *Mind & Society*, 11 (2), 133-148.

Ridley, M. (2003). *Nature via nurture: Genes, experience, and what makes us human*. HarperCollins Publishers.

Saks, A. M., & Uggerslev, K. L. (2010). Sequential and combined effects of recruitment information on applicant reactions. *Journal of Business and Psychology*, 25(3), 351-365.

Simon, H.A. (1985) Human Nature in Politics: The Dialogue of Psychology with Political Science. In: *The American Political Science Review*, Vol. 79, No. 2., 293-304.

Shin, Sun-Joo (1994) *The logical status of diagrams*. Cambridge University Press.

Smith, A.F.M. & Gelfand, A.E. (1992) Bayesian Statistics without Tears: A Sampling–Resampling Perspective. In: *The American Statistician*. Vol. 46, Iss. 2.

Volf, P. (2014) On precision of stochastic optimization based on estimates from censored data. In: Kybernetika, Vol. 50, Iss. 3, 297-309.

Vranas, P.B.M. (1999) Gigerenzer's normative critique of Kahneman and Tversky. In: *Cognition 76*,179±193.

Williamson, O.E. (1985) *The economic intstitutions of capitalism*. The Free Press, A Division of Macmillan, Inc., N.Y.

Gender Wage Differences in V4 Countries: An Analysis Using Intra-Household Specialization

Veronika Hedija[1]

Abstract. The study is devoted to wage differences between man and women in V4 countries. It aims to estimate the gender pay gap cleaned at least partially by effect of intra household specialization on wage. Taking assumption that the child- and family-care take the partner who earns less, we estimate the unexplained gender pay gap using subsample of employees earning more than their partners using EU-SILC data. We use the average treatment effect on the treated estimation to calculate the unexplained part of gender pay gap and matching procedure to get more homogenous sample of employees in each of V4 countries.
We concluded that the estimated unexplained wage differences are similar in V4 countries. The average treatment effect on the treated amounts approximately -0.12. Women earning more than their partners reach 12 percent lower wage in comparison with men in V4 countries.

Keywords: average treatment effect on the treated, gender pay gap, intra-household specialization, labor market, matching.

JEL Classification: J24, J31, J71, M5
AMS Classification: 91G70

1 Introduction

Large number of studies are focused on the influence of the components of intellectual capital on firm performance (for instance Youndt and Snell [13]; Žižlavský and Senichev [14] or Fiala and Borůvková [4]). According to Mertins and Will [7], human capital is the most important component of intellectual capital. People and their creative abilities in using other sources – material and financial – are the decisive factor of each company´s development, both in the entrepreneurial sector and in the area of public sector. For that reason companies should pay attention to issues connected with raising qualification, teambuilding with the emphasis on the flexibility of workers and their high work enthusiasm and loyalty to the employer. (Měrtlová and Dostálová [8]) In this context many questions arise. Is this true even for women? What is the position of women in contemporary society with regard to their employment and the level of earnings? And are there important differences between individual countries?

The existence of wage gap between men and women is a well-known fact. Empirical studies show that part of the existing gender wage differences can be attributed to differences in the characteristics of men and women, especially differences in average company related characteristics such as occupation and sector (Plantenga et al. [10]). However, these factors provide an explanation of only a part of the gender pay gap, the rest stays unexplained.

One of the reasons of women´s lower wage can be the motherhood and the traditional role of women in caring for household and family. A number of empirical studies conclude that working mothers earn less than women with no children. There are more theoretical arguments for existence of family gap: different abilities and preferences of women with children, limited mobility, lower accumulated human capital and discrimination against mothers. (Felfe [3]) Caring for the family and children may not only be the cause of a lower wage for women, it may also influence the earnings of men. The studies show that the family penalty for women ranges from 10 to 15 percent. On the other hand, marriage and a family increase the wage of men by 10-15 percent. (Waldfogel [11])

To cleanse the gender pay gap at least partially of the effect of intra-household specialization on productivity, Hedija [5] estimate the gender pay gap using subsample of employees earning more than their partners assuming that the larger part of care for the household and children is taken up by the partner earning less. Using the European Union Statistics on Income and Living Conditions (EU-SILC) data for 19 member countries of the

[1] Vysoká škola polytechnická Jihlava, katedra ekonomických studií, Tolstého 16, 586 01 Jihlava, veronika.hedija@vspj.cz

European Union, she concludes that the unexplained part of gender pay gap amounts approximately 10 percent working to the disadvantage of women. We follow this approach.

The aim of this study is to estimate the gender pay gap in V4 countries cleansed at least partially of the effect of intra-household specialization on productivity. To achieve this, we use a subsample of employees earning more than their partners assuming that the larger part of care for the household and children is taken up by the partner earning less.

The V4 countries are a culturally, historically and geographically very closed. One might therefore expect that wage differentials are similar in these countries. Nevertheless, the empirical studies show that the raw gender pay gap and its unexplained part varies significantly among V4 countries (for example Christofides et al. [2]; Plantenga et al. [10]; Mysíková [9]). Unlike these studies, we estimate the gender wage differences using the subsample of employees earning more than partners and compare the results with conclusions of these studies.

2 Methods and data

We use data from EU-SILC which contains multidimensional micro-data on income, poverty, social exclusion and living conditions. We use cross-sectional data for 2010, which can be found in EU-SILC 2011 and covers data from 30 European countries. Our study is based on EU-SILC data for 4 member states of the European Union: Czech Republic, Hungary, Poland and Slovakia.

Unfortunately, EU-SILC data does not contain information on hourly gross wages. It was therefore necessary to narrow down the sample to be able calculate the hourly gross wage using the available data. We narrowed down the reference population sample to persons who: were employees during the reference period, worked all twelve months in a full-time job, had no other jobs and earned an income. We excluded the self-employed, as we are interested in wages and the potential different evaluation of male and female employees by the employer.

We use data based on the selected personal and company characteristics of the employee: education level (highest attained education level according to ISCED-97), sickness (temporary inability to work due to sickness during the income reference year), partnership (having a partner living in the same household), dependent children (having at least one economic inactive child under 24 years), occupation (according to the classification ISCO-88), sector (economic activity using the classification NACE Rev.2), company size (number of persons working at the local unit), contract (having a work contract of limited duration), managerial position (having a formal responsibility supervising a group of other employees), age, gross hourly wage and gross hourly wage of the partner. The gross hourly wage is calculated as the employee's cash and non-cash incomes per year divided by the number of hours usually worked per year (including overtime).

To illustrate the gender wage differences in the different group of population, we work with three samples of employees: full sample of employees (with and without partner), employees having partner (living in the some household) and employees earning more than their partner. Figure 1 shows the raw gender pay gap in V4 countries calculated as the difference in logarithm of average female and male gross hourly wage.

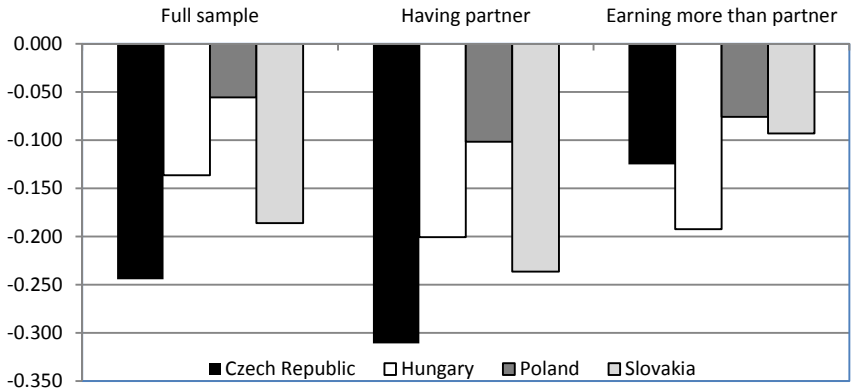


Figure 1 Raw gender pay gap

We can see that the wage difference between men and women earning more than their partner was smaller as compared with group of employees with partner in all V4 countries. This decline could be at least partly attributed to intra-household division of labor. However above data is very raw and rather indicative. The decline of wage disparity or its part could be also due to better average characteristics of women as compared with men in this sample. To adjust the raw gender pay gap for differences that can be explained by variation in personal and firm characteristics of men and women, we will work with unexplained gender pay gap.

There are several methods to estimate the unexplained part of gender pay gap. The most common includes Oaxaca-Blinder decomposition or an estimate of wage function using linear regression model. Like Hedija [5], we use here an estimate of the average treatment effect on the treated which is another alternative to estimate the unexplained part of the gender pay gap. *"The average treatment effect on the treated is the mean effect for those who actually participated in the programme"*. (Wooldridge [12], p. 605) In our case, the ATT will be the mean effect for women in the form of a lower wage resulting from being a woman.

We use this formula for the calculation of the average treatment effect on the treated

$$ATT = E(y_i(1) - y_i(0)|T_i = 1). \quad (1)$$

Where T is the binary treatment indicator, T = 1 denotes treatment and T = 0 otherwise, y(1) is the potential outcome with treatment and y(0) is the potential outcome without treatment. In our case, to be treated means to be a woman. We can rewrite the ATT as

$$ATT = E(y_i(1)|T_i = 1) - E(y_i(0)|T_i = 1). \quad (2)$$

Where ATT represents the gender pay gap, which cannot be explained by the different characteristics of men and women. The term $E(y_i(1)|T_i = 1)$ is the sample average of the logarithm of the gross wage of women and the term $E(y_i(0)|T_i = 1)$ is the sample average of the logarithm of the gross wage of women, should they be men. From our sample we know the first term on the right-side of equation 2, which is the sample average of the logarithm of the gross hourly wage of women. The second term, the average of the logarithm of a woman's gross hourly wage if she were a man, needs to be somehow estimated. There are more ways to estimate this. For more details see Wooldridge [12] and Ho et al. [6]. We estimate this using the regression model.

First, we estimate the coefficients of the wage function of men

$$(y_i|T_i = 0) = \beta_0 . X_i + u_i \ , \ i=1, \ldots, n_m. \quad (3)$$

Where, y_i is the logarithm of the male gross hourly wage, β_0 is the vector of the coefficients of the wage function, X is the vector of the chosen observed characteristics of men and u is a disturbance term. As explanatory variables we use age, age squared, education level, sickness, dependent children, occupation, sector, company size, contract and managerial position.[2] We estimate the coefficients of wage function using OLS estimator robust to disturbance being heteroskedastic.[3]

We then use the estimated coefficients of the male wage function to calculate the average of the logarithm of the gross hourly wage of women, should they be men

$$E(y_i(0)|T_i = 1) = E(\beta_0 . X_i), \quad i=1, \ldots, n_f \quad (4)$$

Where $E(\beta_0 . X_i)$ is the mean of the predicted wages (the logarithm of the gross hourly wage) of every woman in the sample.

Finally, we estimate the average treatment effect on the treated as the difference between the average of the logarithm of the gross hourly wage of women and the average of the predicted values of the wages calculated from the male wage function

$$ATT = E(y_i(1)|T_i = 1) - E(\beta_0 . X_i|T_i = 1). \quad (5)$$

The estimated ATT refers to the amount of the gender wage gap that cannot be explained by differences in observed characteristics of men and women. In our case, differences in age, education level, sickness, having dependent children, occupation, sector, company size, type of contract and managerial position. The part which stays unexplained could not only the wage discrimination against women but also the penalty for responsibility for the household and child-care women. We assume that the partner who earns less takes on a larger responsibility for the household and child-care. We estimate the ATT using the subsample of individuals, who earn more than their partners. This enables us to detect the unexplained part of the gender pay gap, which is at least partially adjusted for the care for the household and children. To be able to make a comparison, we also estimate the ATT for the entire sample of men and women.

[2] We use age as indicator of work experience. The data that indicates the number of years spent in paid work are not available for Hungary.

[3] In the case of subsample after matching procedure (CEM sample), we use WLS estimator with weights from coarsen exact matching procedure and robust to disturbance being heteroskedastic (more Blackwell et al. [1])

To compare the gender wage differences among V4 countries, we estimate the ATT for each V4 country. To minimalize potential bias from too different samples of employees in individual countries, we use matching as a preprocessing procedure to get more homogenous sample of employees for each country. We use coarsened exact matching (CEM), which performs exact matching on coarsened data. It coarsens variables into groups and goes on to exclude untreated units, whose coarsened characteristics do not match with any treated units and vice versa. Finally, it returns uncoarsened data from observations that were matched. (Blackwell et al. [1]) The matching procedure, in our case coarsened exact matching, enables us to create subsamples of employees for every country, these being as homogeneous as possible.

3 Empirical results

We use equation 5 and estimate the average treatment effect on the treated gradually for all V4 countries (Czech Republic, Hungary, Poland and Slovakia). The results and the number of men and women in the sample are shown in the table 2. The aim is to calculate the ATT for the subsample of employees which earns more than their partners. However for the purpose of comparison we also estimate this for the full sample of employees and for the subsample of men and women having partners.

The results show that the ATT which reflects the unexplained part of the gender pay gap differs significantly among V4 countries. Using full sample of employees, the unexplained gender pay gap ranges from -0.196 in the Czech Republic to -0.126 in Hungary.

The legitimate reason for a lower female wage can be attributed to the intra-household specialization implying the dominant role of women in caring for the household and family and men in financially providing for the family. To minimize this effect we calculate the ATT for the subsample of men and women earning more than their partners expecting that the household and family care lies on a partner who earns less. Using this subsample, the differences in unexplained pay gap decrease. The ATT ranges from -0.101 in the Czech Republic to -0.153 in Hungary. The ATT declines significantly in the Czech Republic and Slovakia comparing to sample of full employees. On the other hand it increases in Hungary and Poland. The reason for the growth of unexplained gender pay gap may be difference in composition of samples for individual countries and the fact that wage differences between men and women without partners are significantly lower. For this reason, we also estimated ATT taking the subsample of men and women who have a partner, where the intra-household division of labour plays a role. When we compare the unexplained gender pay gap estimated for subsample of employees having partners and subsample of employees earning more than their partners, we can conclude that the ATT is lower in the second case in all V4 countries except Hungary. Here the ATT amounts -0.153 and estimated wage difference is paradoxically the highest of the V4 countries.

	Full sample	**Having partner**	**Earning more than their partner**	**Earning more than partner (CEM sample)**
Czech Republic	-0.196***	-0.234***	-0.101***	-0.123***
	(0.013)	(0.015)	(0.023)	(0.045)
Men	3492	2434	1306	833
Women	2825	1942	407	174
Hungary	-0.126***	-0.132***	-0.153***	-0.117
	(0.013)	(0.015)	(0.024)	(0.073)
Men	3604	2604	1301	740
Women	3732	2447	643	228
Poland	-0.134***	-0.163***	-0.137***	-0.128**
	(0.014)	(0.016)	(0.024)	(0.060)
Men	3957	3056	1252	744
Women	3456	2528	848	238
Slovakia	-0.165***	-0.185***	-0.115***	-0.105**
	(0.013)	(0.016)	(0.026)	(0.047)
Men	2616	1791	1017	649
Women	2607	1707	388	169

Table 2 Average treatment effect on the treated

***significant at the 1 per cent level, **significant at the 5 per cent level, *significant at the 10 per cent level, robust standard errors in brackets.

We conclude that the women earning more than their partner obtain approximately 10 percent lower wage compared to they were the men in the Czech Republic, 11.5 percent in the Slovakia, 13.7 percent in Poland and 15.3 percent in Hungary. Though, estimated ATT is limited comparable between individual countries. The reason is the fact that results might be biased in some extent due to the differences in final subsamples in each of V4 country. To get more comparable sample of employees in each country, we use coarsen exact matching as preprocessing procedure and apply it to subsample of employees earning more than their partner. We match the countries using only key personal and company observed characteristics: age, education, occupation (1digit) and sector.[4] The sample after matching procedure is referred to as CEM sample. After matching we get narrowed but more homogenous sample of employees in individual countries. Then we estimate the ATT in each of V4 countries. The results are shown in fifth column of the Table 2. The differences in ATT have diminished among V4 countries and the unexplained gender pay gap amounts around 12 percent to the disadvantage of women. These findings are very similar to Hedija [5]. She identifies the unexplained gender pay gap amounting approximately 10 percent using the sample of employees earning more than their partner and data of 19 the European Union member countries.

4 Conclusion

The existence of wage disparity between men and women is well known fact. Part of these differences can be explained by difference in personal and firm characteristics of men and women, the rest stays unexplained. One of the reasons for lower female wage could be intra-household division of labour and dominant role of women in child- and family-care. The aim of this study is to estimate the gender pay gap in V4 countries cleansed at least partially of the effect of intra-household specialization on productivity. To achieve this, we use a subsample of employees earning more than their partners assuming that the larger part of care for the household and children is taken up by the partner earning less.

After adjusting the raw gender pay gap for differences in observed characteristics of men and women, we can conclude that the unexplained gender pay gap vary in individual V4 countries. The women earning more than their partner obtain approximately 10 percent lower wage compared to men in the Czech Republic, 11.5 percent in the Slovakia, 13.7 percent in Poland and 15.3 percent in Hungary. The wage disparity is lower by comparison with sample of employees having partners in all V4 countries except Hungary.

However comparison of gender wage differences among individual V4 countries could be limited due to different composition of samples in individual countries. We find out that the differences in unexplained gender pay gap among V4 countries are reduced using more comparable sample of employees. The ATT amounts approximately -0.12. Women earning more than their partners reach 12 percent lower wage in comparison with men in V4 countries.

Under assumption, that the dominant role in family- and child-care is taken up by the partner earnings a lower wage, then this difference could neither be explained by differences in the observed personal and company characteristics nor by the dominant role of women in care for the household and children and could actually be attributed to wage discrimination against women (Hedija [5]).

Acknowledgements

This paper has been supported by the internal grant „The Relationship Among Firm Size, Firm Growth and Gender Characteristics“ of College of Polytechnics Jihlava.

References

[1] Blackwell, M., Iacus, S., King, G. and Porro, G.: CEM: Coarsened exact matching in Stata. *The Stata Journal* **9** (2009), 524-546.
[2] Christofides, L.N., Polycarpou, A. and Vrachimis, K.: Gender wage gaps, ‘sticky floors’ and ‘glass ceilings’ in Europe. *Labour Economics* **21**(2013), 86–102.
[3] Felfe, C.: The motherhood wage gap: What about job amenities? *Labour Economics* **19** (2012), 59-67.
[4] Fiala, R., Borůvková, J.:. Relationship between Organizational Capital and Firm Performance. In: *Finance and the Performance of Firms in Science, Education, and Practice* (Jircikova, E.; Pastuszkova, E.; Svoboda, J., eds.). Tomas Bata University in Zlín, 2011, 88–95.
[5] Hedija, V.: Do Women Really Face Wage Discrimination on the Labor Market? An Analysis Using Intra-household Specialization. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **62** (2014), 1279-1286.

[4] The variable “age” is coarsened into nine conventional groups (<20, 20-24, 25-29, 30-34, 35-39, 40-44, 45-50, 50-54, 55-60 and 60+). The variables: education, occupation and sector are matched exactly.

[6] Ho, D. E., Imai, K., King, G., Stuart, E.A.: Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* **15** (2007), 199-236.
[7] Mertins, K., Will, M.: Strategic Relevance of Intellectual Capital in European SMEs and Sectoral Differences. InCaS: Intellectual Capital Statement – Made in Europe. In*: 9th European Conference on Knowledge Management* (Harorimana, D.; Watkins, D., eds.). Southampton: Southampton Solent University, 2008, 11-122.
[8] Měrtlová, L., Dostálová, Z.: Human Resources in Public Administration. In: *Proceedings of the 16th International Colloquium on Regional Sciences* (Klimova, V.; Zitek, V., eds.). Masaryk university Brno, 2013, 466-473.
[9] Mysíková, M.: Gender Wage Gap in the Czech Republic and Central European Countries. *Prague Economic Papers* **2012/3** (2012), 328-346.
[10] Plantenga, J., Remery, C. et al.: The Gender Pay Gap. *Origins and Policy Responses. A Comparative Review of Thirty European Countries.* Office for Official Publications of the European Communities, Luxembourg, 2007.
[11] Waldfogel, J.: Understanding the ‘‘Family Gap’’ in Pay for Women with Children. *Journal of Economic Perspectives* **12** (1998), 137–156.
[12] Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 2002.
[13] Youndt, M. A., Snell, S. A.: Human Resource Configurations, Intellectual Capital and Organizational Performance. *Journal of Managerial Issues* **16** (2004), 337–360.
[14] Žižlavský, O., Senichev, V.: The Role of Human Resource Management and Group Factors in Innovation Process. In: *Vision 2020: Innovation, Development Sustainability, and Economic Growth* (Soliman, K.S. ed). IBIMA. 2013, 679-689.

Analysis of Business Accounts in the Czech Republic

Martina Hedvicakova[1], Alena Pozdilkova[2]

Abstract. Due to low interest rates and growing competition from smaller "low-cost" banks, banks are constantly looking for opportunities to increase their revenues and maintain or gain new clients. The highest bank fees for bank accounts are generated by business accounts. For this reason, Banks are looking for new clients via packages and tailor-made services. The target group also includes entrepreneurs, i.e. legal entities as well as individuals. The aim of the paper is a comparison of the seven business accounts offered intended for entrepreneurs and small businesses with no restriction on the type of the client and the company's turnover. The comparison of various business accounts is performed by the height of the bank fees, benefits offered, availability of ATM's and quality of smartbanking using multi-criteria evaluation of the alternatives.
Multi-criteria analysis uses two types of profiles, small and medium entrepreneurs. Given the uncertain boundary between defined types the decision under uncertainty and fuzzy set theory will be used, which allows any client to adapt the results to their own needs.

Keywords: business account, cost, multi-criteria analysis, fuzzy set, uncertainty

JEL Classification: G21 C58, C610
AMS Classification: 91B06, 62C86

1 Introduction

Similarly to personal accounts, there is also high competition in the banking market regarding business accounts. With the entry of new, mostly "low-cost" banks to the market, it is possible to obtain business accounts free of charge (e.g. FIO bank, mBank, Equa bank). It depends on the client whether they choose a basic account, which is free of charge or pay a monthly fee and have all the operations within the package for free. They can also receive various benefits free of charge.

Management of a business account as well as ordinary bank account can be free under certain conditions. For example, Komerční banka offers management of a business account as well as a personal account for free (instead of 169 CZK/month) when at least two of the following activities are met: a regular payment must come to the account, an entrepreneur has at least 100,000 crowns saved in the KB group, has a term or saving account in the bank, uses a credit card, has a loan with the bank.

Banks target their offer at the most business entities operating on the Czech market. They are offering favourable accounts for people who do business under a special law and have no trade licence, such as physicians, lawyers and freelancers.

2 Advantages and disadvantages of business accounts

Self-employed individuals do not have to mandatory set up a special business account in a bank, but in many ways it might be preferable for them. The biggest advantage is the strict separation of personal finance from corporate finance, which helps the most during checks from the tax office. It also eliminates the administration with documentation of individual items, where the money came from, or on what purpose it was used.

The second advantage is easier access to a bank loan. Some banks (e.g. Česká spořitelna and GE Money Bank) offer an overdraft or authorised overdraft upon the initial set-up of a business account.

Another advantage is linking a business account directly with accounting software. Companies can also save on taxes because the cost of business accounts is tax deductible.

Some banks (such as Citibank, Raiffeisenbank) offer free extra assistance and legal services along with a business account.

[1] University of Hradec Kralove, Faculty of Informatics and Management, Department of Economy, Rokitankeho 62, 500 03 Hradec Kralove 3, Czech Republic, e-mail: martina.hedvicakova@uhk.cz
[2] University of Pardubice, Faculty of Electrotechnics and Informatics, Department of Mathematics and Physics, Studentská 95, 532 10 Pardubice 2, Czech Republic, e-mail: alena.pozdilkova@upce.cz

Another additional service is, for example, better interest-bearing savings account (e.g. Equa bank, LBBW Bank, Raiffeisenbank and Sberbank).

When opening a business account some banks offer personal account management free of charge or on more favourable terms. Other benefits, which can be obtained similarly to personal accounts include: the same bank account number (just a different bank code) when changing bank, or the possibility to choose your own account number, selection of a picture on credit cards, etc.

The main disadvantages are the particularly high costs for management of a business account or a poorly chosen services and products package. If an entrepreneur doesn't know the number of their monthly transactions, what additional services they will use, how many operations will be in foreign currencies, etc., their business account management can be more expensive.

3 Objective and methodology

This paper deals with the analysis of accounts for entrepreneurs and small businesses, without limitation of client type and the amount of annual turnover. These business accounts are provided in the Czech Republic by 12 major banks, which will be discussed in the following analysis, and other less well-known institutions, while the majority of these banks provide multiple products. Most banks have their products differentiated either by the fact for whom they are intended (individuals/legal person, field, annual turnover), or according to the monthly fee, with which additional advantages and benefits are associated.

The aim of this paper is to confirm the scientific issue:

Which business account is the best with respect to the defined criteria?

The banks providing business accounts in Czech currency in the Czech Republic include Citibank, Česká spořitelna, ČSOB, Equa bank, ERA Poštovní spořitelna, Expobank, Fio banka, GE Money Bank, Komerční banka, mBank, Raiffeisenbank, Sberbank, UniCreditBank. [1] Business accounts are also provided by less-known banking institutions such as: Expobank, Waldviertler Sparkasse Bank, Citfin, Artesa - savings association, Moravský peněžní ústav, Záložna CREDITAS, Spořitelní družstvo. [2], [22], [23]

For further comparison an account will be chosen from each bank account, which minimises the monthly fees (i.e. the fee for account management with regard to benefits provided), which for many banks for this type of account are not negligible. At the same time an account was selected, which is not limited to start-up entrepreneurs and have no other restrictions. For large businesses with a high turnover and many monthly transactions the decision-making would have been different, about half of these banks offer accounts allowing a high premium service for a high fee.

Given the extent of the processed analysis, we will focus on seven business accounts, selected according to the above requirements, i.e. business accounts from Česká spořitelna [7], ČSOB [8], Fio banka [5], GE Money bank [9], Komerční banka [4], mBank [6] and Raiffeisenbank.

Criteria for multi-criteria evaluation of the alternatives will be the fees and bonuses offered.

Another important consideration when choosing an account may be the availability of ATMs, assuming that the account does not offer withdrawals from any ATM without fees as the benefit (Raiffeisenbank offers this benefit). The best coverage of ATM network belongs to Česká spořitelna, Komerční banka, ČSOB and GE Money bank, all offering over 700 ATMs across the Czech Republic (plus mBank, which charges withdrawal of any ATM at the same rate). [3]

Nowadays, a new very important criterion is the level of Smartbanking. Many people today make a decision on their own account on the basis of whether they will be able to use smartbanking services on their smartphone and conveniently operate their account from anywhere.

We determine the overall evaluation of alternatives using the utility function as follows:

$$u(x) = \sum_{j=1}^{m} v_j \, u_j(x) \tag{1}$$

wherein v_j is standardised criteria importance and $u_j(x)$ are the evaluations of each alternatives. [13], [14], [15]

4 Comparison of business accounts

The following table 1 compares business accounts according to the amount of the monthly fees for account management plus it shows the most important benefits which are free of charge within a monthly fee.

Bank offering a business account	Type of business account	Monthly fees for account management in CZK	Free benefits
Česká spořitelna	ČS business account	149	10 payment transactions per month, debit card, overdraft
ČSOB	Business account	165	None, benefits are only for more expensive products
Fio banka	Fio business account	0	Electronic transactions (incoming and outgoing payments), 10 ATM withdrawals from own bank, withdrawals over 1,000 CZK and deposits at the branch
GE Money Bank	Genius business complete	149	Electronic transactions, deposits and withdrawals, card management, unlimited withdrawal from ATM of own bank
Komerční banka	Profi account	169	10 payment transactions per month
mBank	mKonto business	0	Electronic transactions, 3 withdrawals from any ATM, overdraft up to 300,000 CZK
Raiffeisenbank	Business eKonto Trade	199	Electronic transactions, unlimited withdrawal from any ATM, overdraft, insurance, savings account

Table 1 Comparison of accounts by charges and benefits, custom processing

The availability of ATMs is compared in the following table 2, which reflects the degree of availability using linguistically-defined variables. It is divided into the range of easily available, moderately available, rarely available. The value of withdrawals for free of charge (business account of the bank offering free ATM withdrawals in any bank, i.e. it is not necessary to solve the availability of specific bank ATM) is superior to the previous mentioned values.

Bank	ATM availability
Česká spořitelna	Easily available
ČSOB	Easily available
Fio banka	Moderately available
GE Money	Easily available
Komerční banka	Easily available
mBank	Easily available
Raiffeisenbank	Free withdrawals

Table 2 Comparison of accounts according to the availability of ATMs, custom processing

Using linguistically-defined variables the following table 3 shows whether a bank offers smartbanking and whether it is at average or excellent level (excellent level means the support of smartbanking on all three platforms - Android, iOS, Windows phone).

Bank	Smartbanking level
Česká spořitelna	Average
ČSOB	Excellent
Fio banka	Excellent
GE Money	Average
Komerční banka	Excellent
mBank	Average
Raiffeisenbank	Average

Table 3 Comparison of the accounts by smartbanking level, custom processing

Now let's imagine two model examples of entrepreneurs. The first model example will represent a small entrepreneur who has a monthly total of 30 incoming and outgoing payments, and favours ATM withdrawals compared to smartbanking (which is also used, but only in about a quarter of the cases). The second model example will represent an entrepreneur who has a business run successfully over 5 years, co-operates with many suppliers and customers, and has about 200 incoming and outgoing payments monthly. This entrepreneur almost never withdrawals from ATMs, prefers smartbanking, which he uses very often. This model example will be called a medium-sized entrepreneur.

The aforementioned three comparisons will be considered as criteria for multi-criteria decision-making and determine importance that will correspond to the different types of entrepreneurs.

The importance will be used in a standardised form:

$$v_j \geq 0, j = 1, \dots m, \sum_{j=1}^{m} v_j = 1 \tag{2}$$

For the above types of entrepreneurs we can define the importance as follows (see tab. 4):

Criterion	Standardised importance for a small entrepreneur	Standardised importance for a medium entrepreneur
Fees for services	0.60	0.45
Availability of ATMs	0.30	0.05
Smartbanking level	0.10	0.50

Table 4 Standardised importance of individual alternatives, custom processing

For a small entrepreneur, the largest importance was determined by the charges, because in addition to charges for account management, inconsiderable expense are also fees for ATM withdrawals, whose availability is in second position for them. Smartbanking is used in roughly one-quarter in relation to ATMs, i.e. the importance of this criterion makes up a quarter of 0.4.

For a medium-sized entrepreneur, the most important criterion is the use of smartbanking, without it they almost hardly operate their account. At nearly the same level are service charges, which include almost exclusively incoming and outgoing payments, which they have many. In last place with almost negligible importance is the availability of ATMs, which are used very little by this type of entrepreneur.

It is necessary to evaluate linguistically-defined variables so that the ratings reflect the ratio of preferences. For example, for level of smartbanking, the importance was determined at 1, 0.7 and 0, because the average level indicates the provision of smartbanking for two of the three mobile platforms. The table 5 capturing the evaluation of individual criteria will look at small and medium-sized enterprises as follows:

Bank	Charges + bonuses	Availability of ATMs	Smartbanking level
Česká spořitelna	0.3	0.8	0.7
ČSOB	0.1	0.8	1
Fio banka	1	0.5	1
GE Money	0.8	0.8	0.7
Komerční banka	0.3	0.8	1
mBank	1	0.8	0.7
Raiffeisenbank	0.8	1	0.7

Table 5 Evaluation of individual criteria, custom processing

5 The results of multi-criteria analysis

The results of multi-criteria decision-making are shown in the following table 6.

Bank	Results for a small entrepreneur	Results for a middle-sized entrepreneur
Česká spořitelna	0.49	0.525
ČSOB	0.4	0.585
Fio banka	0.85	0.975
GE Money	0.79	0.75

Komerční banka	0.52	0.675
mBank	0.91	0.84
Raiffeisenbank	0.85	0.76

Tab. 6. Evaluation of individual criteria, custom processing

The first places for small as well as medium-sized entrepreneurs belonged to the products from "low-cost" banks, which gradually gain more and more market share. For small entrepreneurs the best account is in mBank and for medium-sized entrepreneurs in FIO bank. On the contrary, the highest costs are paid by clients of the "Big Three", i.e. ČSOB, Česká spořitelna and Komerční banka (see Fig. 1).

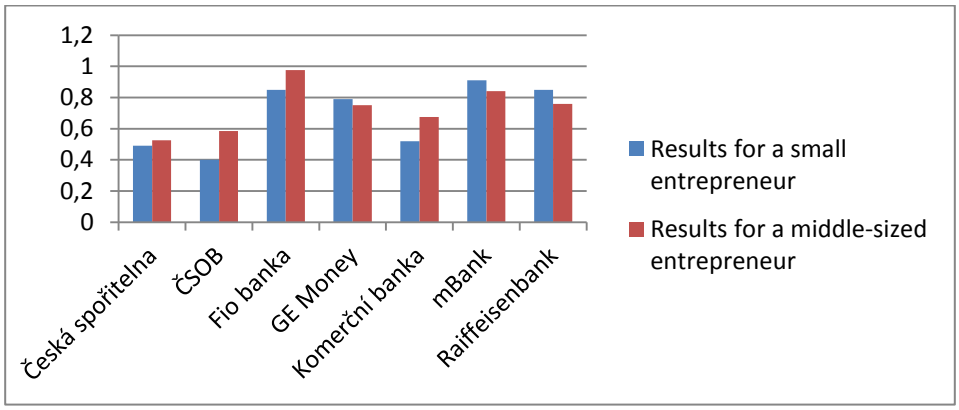


Figure 1 Overall comparison of the alternatives

6 Extension by using fuzzy set theory

The basic building block of the mentioned theory is the concept of fuzzy set. If we are unable to determine the exact boundaries of a class designed by some vague concept, to specify competence of an element to the class we can use a degree from a certain range. Therefore, each element will have an assigned degree which will indicate how much it belongs to that class.

This degree can be called the degree of competence of element into the class, and the class where each element is characterised by a degree of competence will be called a fuzzy set, defined as follows.

Let's consider class U, which is universe. This may be a universal class of all sets, its part or only a particular set. For simplicity, it is designed as U set. Let's suppose $U \neq \emptyset$. Then the fuzzy set for U is defined as follows:

$$\mu_A: U \rightarrow \langle 0.1 \rangle \tag{3}$$

μ_A function is called competence function of A fuzzy set and $\mu_A(x)$ is the degree of competence of x element to A fuzzy set.

Fuzzy set theory can be used in multicriteria decision making under uncertainty. In previous models, it was considered that the client is a small entrepreneur or a middle-sized entrepreneur. But many clients do not belong to exactly one of these profiles. It is therefore very useful to use fuzzy sets.

All considered criteria can be considered as fuzzy sets. To simplify the decision-making situation, two profiles which have been described previously, which are also fuzzy sets, and the client will identify with each of them with any degree of competence. Given the uncertain boundary between defined types allows any client to adapt the results to his own needs.

7 Discussion and Conclusion

The resulting table and graph shows that for a small entrepreneur the most beneficial is a business account from mBank, for reasons of zero monthly fees and the relatively good availability of ATMs. Closely followed by the products from Fio banka and Raiffeisenbank. Conversely, the least favourable is a business account from ČSOB, followed by business accounts from Česká spořitelna and Komerční banka. It is questionable whether the higher pricing of those accounts is sufficiently remedied by other services and products that these traditional big banks offer.

For middle-sized entrepreneurs the most beneficial is a business account from Fio banka, which also has zero monthly fees and bonuses, and Fio banka offers smartbanking for all platforms. In second place is the business

account from mBank. The least advantageous were business accounts from the "Big Four" of the banking market: Česká spořitelna, ČSOB, Komerční banka and GE Money bank. The highest average monthly cost for business account management for medium-sized businesses is generated at Česká spořitelna. For medium-sized enterprises, these banks offer a far greater range of bonuses. It just depends on the particular company, how well they know their needs and can choose the appropriate setting of their requirements for bank packages.

Acknowledgements

This paper is written with financial support from specific university research funds of the Czech Republic Ministry of Education, Youth and Sport allocated to the University of Hradec Králové Faculty of Informatics and management, project no. 2110/2015.

References

[1] Finance, URL: http://www.finance.cz/zpravy/finance/184274-podnikatelsky-nebo-klasicky-ucet/ on-line, [cit. 10.12.2014]
[2] Kompletní přehled bankovních účtů pro podnikatele, URL: http://probyznysinfo.ihned.cz/c1-62689390-kompletni-prehled-bankovnich-uctu-pro-podnikatele-kolik-ktere-stoji-a-co-nabizeji-navic, on-line, [cit. 10.12.2014]
[3] Finparáda, http://www.finparada.cz/Banky-Pojistovny-Sporitelny-Zalozny-Dostupnost-Pobocek-A-Bankomatu.aspx?T=A, on-line, [cit. 10.12.2014]
[4] Komerční banka, (2014), URL: http://www.kb.cz/cs/lide/mladez-a-studenti/index.shtml, on-line, [cit. 10.12.2014]
[5] FIO (2014), URL: http://www.fio.cz/bankovni-sluzby/bankovni-ucty/osobni-ucet/, on-line, [cit. 10.12.2014]
[6] mBank (2014), URL: http://www.mbank.cz/osobni/ucty/mkonto/, on-line, [cit. 10.12.2014]
[7] Česká spořitelna (2014), URL: http://cs.cz/banka/nav/osobni-finance/osobni-ucet-cs-ii-fresh-ucet/o-produktu-d00022361, on-line, [cit. 10.12.2014]
[8] ČSOB, (2014), URL: http://www.csob.cz/cz/Lide/Ucty-a-platby/Stranky/CSOB-Studentske-konto.aspx, on-line, [cit. 10.12.2014]
[9] GE Money Bank (2014) URL: https://www.gemoney.cz/lide/ucty, on-line, [cit. 10.12.2014]
[10] Bankovní poplatky (2014) URL: http://www.bankovnipoplatky.com/kalkulator.html, on-line, [cit. 10.12.2014]
[11] Soukal, I., Hedvicakova, M., (2015) http://www.bankovnipoplatky.com/klientsky-index-iv-ctvrtleti-2014-prumerne-bankovni-poplatky-181-kc-mesicne-26690, on-line, [cit. 15.1.2015]
[12] Draessler, J., Soukal, I., Hedvičáková, M. (2011). Shluková analýza poptávkové strany trhu základních bankovních služeb. In *E+M Ekonomie a Management, 14*(4), 102-114. ISSN 1212-3609
[13] Dostál, P. (2008). *Optimalizační metody. Optimalizační metody.* Kunovice: EPI Kunovice, 2008. ISBN: 978-80-7314-136- 3.
[14] Mikulecký, P., Lenharčík, I., Hynek, J., (2002). *Znalostní technologie II. Expertní systémy. 2. vydání,* GAUDEAMUS, UHK Hradec Králové, 2002, ISBN 80-7041-904-0
[15] Talašová, J. (2003), *Fuzzy metody vícekriteriálního hodnocení a rozhodování*. Olomouc: Vydavatelství Univerzity Palackého, 2003, 179 s., ISBN 80-244-0614-4.
[16] Polouček, S. a kol., (2009*) Bankovnictví, 2. vydání*, C.K.Beck, 2009, 736 s., ISBN 80-7179-462-7.
[17] Revenda, Z., (2011) *Centrální bankovnictví, 3. Aktualizované vydání,* Managementpress, 560 s, 2011, ISBN: 978-80-7261-230-7
[18] Tennant, D., Sutherland, R., (2014) What types of banks profit most from fees charged? A cross-country examination of bank-specific and country-level determinants, In *Journal of Banking & Finance*, Volume 49, Elsevier, December 2014, Pages 178–190, doi:10.1016/j.jbankfin.2014.08.023
[19] European Union Directorate-General for Health and Consumers Protection. Retail financial services to the consumer markets scorecards. 2009, [online], URL: http://ec.europa.eu/consumers/rights/docs/swd_retail_fin_services_en.pdf, cit. 5.3.2015.
[20] Ekel, P.Ya., Martini, J.S.C., Palhares, R.M., (2008) Multicriteria analysis in decision making under information uncertainty, In *Applied Mathematics and Computation*, Volume 200, Issue 2, 1 July 2008, Pages 501-516, doi:10.1016/j.amc.2007.11.024
[21] Jablonský, J.,(2007) *Operační výzkum kvantitativní modely pro ekonomické rozhodování*, Professional Publishing, 2007, 323 p., ISBN: 978-80-86946-44-3
[22] Finace iDnes URL: http://finance.idnes.cz/bankovni-ucty-pro-podnikatele-deg-/ucty.aspx?c=A140613_161036_bank_zuk, [cit. 10.3.2015]
[23] Hedvičáková, M., Pozdílková, A., Srovnání podnikatelských účtů v České republice, str. 96-98, *Aplimat 2015, Book of Abstracts*, February 3-5, 2015, Bratislava, Slovak Republic, ISBN 978-80-227-4315-0

Robustified on-line estimation of the EWMA models: Simulations and applications

Radek Hendrych[1]

Abstract. The exponentially weighted moving average (EWMA) model is a particular modelling scheme used by RiskMetrics for forecasting the current level of volatility of financial returns. The aim of this paper is to introduce and study the self-weighted sequential estimation algorithm, which represents a numerically effective alternative to already established calibration approaches. Firstly, its derivation and theoretical properties are briefly outlined. Secondly, the presented calibration technique is robustified to eliminate destructive influence of eventual additive outliers. Thirdly, both versions are examined by means of Monte Carlo simulations and real financial data.

Keywords: EWMA model, recursive estimation, RiskMetrics, robustification.

JEL classification: C51
AMS classification: 62M10

1 Introduction

The exponentially weighted moving average (EWMA) model is a particular conditional heteroskedasticity modelling scheme. This theoretical approach is frequently linked to investigating financial time series, more specifically to monitoring volatility (i.e. the conditional standard deviation of financial returns). The EWMA model has been primarily developed as a simple alternative to the GARCH models. The name of this concept originates from the fact that the conditional variance is an exponentially weighted sum of historical squared financial returns with the geometrically declining weights going back in time. Therefore, this model is easily capable to track changes in the conditional variance and volatility. Since its introduction in [5], it has been investigated from various theoretical and practical perspectives. It has been successfully applied in many empirical studies. For example, one may employ the EWMA model to predict volatility, to calculate distinct risk measures (e.g. Value at Risk), or to define a trading rule. Moreover, the EWMA framework is regarded as the benchmark by many practitioners.

The value of the only parameter of the EWMA model defining the discussed geometrically declining weights is conventionally prescribed by experts or users (e.g. by RiskMetrics). Alternatively, it can be calibrated employing standard (off-line) statistical inference procedures (e.g. the conditional maximum likelihood method). However, it is indeed rarely estimated recursively (i.e. sequentially or on-line). On the other hand, it might be advantageous to adopt a numerically effective technique that could be able to estimate and control this parameter (or model) in real time. For instance, one can employ this approach in the case of high-frequency data. Consequently, the aim of this contribution is to introduce and study the one-stage self-weighted recursive estimation method for calibrating the EWMA model on-line, jointly with its robustified variant. The suggested algorithm has been derived by using standard recursive identification instruments outlined, e.g., in [4]. It has demonstrated its numerical capabilities by means of simulations and an empirical application.

This paper is organized as follows. Section 2 reviews the EWMA modelling framework and its fundamental features. It shortly discusses corresponding (standard) off-line estimation procedures. Section 3 derives and briefly comments the one-stage self-weighted recursive algorithm for calibrating the EWMA model. Section 4 introduces its robustified version. Section 5 analyses this estimation procedure by Monte Carlo experiments. Section 6 considers an empirical application of this methodology. The key points of this paper are summarized by conclusions presented in Section 7.

[1]Charles University in Prague, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, Radek.Hendrych@mff.cuni.cz.

2 EWMA model

Formally, the EWMA model of financial returns $\{y_t\}$ is commonly defined as in [5], i.e.:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2, \tag{1}$$

where the only modelling parameter λ lies in the interval $(0,1)$, $\{\varepsilon_t\}$ is a sequence of i. i. d. random variables with zero means and unit variances, and σ_t^2 is $\mathcal{F}_{t-1}$-measurable. Remind that $\mathcal{F}_t$ denotes the smallest σ-algebra with respect to which y_s is measurable for all $s \leq t$.

One can readily compute the following conditional moments:

$$\mathbb{E}(y_t|\mathcal{F}_{t-1}) = 0, \quad \mathsf{var}(y_t|\mathcal{F}_{t-1}) = \mathbb{E}(y_t^2|\mathcal{F}_{t-1}) = \sigma_t^2. \tag{2}$$

Apparently, positivity of the conditional variance σ_t^2 is ensured by construction of (1); volatility is defined as was mentioned before, i.e. as $\sqrt{\sigma_t^2}$. The one-step ahead prediction of σ_t^2 is expressed as:

$$\sigma_{t+1|t}^2 := \mathbb{E}(\sigma_{t+1}^2|\mathcal{F}_t) = (1-\lambda) y_t^2 + \lambda \sigma_t^2 = \sigma_{t+1}^2. \tag{3}$$

Similarly, the k-step ahead forecast of σ_t^2 is given by (for $k > 1$):

$$\sigma_{t+k|t}^2 := \mathbb{E}(\sigma_{t+k}^2|\mathcal{F}_t) = \sigma_{t+1|t}^2. \tag{4}$$

To calibrate the EWMA model (1) using T observations $\{y_1, \ldots, y_T\}$, we usually employ one of these methods: (i) the value of λ is prescribed by experts or users (e.g. the choice 0.94 is obviously recommended for daily data by RiskMetrics); (ii) λ is estimated minimizing the root mean squared error of the forecast inaccuracies $(y_t^2 - \sigma_t^2)$ assuming that y_0 and σ_0^2 are either defined or observed; (iii) supposing a certain probability distribution of ε_t (the Gaussian innovations are preferred in regarding to the consistency of estimates), one may calibrate the parameter λ maximizing the conditional log-likelihood function (similarly as before, y_0 and σ_0^2 are supposed to be known).

3 On-line estimation of the EWMA model

In this section, we shall introduce the one-stage self-weighted recursive estimation algorithm that can calibrate the parameter of the EWMA model (1) in real time. In many instances, this approach may be truly advantageous. For example, it is possible to monitor or predict volatility sequentially in the high-frequency financial data context. Recursive estimation methods are also effective in terms of memory storage and computational complexity since the current parameter estimates are evaluated using the previous estimates and actual measurements. Incidentally, they can be used to detect structural changes.

Applying general recursive prediction error method, see e.g. [4], one can derive the recursive scheme for on-line estimating the parameter λ of the EWMA model (1). Note that the conditional log-likelihood criterion is supposed (assuming normally distributed innovations ε_t). The resulting algorithm can be concisely formulated as follows:

$$\begin{aligned}
\widehat{\lambda}_t &= \widehat{\lambda}_{t-1} + \frac{\widehat{p}_{t-1}(y_t^2 - \widehat{\sigma}_t^2)\widehat{\sigma}_t^{2'}}{\alpha_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2 \widehat{p}_{t-1}}, \\
\widehat{p}_t &= \frac{1}{\alpha_t}\left\{\widehat{p}_{t-1} - \frac{\widehat{p}_{t-1}^2(\widehat{\sigma}_t^{2'})^2}{\alpha_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2 \widehat{p}_{t-1}}\right\}, \\
\widehat{\sigma}_{t+1}^2 &= (1-\widehat{\lambda}_t) y_t^2 + \widehat{\lambda}_t \widehat{\sigma}_t^2, \\
\widehat{\sigma}_{t+1}^{2'} &= -y_t^2 + \widehat{\sigma}_t^2 + \widehat{\lambda}_t \widehat{\sigma}_t^{2'}, \; t \in \mathbb{N},
\end{aligned} \tag{5}$$

where $\widehat{\lambda}_t$ denotes the recursive estimate of the parameter λ at time t. We recommend initializing the foregoing procedure under these conditions (similarly as it is outlined in [3]): (i) $\widehat{p}_0$ is a large positive number, e.g. $\widehat{p}_0 = 10^5$; (ii) $\widehat{\lambda}_0$ should be taken from the interval $(0,1)$, e.g. as 0.94 as it is usually preferred for daily data; (iii) $\widehat{\sigma}_1^2$ is a positive number (e.g. the sample variance of several first measurements) and $\widehat{\sigma}_1^{2'}$ equals zero; (iv) $\{\alpha_t\}$ is a deterministic sequence of real positive numbers smaller or equal to one that either accelerates convergence or allows tracking parameter changes (see below and Section 5).

At each time t, it is necessary to check whether the current recursive estimate belongs to the interval $(0, 1)$ before evaluating other quantities in (5). If not, one should artificially set the actual estimate as the previous one to avoid eventual specification problems. This simple projection ensures positivity of the conditional variance since $\widehat{\lambda}_0$ lies inside the interval. The sequence $\{\alpha_t\}$, the so-called *forgetting factor*, may be selected as follows: (i) α_t gradually grows to one as t goes to infinity, e.g. $\alpha_t = 0.99\alpha_{t-1} + 0.01$, $\alpha_0 = 0.95$; (ii) $\alpha_t \equiv \alpha$ for some $\alpha \in (0, 1)$, e.g. $\alpha = 0.997$, and all t. The first option corresponds to estimating the model (1) supposing time-invariant λ. The increasing forgetting factor improves the convergence speed of the algorithm during the transient phase. The second case is associated with the eventuality that λ can vary over time. The constant forgetting factor less than one progressively reduces the influence of historical measurements, and thus enables to detect parameter changes.

Theoretical properties of the suggested recursive estimation algorithm coincide with the off-line case (as t goes to infinity), where the corresponding conditional log-likelihood criterion is maximized. Namely, convergence and asymptotic distributional features are identical for a sufficiently large portfolio of observations. Refer to [4].

4 Robustification of the suggested estimation algorithm

In practice, it is necessary to be concerned with abnormal observations which may occur in data. They can be caused by many reasons, e.g. by additive innovations, measurement failures, etc. As is evident from (5), abnormal data points (the so-called *outliers*) will influence the model estimation considerably if no specific action is taken. Therefore, if such defects are expected in the data set, one should modify the estimation algorithm to make it more robust. The outliers tend to appear as spikes in the sequence of $\{y_t/\widehat{\sigma}_t\}$, which obviously result in large contributions to the estimation procedure (5).

There exist various ways how to robustify recursive estimation algorithms. For instance, the criterion function may be selected to become less sensitive to large errors (e.g. by using the Huber functions). Another way of handling single outliers is based on testing a measurement at each time t. If it is large compared with a given limit, it will be indicated as erroneous and will be substituted by another value. See, e.g., [1] for further insights.

Assume the following criterion applicable in the context of the proposed estimation algorithm (5):

$$\left|\frac{y_t}{\widehat{\sigma}_t}\right| \leq a_t, \tag{6}$$

where $\{a_t\}$ is a deterministic sequence of positive real numbers that reflects the character (distribution) of the data set (usually $a_t \equiv a > 0$ for all $t \in \mathbb{N}$). If this condition is satisfied, then the estimation algorithms will remain unchanged. Otherwise, the original measurement y_t will be replaced by $a_t\widehat{\sigma}_t\text{sign}(y_t)$. In Section 5, this concept is further investigated by means of Monte Carlo experiments.

5 Monte Carlo experiments

This section briefly examines the proposed recursive estimation technique by means of Monte Carlo simulations. Various numerical experiments have been performed with almost analogical results. Therefore, only two representative instances are reviewed here. Particularly, we replicated two EWMA processes (1) with Gaussian disturbances of the length 10000 with two distinct parameters λ (i.e. 0.94 and 0.99) in order to study convergence properties of the suggested estimation method. We generated one thousand repetitions. The chosen length corresponds to an approximately three-hour dataset working with one-second data. All computations were conducted in the statistical software `R` by implementing original procedures for simulating and estimating the EWMA models.

Figure 1 illustrates numerical behaviour of the one-stage self-weighted recursive prediction error procedure (5) specified by the consequent recommendations (with $\alpha_t = 0.99\alpha_{t-1} + 0.01$, $\alpha_0 = 0.95$). The estimation process was stopped at the times $T_a = 1000$, $T_b = 3000$, $T_c = 5000$, and $T_d = 10000$; the current estimates were always stored. Figure 1 summarizes sample characteristics of these estimates using the standard box-plots for each stopping time. It is apparent that the estimates converge to the true values jointly with decreasing variances. Thus, one might conclude that the suggested self-weighted recursive method (5) is capable to estimate the EWMA parameter in accordance with [3] and [4].

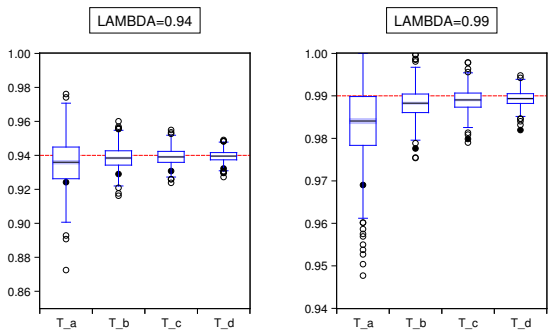


Figure 1 Results of Monte Carlo simulations (estimating EWMA modelling parameters)

Additionally, Section 4 has proposed a simple concept, which should handle eventual problems with additive outliers in data. This approach is studied by means of a demonstrative numerical example, which can establish its adequacy. In particular, we replicated one thousand EWMA processes of the length 10000 with normally distributed innovations and $\lambda = 0.94$. After each simulation, we incorporated an additive outlier into each generated time series; namely, we set $y_{2500} = 10$. Other experimental conditions remained as before. We have studied two versions of the estimation scheme (5), i.e. the first with the forgetting factor increasing to one (defined as before) and the second with the constant forgetting factor $\alpha_t \equiv 0.997$ for all t. Moreover, we have distinguished between three estimation variants of both approaches (without any robustification, with the robustification as was described in Section 4 using $a_t \equiv u_{0.99}$ and $a_t \equiv u_{0.9999}$ for all t, respectively). Note that u_α denotes the corresponding α-quantile of the standard normal distribution.

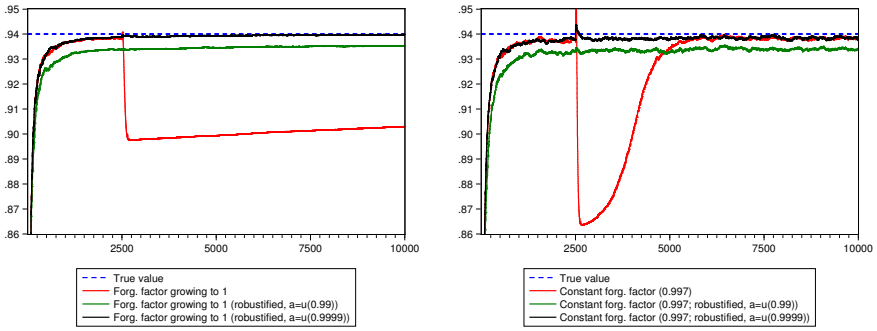


Figure 2 Medians of the EWMA parameter estimates (the outlier at $t = 2500$)

Figure 2 displays medians of the distinct on-line estimates (calculated at each time step). Apparently, the additive outliers have had absolutely destructive influence on estimation, when no robustification has been applied. Employing the recommended robustification, the results look more favourably. However, one should use a less conservative a_t (e.g. $u_{0.9999}$) in (6) not to restrict on-line estimation excessively (especially during the transient phase). Note that using the constant forgetting factor results in the more volatile estimates since it reduces influence of past measurements assigning the geometrically decreasing weights. On the other hand, it is able to adapt (partly) to the inserted outlier.

6 Empirical analysis of the PX index

The PX index (ISIN XC0009698371) is an official market-cap weighted stock index launched on 5th April 1994 composed of the most liquid shares traded on the Prague Stock Exchange. In particular, it is a price index of blue chips issues, which is calculated in real-time and weighted by market capitalization. Dividends are not considered. A new value of the PX index is delivered by a particular formula; it reflects each single price change of index constituents. The maximum weight for a share issue is 20% on a decisive

day. A portfolio of core issues is variable and it can be restructured quarterly. Further details (including historical data) can be found on the official web pages of the Prague Stock Exchange.

Figure 3 presents all historical daily closing quotes of the PX index until 31st March 2015 (i.e. 5248 observations). The minimal value 316 occurred on 8th October 1998 after the Russian financial crisis. The maximal observation 1936 was achieved 29th October 2007. It is visible that the crisis year 2008 was truly exceptional. See also [2].

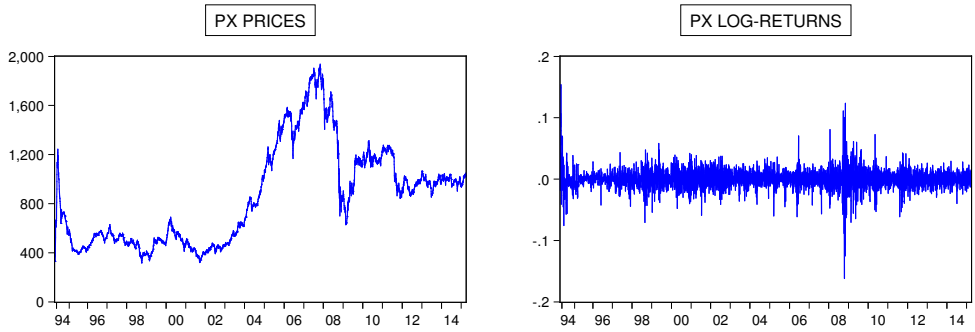


Figure 3 Historical closing quotes and log-returns of the PX index

Figure 4 introduces different off-line and on-line estimates of the EWMA modelling parameter for the PX index log-returns. Namely, the given financial measurements are studied by means of the following methods: (i) the recursive algorithm (5) with the forgetting factor growing to 1 (defined as before) with the robustification (6) using $a_t \equiv u_{0.9999}$ and without it; (ii) the recursive algorithm (5) with the constant forgetting factor 0.997 with the robustification (6) using $a_t \equiv u_{0.9999}$ and without it; (iii) the off-line conditional maximum likelihood method. The on-line estimation has been initialized as was described in Section 3.

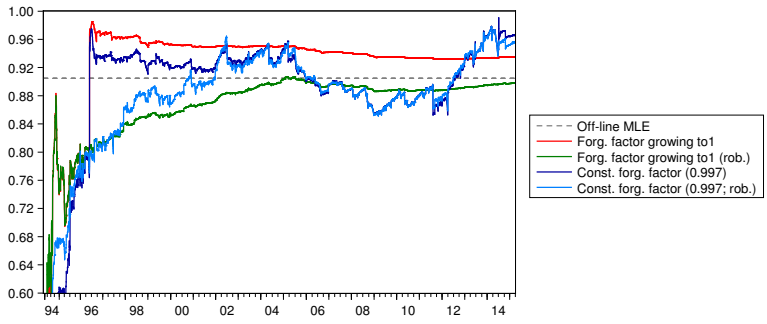


Figure 4 Different estimates of the EWMA modelling parameter for the PX log-returns

At first glance, the accepted robustification has eliminated influence of additive outliers (especially the one occurring in the first part of the dataset). Therefore, the robustified methods are more reliable from this perspective. Moreover, the recursive estimates calculated using the constant forgetting factor apparently fluctuate around the off-line one. One can discover several shared trends, which obviously correspond to the overall development of the PX index closing quotes (see Figure 3). The on-line estimate evaluated by the growing forgetting factor is more rigid. Note that all mentioned recursive estimates are less reliable in the beginning of the observed time series since these methods are initialized therein.

Figure 5 surveys the estimated conditional volatilities of the logarithmic returns of the PX index. According to the preceding discussion, we have examined the off-line method and two robustified recursive procedures for calibrating the EWMA parameter only. At first sight, one could conclude that all introduced outputs follow analogical trends. The estimates based on the off-line conditional likelihood procedure and the suggested recursive algorithm seem to be closely related. One may also compare the achieved values of the (conditional) log-likelihood function associated with this particular estimation

problem (or equivalently contrast some information criteria). The calculated log-likelihoods are recapitulated in Table 1. The complete and truncated samples are considered in computing the presented numbers. Namely, first 10%, 30%, and 50% observations were cut off in order to verify the adequacy since the recursive methods are less stable during the initial phase of estimation (see also above). Consequently, the calibration algorithm (5) introduced is visibly competitive and may be undoubtedly employed in the PX index daily data context.

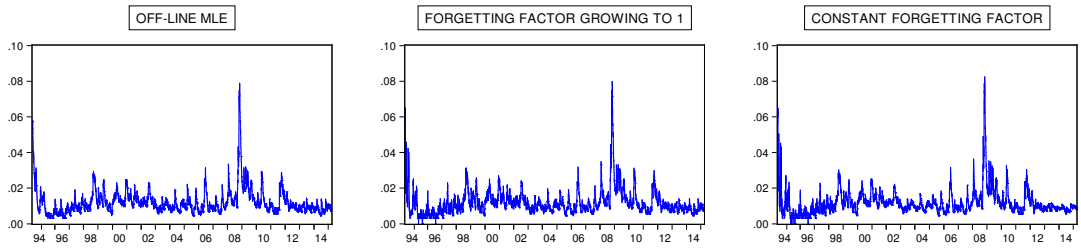


Figure 5 Different estimates of volatility of the PX log-returns

Sample/Method	**Off-line MLE**	**On-line** ($\alpha_t \nearrow 1$)	**On-line** ($\alpha_t \equiv 0.997$)
Complete	**15819.37**	15751.06	15755.63
Truncated (10%)	14191.75	14178.27	**14191.80**
Truncated (30%)	11001.93	10994.66	**11007.17**
Truncated (50%)	7858.61	7856.11	**7862.97**

Table 1 Values of log-likelihood functions corresponding to estimates in Figure 5

7 Conclusion

In this paper, we have introduced the one-stage self-weighted recursive estimation algorithm for calibrating the EWMA process employing the general recursive identification instruments. The procedure has been further robustified applying a simple truncation. The qualities of both proposed methods were demonstrated by means of simulations. The accepted modelling framework was further examined in investigating volatility of the PX index. It has proved its competitiveness. These findings indeed motivate further research of on-line estimation of the conditional heteroskedasticity models.

Acknowledgements

This research was supported by the grants GA P402/12/G097 and SVV 2015 No. 260225.

References

[1] Cipra, T., and Hanzák, T.: Exponential smoothing for time series with outliers, *Kybernetika* **47** (2011), 165–178.

[2] Hendrych, R.: On comparing various modelling schemes: The case of the Prague Stock Exchange index. In: *The 8th International Days of Statistics and Economics: Conference Proceedings* (T. Löster, and T. Pavelka, eds.), Melandrium, Slaný, 2014, 456–465.

[3] Ljung, L.: *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River (NJ), 1999.

[4] Ljung, L., and Söderström, T. S.: *Theory and Practice of Recursive Identification*. MIT Press, Cambridge (MA), 1987.

[5] Morgan, J.: *RiskMetrics - Technical document*. (4th edition). Morgan Guaranty Trust Company, New York, 1996.

Optimization model with nonconvex monotone piecewise linear objective function

Robert Hlavatý[1], Helena Brožová[2]

Abstract. The paper describes new approach to solution of some optimization models with non-convex piecewise linear objective function with monotone course. The matter of non-convexity complicates searching for an optimal solution of an optimization problem. We propose computationally tractable algorithm based on multiobjective optimization, capable of solving some non-convex optimization problems. The class of the optimization problems we deal with is specified and written as a mathematical model. The model is then transformed into appropriate form with the help of additional linear constraints. When piecewise linear functions are present, the search space represented here by convex polytope has some interesting properties we describe before the algorithm itself. The algorithm is proposed on the basis of multiobjective simplex algorithm.

Keywords: Piecewise linear functions, nonconvex optimization, monotone functions, multiobjective programming.

JEL Classification: C61
AMS Classification: 90C26, 90C29, 90C30

1 Introduction

The paper deals with the specific class of mathematical programming problems. We consider a mathematical model with linear constraints, constraints of non-negativity where the objective function is piecewise linear and non-convex. Certainly, a way of searching for the optimal solution of such programming problem has been already known. Here, we would like to point out an interesting approach that, to our knowledge, has not yet been used for solving of this class of optimization problems. Our new approach would create other perspective to solutions of these models which can be otherwise solved by the means of combinatorial approaches, heuristics or mixed integer linear programming (MILP).

Already in 1969 the algorithm for separable nonconvex programming problems based on *branch and bound* principles was introduced by Falk and Soland (1969). Later on, more specific approach based on the same principle was presented by Benchekroun and Falk (1991) where authors present the way of solution for nonconvex piecewise linear optimization problem. Polisetty and Gatzke (2005) show how to utilize a *decomposition algorithm* for nonconvex MILP using piecewise linear relaxation. Another means of solution of the problem using MILP are proposed by Kameshwaran and Narahari (2009) who demonstrated their approach on the nonconvex *knapsack problem*. Van Bokhoven and Leenaerts (2010) describe some pivotization algorithm of Katzenelson (1965) and Lemke (1967) and also the relation to linear programming where the *linear complementary problem* is taken into consideration. These methods are considerably difficult and would usually require an appropriate software to be utilized.

We would like to present our original method that is based on multiobjective programming approach described by Zelený (1976) or Cohon (1978) and the software solution is not required for smaller problems. The concept of *boundary points* proposed earlier by Houška and Brožová (2002) will be used for decomposition of the complex problem. Houška (2003) has shown some heuristics for the solution of nonconvex piecewise linear optimization problems which turned out to be far too complex to be used for the solution of the general problems of larger scale. Here, we present a general algorithm for nonconvex piecewise linear optimization problems, especially suitable for some particular models that will be described further in the text.

2 Problem description

Let us first introduce the general piecewise linear programming problem (PWLP) in the form we will further consider within this paper:

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, hlavaty85@gmail.com
[2] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, brozova@pef.czu.cz

$$\text{maximize } Z(x) = Z_1(x_1) + Z_2(x_2) + \cdots + Z_n(x_n)$$
$$\text{s.t.}$$
$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i, i = 1,2,\dots,m\,, j = 1,2,\dots,n$$
$$x_j \geq 0, j = 1,2,\dots,n$$

where (1)

$$Z_j(x_j) = \begin{cases} Z_{j1}(x_j), x_j \in \langle 0\,, k_{j1}\rangle \\ Z_{j2}(x_j), x_j \in \langle k_{j1}\,, k_{j2}\rangle \\ \vdots \\ Z_{jp_j}(x_j) \in \left\langle k_{jp_j-1}, \infty\right) , j = 1,2,\dots,n \end{cases}$$

where $Z(x)$ is continuous smooth piecewise linear function, possibly nonconvex, consisting of *j* partial piecewise linear functions, subjected to the linear constraints. The objective function here can be viewed from four perspectives:

- Z**(x)** …**aggregate objective function** – is the separable objective function. It value expresses the real value of objective of the given optimization problem.
- $Z_1(x_1), Z_2(x_2), .., Z_n(x_n)$…**partial piecewise linear functions** are the functions of one variable whose sum completes the value of **aggregate objective function**. We will refer to these functions simply as partial objective functions.
- $Z_{11}(x_1), Z_{12}(x_1), \dots Z_{jp_j}(x_j)$…**linear segments of partial objective function** are individual linear segments of **partial objective functions** on given intervals from $\langle 0; k_{jk}\rangle \cup \langle k_{jp_j-1}; \infty\rangle$. We will refer to them shortly as **linear segments**.
- $Z_{h_1,h_2,\dots,h_n}$…**fragments of objective function.** The aggregate objective function can be also expressed as so called **fragments of objective function**. These fragments are defined on multi-dimensional intervals and are received upon combining different linear segments of **partial objective functions** (see section 3).

Relations between these four perspectives are expressed using the following schematics:

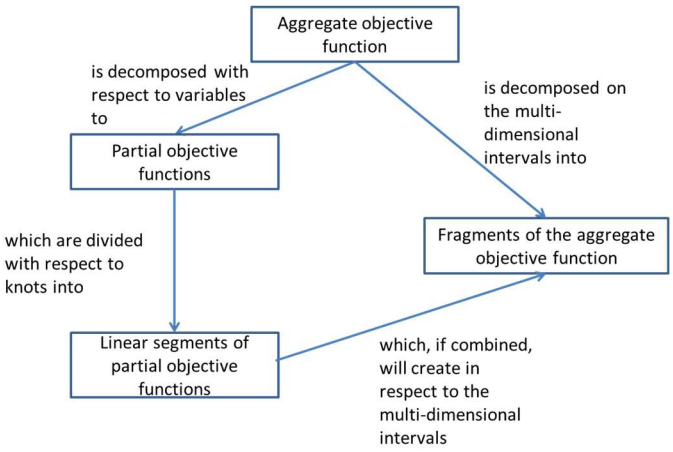


Figure 1 Four perspectives on the objective function.

As stated in the introduction, the concept of *boundary points* will be used for the solution of the model. In the next step, we enrich the general model (1) with additional constraints. Let us first briefly introduce the concept of boundary points.

The partial objective functions are piecewise linear (or at least one of them) and the linear segments are connected in knots in which these segments also change its slope. We will introduce a new variable $r_{jk}^{-/+}$ where *j* is the index of partial objective function, *k* is the index of a knot and the upper index -/+ expresses the under-achievement or over-achievement of an individual boundary point. In general it expresses the distance of the

current basic solution of the problem from the boundary point in the solution space. The distance from the boundary point can be in general written as a new constraint:

$$x_j + r_{jk}^- - r_{jk}^+ = k_{jk} \tag{2}$$

for j-th variable x_j and k-th knot of the partial objective function k_{jk}. The constraint is originally based on the concept of goal programming and its deviations from a certain objective. It is desired that at most one of these variables r_{jk}^-, r_{jk}^+ is in the basic solution at the same time as the solution can lie only on one side from the k_{jk} in the solution space. Thus, we introduce a nonlinear constraint:

$$r_{jk}^- . r_{jk}^+ = 0 \tag{3}$$

It can be proved this constraint does not have to be taken into consideration in terms of model solvable by simplex algorithm or multiobjective simplex algorithm since this condition is met automatically due the nature of column vector of these variables.

Now, let us transform the model into the new form with the help of constraints (2). Let the new model be referred to as PWLP-B (boundary points). The model goes as follows:

$$\begin{gathered}
\text{maximize } Z(x) = Z_1(x_1) + Z_2(x_2) + \cdots + Z_n(x_n) \\
\text{s.t.} \\
\sum_{j=1}^{n} a_{ij} x_j \leq b_i, i = 1,2,\dots,m\,, j = 1,2,\dots,n \\
x_j + r_{jk}^- - r_{jk}^+ = k_{jk}, j = 1,2,\dots,n; k = 1,2,\dots,p_j \\
r_{jk}^- . r_{jk}^+ = 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j \\
x_j \geq 0, j = 1,2,\dots,n
\end{gathered} \tag{4}$$

where

$$Z_j(x_j) = \begin{cases} Z_{j1}(x_j), x_j \in \langle 0\,, k_{j1} \rangle \\ Z_{j2}(x_j), x_j \in \langle k_{j1}\,, k_{j2} \rangle \\ \vdots \\ Z_{jp_j}(x_j) \in \left\langle k_{jp_j-1}, \infty \right), j = 1,2,\dots,n \end{cases}$$

2.1 Specifics of the solution space

Due to presence of the boundary points representing the knots of piecewise linear function, the solution space has some interesting properties. According to (1) the solution space is given by

$$\begin{gathered} \boldsymbol{Ax} \leq \boldsymbol{b} \\ \boldsymbol{x} \geq \boldsymbol{0} \end{gathered} \tag{5}$$

Considering the boundary points of each of j partial piecewise linear function with p linear segments (each with the same number of segments), the solution space will split into p^j multi-dimensional intervals. Let us show the illustrative example for $\boldsymbol{Z}(\boldsymbol{x}) = Z_1(x_1) + Z_2(x_2)$ where each of the partial piecewise linear function has 3 linear segments. Then:

$$\begin{aligned} Z_1(x_1) &= \begin{cases} Z_{11}(x_1), x_j \in \langle 0\,, k_{11} \rangle \\ Z_{12}(x_1), x_j \in \langle k_{11}\,, k_{12} \rangle \\ Z_{13}(x_1) \in \langle k_{12}, k_{13} \rangle \end{cases} \\ Z_2(x_2) &= \begin{cases} Z_{21}(x_2), x_j \in \langle 0\,, k_{21} \rangle \\ Z_{22}(x_2), x_j \in \langle k_{21}\,, k_{22} \rangle \\ Z_{23}(x_2) \in \langle k_{22}, k_{23} \rangle \end{cases} \end{aligned} \tag{6}$$

The situation can be depicted in $\mathbb{R}^2$ as shown in figure 2:

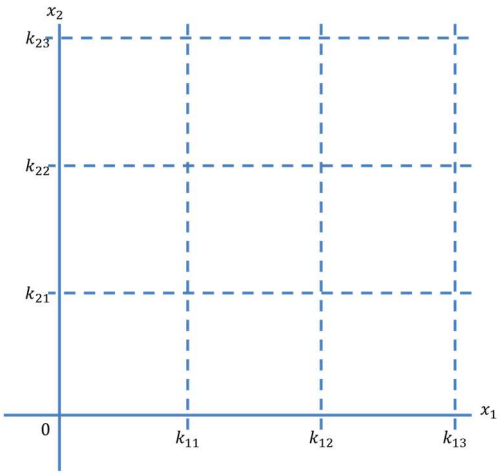


Figure 2: Solution space split into multi-dimensional intervals.

Reader can note that the range $\langle k_{jk}; k_{jk+1}\rangle$ of these intervals does not have to be necessarily the same as is in the picture and each partial objective function can have a different number of linear segments. The value of aggregate objective function is unique on each of these multi-dimensional intervals. It is generally well known that if the optimal solution of the convex optimization problem exists it lies in one of the vertices of the convex polytope given by the constraints of the problem. Due to the nonconvex nature of our problem the optimal solution can also possibly lie in vertices of the multi-dimensional intervals, i.e. all vertices of all squares in the figure one can represent candidate solutions of the optimization problem.

2.2 Solution design

Some eligible ways of solution come under consideration. For the smaller tasks, at most in $\mathbb{R}^2$, one could possibly find the aggregate objective function value for all candidate solutions (if capable to identify them all) and choose the optimal one or more optimal solutions if possible. This approach is merely combinatorial and does not require any algorithmic approach. It would be also possible to decompose the solution of the problem to subproblems. One could find the optimal solution by solving at most $p^2 - 2p + 1$ subproblems

$$\begin{gathered}\text{maximize } Z_{h_1,h_2,\dots,h_n}(\boldsymbol{x}) \\ \text{s.t.} \\ \sum_{j=1}^{n} a_{ij}x_j \le b_i, i = 1,2,\dots,m;\ j = 1,2,\dots,n \\ x_j \le k_{jh_j}, j = 1,2,\dots,n;\ h_j = 1,2,\dots,p_j \\ x_j \ge k_{jh_j-1}, j = 1,2,\dots,n;\ h_j = 1,2,\dots,p_j \\ \boldsymbol{x} \ge \mathbf{0}\end{gathered} \tag{7}$$

The decomposition approach does not seem to be effective either since it mostly depends on the number of knots of the partial objective functions.

3 Decomposition of aggregate objective function to fragments

Both approaches mentioned in the previous sections show serious drawbacks when the optimization problem is of larger scale. Thus we propose utilization of objective function fragments that can be defined on all multi-dimensional intervals received on the basis of boundary points present in the PWLP-B model. On all of these intervals, the aggregate objective function increases or decreases with a different intensity. It means that on each of these intervals we are heading towards optimal solution with a different tempo.

Let us consider the growth rate of fragments of aggregate objective function for individual multi-dimensional intervals to be a criterion of reaching the optimal solution of the problem. We can then consider this problem to be multiobjective optimization problem. The solution of the original model (PWLP) is one of the non-dominated solutions because the optimal value of the objective function is certainly the optimal value of the given segment of the function a thus it cannot be dominated.

In that case, each fragment of the aggregate objective function can be extended to the entire solution space. Having these assumptions in mind, we can reformulate the modified problem (PWLP-B) to a problem of multiobjective programming (PWLP-MC). We define the PWLP-MC model in the following way:

$$\text{maximize } \boldsymbol{Z}(\boldsymbol{x}) = \begin{pmatrix} Z_{1,1,\dots,1} \\ \vdots \\ Z_{h_1,h_2,\dots,h_n} \\ \vdots \\ Z_{p_1,p_2,\dots,p_n} \end{pmatrix}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij}x_j \le b_i, i = 1,2,\dots,m, j = 1,2,\dots,n \tag{8}$$

$$x_j + r_{jk}^- - r_{jk}^+ = k_{jk}, j = 1,2,\dots,n; k = 1,2,\dots,p_j$$

$$r_{jk}^-.r_{jk}^+ = 0, j = 1,2,\dots,n; k = 1,2,\dots,p_j$$

$$x_j, r_{jk}^-, r_{jk}^+ \ge 0$$

where $Z_{1,1,\dots,1}$ is the fragment of the aggregate objective function corresponding to the first multi-dimensional interval that is given for each variable as $\langle 0; k_{j1}\rangle$. $Z_{h_1,h_2,\dots,h_n}$ is a fragment of the aggregate objective function corresponding to individual multi-dimensional intervals $\langle k_{jh_j-1}; k_{jh_j}\rangle, \forall j = 1,2,\dots,n$. And $Z_{p_1,p_2,\dots,p_n}$ is a fragment of the aggregate objective function corresponding to the last multi-dimensional interval that is given for each variable as $\langle k_{jp_j-1}; K_j\rangle$ where K_j is the value $K_j = \underset{x_j}{\text{argmax}}(\sum_{j=1}^{n} \boldsymbol{a}_j x_j \le \boldsymbol{b})$.

Having the optimization problem in the form of PWLP-MC, it is possible to utilize multiobjective simplex algorithm for finding of the optimal solution. This algorithm finds all non-dominated solutions of the multiobjective optimization problem. In our problem (PWLP), set of all candidate solutions is subset of the set of all non-dominated solutions. If we search through all non-dominated solutions of the problem then surely on these solutions is the optimal solution of the problem PWLP.

4 Results

In our paper we consider that all partial objective functions are monotone. This fact will ensure in multiobjective simplex algorithm that each iteration of the algorithm will necessarily bring better value of the aggregate objective function than the previous iteration. It can be safely said that in case of monotone partial objective functions is the found candidate solution truly optimal. Although it is possible to reach this solution in different ways through the solution space. The problem is that the optimal solution might not be the only one and one could also find the alternative optimal solution(s) in a different section of the solution space. It is then necessary to inspect all unvisited branches of the algorithm to find out if there is, perhaps, the alternative optimal solution. We should also add that our approach can be also applied for the problems with minimizing objective function. The general approach can be described by the schematics in the figure 3.

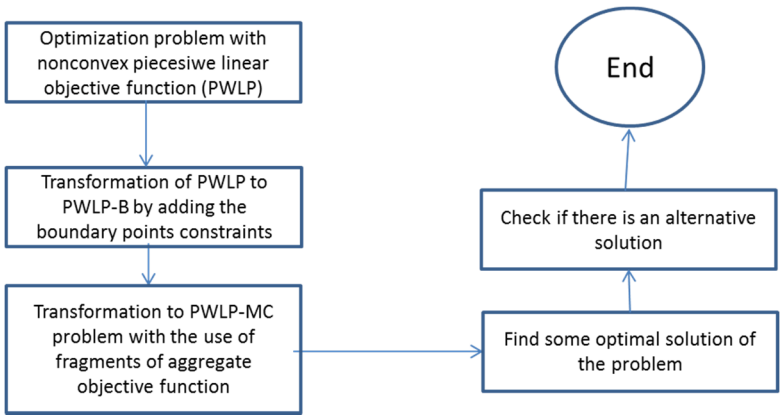


Figure 3: The algorithm flowchart.

5 Conclusion

In our paper, we proposed a different approach to the solution of optimization problems with nonconvex monotone piecewise linear functions. Our approach is based on decomposition of the objective function to its fragments according the multi-dimensional intervals within the solution space. Multiobjective simplex algorithm is known to be (in most cases) computationally tractable and deterministic after finite number of steps. Due to the monotone nature of the partial objective functions, the algorithm heads towards the optimal solution in unambiguous way. In our research, we performed a number of calculations on chosen case problems. These case studies are not presented due to the lack of space. This approach can be especially applicable for solution of the optimization models where the objective function represents the cost or production function. These functions, if approximated, can be often non-convex piecewise linear. We are currently developing the general algorithm that will be able to find the optimal solution also for non-monotone objective functions while we consider the computational difficulties caused by degeneracy or symmetry of the optimization problem.

References

[1] Benchekroun, B. and Falk, J.E.: A nonconvex, piecewise linear optimization problem. *Computers & Mathematics with Applications* **(21)** (1991), 77-85.

[2] Cohon, J.L.: *Multiobjective programming and planning*. New York: Academic press, 1978

[3] Falk, J.E. and Soland R.M.: An algorithm for separable nonconvex programming problems, *Management Science* **(15)** (1969), 550-569.

[4] Houška, M.: Modified simplex algorithm for solving nonconvex PWLP models, *In Proceedings of Mathematical Methods in Economics, PEF ČZU Prague*, 2003.

[5] Houška, M. and Brožová, H.: The method of cost evaluated distance variables for solving piecewise linear programming models. *Proceedings of the 20th international conference MME, Ostrava*, 2002.

[6] Kameshwaran, S. and Narahari, Y.: Nonconvex piecewise linear knapsack problems. *European Jounral of Operational Research* **(192)** (2009), 56-68.

[7] Katzenelson, J.: An algorithm for solving nonlinear resistor networks. *Bell System Technical Journal* **(44)** (1965), 1605-1620.

[8] Lemke, C.E.: On complementary pivot theory. *Proceedings og 5th summer seminar Mathematics of the Decision Sciences*, (1967).

[9] Polisetty P.K. and Gatzke E.P.: Decomposition based algorithm for nonconvex mixed-integer nonlinear programs using piecewise linear relaxations. *International Transactions in Operations Research* (2005).

[10] Van Bokhoven, W.M.G. and Leenaerts, D.M.W.: *Piecewise linear modeling and analysis*. New York: Springer, 2010.

[11] Zelený, M.: *Multiple criteria decision making*. Berlin and New York: Springer-Verlag, 1976.

Generalized EOQ Type Inventory Models with Time Dependent both Demand Rate and Holding Cost Rate

Jiří Hofman[1], Ladislav Lukáš[2]

Abstract. The paper deals with some generalized inventory control models which are linked to the well-known EOQ model. There are dynamic periodic and deterministic models with common feature being represented by instantaneous replenishment in constant time periods. First, we present models assuming another than constant rate of demand within an inventory cycle. The demand rate takes polynomial form which is specified by particular interpolation conditions. These conditions give a chance to include various intermediate data to describe inventory level during the cycle with particular details. Second, we assume the holding cost rate to be time dependent as well, though of the first algebraic degree basically. However, the generalization to higher degrees is quite straightforward. The results, in particular – optimal replenishment lot size and optimal total cost per unit of time, yielded from these models with different interpolation conditions are compared with similar quantities provided by classical EOQ model. Sensitivity analysis of optimal quantities upon selected interpolation parameters is presented, too.
Procedure used for symbolic derivation of optimal order quantity is elaborated in detail, and hints for its computer implementation are presented, as well. We use Mathematica not only for numerical calculations but also for symbolic derivations of analytic formulae giving optimal quantities provided by all models discussed. Some code snippets are also presented and discussed thus enabling us to inspect a particular symbolic versatility and performance power of Mathematica.

Keywords: generalized EOQ model, inventory control model, time dependent demand rate, time dependent holding cost rate.

JEL Classification: C54, D29
AMS Classification: 90B05

1 Introduction

In general, the EOQ model is still frequently implemented in many practical inventory control systems regardless its simplicity and limits. This model seeks minimum of total cost under an objective to optimize lot sizing in order to reduce the cost of satisfying demand. Hence, the concept of the EOQ is that there is a trade off between the fixed order cost and the holding cost assuming demand rate being constant exclusively.

Before presenting our concept of generalized EOQ type models, we mention some other closely related models. We know, there are many various inventory models at disposal in literature, now. EOQ model for deteriorating items with exponentially dependent demand rate in which inflation and time value of money are taken into account is presented in [6]. Inventory model presented in [5] is based upon classical EOQ model with continuous reviewing of inventory level and facing deterministic constant demand rate, but assuming supplier side to undergo random supply disruptions causing uncertainty of replenishments. In [2], various lot sizing algorithms applied for inventory control and emphasizing the financial implications of corresponding inventory policies in supply-chain management are discussed. While in [3], good survey of literature reviews in the area of lot sizing problems is presented. There is lot of textbooks covering the topic of inventory control, e.g. [1], and [4], too.

2 Generalized EOQ type models

Classical EOQ model assumes constant demand rate, infinite replenishment rate, zero lead-time, and minimization of total inventory cost per unit time $C(q)$. Let $z(\tau)$ denote an inventory level during a periodic inventory cycle t, with $\tau \in [0, t]$. Further, c_1, and c_3 denotes unit holding cost and fixed cost incurred per order called ordering cost, respectively, T gives total inventory control period, and Q gives aggregate product demand during T,

[1] University of West Bohemia/Faculty of Economics, Dept. of Economics and Quantitative Methods, Husova 11, 306 14 Pilsen, hofman@kem.zcu.cz.
[2] University of West Bohemia/Faculty of Economics, Dept. of Economics and Quantitative Methods, Husova 11, 306 14 Pilsen, lukasl@kem.zcu.cz

thus defining the number of cycles by $n = Q/q = T/t$, and finally, $N(q)$ stands for total inventory cost per inventory cycle.

There is well-known, the optimal replenishment lot size q_{opt}, or order quantity, is solution of the optimization problem

$$\min C(q), \quad q \in]0,\infty[$$
$$C(q) = N(q)/t, \; N(q) = c_3 + \int_0^t c_1 z(\tau)\mathrm{d}\tau, \; z(\tau) = q(1 - \tau/t), \; t = qT/Q. \tag{1}$$

It yields the extremely popular known results $q_{opt} = \operatorname{argmin} C(q) = \sqrt{(2c_3Q/(c_1T))}$, $C(q_{opt}) = \sqrt{(2c_1c_3Q/T)}$, assuming quantities c_1, c_3, T, and Q are given.

In [4], suitable general framework for large family of inventory models is presented in the following form

$$\text{Find } s(t) \in U, \; \text{extrem} \left(\int_0^T \varphi(z(\tau), \lambda)\mathrm{d}\tau \right), \; \forall z(t) \in V, \tag{2}$$

where U, and V denotes set of feasible control, and set of feasible solutions, respectively.

As demand function $d(\tau)$ being given, a feasible solution will be any continuous function $z(\eta)$ satisfying both inventory balance condition (3) and backordering condition (4), for suitably selected set of feasible control U which represents collection of supply processes $s(\tau)$ acceptable from managerial point of view, in practice.

$$z(\eta) = z_0 + \int_0^\eta s(\tau)\mathrm{d}\tau - \int_0^\eta d(\tau)\mathrm{d}\tau, \; \text{at } \eta \in [0,T], \tag{3}$$

$$\mu(\mathrm{ind}\{z(\eta) \le 0 \mid \eta \in [0,T]\}) = \kappa. \tag{4}$$

The quantity κ introduced in (4) enables to distinguish shortages of stock, and prospective backorders, as well. If $\kappa > 0$ then shortage exists and backordering may be accepted, else $\kappa = 0$ the opposite holds, assuming measure $\mu(.)$ being suitably defined, and ind(.) is a set indicator function.

The integral $\int_0^T \varphi(z(\tau), \lambda)\mathrm{d}\tau$ gives an objective functional, e.g. total inventory cost per unit time, depending upon parameters λ given, e.g. c_1, c_3. Initial inventory level $z(0) = z_0$ is given, too.

There is well-known that inventory balance condition could take either the integral form (3), or a differential form (5), which is called local balance condition

$$\mathrm{d}z(\tau)/\mathrm{d}\tau = s(\tau) - d(\tau), \quad z(0) = z_0. \tag{5}$$

The EOQ model assumes $d(\tau) = q/t = Q/T = \text{const}$, and $s(\tau) = q \sum_{i=0}^{\vartheta} \delta(\tau - it)$, where $\delta(.)$ denotes the Dirac function, and ϑ represents the greatest integer obeying relation $\vartheta\, t < \eta$. In such case equation (5) takes very simple form

$$\mathrm{d}z(\tau)/\mathrm{d}\tau = -q/t, \quad z(0) = q, \tag{6}$$

which yields the EOQ classical solution $z(\tau) = q(1 - \tau/t)$, with $z(t) = 0$.

Keeping line with [4] Ch.3.2.4, the generalized EOQ type models release restrictive assumptions of constant demand rate, and constant unit holding cost c_1, both being thus required by EOQ model exclusively. Hence, the total inventory cost per cycle $N(t)$ will take the following form

$$N(t) = c_3 + \int_0^t c_1(\tau)\, z(\tau)\mathrm{d}\tau. \tag{7}$$

However, question arises, which functional form to select for functions $c_1(\tau)$ and $z(\tau)$, in general. We have proposed the simplest possible form – algebraic polynomials. Let denote $p_n(\tau)$ a polynomial of n-th degree. We will assume $c_1(\tau) = p_i(\tau)$, $i \ge 1$, and $z(\tau) = p_j(\tau)$, $j \ge 2$, while linear function $p_1(\tau)$ suits exactly to describe inventory level within the EOQ model.

Advantages of such generalization are following:

- variable unit holding cost $c_1(\tau)$ enables to handle inventory cases of storing deteriorating items, perishable goods, or any other products with their holding costs sensitive to elapsing time during an inventory cycle;
- higher degree of demand rate $d(\tau)$ starting even with the first degree, which corresponds actually to inventory level $z(\tau)$ of the second degree, enables us both to cope more flexibly with time dependent demand during inventory cycle, and also to use prospectively additional intermediate information of inventory level being detected by inventory checks during inventory cycle course;
- these versatile inventory models provide rational platform for building adaptive inventory control systems.

We have yet elaborated the generalized EOQ type inventory models with $i = 0, 1$ and $j = 1, 2, 3$, with various interpolation conditions, which we suppose to denote GEOQ(i, j) in unique fashion. Derivation of analytic formula for q_{opt} for corresponding GEOQ(i, j) model follows eight steps, in general, which are realized in Mathematica utilizing its symbolic computation power:

1. express $c_1(\tau)$ being represented by $p_i(\tau)$ in analytic form using its interpolation data given in symbols;
2. express $z(\tau)$ being represented by $p_j(\tau)$ in pure analytic form using its interpolation data given in symbols, too;
3. calculate $N(t)$ using (7), which is possible in full analytic form because the integrand $c_1(\tau)\, z(\tau)$ having been assumed as product of polynomials takes polynomial form as well;
4. use the substitution $t = qT/Q$ to recast $N(t)$ into $N(q)$;
5. express objective function, i.e. the total inventory cost per unit time $C(q)$, using $C(q) = N(q)/t$;
6. calculate $\mathrm{d}C(q)/\mathrm{d}q$ in pure symbolic form using Mathematica function D;
7. solve the equation $\mathrm{d}C(q)/\mathrm{d}q = 0$ which expresses necessary condition of optimality to get q_{opt} formula desired by function Solve;
8. finally, express formula for minimal value of total inventory cost per unit time $C_{opt} = C(q_{opt})$ by substituting q_{opt} into $C(q)$ in symbolic form.

The proposed procedure is quite straightforward. However, it is rather tedious technically and very error-prone one when doing by hand. So, we decided to use symbolic calculation power of Mathematica to perform such challenging task.

Assuming $\tau \in [0, t]$, the analytic details of elaborated GEOQ(i, j) models are following:

- GEOQ(1,1): $c_1(\tau) = p_1(\tau) = \gamma_h + (\gamma_d - \gamma_h)\,\tau/t$, $c_1(0) = \gamma_h$, $c_1(t) = \gamma_d$,
  $z(\tau) = q(1 - \tau/t)$, $z(0) = q$, $z(t) = 0$;

- GEOQ(0,2): $c_1(\tau) = c_1$,
  $z(\tau) = a_0 + a_1\tau + a_2\tau^2$, $z(0) = q$, $z(t) = 0$, plus one additional condition:
  a) $z(s) = w$, $s = \theta t$, $w = \omega q$, $\theta, \omega \in]0,1[$,
  b) $z'(0) = -\varphi_h\, q/t$, $\varphi_h \in [0,2]$,
  c) $z'(t) = -\varphi_d\, q/t$, $\varphi_d \in [0,2]$;

- GEOQ(1,2): $c_1(\tau) = p_1(\tau) = \gamma_h + (\gamma_d - \gamma_h)\,\tau/t$, $c_1(0) = \gamma_h$, $c_1(t) = \gamma_d$,
  $z(\tau) = a_0 + a_1\tau + a_2\tau^2$, with interpolation conditions like in GEOQ(0,2) model;

- GEOQ(0,3): $c_1(\tau) = c_1$,
  $z(\tau) = a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3$, $z(0) = q$, $z(t) = 0$, plus two additional conditions:
  a) $z(s_k) = w_k$, $s_k = \theta_k t$, $w_k = \omega_k q$, $\theta_k, \omega_k \in]0,1[$, $k = 1,2$, $\theta_1 < \theta_2$, $\omega_1 > \omega_2$,
  b) $z'(0) = -\varphi_h\, q/t$, $z(s_2) = w_2$, $s_2 = \theta_2 t$, $w_2 = \omega_2 q$, $\theta_2, \omega_2 \in]0,1[$, $\varphi_h \in [0,3]$,
  c) $z'(t) = -\varphi_d\, q/t$, $z(s_1) = w_1$, $s_1 = \theta_1 t$, $w_1 = \omega_1 q$, $\theta_1, \omega_1 \in]0,1[$, $\varphi_d \in [0,3]$,
  d) $z'(0) = -\varphi_h\, q/t$, $z'(t) = -\varphi_d\, q/t$, $\varphi_h, \varphi_d \in [0,3]$;

- GEOQ(1,3): $c_1(\tau) = p_1(\tau) = \gamma_h + (\gamma_d - \gamma_h)\,\tau/t$, $c_1(0) = \gamma_h$, $c_1(t) = \gamma_d$,
  $z(\tau) = a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3$, with interpolation conditions like in GEOQ(0,3) model;

Together with EOQ model corresponding to GEOQ(0,1), we have got complete family of sixteen inventory models of EOQ type at our disposal, in general. However, we need to emphasize that the bounds given for interpolation parameters introduced are approximate or upper bounds only, since their particular admissible value combinations must be checked accordingly to necessary condition being imposed upon function $z(\tau)$ to be decreasing, more precisely non-increasing function, on whole interval $[0,t]$. The simplest way how to do it in particular case, it is to inspect plot of $z(\tau)$ directly, which is to have been drawn by Mathematica plotting functions.

The simples case of model GEOQ(0,2) with additional interpolation condition $z(t/2) = \omega q$, $\omega \in [0.25,0.75]$ yields exactly the results given in [4], p.64, as follows

$$q_{opt} = \sqrt{(6c_3Q/(c_1(1+4\omega)T))}, \quad C_{opt} = C(q_{opt}) = c_1(1+4\omega)\sqrt{(2c_3Q/(3c_1(1+4\omega)T))}). \tag{8}$$

3 Some numerical results

First, we shall describe implementation of GEOQ(i, j) inventory models in Mathematica. On the very beginning, we need to derive corresponding formulae for q_{opt}, and C_{opt} in symbolic forms. Then having such formulae, we may substitute particular numerical values of corresponding parameters therein and calculate quantities desired, and plot them prospectively.

In previous paragraph, we have already sketched the steps of procedure for symbolic derivation of such formulae, in general. Here, we select just one model to illustrate it:

- model GEOQ(1,3) case a), i.e. assuming linear function for unit holding cost rate $c_1(\tau)$, and cubic polynomial for inventory level $z(\tau)$ with two additional interpolation conditions of Lagrangean type, in particular.

Since the whole code for all GEOQ(i, j) models developed is rather long, we give some Mathematica snippets, only

– perform symbolic interpolation of $z(\tau) = a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3$:

```
Solve[t*a1c+t^2*a2c+t^3*a3c ==–q && sh*a1c+sh^2*a2c+sh^3*a3c ==–q(1–wh) &&
sd*a1c+sd^2*a2c+sd^3*a3c ==–q(1–wd),{a1c,a2c,a3c},Reals]
```

– extract a_1, and make symbolic substitution $s_k = \theta_k t$, $k = 1,2$:

```
sol=%//Flatten;
a1w=a1c/.sol[[1]];
a1w//FullSimplify;
a1=%/.{s1->θ1*t,s2->θ2*t}//FullSimplify
```

programming notes:

1) the symbol % represents in Mathematica programming language the result of previously calculated expression.

2) command function FullSimplify makes the best symbolic simplification of any symbolic expression, e.g. algebraic one, which Mathematica knows and is able to do.

– integrate $\int_0^t c_1(\tau)\, z(\tau) \mathrm{d}\tau$ with $c_1(\tau) = p_1(\tau) = \gamma_h + (\gamma_d - \gamma_h)\tau/t$, and calculate $C(t) = N(t)/t$:

```
Integrate[(gH+(gD-gH)v/t)*(q+a1 v+a2 v^2+a3 v^3),{v,0,t}]//FullSimplify
c3+%;
%/t//FullSimplify;
```

– recast $C(t)$ into $C(q)$, calculate the derivative in symbolic form, and solve $\mathrm{d}C(q)/\mathrm{d}q = 0$:

```
cqf13=%/.t->q T/Q//FullSimplify
dcqf13dq=D[%,q]//FullSimplify
Solve[dcqf13dq==0,q]//Flatten
qmin13=q/.%[[1]]
```

– set numerical values, make replacement, i.e. substitution of symbolic parameters of the model considered by the numerical values given, and calculate both q_{opt} and $C_{opt} = C(q_{opt})$, e.g. for GEOQ(1,3) model as follows:

```
wgh=2; wgd=0; wθ1=.3; wθ2=.7;
wq=qmin13/.{gH->wgh, gD->wgd,θ1->wθ1,θ2->wθ2}//Re;
wcq=cqf13/.{gH->wgh, gD->wgd,θ1->wθ1,θ2->wθ2,q->wq}//Re;
```

However, using generalized EOQ type models should be made carefully. Variety of possible cases is large, and flexibility, how to describe inventory level function $z(\tau)$ by various interpolation conditions, is great as well.

As an illustration, we select the model GEOQ(0,3) with different interpolation conditions, and in **Fig. 1**, and **Fig. 2** the calculated normalized inventory levels $\zeta(\tau)$, defined by ratios $\zeta(\tau) = z(\tau)/q$, are plotted on normalized replenishment cycle [0,1].

Data for additional interpolation conditions used in **Fig. 1** are following:

- on the left: GEOQ(0,3) case a): (θ_1, θ_2) = {(.15,.55), (.20,.60), (.25,.65), (.30,.70), (.35,.75), (.40,.80), (.45,.85)}, and $\omega_1 = 0.7$, $\omega_2 = 0.3$, constant in all cases;
- on the right: GEOQ(0,3) case d): (φ_h, φ_d) = {(0,2),(1/3,5/3),(2/3,4/3),(1,1),(4/3,2/3),(5/3,1/3),(2,0)};

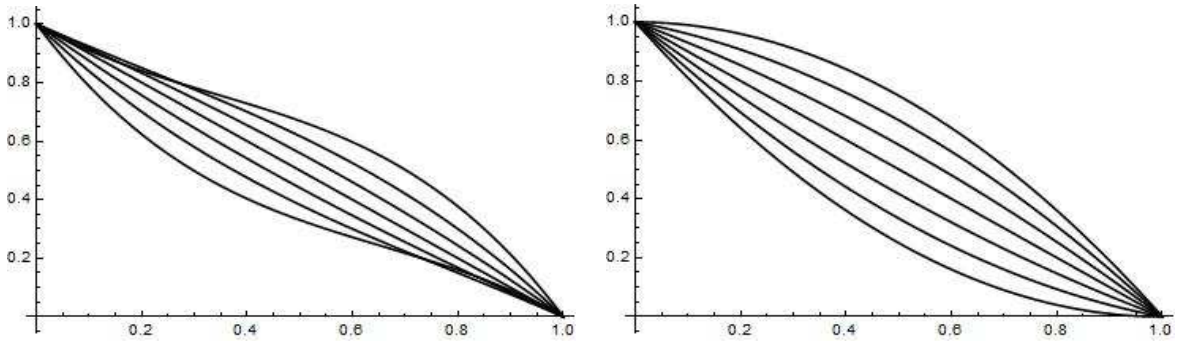


Figure 1 GEOQ(0,3) case a) ~ (left), GEOQ(0,3) case d) ~ (right)

Data for additional interpolation conditions used in **Fig. 2** are following:

- GEOQ(0,3) case b): φ_h = (0, 1/3, 2/3, 1, 4/3, 5/3, 2), θ_2 = (.35, .40, .45, .50, .55, .60, .65), and $\omega_2 = 0.5$, constant in all cases;

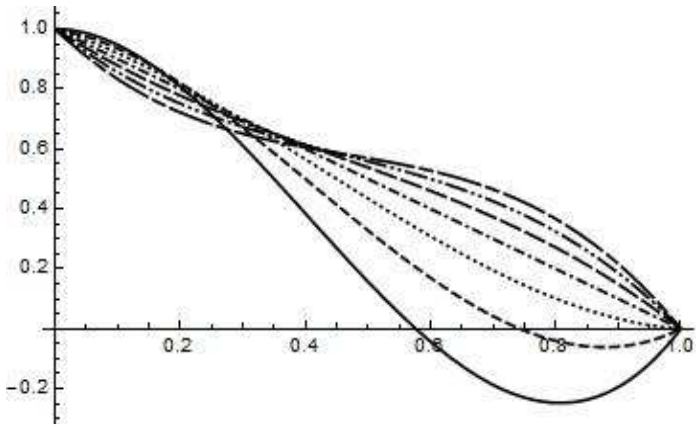


Figure 2 GEOQ(0,3) case a)

First, in both **Fig. 1**, and **Fig. 2**, we may detect linear functions which correspond to classical EOQ model. In **Fig. 1**, all normalized inventory levels $\zeta(\tau)$ are acceptable as being decreasing functions on whole [0,1]. However, in **Fig. 2**, we see that two of constructed functions $\zeta(\tau)$ depicted with full and dashed lines in particular, are not acceptable to represent any inventory level for EOQ type model because both violate the necessary condition, i.e. being non-increasing function on whole replenishment cycle. On the contrary, the others are still acceptable functions to constitute prospective inventory levels.

Next, we present two series of calculated examples with GEOQ(1,3) model which enables linear approximation of unit holding cost rate, and cubic polynomial approximation of inventory level, in gene Main purpose of these calculations, the is sensitivity analysis of optimal quantities q_{opt}, and $C(q_{\text{opt}})$, respectively, upon selected interpolation parameters.

In the first case, we concern ourselves with influence of changing unit holding cost rate $c_1(\tau)$, while $z(\tau)$ to degenerate to $p_1(\tau)$ by properly selected interpolation conditions thereat, i.e. GEOQ(1,3) → GEOQ(1,1), setting $(\theta_1, \omega_1) = (0.3,0.7)$, $(\theta_2, \omega_2) = (0.7,0.3)$. Results are summarized in **Tab. 1**, and depicted in **Fig. 3** on the left.

γ_h	2	1.75	1.5	1.25	1	0.75	0.5	0.25	0
γ_d	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2
q_{opt}	34.64	35.78	37.03	38.43	40	41.78	43.82	46.19	48.99
C_{opt}	46.19	44.72	43.20	41.63	40	38.30	36.51	34.64	32.66

Table 1

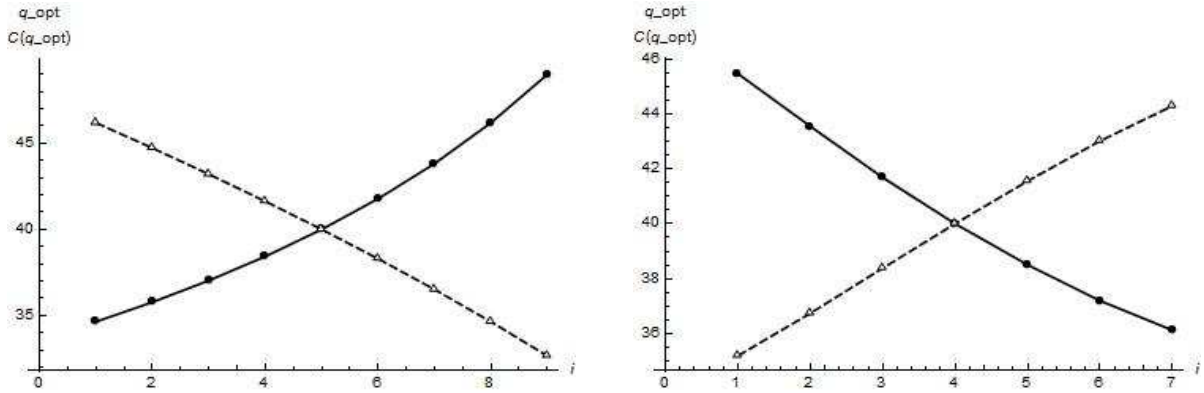


Figure 3 q_{opt}, and $C(q_{\text{opt}})$: Tab.1 GEOQ(1,3) → GEOQ(1,1) ~ (left), Tab.2 GEOQ(1,3) → GEOQ(0,3) ~ (right)

In the second case, we concern ourselves with influence of changing $z(\tau)$ as $p_3(\tau)$, while keeping $c_1(\tau) = c_1$, constant, i.e. GEOQ(1,3) → GEOQ(0,3), setting $\omega_1 = 0.7$, $\omega_2 = 0.3$, $\gamma_h = 1$, $\gamma_d = 1$. Results are summarized in **Tab. 2**, and depicted in **Fig. 3** on the right.

θ_1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
θ_2	0.55	0.6	0.65	0.7	0.75	0.8	0.85
q_{opt}	45.46	43.55	41.68	40	38.51	37.20	36.13
C_{opt}	35.19	36.74	38.38	40	41.55	43.01	44.29

Table 2

4 Conclusions

- Framework of generalized EOQ-type models has been developed which enable both time dependent unit holding costs and demand rate. Their descriptions are based upon additional interpolation conditions expressing intermediate information during replenishment cycle.
- Procedure for symbolic derivation of optimal quantity order for submitted GEOQ(*i*, *j*) models and its implementation in sw Mathematica is described in detail.
- Near future research will be focused on thorough numerical experiments with discussed models and their practical use. Further, we would like to concentrate ourselves also upon a role of inventory, its financial issues in particular, within a broad framework of valuation of firms, as an amount of firm asset allocated in inventory might cause awkward and unexpected effects thereon.

Acknowledgements

The research project was supported by the grant no. 15-20405S of the Grant Agency, Prague, Czech Republic, and the grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Axsaeter, S.: *Inventory Control*. 2-nd ed., Springer, New York, 2006.
[2] Beullens, P.: Revisiting foundations in lot sizing – Connections between Harris, Crowther, Monahan, and Clark. *Int. J. Production Economics* **vol. 155** (2014), 68–81.
[3] Glock, Ch. H., Grosse, E. H., Ries, J. M.: The lot sizing problem: A tertiary study, *Int. J. Production Economics*, 2014, Vol.155, pp. 39-51.
[4] Lukáš, L.: *Probabilistic models in management – Theory of inventory and statistical description of demend* (in Czech: *Pravděpodobnostní modely v managementu – Teorie zásob a statistický popis poptávky*). Academia, Praha 2012.
[5] Snyder, L.V.: A tight approximation for an EOQ model with supply disruptions. *Int. J. Production Economics* **vol. 155** (2014), 91–108.
[6] Tripathi, R. P., Singh, D., Mishra, T.: EOQ Model for Deteriorating Items with exponential time dependent Demand Rate under inflation when Supplier Credit Linked to Order Quantity, *Int. J. Supply and Operations Management*, 2014, Vol.1, Iss.1, pp. 20-37.

Backtesting of value at risk with fuzzy-stochastic variables

Michal Holčapek[1], Tomáš Tichý[2]

Abstract. Since financial quantities – opposed to natural processes – depend on human activity, their modeling is often very challenging. Many scholars therefore suggest for some cases, such as illiquid positions, newly introduced securities or instable market condition, to specify some parts of financial models by means of fuzzy set theory. In this paper we formulate a fuzzy-stochastic model of risk estimation to be solved by simulation approach. The application possibility is shown for the case of market (FX rate) risk and its subsequent backtesting. Although a single position is considered, it can be easily generalized to deal with more complex problem.

Keywords: fuzzy random variable, fuzzy quantile, fuzzy stochastic process, Value at Risk estimation, backtesting

JEL classification: C44
AMS classification: 90C15

1 Introduction

A crucial step of market risk modelling is the estimation of parameters and feeding up of the selected model. This issue is challenging especially for risk sources, which have poor market history (untraded, illiquid, newly introduced assets, etc.). In case that a complex portfolio is in a question, the mutual dependencies and tail dependencies can change rapidly even for assets with quite good liquidity and under normal market conditions.

Assuming a short horizon and a single position, the most important parameter is the volatility. Wrong estimation of the volatility can obviously have serious impact on the risk management and indirectly also on the credibility of the entity (e.g., bank).

For instance, if the estimate is too high, the bank is required to keep higher capital – this can be too costly and, what is worst, this can support its rivals at the market when, e.g., acquiring new clients and more competitive price. Moreover, if too high estimates persists for a longer time, the users start to consider it as a standard feature of the risk model and might, in turn, potentially underestimate the true risk.

On the other hand, if the risk estimation is too low, the capital will not be sufficient to cover losses that probably occur in the near future. Clearly, the entity can call equity holders to provide additional capital, so that the situation, most probably, will not lead to default, but the confidence (into the entity, its risk model, management) will already be lost.

In this contribution we aim on providing a robust model based on fuzzy-random variables and fuzzy-quantiles as introduces by us previously [8, 17] and show its performance within backtesting procedure. Such model can be suitable for the market risk estimation especially for assets newly introduced to the market or with low liquidity or for periods of instability at the market.

The paper is organized as follows. Section 2 is focused on elements of fuzzy sets theory and, especially, on an LU–fuzzy random variable that are used in the contribution. In Section 3 we provide an illustrative example of the market risk modelling related to the gold price. Due to the lack of space we do not pay

[1]VŠB – Technical University of Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic & University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, michal.holcapek@osu.cz

[2]VŠB – Technical University of Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz

special attention to the models of market risk and the risk management process – instead we refer the reader to [15, 16].

2 Fuzzy sets theory

Let $\mathbb{R}$ denote the set of real numbers. A fuzzy number is usually called a mapping $A : \mathbb{R} \to [0,1]$ which is normal (i.e., there exits an element x_0 such that $A(x_0) = 1$), convex (i.e., $A(\lambda x + (1-\lambda)y) \geq \min(A(x), A(y)$ for any $x, y \in \mathbb{R}$ and $\lambda \in [0,1]$), upper semicontinuous and $\mathrm{supp}(A)$ is bounded, where $\mathrm{supp}(A) = cl\{x \in \mathbb{R} \mid A(x) > 0\}$ and cl is the closure operator (see, e.g., [9, 4]). The most popular models of fuzzy numbers are the triangular and trapezoidal models investigated by Dubois and Prade in [5]. Their popularity follows from the simple calculus as the addition or the multiplication of fuzzy numbers. This is also a reason why we can find many recent papers on the approximation of fuzzy numbers by the mentioned models (see, e.g., [1, 2] and the references therein).

2.1 LU–fuzzy numbers

In order to model fuzzy numbers we will use a more advanced model of fuzzy numbers based on the interpolation of given knots using rational splines that was proposed by Guerra and Stefanini in [7] and developed in [14]. This model generalizes the triangular fuzzy numbers and gives a broad variety of shapes enabling more accurate representation of fuzzy real data, nevertheless, the calculus remains still very simple.

It is well known that each fuzzy number has a representation using α-cuts. Recall that the α-cut of a fuzzy number A is the common set $A_\alpha = \{x \in \mathbb{R} \mid A(x) \geq \alpha\}$ for $\alpha \in [0,1]$ and

$$A(x) = \sup\{\alpha \mid \alpha \in [0,1] \ \& \ x \in A_\alpha\}. \tag{1}$$

Since the fuzzy numbers are upper semicontinuous real functions, then each α-cut may be replaced by its endpoints, say u_α^- for the left endpoint and u_α^+ for the right endpoint. Hence, each fuzzy number can be completely represented by two functions $u, v : [0,1] \to \mathbb{R}$ ($u^-(\alpha) = u_\alpha^-$ and $u^+(\alpha) = u_\alpha^+$) such that

1. u^- is a bounded monotonic non-decreasing function which is left-continuous on $(0,1]$ and the right-continuous for $\alpha = 0$,
2. u^+ is a bounded monotonic non-increasing function which is left-continuous on $(0,1]$ and the right-continuous for $\alpha = 0$,
3. $u^-(\alpha) \leq u^+(\alpha)$ for any $\alpha \in [0,1]$.

The arithmetic operations between two fuzzy numbers A and B represented by pairs of functions (u_A^-, u_A^+) and (u_B^-, u_B^+), respectively, can be introduced using a suitable manipulation of the functions u and v. For example, $A + B$ can be simply obtained by $(u_A^- + u_B^-, u_A^+ + u_B^+)$. For further definitions of arithmetic operations, we refer to [14]. Since the modeling of fuzzy numbers and the manipulation with them is not so simple in general, we use the parametric representation of fuzzy numbers proposed by Stefanini et al. in [14] and in particular their matrix form:

$$\mathbf{U}^- = \begin{pmatrix} \mathbf{f}^- \\ \mathbf{d}^- \end{pmatrix} = \begin{pmatrix} f_{\alpha_0}^- & \dots & f_{\alpha_n}^- \\ d_{\alpha_0}^- & \dots & d_{\alpha_n}^- \end{pmatrix} \quad \text{and} \quad \mathbf{U}^+ = \begin{pmatrix} \mathbf{f}^+ \\ \mathbf{d}^+ \end{pmatrix} = \begin{pmatrix} f_{\alpha_0}^+ & \dots & f_{\alpha_n}^+ \\ d_{\alpha_0}^+ & \dots & d_{\alpha_n}^+ \end{pmatrix}$$

where $\alpha_0 = 0 < \alpha_1 < \cdots < \alpha_n = 1$, $\{f_{\alpha_0}^-\}_{i=0}^n$ ($\{f_{\alpha_0}^+\}_{i=0}^n$) and $\{d_{\alpha_n}^-\}_{i=0}^n$ ($\{d_{\alpha_n}^-\}_{i=0}^n$) denote the sequences of knots and slopes, respectively, from which u^- (u^+) is derived as a piecewise rational cubic Hermite parametric function (see [6, 14]). The set of all such LU–fuzzy numbers will be denoted by $\mathcal{F}_{LU}$.

3 Fuzzy random variable and its presentation

Here, we follow the approach to fuzzy random variable proposed by Kwakernaak [11, 12] and later formalized in a clear way by Kruse and Meyer [10]. In this approach, a fuzzy random variable is viewed as a fuzzy perception/observation/report of a classical real-valued random variable. Moreover, fuzzy random variables require special tools of their presentation – we present some alternatives in the second part of this section.

3.1 Fuzzy random variable

A fuzzy random variable is simply defined (in Kwakernaak and Kruse-Meyer sense) as a mapping that assigns a fuzzy number from a certain set of fuzzy numbers to each elementary event in such a way that α-cuts (over all elementary events) define random intervals. Here, we use the set of LU–fuzzy numbers from $\mathcal{F}_{LU}$ to represent the values of fuzzy random variables. Following the Kwakernaak and Kruse-Meyer idea, we propose an appropriate definition of fuzzy random variable for parametrically defined LU–fuzzy numbers based on random matrices.

Definition 1. Given a probability spaces $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathcal{F}_{LU}$ is said to be a *fuzzy random variable* (or FRV for short) if there exists a partition $0 = \alpha_0 < \cdots < \alpha_n = 1$ of the interval $[0, 1]$ and mappings

$$\mathbf{F}^-, \mathbf{F}^+, \mathbf{D}^-, \mathbf{D}^+ : \Omega \to \mathbb{R}^{n+1}$$

such that $p_i \circ \mathbf{F}^-$, $p_i \circ \mathbf{F}^+$, $p_i \circ \mathbf{D}^-$, $p_i \circ \mathbf{D}^+$, where p_i denotes i-th projection, are real-valued random variables for any $i = 0, \ldots, n$ and $X(\omega)$ is determined by random matrices

$$\begin{aligned} \mathbf{U}^-(\omega) &= \begin{pmatrix} \mathbf{F}^-(\omega) \\ \mathbf{D}^-(\omega) \end{pmatrix} = \begin{pmatrix} p_0 \circ \mathbf{F}^-(\omega) & \ldots & p_n \circ \mathbf{F}^-(\omega) \\ p_0 \circ \mathbf{D}^-(\omega) & \ldots & p_n \circ \mathbf{D}^-(\omega) \end{pmatrix} \\ \mathbf{U}^+(\omega) &= \begin{pmatrix} \mathbf{F}^+(\omega) \\ \mathbf{D}^+(\omega) \end{pmatrix} = \begin{pmatrix} p_0 \circ \mathbf{F}^+(\omega) & \ldots & p_n \circ \mathbf{F}^+(\omega) \\ p_0 \circ \mathbf{D}^+(\omega) & \ldots & p_n \circ \mathbf{D}^+(\omega) \end{pmatrix}. \end{aligned} \tag{2}$$

Note we assume in the definition of the fuzzy random variable the same probability space for knots as well as slopes from which LU–fuzzy numbers are derived. Nevertheless, one could admit that the randomness of knots and slopes may be modeled in different probability spaces.

Definition 2. We say that two FRVs X and Y are *independent (identically distributed)* provided that $p_i \circ \mathbf{F}_X^-$, $p_i \circ \mathbf{F}_X^+$, $p_i \circ \mathbf{D}_X^-$, $p_i \circ \mathbf{D}_X^+$ and $p_i \circ \mathbf{F}_Y^-$, $p_i \circ \mathbf{F}_Y^+$, $p_i \circ \mathbf{D}_Y^-$, $p_i \circ \mathbf{D}_Y^+$ are independent (identically distributed), respectively, for any $i = 0, \ldots, n$.

Note that using the interpolation the resulting FRVs are the FRVs in the Kruse-Meyer sense, i.e., all α-cuts are results of random intervals.

3.2 Presentation of fuzzy random sample

One of the simplest methods to present data from a random sample is to construct a histogram. In order to present fuzzy data represented by the LU-fuzzy numbers obtained from a random procedure we will extend the concept of histogram. Our approach is more or less straightforward and uses α-cuts for the construction.

Let us suppose that $A_1, \ldots, A_m$ are LU–fuzzy numbers which are values of FRVs $X_1, \ldots, X_m$ which are independent and identically distributed.[1] Let the suitable interval be divided in M mutually disjoint bins B_j of a bin width h covering the considered interval. For each $\alpha \in [0, 1)$ we can determine the histogram as follows:

$$\hat{f}_{h,\alpha}(x) = \frac{1}{mh} \sum_{i=1}^{m} \sum_{j=1}^{M} I(x \in B_j) \frac{\int_{u_{A_i}^-(\alpha)}^{u_{A_i}^+(\alpha)} \chi_{B_j}(y) dy}{u_{A_i}^+(\alpha) - u_{A_i}^-(\alpha)}, \tag{3}$$

and similarly for $\alpha = 1$ we get:

$$\hat{f}_{h,1}(x) = \frac{1}{mh} \sum_{i=1}^{m} \sum_{j=1}^{M} I(x \in B_j) I(x_i \in B_j), \tag{4}$$

[1] Let us stress that the parameters of LU–fuzzy numbers need not to be determined under the same probability distribution in general.

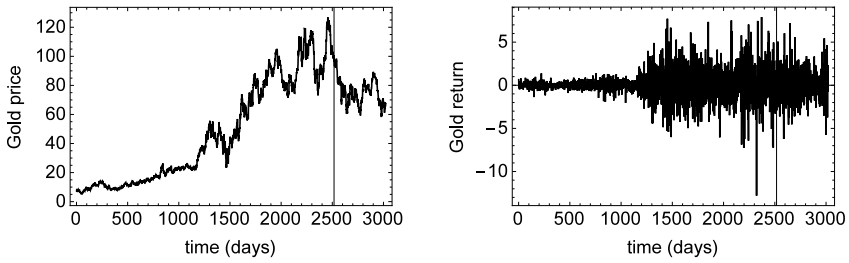


Figure 1 Evolution of gold prices (left) and thier returns (right) over 2003–2014

where $x_i = u^-_{A_i}(1) = u^+_{A_i}(1)$ and $(\cdot)$ denotes the indicator function:

$$I(A) = \begin{cases} 1, & \text{if } A \text{ holds,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

A collection $\{f_{h,\alpha} \mid \alpha \in [0,1]\}$ of histograms defined above will be called the *fuzzy histogram*. Given a fuzzy random sample $A_1, \ldots, A_m$ and a number $q \in (0,1)$, a *q-th fuzzy quantile* is an LU-fuzzy number determined by the matrices

$$\mathbf{U}^-_p = \begin{pmatrix} f^-_{\alpha_0,q} & \cdots & f^-_{\alpha_n,q} \\ d^-_{\alpha_0,q} & \cdots & d^-_{\alpha_n,q} \end{pmatrix} \quad \text{and} \quad \mathbf{U}^+_p = \begin{pmatrix} f^+_{\alpha_0,q} & \cdots & f^+_{\alpha_n,q} \\ d^+_{\alpha_0,q} & \cdots & d^+_{\alpha_n,q} \end{pmatrix},$$

where $f^{\mp}_{\alpha_i,q}$ denotes the common q-th quantile for the values $f^{\mp}_{A_1,\alpha_i}, \ldots, f^{\mp}_{A_m,\alpha_i}$ and $d^{\mp}_{\alpha_i,q}$ the corresponding slope for the knot $f^{\mp}_{\alpha_i,q}$.[2]

4 Illustrative example

In order to present the power of the fuzzy–quantiles in more details, let us assume a price of gold. Although gold can be regarded as highly liquid asset, its price as well as the volatility of its returns can be strongly affected by market distress situation, changes of believes of large investors, including governments, and as well as their fears.

Let us assume a time series of gold prices since January 2, 2003 until December 31, 2014 (daily closing prices were downloaded from www.yahoo.com). First ten years, i.e., 2003–2012 will be used for the construction of the LU-fuzzy volatility. Next, the last two years data will be used for backtesting purposes.

Figure 1 shows the evolution of gold prices over given period (on the left) and discretely calculated returns (on the right). In both cases the vertical line separates charts into estimation (2003-2012) and backtesting subperiods (2013–2014). Apparently, the price for sharply rising starting with year 3, which was, however, interrupted by several large drops. Such behaviour results into much higher volatility in recent years, or even instability of the price. Note that the backtesting subperiod shows rather decrease in the price, while volatility is kept still high.

The estimation period shows several distinct regimes of the price variability, which can be easily used for construction of LU-fuzzy volatility in line with [3]. Obviously, the shape of the LU-fuzzy volatility can vary a little during the time as we gain new information. Notwithstanding, the LU-fuzzy volatility will look more or less as pictured on Figure 2 with vertical line showing the crisp volatility number.

Knowing the volatility we can now follow the simulation procedure suggested in [8] and estimate the Value at Risk, obviously, as an LU–fuzzy quantile, see [8] for more details. Now we can use the remaining time series of gold prices and run the backtesting procedure, i.e., comparing the estimates of VaR (in terms of LU–fuzzy number) with the true losses. Obviously, the number of exceptions will be closed to the assumed number only if our estimate of the price returns volatility will match the true behaviour.

[2] We consider here the simplest technique to derive q-th quantile when the value Mq is round up (if it is necessary) to the next integer to obtained the appropriated index for the quantile (no estimation is used).

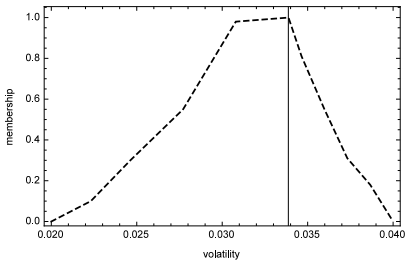


Figure 2 Shape of the LU-fuzzy volatility of gold returns as obtained over 2003–2012

Moreover, since each degree of uncertainty comes with interval volatility, which, in turn, implies interval VaR, the capital to cover the risk for a give level of uncertainty is also an interval. However, in reality, the manager should select either the minimal or maximal value, depending on the kind of the position and policy (risk, risk averse) followed by the entity.

	α			
volatility	1%	5%	10%	0.1%
σ_{min}	15	39	75	4
$\sigma_1(-)$	11	32	62	2
$\sigma_2(-)$	9	27	49	2
$\sigma_3(-)$	7	20	47	1
$\sigma_4(-)$	4	17	39	1
$\bar{\sigma}$	4	13	30	1
$\sigma_4(+)$	2	12	23	1
$\sigma_3(+)$	2	8	19	0
$\sigma_2(+)$	1	6	16	0
$\sigma_1(+)$	0	4	15	0
σ_{max}	0	3	9	0
assumption	5	25	50	1/2

Table 1 Observed exceptions for various levels of volatility

As an example, in Table 1 we provide observed number of exceptions of daily VaR estimates over 2013-2014 (about 505 observations in total) assuming volatility at different nodes of the LU-fuzzy number together with the assumed number of exceptions. The backtesting procedure is evaluated for four distinct probability levels.

Let us assume the standard level for risk management in bank (Basel Accord), which is 1% probability level. The crisp number of volatility would lead to almost the same number of exceptions (4 v. 5); the same is true for LU-fuzzy volatility with membership one. If we decrease the membership, i.e. the level of certainty, we get interval volatility, which leads to two extreme VaR's – each showing a distinct number of exceptions. Obviously, maximal volatility is the most conservative estimate and indicates much lower number of exceptions as should be assumed, as well as much higher capital then would be efficient.

However, if we increase the probability level to 5% or even 10% we can see that the crisp level will on average lead to overestimation of the risk level. Adequate risk level would be obtained with rather aggressive $\sigma_2(-)$. By contrast, when we are interested in risk levels in tails, such as $\alpha = 0.1\%$, see the last column, we should rather select conservative value of $\sigma_2(+)$. The reason of such behaviour might be that the underlying distribution of gold price returns is far from the assumptions of normality.

5 Conclusion

Market risk estimation is a crucial part of the management process in financial institutions. In this paper we have suggested utilization of a methodology based on fuzzy numbers as a tool to estimate VaR and subsequently the suitable amount of capital to be kept to cover potential future risk and we have also showed what might be the impact of backtesting of VaR specified as an LU-fuzzy number.

Obviously, if we can detect which processes is followed by the price returns, we should used it. However, in case that we do not have enough information for its reliable estimation, application of Fuzzy sets theory or even Uncertainty theory [13] can provide us complex picture about potential future impacts. In the future, the methodology should be extended to deal with more variables (portfolio risk).

Acknowledgements

The authors acknowledge support provided by the Czech Science Foundation (GACR) under project 13-13142S and through SP2015/15, an SGS research project of VSB-TU Ostrava, and furthermore by the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296 and the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] Ban, A.I. On the nearest parametric approximation of a fuzzy number – revisited. *Fuzzy Sets and Systems* 160(21), 3027–3047, 2009.

[2] Ban, A.I. Triangular and parametric approximations of fuzzy numbers – inadvertences and corrections. *Fuzzy Sets and Systems*, 160(21), 3048–3058, 2009.

[3] Czudek, P. *Assessment of Alternative Methods for Parameter Estimation of Investment Instruments Price Models.* Master thesis, Faculty of Economics, VSB-TU Ostrava, 2014.

[4] Dubois, D., Prade, H. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9, 613–626, 1978.

[5] Dubois, D., Prade, H. *Fuzzy sets and systems. Theory and applications.* Mathematics in Science and Engineering, Vol. 144. New York etc.: Academic Press. XVII, 393 p., 1980.

[6] J. A. Gregory. Shape preserving spline interpolation. *Comput. Aided Design*, 18: 53-57, 1986.

[7] Guerra, M. L., Stefanini, L. Approximate fuzzy arithmetic operations using monotonic interpolations. *Fuzzy Sets and Systems* 150(1), 5–33, 2005.

[8] Holčapek, M., Tichý, T. Option Pricing with fuzzy parameters via Monte Carlo simulation. *Communications in Computer and Information Science* **211**(1), 25–33, 2011.

[9] Klir, G.J., Yuan, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, New Jersey, 1995.

[10] Kruse, R., Meyer, K. D. *Statistics with vague data.* Theory and Decision Library. Series B: Mathematical and Statistical Methods, 6. Dordrecht etc.: D. Reidel Publishing Company, a member of the Kluwer Academic Publishers Group. VII, 279 p., 1987.

[11] Kwakernaak, H. Fuzzy random variables – I. Definitions and theorems. *Information Sciences* 15, 1–29, 1978.

[12] Kwakernaak, H. Fuzzy random variables – II. Algorithms and examples for the discrete case. *Information Sciences* 17, 253–278, 1979.

[13] Liu, B. *Uncertainty Theory.* 5th edition. Berlin: Springer, 2015.

[14] Stefanini, L., Sorini, L., Guerra, M. L. Parametric representation of fuzzy numbers and application to fuzzy calculus. *Fuzzy Sets Syst.* 157(18), 2423–2455, 2006.

[15] Stiborová, E., Sznapková, B., Tichý, T. Comparison of market risk models with respect to suggested changes of Basel Accord. *Acta Oeconomica* **64** (S2), 257–274, 2014.

[16] Tichý, T. *Lévy Processes in Finance: Selected applications with theoretical background.* SAEI, vol. 9. Ostrava: VB-TU Ostrava, 2011.

[17] Tichý, T., Holčapek, M. (2011). Simulation methodology for financial assets with imprecise data. In *Proceedings of Mathematical Methods in Economics*, Praha: VSE Praha, 2011.

Strong Markov perfect general equilibrium with innovation

Milan Horniaček[1], Ľubica Šimková[2]

Abstract. We analyze general equilibrium in an oligopolistic economy with innovation. The model has the form of a stochastic game. Firms, labor unions, and commercial banks are players in it. Single period production possibility sets of firms and banks can be changed by innovation. A strong Markov perfect general equilibrium is a solution concept in the analyzed game. It is a profile of Markov strategies such that no coalition of players in no subgame can increase average discounted expected real wealth of its stakeholders. We show that it exists in our model for any discount factor.

Keywords: general equilibrium, innovation, oligopoly, strong Markov perfect equilibrium.

JEL Classification: C73, D43, D59
AMS Classification: 91A10, 91A80

1 Introduction

In the present paper, we analyze a stochastic game modelling general equilibrium in an oligopolistic economy with innovation changing single period production possibility sets of firms and banks. Innovation results from a successful research and development (henceforth, R&D). The latter has stochastic nature and is the only source of uncertainty in our model. Firms, banks, and labor unions are players in the game. A strong Markov perfect general equilibrium (henceforth, SMPGE) is the solution concept that we use. It is an application of Rubinstein's [4] strong perfect equilibrium to Markov strategies in our game. (See [3] for characterization of Markov strategies.) It is a profile of Markov strategies with the property that no coalition of players in no subgame can increase average discounted expected real wealth of its stakeholders (i.e., shareholders or members).[3] A profile of strategies of members of a coalition maximizes average discounted real wealth of its stakeholders if there does not exist another profile of strategies that (with unchanged strategies of the players outside the coalition) allows its stakeholders to increase sum of average discounted expected consumed quantities of each consumption good and to decrease sum of average discounted expected supplies of each labor service to each firm and bank (that is not yet zero). We show that an SMPGE exists in our model for any discount factor. We analyze its welfare properties in our accompanying paper [5].

2 Model

Throughout the paper, N is the set of positive integers and R denotes the set of real numbers. We endow each finite dimensional real vector space with the Euclidean topology and each infinite dimensional Cartesian product of finite dimensional real vector spaces with the product topology. We use the following symbols for relations between vectors. Let $a \in R^n$ and $b \in R^n$. Then $a \geq b$ implies that $a_j \geq b_j$ for each $j \in \{1,\ldots,n\}$, $a > b$ implies that $a \geq b$ but $a \neq b$, and $a \gg b$ implies that $a_j > b_j$ for each $j \in \{1,\ldots,n\}$. For any subset A of a vector space, $con(A)$ denotes its convex hull.

$J \cup B \cup L$ is the finite set of players in the analyzed game.[4] J (B, L) is the set of firms (commercial banks, labor unions). We denote the analyzed stochastic game with discount factor $\delta \in (0,1)$ by $\Gamma(\delta)$. It is played in

[1] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 82105 Bratislava, Slovakia, milan.horniacek@fses.uniba.sk.
[2] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 82105 Bratislava, Slovakia, lubica.simkova@fses.uniba.sk.
[3] Maximization of the real wealth of shareholders of a firm was introduced into the literature by [2].
[4] Taking into account space limitations, we describe here only basic features of the model that are essential for the proof of the existence of an SMPGE. Further details of the model can be found in our paper presented at SAET 2015 conference in Cambridge, available at the conference website http://saet.uiowa.edu/papers/2015/horniaceksimkovaSAET.pdf

periods numbered by positive integers. Each player observes past actions of all other players. $G = G_C \cup G_J \cup G_B \cup G_L$ is the set of goods in the model. G_C (G_J, G_B, G_L) is the countable infinite set of new consumption goods (producer goods, banking products, labor services). Since product innovation leads to emergence of new goods and $\Gamma(\delta)$ has infinite time horizon, we allow countable infinite set of goods. Nevertheless, there exists $\bar{n}_C \in N \setminus \{1\}$ ($\bar{n}_J \in N \setminus \{1\}$, $\bar{n}_B \in N \setminus \{1\}$, $\bar{n}_L \in N \setminus \{1\}$) such that, in each single period, at most $\bar{n}_C$ consumption goods ($\bar{n}_J$ producer goods, $\bar{n}_B$ banking products) can be produced and at most $\bar{n}_L$ labor services can be provided. We assume that all consumption goods are differentiated. For each $g \in G$, $y_g^{\max} > 0$ is the maximal amount of good g that can be produced (provided) in a single period, $p_g^{\max}$ is the upper bound and $p_g^{\min}$ the lower bound on its price. We assume that $\sup_{g \in G} y_g^{\max}$ and $\sup_{g \in G} p_g^{\max}$ are finite and $\inf_{g \in G} y_g^{\max} > 0$, $\inf_{g \in G} p_g^{\min} > 0$.

I is the finite set of infinitely living consumers. Their endowments by labor services depend on the set of goods that can be produced. Each consumer makes his decisions (on consumption, savings, borrowing from banks, deposits to and withdrawals from bank accounts, and labor supply) on the basis of maximization of his average discounted expected utility in the infinite horizon. For each $i \in I$ his average discounted expected utility is computed using his single period vonNeumann - Morgenstern utility function u_i (which is the same one in each period). Its argument is a vector containing consumption of non-durable consumption goods, available quantities of durable consumption goods of all vintages, and provision of each labor service to each firm and bank. We assume that it is generic, strictly concave, and such that each consumer's maximization problem has the unique solution.

A single period production possibility set of $k \in J \cup B$, denoted by Y_k, contains all netput vectors that are technologically feasible for k. (In a netput vector, inputs have a negative sign and outputs have a positive sign.)

Assumption 1. For each $k \in J \cup B$, (i) $0 \in Y_k$, (ii) if $y_k \in Y_k$ and $y_{kg} = 0$ for each $g \in G_L$, then $y_{kg} \le 0$ for each $g \in G$, (iii) if $y_k \in Y_k$ and $y_k' \le y_k$, then $y_k' \in Y_k$, (iv) Y_k is closed, (v) the upper boundary of Y_k is convex, and (vi) there exists $\varphi > 0$ such that for each $y_k \in Y_k$ with at least one positive component $\dfrac{\sum_{g \in G} \max\{y_{kg}, 0\}}{-\sum_{g \in G_L} y_{kg}} < \varphi$.

Firms and banks can change their production possibility sets in the following periods by successful R&D. New production possibility sets also satisfy Assumption 1. Individual firms and banks, as well as their coalitions, can conduct R&D. An R&D project can last at most $\bar{T}_{R\&D} \in N$ periods.

In order to keep the model as simple as possible, we assume that - except for credit contracts - all contracts are concluded for one period. A contract for delivery of a producer good specifies the delivered quantity and the unit price. It can be concluded between a firm as the seller and a firm or a bank as the buyer. It is concluded if and only if the seller's and the buyer's proposal coincide.[5] (If a seller and a buyer coincide, it is a planned production of a producer good for own use.) A credit contract specifies the lent amount, maturity of the loan, and the interest rate per period. It can be concluded between a bank as the lender and a firm or another bank as the borrower. A credit contract is concluded if and only if the bank's proposal and the customer's proposal coincide. For consumers, each bank $b \in B$ specifies in each period $t \in N$ the interest rate for deposits and the interest rate per

[5] This is a shortcut for modelling of the bargaining between a seller and a buyer. An attempt to model this bargaining would make the model much more complicated and longer. Moreover, there does not exist a widely accepted model of finite horizon bargaining that could be applied to model bargaining between a seller and a buyer. Even if there were such model, taking into account strategic linkages between periods, the players could agree to deviate from the behavior consistent with it.

If a potential supplier is not interested in delivering good *g* to some potential buyer, or a potential buyer is not interested in buying good *g* from some potential supplier, he proposes zero quantity and price $p_g^{\min}$ for *g*. The same convention applies also to proposals of credit contracts and labor contracts.. A bank cannot completely discourage a firm or another bank from having a deposit in it, but it can set the lowest possible interest rate for undesirable customers.

period for consumer credit provided in period *t*. Each $b \in B$ in each period $t \in N$ specifies for each $k \in J \cup (B \setminus \{b\})$ interest rate for deposits made by *k* in *b* in period *t*.

A labor contract for period $t \in N$ can be concluded in period $t-1$ between each labor union as the seller and each firm and each bank as the buyer. It specifies hourly wage and the upper bound on the employment (in hours) of a labor service. It is concluded if and only if the seller's and the buyer's proposal coincide. Wage rates specified in labor contracts concluded by labor unions are used by consumers in decisions on their labor supply.

Each period $t \in N$ is divided into eight phases.[6] Each firm and bank makes a decision on deposits to banks in each phase in which it has revenue. It makes a decision on withdrawal from its bank accounts in each phase in which it incurs expenditures.

In the first phase (of each period other than the first one), firms and banks learn whether R&D projects, in which they participated in the preceding period, were successful. R&D projects that were successful, as well as those that were not successful but have already been carried out for $\overline{T}_{R\&D}$ periods, are terminated. Further, firms make proposals of contracts for delivery of producer goods to other firms and banks, and firms and banks make proposals of contracts for purchase of producer goods to other firms. Firms capable of production of consumption goods make decisions on production of these goods. Firms and banks make decision on starting new individual R&D projects (on number of started projects and for each project on planned inputs and desired production possibility set). They also make proposals of contracts on joint R&D to other firms and banks. Proposals of contracts for delivery of producer goods and decisions on production of consumption goods by each firm should be technologically feasible with respect to its input vector (which contains produced inputs purchased in the preceding periods and labor services secured by the contracts concluded in the preceding period). In evaluating input requirements, a firm or bank should take into account also input requirements of R&D projects continuing from the preceding periods, R&D projects it decided to start in the current period, and proposals of joint R&D projects made in the current period. Banks announce the deposit interest rates for consumers' deposit accounts, the individual interest rates for deposit accounts of firms and other banks, and the interest rates for consumer credits. Banks pay the interest on deposits made in the preceding period. Firms and banks choose dividends per share and pay the dividends. The dividend per share of $k \in J \cup B$ should not exceed *k*'s profit in the preceding period minus the sum appropriated in the preceding period as a consequence of unpaid debts divided by number of *k*'s shares. (The profits in period zero are given.)

In the second phase, production and delivery of producer goods by firms according to concluded contracts and production of consumption goods and announcement of their prices by firms capable of producing them takes place. If firm $j \in J$ buys producer goods from firm $k \in J \setminus \{j\}$ and the value of its purchases exceeds the value of its sales to *k*, it pays to *k* only the difference between these values. Firms and banks carry out R&D under all new projects (that they decided to start in the preceding phase, either individually or on the basis of contracts), as well as under projects continuing from the preceding periods.

In the third phase, wages are paid.

In the fourth phase, sale of consumption goods takes place.

In the fifth phase, banks make proposals of contracts for granted credit to firms and other banks and firms and banks make proposals of contracts for received credit to banks.

In the sixth phase, installments (including interest) of the credits granted in the preceding periods are paid. If a firm or a bank is unable to pay installments of its debt, the creditor banks appropriate its profits in the current and following periods until the owned sum is repaid. If there are two or more creditor banks, they divide the debtor's profit in the ratios given by the unpaid sums.[7] If a consumer is unable to pay his debt, creditor banks divide between them his financial resources in the ratios given by the unpaid sums.

[6] Description of this division is needed in order to make the specification of the model complete. Nevertheless, we do not give a detailed description here because we do not need it to prove our results. It is enough to ensure that the activities within a period and in successive periods take place in the order compatible with technological and logical requirements (e.g., production and sale of produced inputs precede their use in production). We do not view the phases of a period as calendar weeks or months. Therefore, the interest on deposits made in period *t* does not depend on the phase of period *t* in which it was made.

[7] For low values of the discount factor this appropriation rule can lead to moral hazard on the part of firms and banks. Nevertheless, in equilibrium (where each bank maximizes real wealth of its shareholders with respect to given strategies of all other players) banks take this danger into account in formulating their proposals of credit contracts. Also, in an SMPGE all consumers pay all installments of all their debts

In the seventh phase, labor unions make proposals of contracts for delivery of labor services (in the following period) to firms and banks and firms and banks make proposals of contracts for purchase of labor services to labor unions.

In the eighth phase, members of each labor union apply for jobs in the firms with which their labor union concluded a labor contract in the seventh phase.

In $\Gamma(\delta)$ we restrict attention to pure Markov strategies. We denote the set of pure Markov strategies of player $k \in J \cup B \cup L$ by $\tilde{S}_K$ and set $\tilde{S} = \Pi_{k \in J \cup B \cup L} \tilde{S}_K$. A pure Markov strategy of a player depends only on payoff relevant elements of a non-terminal history. We denote by S the subset of $\tilde{S}$ that contains only pairs of contract proposals of a buyer and a seller that coincide. (Starting from any $s \in \tilde{S}$, the latter property can be secured by replacing pairs of contract proposals that do not coincide by a pair of proposals that coincide and specify zero delivered quantity.) S^+ is the subset of S containing only profiles of continuous strategies (i.e., strategies that are continuous in payoff relevant elements of a non-terminal history). Since the outcome of a strategy profile (i.e., inputs, outputs, prices, and R&D activities) depends only on concluded contracts with positive traded quantity to which it leads and on its prescriptions after histories consistent with it, we can restrict attention to S^+. SMPGE on S^+ is immune to deviations by any coalition to any profile of its strategies (even non-Markov one).

We denote by $\Im$ the set of states. Each $\sigma \in \Im$ contains all non-terminal histories that have the same payoff relevant elements. $\Im_1$ is the set of states in the first phase of a period. We denote the initial state by σ_0. $\Gamma_{(\sigma)}(\delta)$ is the class of subgames following histories belonging to σ. Restriction of any set or function defined for $\Gamma(\delta)$ to $\Gamma_{(\sigma)}(\delta)$ is indicated by subscript "(σ)". For each $D \in 2^{J \cup B \cup L} \setminus \{\varnothing\}$, function γ_D assigns to each $s \in con(\tilde{S})$ the vector of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks multiplied by minus one by stakeholders of players in D. It is continuous on S^+. For $D \in 2^{J \cup B \cup L} \setminus \{\varnothing\}$ and $s \in con(\tilde{S})$, $A_D(s)$ is the set of all vectors of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks (multiplied by minus one) that are feasible for stakeholders of players in D when the players follow s.

For each $s \in \tilde{S}$, $\Im_1(s)$ is the subset of $\Im_1$ containing states that occur with a positive probability when the players follow s. For each $s \in \tilde{S}$, each $\sigma \in \Im_1(s)$, and each $t \in N$, $\pi(s, \sigma, t)$ is the probability of occurrence of state σ in period t when the players follow s. For each $s \in \tilde{S}$ and each $\sigma \in \Im_1(s)$, let $N(s, \sigma) \subset G$ be the finite set of goods that can be produced (in the case of labor services provided) at state σ when the players follow s.

Set $\tilde{\Psi}$ expresses relations between average discounted expected used quantities of labor services with minus sign in each firm and bank and average discounted expected outputs of consumption goods at each $\sigma \in \Im_1$. It is generated by some $s \in S$. We denote by Ψ the upper boundary of $\tilde{\Psi}$. Part (v) of Assumption 1 implies that it is convex. Function γ_ψ assigns to each $\psi \in \Psi$ the vector of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks multiplied by minus one by all consumers.

For each $\psi \in \Psi$, denote by $\tilde{S}(\psi)$ the set of Markov strategy profiles that generate ψ.

Denote by $\wp(\psi)$ the set of feasible vectors of prices of consumption goods and labor services in each firm and bank at states in $\Im_1$ for which there exists $K(\psi, p)$ such that the subspace of

$$\bigcup_{s \in \tilde{S}(\psi)} \left(\Pi_{\sigma \in \Im_1(s)} \left(\begin{array}{l} \Pi_{g \in G_L \cap N(s,\sigma)} \left[-\sum_{t \in N} \delta^{t-1} \pi(s, \sigma, t) y_g^{\max}, 0 \right]^{\#(J \cup B)} \\ \times \Pi_{g \in G_C \cap N(s,\sigma)} \left[0, \sum_{t \in N(s,\sigma)} \delta^{t-1} \pi(s, \sigma, t) y_g^{\max} \right] \end{array} \right) \right)$$

defined by $pz = K(\psi, p)$ is tangent to Ψ at ψ. $\tilde{\Psi}_{(\sigma)}$, $\Psi_{(\sigma)}$, and $\wp_{(\sigma)}(\psi)$ are defined for $\Gamma_{(\sigma)}(\delta)$ in an analogous way as $\tilde{\Psi}$, Ψ, and $\wp(\psi)$ for $\Gamma(\delta)$. We define $\wp_{(\sigma)}(\psi)$ in an analogous way also for each $\kappa \in (0,1)$ and each $\psi \in \kappa\Psi_{(\sigma)}$.

Assumption 2. $\wp_{(\sigma)}(\psi)$ is non-empty for each $\sigma \in \Im_1$, each $\kappa \in (0,1]$, and each $\psi \in \Psi_{(\sigma)}$.

Definition 1. *A strategy profile $s^* \in \tilde{S}$ is an SMPGE of $\Gamma(\delta)$ if there do not exist $\sigma \in \Im$, $D \in 2^{J \cup B \cup L} \setminus \{\varnothing\}$, $s_D \in S_{D(\sigma)}$, $x_D \in A_{D(\sigma)}(s^*_{-D(\sigma)}, s_D)$,* [8] *and $z_D \in \Pi_{g \in G_C}(0,\infty) \times \Pi_{k \in J \cup B} \Pi_{g \in G_L}(0,\infty)$ such that $x_{Dg} = \gamma_{Dg}(s^*_{(\sigma)}) + z_{Dg}$ for each $g \in G_C$ and $x_{Dkg} = \gamma_{Dkg}(s^*_{(\sigma)}) + z_{Dkg}$ for each $(k,g) \in (J \cup B) \times G_L$ with $\gamma_{Dkg}(s^*_{(\sigma)}) > 0$.*

3 Existence of an SMPGE

Proposition 1. *For each $\delta \in (0,1)$ and each $\psi^* \in \Psi$ there exists $s^* \in S^*$ that (i) is an SMPGE of $\Gamma(\delta)$, (ii) it generates a vector of prices belonging to $\wp(\psi)$, and (iii) $\gamma_{J \cup K \cup L}(s^*) = \gamma_\psi(\psi^*)$.*

Proof. Take arbitrary $\delta \in (0,1)$, $\psi^* \in \Psi$, [9] and let Θ be the set of all distributions of $\gamma_\psi(\psi^*)$ among consumers. We define function $\beta : con(S^+) \times \Theta \to [0,1]$ in such a way that $\beta(s,\theta) = 0$ if and only if the requirement of Definition 1 is satisfied for each $D \in 2^{J \cup B \cup L} \setminus \{\varnothing, J \cup B \cup L\}$, $\gamma_{J \cup K \cup L}(s) = \gamma_\psi(\psi^*)$, s generates price vector belonging to $\wp(\psi^*)$, and $s \in S^+$. Fix (arbitrary) $i^{(1)} \in I$, $i^{(2)} \in I \setminus \{i^{(1)}\}$, and define $\theta^{(1)} \in \Theta$ by assigning whole $\gamma_\psi(\psi^*)$ to $i^{(1)}$. Choose (arbitrary) $g^{(1)} \in G_C$ such that $\gamma_{\psi g^{(1)}}(\psi) > 0$. Further, fix $\varepsilon^{(1)} \in \left(0, \dfrac{\theta^{(1)}_{g^{(1)}}}{\max_{\theta \in \Theta} \|\theta - \theta^{(1)}\|}\right)$ and $\varepsilon^{(2)} \in \left(0, \min\left\{\theta^{(1)}_{g^{(1)}} - \varepsilon^{(1)} \max_{\theta \in \Theta} \|\theta - \theta^{(1)}\|, 0.25\right\}\right)$.

Let $X = \left(\Pi_{g \in G_L} R^{\#(J \cup B)} \times \Pi_{g \in G_C} R\right)^{\#(I)}$. Define function $f^1 : con(S^+) \times \Theta \to X$ by

$$f^{(1)}(s,\theta) = \theta^{(1)} + [1 + \beta(s,\theta)](\theta - \theta^{(1)}). \tag{1}$$

Take the coordinates $q^{(1)}$ and $q^{(2)}$ in X such that $q^{(1)}$ corresponds to consumption of $g^{(1)}$ by consumer $i^{(1)}$ and $q^{(2)}$ corresponds to consumption of $g^{(1)}$ by consumer $i^{(2)}$. For each $n \in \{1,2\}$ let $v^{(n)}$ be the unit vector in X with coordinate $q^{(n)}$ equal to one. Define function $f^{(2)} : con(S^+) \times \Theta \to X$ by

$$f^{(2)}(s,\theta) = f^{(1)}(s,\theta) + \beta(s,\theta)\left(\varepsilon^{(2)} + \varepsilon^{(1)}\|\theta - \theta^{(1)}\|\right)\left(v^{(1)} - v^{(2)}\right) \tag{2}$$

Set $f^{(1)}_{(\sigma_0)} = f^{(1)}$ and $f^{(2)}_{(\sigma_0)} = f^{(2)}$. For each $\sigma \in \Im \setminus \{\sigma_0\}$ we define set $\Theta_{(\sigma)}$ and functions $f^{(1)}_{(\sigma)}$ and $f^{2}_{(\sigma)}$ in an analogous way as set Θ and functions $f^{(1)}$ and $f^{(2)}$, using $\theta^{(1)}_{(\sigma)}$, $\varepsilon^{(1)}_{(\sigma)}$, $\varepsilon^{(2)}_{(\sigma)}$, $v^{(1)}_{(\sigma)}$, and $v^{(2)}_{(\sigma)}$ chosen in analogous way as $\theta^{(1)}$, $\varepsilon^{(1)}$, $\varepsilon^{(2)}$, $v^{(1)}$, and $v^{(2)}$. Finally, define function $f : con(S^+) \times \prod_{\sigma \in \Im} \Theta_{(\sigma)} \to con(S^+) \times \prod_{\sigma \in \Im} X$ by

$$f\left(s, (\theta_{(\sigma)})_{\sigma \in \Im}\right) = \left(s, \left(f^2_{(\sigma)}(s, \theta_{(\sigma)})\right)_{\sigma \in \Im}\right). \tag{3}$$

It can be shown that f satisfies all conditions of Theorem 2 in [1]. Therefore, it has fixed point $\left(s^*, (\theta^*_{(\sigma)})_{\sigma \in \Im}\right) \in con(S^+) \times \prod_{\sigma \in \Im} \Theta_{(\sigma)}$. By construction of f, $\beta_{(\sigma)}(s^*_{(\sigma)}, \theta^*_{(\sigma)}) = 0$ for each $\sigma \in \Im$. This implies that s^* satisfies the requirement of Definition 1 for each $D \in 2^{J \cup B \cup L} \setminus \{\varnothing, J \cup B \cup L\}$ in each subgame, $\gamma_{J \cup B \cup L}(s)^* = \gamma_\psi(\psi^*)$, s^* generates price vector belonging to $\wp(\psi^*)$ and $s^* \in S^+$. Using the fact that a fixed point with analogous properties exists also for other elements of Ψ and consumers' utility functions are generic, we

[8] We replace $A_{D(\sigma)}(s^*_{-D(\sigma)}, s_D)$ by $A_{J \cup B \cup L(\sigma)}(s)$ if $D = J \cup B \cup L$.

[9] Due to space limitations, we give here only an outline of the proof.

can show that s^* satisfies also the requirement of Definition 1 for $J \cup B \cup L$ in each subgame of $\Gamma(\delta)$, which completes the proof. *Q.E.D.*

In our paper, an SMPGE is a fixed point of a function. Its functional values depend on the extent to which equilibrium conditions and some other requirements, which should be satisfied in an equilibrium (e.g., that a strategy profile should belong to the set of feasible continuous strategy profiles, not just to the convex hull of the latter), are violated by its vector argument. An argument, at which none of them is violated, is a fixed point of the function. This approach can be used if and only if for each of the equilibrium conditions and other requirements there exists an argument of the function at which this condition or requirement considered separately is satisfied. (In our paper, we cannot establish separately that there exists an argument of the function that does not allow the coalition of all players to gain by a deviation in any subgame. Therefore, we had to prove the immunity of a fixed point to deviations by the coalition of all players separately.) Then the proof of the existence of a fixed point is tantamount to the proof that all these conditions and requirements can be satisfied simultaneously.

4 Conclusions

It can be surprising for a reader that an SMPGE exists for any $\delta \in (0,1)$. The explanation lies in the difference between our general equilibrium infinite horizon model of an oligopolistic economy and infinite horizon games modeling one oligopolistic industry (without binding capacity constraints in equilibrium). In the latter, firms can increase their outputs and sales. This enables them, as well as coalitions of them other than the grand one, to violate a collusive agreement. A punishment can wipe out gains from a single period deviation only if discount factor is sufficiently close to one. In the case of our general equilibrium oligopolistic model, inputs are determined by concluded contracts. Thus, in the second phase of any period, a firm's ability to increase outputs of consumption goods is constrained by inputs purchased (including labor services hired) in the preceding periods and contracts for delivery of producer goods concluded in the first phase of the current period. Also, ability of any coalition of firms other than the grand one in any period to increase outputs of consumption goods by its members is limited by inputs purchased in the preceding periods and contracts on delivery of producer goods with firms outside the coalition concluded in the current period. A coalition of firms other than the grand one contemplating a deviation in the first phase of a period can change contracts for delivery of producer goods among its members. Nevertheless, in the case of contracts with firms outside the coalition it can only refrain from concluding them. Moreover, inputs contracted in the current period will be used for production of consumption goods only in (in the case of durable inputs starting from) the following period, in which players outside the deviating coalition can already react to a deviation by their contract proposals and production of consumption goods. Therefore, an SMPGE exists also for low discount factors.

Acknowledgements

The research reported in the present paper was financially supported by grant APVV-14-0020.

References

[1] Browder, F. E.: A new generalization of the Schauder fixed point theorem. *Mathematische Annalen* **174** (1967), 285-290.

[2] Dierker, E., and Grodal, B.: The price normalization problem in imperfect competition and the objective of the firm. *Economic Theory* **14** (1999), 257-284.

[3] Maskin, E., and Tirole, J.: A theory of dynamic oligopoly II: Price competition, kinked demand curves, and Edgeworth cycles. *Econometrica* **56** (1988), 571-599.

[4] Rubinstein, A.: Strong perfect equilibrium in supergames. *International Journal of Game Theory* **9** (1980), 1-12.

[5] Šimková, Ľ., and Horniaček, M.: Pareto efficiency of consumers' allocations in oligopolistic general equilibrium with innovation. In: *Proceedings of the Mathematical Methods in Economics 2015*. University of West Bohemia, Cheb, 2015.

The Regularization of the Random Walk

Richard Horský[1]

Abstract. The non-stationary time series seem to be a serious problem for their instability which leads to poor results in forecasting. Unfortunately they occur frequently in economic and financial practice. A typical example of a non-stationary process is a random walk. It can be defined as the solution to the stochastic difference equation of the first order. Lool at this equation as an operator equation allows us to employ the access of the functional analysis.
It turns out that the equation describing the random walk is an ill-posed problem. The term of ill-posedness was introduced first by J. Hadamard (1902). It means that the solution to the given operator equation either does not exist at all or it is not unique or it is unstable. The ill-posed problem has to be regularized if we wish to obtain a reasonable result. There exist a lot of methods of regularization. One of them is the well-known Tikhonov regularization method named by the outstanding Russian mathematician A. N. Tikhonov, who introduced it in early 1960's.

Keywords: random walk, symmetrization, ill-posed problem, regularization.

JEL Classification: C220
AMS Classification: 91B70

1 Ill-posed problem and its regularization

The term *ill-posed problem (improperly posed problem)* has appeared at the beginning of the 20th century. J. Hadamard formulated (1902) what means that a problem is well-posed. He introduced the term of well-posedness in the context of his study of the partial differential equations. Archetypal well-posed problems are the Dirichlet problem for Laplace's equation or the Cauchy problem for wave equation. Hadamard also gave the example (classical today) of the problem which is not well-posed: the Cauchy problem for Laplace's equation. Hadamard's concept reflected the idea that any mathematical model of a physical phenomena should have all the following properties:

- There exists a solution to the problem (existence);
- There is at most one solution to the problem (uniqueness);
- The solution depends continuously on the data (stability).

A problem which does not satisfy at least one of these conditions was called ill-posed problem. It turned out later that many branches of not only natural sciences involve ill-posed problems (astronomy, geophysics, medicine and others).

In 1950's a great attention was drawn to ill-posed problem. Here we have to recall the names of outstanding Russian mathematicians A. N. Tikhonov, M. M. Lavrentiev, V. K. Ivanov and their disciples, e.g. V. A. Morozov. They contributed much to the development of the theory and appeared new techniques which became a fruitful area in numerical analysis. The foundations of approximate methods for solving ill-posed problems were laid by A. N. Tikhonov. He gave a generalization of the classical Hadamard's concept of well-posedness of the problem.

Many problems in mathematics and applied sciences may be formulated as an operator equation

$$Ax = b, \tag{1}$$

where $A: V \to W$ is a mapping defined on a normed linear space V with values in a normed linear space W, b in W is given. If V and W are Banach spaces and A is a linear, bounded and injective operator with a closed range $R(A)$, the inverse A^{-1} is bounded as well (see [6], p.176). Then the problem (1) is well-posed. The typical example of the well-posed equation (1) is the Fredholm integral equation of the second kind (see [5], p. 88).

The equation (1) is ill-posed if either A is not bounded below (see [4], p. 73) or its right side is not in the range $R(A)$, where V and W are Banach spaces. The typical example of the operator which is not bounded below is a compact operator of infinite rank. In the case of an ill-posed problem some steps has to be carried out in order to obtain a reasonable solution to the given ill-posed problem.

[1] University of Economics Prague, department of mathematics, Ekonomická 957, Prague 4, rhorsky@vse.cz.

1.1 General concept of regularization

Suppose (1) is an ill-posed problem, *V* and *W* are normed linear spaces, *A* is a bounded linear operator. In the following considerations we will deal with the case which often arises in practical situations: the inverse of *A* is not bounded. Then the stability of the solution is violated and may be restored by an approximation of the given ill-posed problem by a well-posed one. The procedures for this replacement are called *regularization methods*.

The *regularization family for A* is a set of bounded linear operators $R_\alpha : W \to V$, $\alpha > 0$, for which $\lim_{\alpha \to 0+} \|R_\alpha A x - x\| = 0$ for any $x \in V$. The regularization family may not be uniformly bounded. In fact, if there were a constant *C* such that for any $\alpha > 0$ we would have $\|R_\alpha\| < C$, then for any $y \in R(A)$ we would obtain the following inequalities

$$\|A^{-1} y\| \le \|A^{-1} y - R_\alpha y\| + \|R_\alpha y\| \le \|x - R_\alpha A x\| + \|R_\alpha\| \|y\| \le \|x - R_\alpha A x\| + C\|y\|. \tag{2}$$

The first term on the right side in (2) tends to zero as $\alpha \to 0$ (the definition of the regularization family) but it yields A^{-1} is bounded (contradiction).

Another fact within the regularization is that the right hand side in (1) involves some noise. Suppose that $b_\delta \in W$ is a perturbation of *b*, where $\|b - b_\delta\| \le \delta$. We define

$$x_{\alpha,\delta} = R_\alpha b_\delta . \tag{3}$$

The vector (3) is called *the regularized solution to (1).*

Let $x^* = A^{-1} b$ be the exact solution to the problem (1). The following fundamental estimate for the error of the regularized solution shows what is the essential problem:

$$\|x^* - x_{\alpha,\delta}\| \le \|x^* - R_\alpha b\| + \|R_\alpha b - R_\alpha b_\delta\| \le \|x^* - R_\alpha A x^*\| + \|R_\alpha\| \delta. \tag{4}$$

If $\alpha \to 0$ then the first term on the right side in (4) tends to zero (*regularization effect*) whereas the second term grows to infinity (*ill-posedness effect*). We observe two competing effects which enter (3). Thus we are forced to make a trade off between the accuracy and the stability. The question of the greatest importance is how to choose the value of the parameter $\alpha > 0$ which is dependent on the given $\delta > 0$ (and maybe on $b_\delta \in W$). We notice this problem later in the chapter 4 within the discussion of the Tikhonov regularization method.

1.2 Least square solution as the generalized solution

We can notice that the exact or rather classical solution x^* to the equation (1) need not exist. If (1) has no solution, i.e. $b \notin R(A)$, one can try to look for a solution in some generalized sense. The natural requirement for such a solution is to minimize the residual norm $\|Ax - b\|$. The usual framework for the following considerations and notions is that *V* and *W* are Hilbert spaces. If there exists such a vector $x_0 \in V$ for which

$$\|Ax_0 - b\| = \inf_{x_0 \in V} \|Ax - b\|, \tag{5}$$

x_0 is called the *least square solution* or the *generalized solution to the equation (1) in the sense of least squares* (GSLS). However GSLS need not exist. A sufficient condition for the existence of GSLS is that the range *R(A)* is a closed subspace of the space *W*. This is the case if e.g. *R(A)* is of finite dimension. On the other hand if *R(A)* is a dense subspace of *W*, different from *W*, then there exists $b \in R(A)$ such that the GSLS does not exist.

$x_0 \in V$ is the GSLS to (1) if and only if $x_0 \in V$ is the solution to the *normal form of the equation (1)*:

$$A^* A x = A^* b. \tag{6}$$

The GSLS (if it exists) is thus a solution to linear equation (6) and in general need not be unique. In any case all the GSLS's form a linear set $L_b = \{x \in V : A^* Ax = A^* b\} = x_0 + N(A)$, where *N(A)* is the kernel of the operator *A*. To assure the uniqueness of the GSLS we have to impose some additional condition. We usually select the GSLS with the least norm: if L_b is non-empty there exists the unique element $\tilde{x} \in L_b$ for which $\|\tilde{x}\| = \inf_{x \in L_b} \|x\|$.

2 Stochastic process and its basic characteristics

Stochastic process is a mapping $\mathbf{X} : T \to L_2(\Omega, \pi)$, where *T* is a time domain and $L_2(\Omega, \pi)$ is the space of all functions (random variables) defined almost everywhere on the measurable space Ω, functions, that are square integrable over the space Ω with respect to a probability measure π. This space is Hilbert space with the norm derived from the scalar product $E(XY) = \int_\Omega XY d\pi$. The mean and variance of any random variable $X \in L_2(\Omega, \pi)$, are finite numbers, particularly the mean $\mu = EX = \int_\Omega X d\pi$ and the variance $\sigma^2 = DX = EX^2 - \mu^2$. The time domain will be the set of all integers here. It means that the values of a stochastic process are random variables $X_t, t = 0, \pm 1, \pm 2, \ldots$ We will write $\mathbf{X} = (X_t)$ to emphasize that the stochastic process is a sequence and talk about a stochastic sequence.

The basic characteristics of the stochastic process are:

- the function of means $\mu_t = EX_t$;
- the function of variances $\sigma_t^2 = DX_t = EX_t^2 - \mu_t^2$;
- the autocovariance function $C(t, s) = \mathrm{cov}(X_t, X_s)$.

2.1 Stationary process and white noise

The concept of the stationary process is well-known (see [1], p. 16). Briefly said a stationary process is such that the functions of means and variances respectively are constant and the values of its autocovariance function are dependent only on the distance (time lag) of the random variables within the process: $C(t, s) = C(t - s, 0) = \gamma_{t-s}$. The stationarity of a process ensures its stochastic stability.

A typical example of the stationary process is the white noise. The *white noise* $\mathbf{W} = (\varepsilon_t)$ is a stochastic process with $E\varepsilon_t = 0$, $D(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2 > 0$ and the autocovariance $\gamma_k = \mathrm{cov}(\varepsilon_t, \varepsilon_{t-k}) = E(\varepsilon_t \varepsilon_{t-k})$ for any integer *k*. In other words the white noise is an orthogonal system in the Hilbert space $L_2(\Omega, \pi)$. It can be transformed to be an orthonormal system, not necessary complete. The white noise may be used as a base for the definition of other processes, mainly for the definition of the general linear process.

2.2 General linear process

The *general linear process* (GLP) is the process of the form

$$X_t = \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\,, \tag{7}$$

where μ is a given scalar (mean of the process) and (ψ_k) is a given sequence of scalars (weights of the process), $\psi_0 = 1$. The convergence of the series is taken in the sense of the convergence in the mean square which is the same as the convergence in the norm topology in $L_2(\Omega, \pi)$. The convergence of the series in (7) is equivalent to the stationarity of the (GLP). The necessary and sufficient condition for the convergence of (7) in the mean square is $(\psi_k) \in \ell_2$ (the space of all square summable scalar sequences). Another conditions

$$\sum_{k=0}^{\infty} |\psi_k| < \infty \tag{8}$$

is only sufficient for the convergence in (7) (see [2]).

3 Lag operator and the settings of convergence structures

The lag operator plays an essential role in the theory of stochastic processes. It simplifies formal writings and at the same time it is a linear operator. It is necessary to specify the spaces which stand for its domain so that we may analyze its properties, especially its spectrum. Let us recall the definition of the lag operator which is given e.g. in [3] including its matrix representation in suitable spaces of stochastic sequences:

$$B:V\to V,\quad B(X_t)=(X_{t-1})\,, \tag{9}$$

where V is a space of stochastic sequences. In the following section it will be particularly specified.

3.1 Stochastic process in Banach or Hilbert space

In what follows we will consider a stochastic process (as a sequence of random variables) in the form

$$\mathbf{X}=(X_t, X_{t-1},..., X_{t-k},...) \tag{10}$$

for any fixed integer t, $X_{t-k}\in L_2(\Omega,\pi), k\geq 0$. The sequence (10) will be regarded as an element in two different spaces. The first one is $\ell_\infty(L_2(\Omega,\pi))$, the space of all bounded stochastic sequences with the norm

$$\|\mathbf{X}\|_\infty=\sup_{k\geq 0}\sqrt{EX_{t-k}^2}\,. \tag{11}$$

The space $\ell_\infty(L_2(\Omega,\pi))$ with the norm (11) is the Banach space like the classical ℓ_∞ (see [6], p. 94). The stationary sequences are contained in the space $\ell_\infty(L_2(\Omega,\pi))$.

The second space is $\ell_2(L_2(\Omega,\pi))$, the space of all square summable stochastic sequences. This space is endowed by the scalar product $\langle\mathbf{X},\mathbf{Y}\rangle=\sum_{k=0}^{\infty}E(X_{t-k}Y_{t-k})$ and it is the Hilbert space with respect to the norm derived from this scalar product

$$\|\mathbf{X}\|_2=\left(\sum_{k=0}^{\infty}EX_{t-k}^2\right)^{1/2}. \tag{12}$$

Unfortunately the space $\ell_2(L_2(\Omega,\pi))$, contains no stationary or constant stochastic sequences (except the trivial sequence). It means for instance that there is no white noise in this space. In spite of this apparent drawback the space $\ell_2(L_2(\Omega,\pi))$, has more comfortable structure for our intentions than the $\ell_\infty(L_2(\Omega,\pi))$ .We may overcome the given deficiency if we adopt slightly modified concept of stationarity and white noise, respectively. As to the mean we will take only centered processes, i.e. those with mean zero. We allow the process to be stationary as far as to the past as one wishes then the norm (and hence the variance) of the random variables within the process has to fall to zero (otherwise the series in (12) cannot be convergent).

Now we set in (9) $V=\ell_\infty(L_2(\Omega,\pi))$. It is shown in [2] that $\|B\|=\sup_{\|\mathbf{X}\|_\infty=1}\|B\mathbf{X}\|_\infty=1$ and $\|B^k\|=1$ for any $k\geq 0$.

The Banach algebra $\Lambda(\ell_\infty(L_2(\Omega,\pi)))$ of all bounded linear operators on the space $\ell_\infty(L_2(\Omega,\pi))$ contains all the polynomials in B. We can also take a power series in B and ask whether it defines some element of $\Lambda(\ell_\infty(L_2(\Omega,\pi)))$:

$$\psi(B)=\sum_{k=0}^{\infty}\psi_k B^k\,. \tag{13}$$

The series on the right side of (13) converges in the operator norm if (8) is satisfied (see [2]). Its sum is called the *linear filter*. It transforms the white noise to a GLP since (7) may be rewritten (by (13)) in the form

$$\mathbf{X}=\psi(B)\mathbf{W}\,. \tag{14}$$

Example 1. There is an important class of stochastic processes which are described by the stochastic difference equation

$$\Phi(B)\mathbf{X} = \Theta(B)\mathbf{W}\ , \tag{15}$$

where $\Phi(B)$ and $\Theta(B)$ are polynomials (in B) of the order p and q respectively. If $\Phi(z) \neq 0$ for $|z| \leq 1$, then (15) is well-known ARMA(p,q) process. In this case the equation (15) is well-posed and (15) has the unique solution in the form (14), where the linear filter $\psi(B) = \Phi(B)^{-1}\Theta(B)$. As we will see in the next section if $\Phi(1) = 0$, the equation (1) is ill-posed. Then we encounter a non-stationary process denoted as ARIMA(p,d,q), where $d > 0$ is the multiplicity of the unit as a root of the $\Phi(z)$.Then there is a question: what is a solution to (15)?

The norm of (9) in the case $V = \ell_2(L_2(\Omega,\pi))$ is equal to 1 as well as in the previous case.

3.2 Spectral properties of the lag operator

The contribution [3] deals with the spectral properties of (9). The spectrum of B is the closed unit circle in the complex plane: $\sigma(B) = \{z \in \mathbf{C} : |z| \leq 1\}$. The main fact is that for any $|\lambda| < 1$ the sequence $\mathbf{E}_\lambda = (1, \lambda, \lambda^2, ...)$ is the eigenvector of (9) assigned to λ. The boundary of unit circle is also in the point spectrum but only for $B \in \Lambda(\ell_\infty(L_2(\Omega,\pi)))$. If $B \in \Lambda(\ell_2(L_2(\Omega,\pi)))$ and $|\lambda| = 1$ the operator $(\lambda I - B)^{-1}$ exists but it is not bounded and therefore it may not be defined on the whole space $\ell_2(L_2(\Omega,\pi))$ or its some closed subspace. In fact $R(\lambda I - B)$ is a proper subspace of $\ell_2(L_2(\Omega,\pi))$ which is dense in $\ell_2(L_2(\Omega,\pi))$. If we set e.g. $\lambda = 1$ the corresponding eigenvector in $\ell_\infty(L_2(\Omega,\pi))$ is $\mathbf{E} = (1,1,1,...)$, however $\mathbf{E} \notin \ell_2(L_2(\Omega,\pi))$.

The spectral analyse of (9) provides the answer to the problem of solving the equation (15) when $\Phi(B) = I - B$.

4 The random walk as the solution to an ill-posed problem

If we set $\Phi(B) = I - B$, i.e. $\Phi(B)$ is the difference operator, and $\Theta(B) = I$ in (15), then the equation (15) is the stochastic difference equation of the first order

$$(I - B)\mathbf{X} = \mathbf{W}. \tag{16}$$

We will deal with a solution to (16) in the Hilbert space $\ell_2(L_2(\Omega,\pi))$. This solution is called the *random walk process.*

4.1 The least square solution to the stochastic difference equation of the first order

In [3] we tried to obtain least square solution to (16). However the spectral analyse of the adjoint operator B^* and the well-known relations between ranges and kernels respectively of the original operator A and the non-negative and self-adjoint operator A^*A (particularly, $\overline{R(A^*A)} = \overline{R(A)} = \ell_2(L_2(\Omega,\pi))$, $N(A^*A) = N(A) = \{o\}$) show that the normal form of (16) is an ill-posed problem as well. It is necessary to emphasize that the GSLS to (16), i.e. the solution to the equation

$$(I - B^*)(I - B)\mathbf{X} = (I - B^*)\mathbf{W} \tag{17}$$

need not exist and really does not exist for some right side from the space $\ell_2(L_2(\Omega,\pi))$. It has been already mentioned in the section 1.2.

4.2 Tikhonov regularization

The Tikhonov regularization consists in the regularization of the normal form of the equation (16). The normal form (6) of the equation (16) is in this case the equation (17). To simplify the writings we set $A = I - B$ .The Tikhonov regularization consists in addition of some positive multiple of the identity αI to the operator A^*A. In

this way we get a bounded linear bijection $A^*A+\alpha I$ on the space $\ell_2(L_2(\Omega,\pi))$. The regularized form of the equation (17) is

$$\left(A^*A+\alpha I\right)\mathbf{X}=A^*\mathbf{W},\ \alpha>0. \tag{18}$$

The equation (18) is well-posed. The operators $R_\alpha=\left(A^*A+\alpha I\right)^{-1}A^*$ form a regularization family for A^*A. In the intention of the section 1.1 we replace the white noise $\mathbf{W}$ by its perturbation $\mathbf{W}_\delta$, $\left\|\mathbf{W}-\mathbf{W}_\delta\right\|_2\le\delta,\delta>0$ in (18) and thus we get the regularized solution $\mathbf{X}_{\alpha,\delta}=R_\alpha\mathbf{W}_\delta$.

As to the ill-posedness effect it can be seen (with the help of polar decomposition and Riesz functional calculus, see [4] p. 87) that the regularization family satisfies

$$\left\|R_\alpha\right\|\le\frac{1}{2\sqrt{\alpha}},\ \alpha>0. \tag{19}$$

In sequel we will suppose GSLS $\tilde{\mathbf{X}}$ of (16) exists and $\tilde{\mathbf{X}}=A^*\mathbf{Z}$ for some $\mathbf{Z}\in\ell_2(L_2(\Omega,\pi))$. We get

$$\left\|R_\alpha A\tilde{\mathbf{X}}-\tilde{\mathbf{X}}\right\|_2\le\alpha\left\|R_\alpha\right\|\left\|\mathbf{Z}\right\|_2\le\frac{\sqrt{\alpha}}{2}\left\|\mathbf{Z}\right\|_2. \tag{20}$$

In the end we substitute the estimations (19) and (20) to (4) in order to get the fundamental estimation for the Tikhonov regularization applied on the stochastic difference equation (16):

$$\left\|\tilde{\mathbf{X}}-\mathbf{X}_{\alpha,\delta}\right\|_2\le\frac{\sqrt{\alpha}}{2}\left\|\mathbf{Z}\right\|_2+\frac{\delta}{2\sqrt{\alpha}}. \tag{21}$$

We have to ballance the parameters α and δ so that $\frac{\delta}{2\sqrt{\alpha}}\to0$ for $\alpha\to0$ and $\delta\to0$. One of the well-known strategy for the choice of the parameter α is the Morozov's discrepancy principle (see [5], p. 172).

Conclusion

The stochastic difference equation (16) is an ill-posed equation in the framework of the structure of Hilbert space $\ell_2(L_2(\Omega,\pi))$. If we try to find its solution as the generalized solution in the sense of least squares then we come again to an ill-posed problem (equation (17)). However this approach allows us to use the well-known Tikhonov regularization method and due to this method to find at least a regularization solution. It remains an open problem how to get a (generalized) solution to the original problem (16) or to the symmetrized problem (17), probably in some more suitable structure than that of Hilbert or Banach space.

References

[1] Arlt, J.: *Moderní metody modelování ekonomických časových řad*. Grada Publishing 1999.

[2] Horský, R.: The lag operator and its spectral properties. In: *Proceedings I of the Mathematical Methods in Economics 2013*, Jihlava, 2013, 285–290.

[3] Horský, R.: Random walk process and its invertibility. In: *Conference Proceedings of the Mathematical Methods in Economics 2014*, Olomouc, 2014, 321–326.

[4] Lukeš, J.: *Zápisky z funkcionální analýzy*. Univerzita Karlova v Praze. Nakladatelství Karolinum. 2012.

[5] Nair, M. T.: *Linear operator equations: approximation and regularization.* World Scientific Publishing Company, 2009.

[6] Taylor, A.E.: *Úvod do funkcionální analýzy.* Academia, Praha, 1973.

A discontinuous Galerkin method for pricing of two-asset options

Jiří Hozman[1], Tomáš Tichý[2]

Abstract. A standard assumption of option pricing models is Gaussian distribution. Despite its simplifying features, it can lead to analytical solution only in some special cases, such as plain vanilla options, continuously monitored price of path dependent options, etc. In other cases, a numerical procedure should be involved to get the option price. In this paper we focus on options whose payoff is given by several (dependent) risk sources – a so called basket options – and develop a numerical scheme to their pricing via discontinuous Galerkin (DG) approach combining the advantages of the finite element methods together with the discontinuous approach. In order to illustrate the potency of the presented numerical scheme we carry out numerical experiments with real data.

Keywords: option pricing, discontinuous Galerkin method, 2D Black-Scholes equation, basket options, numerical solution.

JEL classification: C44, G13
AMS classification: 35, 90C15

1 Introduction

Basket options [2] are mostly traded at OTC (over-the-counter) markets and play important role in designing simple but efficient hedging schemes in corporate finance as well as life insurance. While basket options exhibit approximately the same payoff as a portfolio of individual options, the managing of one basket option, including potential exercising, is much easier than dealing with several individual options – therefore the attractiveness for hedging purposes.

Since the underlying value is a weighted sum of individual asset prices, basket options can be regarded as an option on a portfolio. Thus, it is apparent that besides individual risk factors also a dependency structure among them plays an important role – and makes the pricing very challenging and we can rarely find explicit closed form solution to the basket option pricing problem. Mostly, we need to apply some kind of numerical technique or develop an analytical approximation. While in the past the latter approach was often preferred, the increasing capacity and speed of computers currently support the use of the former.

In this paper we focus on options whose payoff is given by several (for simplicity, let us assume just 2) dependent risk sources and develop a numerical scheme to their pricing via discontinuous Galerkin (DG) method combining the advantages of the finite element methods together with the discontinuous approach. The DG concept is based on piecewise polynomial, generally discontinuous approximations, for a survey, see [6]. From this point of view, DG methods seem to be very promising tool for the numerical simulation of option pricing problems and provides robust and high-order accurate approximations of solutions resulting from the two-dimensional Black-Scholes partial differential equation.

While in Section 2 the mathematical problem is formulated, in Section 3 discontinuous Galerkin scheme is developed. In Section 4 we illustrate the problem assuming a 2-stock option.

[1]Technical University of Liberec, Department of Mathematics and Didactics of Mathematics, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz

[2]VŠB - Technical University of Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz

2 European two-asset options

As a model problem, we consider the two-dimensional Black-Scholes (BS) partial differential equation describing the value of the European option with two assets $V = V(S_1, S_2, t)$ with the instantaneous constant interest rate r and volatility of the i-th asset σ_i, $i = 1, 2$. According to [1], we solve the following parabolic problem (with reversal time):

$$\frac{\partial V}{\partial t} + \mathcal{L}_{BS}(V) = 0, \quad S_1 > 0,\ S_2 > 0,\ t \in (0, T] \tag{1}$$

with a linear differential (Black-Scholes) operator

$$\mathcal{L}_{BS}(V) = -\frac{1}{2}\sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} - \rho\sigma_1\sigma_2 S_1 S_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} - \frac{1}{2}\sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2} - rS_1 \frac{\partial V}{\partial S_1} - rS_2 \frac{\partial V}{\partial S_2} + rV, \tag{2}$$

where S_i are prices of the i-th asset and $\rho \in [-1; 1]$ the correlation factor. The evolution problem (1)–(2) is prescribed on a time interval with the expiration date T (i.e. t denotes the time to maturity) and it has to be equipped with the appropriate initial condition as payoff function

$$V(S_1, S_2, 0) = V_0(S_1, S_2) := \begin{cases} \max(\alpha_1 S_1 + \alpha_2 S_2 - K, 0), & \text{if } V \text{ is call,} \\ \max(K - \alpha_1 S_1 - \alpha_2 S_2, 0), & \text{if } V \text{ is put,} \end{cases} \quad S_1 > 0,\ S_2 > 0, \tag{3}$$

with the strike price K and weights of the i-th asset α_i, $i = 1, 2$.

In order to numerically solve Cauchy problem (1)–(3), the unbounded domain in the variables S_1 and S_2 has to be truncated at first, and then this initial problem should be supplied with suitable choices of additional boundary conditions on appropriate parts of the boundary of the computational domain $\Omega := (0, S_1^{max}) \times (0, S_2^{max})$. We distinguish three parts of the boundary $\partial\Omega$ satisfying $\overline{\partial\Omega} = \overline{\Gamma_1} \cup \overline{\Gamma_2} \cup \overline{\Gamma_3}$, $\Gamma_1 \cap \Gamma_2 \cap \Gamma_3 = \emptyset$ and defined as

$$\Gamma_1 = (0, S_1^{max}) \times \{0\}, \quad \Gamma_2 = \{0\} \times (0, S_2^{max}), \quad \Gamma_3 = \partial\Omega \cap \mathbb{R}^2_+. \tag{4}$$

The Dirichlet boundary conditions for this two-asset option pricing problem may not be so easy to set up. Actually, the boundary conditions on Γ_1 and Γ_2 (i.e. one of the asset prices is zero) are obtained by using Black-Scholes formula. On the other hand the boundary conditions on Γ_3 arising from the asymptotic behaviour of vanilla options at infinity. In the spirit of [5], the boundary conditions are given by

$$V(S_1, S_2, t)\Big|_{\partial\Omega} = \begin{cases} \alpha_i S_i\, \Phi(d_i) - Ke^{-rt}\Phi(d_i - \sigma_i\sqrt{t}) & \text{on } \Gamma_i,\ i = 1, 2, \\ \alpha_1 S_1 + \alpha_2 S_2 - Ke^{-rt} & \text{on } \Gamma_3, \end{cases} \quad \text{if } V \text{ is call,} \tag{5}$$

or

$$V(S_1, S_2, t)\Big|_{\partial\Omega} = \begin{cases} Ke^{-rt}\Phi(-d_i + \sigma_i\sqrt{t}) - \alpha_i S_i \Phi(-d_i) & \text{on } \Gamma_i,\ i = 1, 2, \\ 0 & \text{on } \Gamma_3, \end{cases} \quad \text{if } V \text{ is put,} \tag{6}$$

where

$$d_1 = \frac{\ln(\alpha_1 S_1/K) + (r + \sigma_1^2/2)t}{\sigma_1\sqrt{t}} \qquad \text{and} \qquad d_2 = \frac{\ln(\alpha_2 S_2/K) + (r + \sigma_2^2/2)t}{\sigma_2\sqrt{t}}$$

and Φ denotes the standard normal cumulative distribution function.

2.1 Problem reformulation into divergence form

Next, the equation (1) with the operator (2) will be rewritten as a self-adjoint second-order equation, which is more suitable for the later numerical treatment. Indeed by substitution

$$x_1 = S_1, \quad x_2 = S_2, \quad u(x, t) \equiv u(x_1, x_2, t) = V(S_1, S_2, t)$$

the value of the basket option on two assets is given by a price function $u(x, t) : Q_T = \Omega \times (0, T) \to \mathbb{R}^+_0$ satisfying that

$$\frac{\partial u}{\partial t} - \operatorname{div}(\mathbb{D}(x) \cdot \nabla u) + \nabla \cdot \vec{f}(x, u) + \kappa u = 0 \quad \text{in } Q_T, \tag{7}$$

where $\{D(x)_{ij}\}_{i,j=2}^{2}$ denotes the symmetric positive semi-definite matrix

$$\mathbb{D}(x) := \{D(x)_{ij}\}_{i,j=1}^{2} = \frac{1}{2}\begin{pmatrix} \sigma_1^2 x_1^2 & \rho\sigma_1\sigma_2 x_1 x_2 \\ \rho\sigma_1\sigma_2 x_1 x_2 & \sigma_2^2 x_2^2 \end{pmatrix},$$

the vector-valued function $\vec{f}(x,u) = (f_1(x,u), f_2(x,u))$ plays a role of the so-called physical flux, component-wisely written as $f_i(x,u) = (\sigma_i^2 + \frac{1}{2}\rho\sigma_1\sigma_2 - r)x_i u$, $i = 1,2$, and $\kappa = 3r - \sigma_1^2 - \rho\sigma_1\sigma_2 - \sigma_2^2$ is a reaction coefficient.

From the mathematical point of view the problem (7) represents a convection-diffusion-reaction equation equipped with the initial condition u^0 and the set of three Dirichlet boundary conditions u_D, defined in the spirit of (3) and (5)–(6), respectively. It is important to note that although equation (7) is formally parabolic, when it is convection dominated the numerical approximation behaves as if it was hyperbolic and this fact is the source of many numerical difficulties. Therefore, is should be convenient to follow novel numerical option pricing schemes in order to solve this problem more accurately.

3 Discretization

In a wide class of problems resulting to a solution of a partial differential equation, the DG method is rather popular. The standard DG method uses piecewise polynomial, generally discontinuous, approximation of the p-th order describing a global solution on the whole domain. Therefore, the approximate solution is sought in a finite dimensional space $u_h \in S_{hp} := \{v; v|_T \in P_p(T)\ \forall\, T \in \mathcal{T}_h\}$, where $\mathcal{T}_h = \{T\}_{T\in\mathcal{T}_h}$ is a partition of $\overline{\Omega}$ into elements T with length h representing the mesh size, and $P_p(T)$ denotes the space of all polynomials of order less or equal to p defined on T.

By $\mathcal{F}_h$ we denote the set of all open edges of all elements $T \in \mathcal{T}_h$. Further, the symbol $\mathcal{F}_h^I$ stands for the set of all $\Gamma \in \mathcal{F}_h$ that are contained in Ω (inner edges) and the symbol $\mathcal{F}_h^D$ for the set of all $\Gamma \in \mathcal{F}_h$ such that $\Gamma \subset \partial\Omega$ (boundary edges). Obviously, $\mathcal{F}_h = \mathcal{F}_h^I \cup \mathcal{F}_h^D$. Finally, for each $\Gamma \in \mathcal{F}_h$, we define a unit normal vector $\vec{n}_\Gamma$. We assume that $\vec{n}_\Gamma$, $\Gamma \subset \partial\Omega$, has the same orientation as the outer normal of $\partial\Omega$. For $\vec{n}_\Gamma$, $\Gamma \in \mathcal{F}_h^I$, the orientation is arbitrary but fixed for each edge. For each $\Gamma \in \mathcal{F}_h^I$ there exist two neighbouring elements T_L and T_R. We use a convention that T_R lies in the direction of $\vec{n}_\Gamma$ and T_L in the opposite direction of $\vec{n}_\Gamma$. Since we deal with the discontinuous functions along edges Γ, it is suitable to introduce two new operators – jump $[v] := v|_\Gamma^{(L)} - v|_\Gamma^{(R)}$ and mean value $\langle v\rangle := \frac{1}{2}\left(v|_\Gamma^{(L)} + v|_\Gamma^{(R)}\right)$, where $v|_\Gamma^{(L)}$ is the trace of $v|_{T_L}$ on Γ and $v|_\Gamma^{(R)}$ is the trace of $v|_{T_R}$ on Γ, which are different in general. For $\Gamma \in \partial\Omega$, we simply we put $\langle v\rangle = [v] = v|_\Gamma^{(L)}$.

3.1 Space semi-discrete solution

In order to obtain a semi-discrete formulation, we follow the concept from [4] and multiply the corresponding original equation (7) by a test function $v_h \in S_{hp}$, integrate over an element $T \in \mathcal{T}_h$ and apply Green's theorem on the diffusion term as well as on the convection term. Further we sum over all $T \in \mathcal{T}_h$ and add some artificial terms vanishing for the exact solution such as penalty and stabilization terms which guarantee the numerical stability of the introduced method. Consequently, we employ a concept of an upwind numerical flux for a discretization of the convection term to suppress the undesirable influence of this term (if it is dominated). Finally, we end up with the following DG formulation for the semi-discrete solution $u_h = u_h(t)$ represented with the system of ordinary differential equations

$$\frac{d}{dt}(u_h, v_h) + \underbrace{a_h(u_h, v_h) + b_h(u_h, v_h) + J_h(u_h, v_h) + (\kappa u_h, v_h)}_{=:\mathcal{B}_h(u_h, v_h)} = l_h(v_h)(t) \quad \forall\, v_h \in S_{hp},\ \forall\, t \in (0,T) \tag{8}$$

with the initial condition $u_h(0) = u_h^0$, where u_h^0 denotes the S_{hp}-approximation of the initial condition u^0, and the notation $(\cdot,\cdot)$ is the inner product in $L^2(\Omega)$. The form $a_h(\cdot,\cdot)$ represents the diffusion terms, whose discretization involves an extra adding stabilization in order to guarantee the stability of the numerical scheme.

$$a_h(u,v) = \sum_{T\in\mathcal{T}_h}\int_T \mathbb{D}(x)\,\nabla u\cdot\nabla v\,\mathrm{d}x - \sum_{\Gamma\in\mathcal{F}_h}\int_\Gamma \langle \mathbb{D}(x)\,\nabla u\cdot\vec{n}\rangle\,[v]\,\mathrm{d}S + \sum_{\Gamma\in\mathcal{F}_h}\int_\Gamma \langle \mathbb{D}(x)\,\nabla v\cdot\vec{n}\rangle\,[u]\,\mathrm{d}S \tag{9}$$

A particular attention should be also paid to the treatment of the convection terms. Therefore, the form $b_h(\cdot,\cdot)$ is treated as in the finite volume method with the aid of the numerical flux $H(\cdot,\cdot,\cdot)$ approximating the original physical flux at partition edges Γ as

$$b_h(u,v) = -\sum_{T\in\mathcal{T}_h}\int_T \vec{f}(x,u)\cdot\nabla v\,\mathrm{d}x + \sum_{\Gamma\in\mathcal{F}_h}\int_\Gamma H\left(u|_\Gamma^{(L)}, u|_\Gamma^{(R)}, \vec{n}_\Gamma\right)[v]\,\mathrm{d}S \tag{10}$$

where

$$H\left(u|_\Gamma^{(L)}, u|_\Gamma^{(R)}, \vec{n}_\Gamma\right) = \begin{cases} \vec{f}\left(x, u|_\Gamma^{(L)}\right)\cdot\vec{n}_\Gamma, & \text{if } A > 0 \\ \vec{f}\left(x, u|_\Gamma^{(R)}\right)\cdot\vec{n}_\Gamma, & \text{if } A \le 0 \end{cases}, \quad \text{for } A = \sum_{i=1}^{2}\left(\sigma_i^2 + \frac{1}{2}\rho\sigma_1\sigma_2 - r\right)x_i n_i. \tag{11}$$

Furthermore, the penalty terms $J_h(\cdot,\cdot)$ are artificially added to the formulation of the DG semi-discrete problem and their role is to replace the inter-element discontinuities, a priori arising from DG treatment.

$$J_h(u,v) = \sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma \frac{1}{|\Gamma|}[u][v]\,\mathrm{d}S + \sum_{\Gamma\in\mathcal{F}_h^D}\int_\Gamma \frac{1}{|\Gamma|}u\,v\,\mathrm{d}S, \tag{12}$$

Finally, the right-hand side form $l_h(\cdot)$ arises only from boundary conditions.

$$l_h(v)(t) = \sum_{\Gamma\in\mathcal{F}_h^D}\int_\Gamma \left(I\!D(x)\,\nabla v\cdot\vec{n}\,u_D(t) + \frac{1}{|\Gamma|}u_D(t)\,v\right)\mathrm{d}S \tag{13}$$

In the next, to simplify the notation, the bilinear form $\mathcal{B}_h(\cdot,\cdot)$ representing the semi-discrete variant of the linear differential operator (2) is introduced in (8).

3.2 Fully discrete scheme and linear algebraic problem

In order to obtain the discrete solution, it is necessary to equip scheme (8) with suitable solvers for the time integration. Therefore, we employ the implicit scheme with respect to the time coordinate, giving the first order convergence in time with the unconditional stability and no limitations on the length of the time step. For simplicity, we consider a partition of the interval $[0,T]$ with a constant time step $\tau = T/M$. Let u_h^l stand for the approximate solution of $u(l\tau), l = 0,\ldots,M$, of the problem (8). The values u_h^l at given time levels are computed according to the following formula

$$\frac{1}{\tau}\left(u_h^l - u_h^{l-1}, v_h\right) + \mathcal{B}_h\left(u_h^l, v_h\right) = 0 \qquad \forall\, v_h \in S_{hp},\ \ l = 1,2,\ldots,M \tag{14}$$

with starting data u_h^0. Since the problem (7) is linear, the implicit treatment in (8) also leads to a linear algebraic problem (14) with a sparse matrix at each time level. Finally, the option value u is uniquely determined by the solution vector of the sparse system above.

4 Practical examples

This section is devoted to the illustration of the potency of the derived numerical scheme (14). Let us assume a life insurance company with funds invested at the stock market. Since the company assumes an obligation to pay benefits (pensions) to its clients in the near future, it might be afraid of a large drop in its investment portfolio value – for simplicity, however, let us consider only a 2-stock basket put option with 60% Allianz ($\alpha_1 = 0.6$) and 40% Deutshce Bank ($\alpha_2 = 0.4$), strike at 40 Euro, and maturing in 94 calendar days. Current date is September 13, 2011 and the two stocks trades at $S_1^{ref} = 59.79$ and $S_2^{ref} = 23.40$ Euro, respectively, with last year estimate of the Pearson linear correlation $\rho = 0.88$ and market implied riskless rate for a given horizon $r = 0.01557$ p.a. The shape of the computational domain Ω depends on the maximal stock prices. Here, we take $S_1^{max} = 130.0$ and $S_2^{max} = 220.0$ in order to suppress the influence of boundary conditions.

Numerical experiments are carried out with the use of DG method with the piecewise linear approximations ($p = 1$) on adaptively refined domain Ω with the constant time step $\tau = 1/365$ (i.e. 1 day).

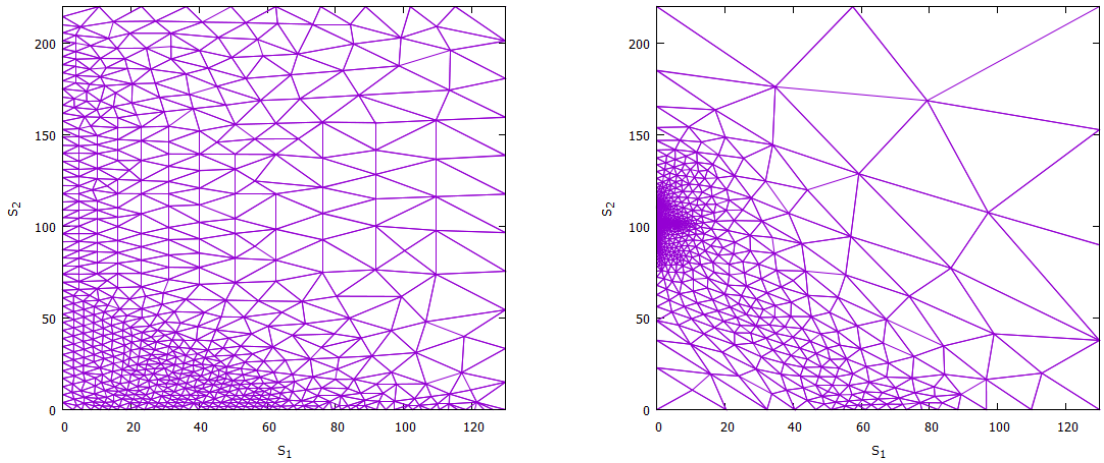


Figure 1 Initial (left) and adapted triangulation (right)

This mesh adaptation is done every 15 iterations and Figure 1 shows initial and final grids. The whole algorithm is implemented in Freefem++, for more details see [3], and uses UMFPACK as a sparse solver.

For the numerical implementation, we consider four distinct ways of volatility determination using market data of individual options (i.e. volatilities implied by the market prices of relevant put options) – (A): constant volatilities $\sigma_1 = 0.6392, \sigma_2 = 0.9461$; (B): piecewise constant implied volatility as a function of x_i, i.e. $\sigma_1 = \sigma_1(x_1), \sigma_2 = \sigma_2(x_2)$; (C): piecewise constant implied volatility as a function of $\alpha_i x_i$, i.e. $\sigma_1 = \sigma_1(\alpha_1 x_1), \sigma_2 = \sigma_2(\alpha_2 x_2)$; (D): piecewise constant implied volatility as a function of moneyness, i.e.

$$\sigma_1 = \sigma_1\left(\frac{\alpha_1 x_1}{\alpha_1 S_1^{ref} + \alpha_2 S_2^{ref}} \exp(-rT)\right), \quad \sigma_2 = \sigma_2\left(\frac{\alpha_2 x_2}{\alpha_1 S_1^{ref} + \alpha_2 S_2^{ref}} \exp(-rT)\right).$$

The results for particular volatility estimations at the reference node $[S_1^{ref}, S_2^{ref}]$ look as shown below:

(A) 15.0376 (B) 15.1697 (C) 15.0376 (D) 15.0393

Clearly, approaches (A) and (C) lead to the same results (volatility is locally the same); also approach (D), which takes into account the option moneyness relative to prices of both assets. By contrast, approach (B) shows completely different prices as might be expected due to ignoring portfolio impact. More extend results showing the impact of particular approaches are depicted in Figure 2.

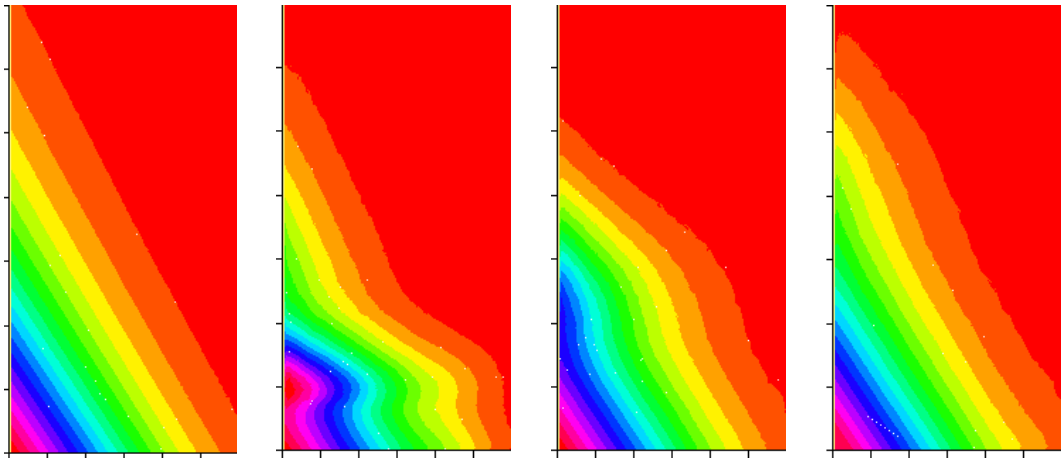


Figure 2 Comparison of approaches, isolvalues of solutions ($\rho = 0.88$), zoom on $[0, 78] \times [0, 154]$

A specific parameter for basket options is the dependency among particular risk sources, which is in our case of Black-Scholes model the Pearson coefficient of linear correlation – the results of its sensitivity are presented in Figure 3.

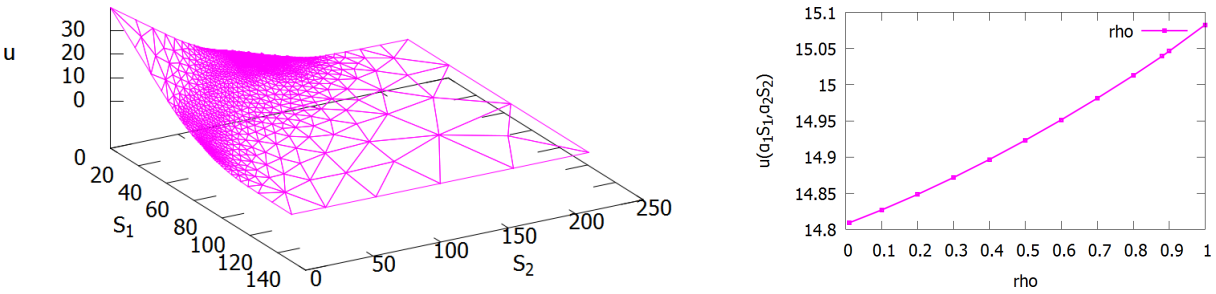


Figure 3 Space-time plot of solution with approach (D) with $\rho = 0.88$ (left) and basket option vs. correlation plot (right)

5 Conclusion

Basket options belong to the family of exotic options for which explicit analytical solution is mostly not available. In this contribution we have developed efficient DG approach for pricing of 2-asset basket options. In the numerical example results for whole range of correlations have been showed. Moreover, results obtained for various specifications of implied volatilites stress the effort to estimate this parameter as efficient as possible. The approach can be subsequently extended for more than 2 dimensions so that pricing of real-life options would be possible.

Acknowledgements

The first author acknowledges support provided within SGS Project No. 21107: Modern numerical methods III, financed by TU Liberec. The second author was supported through the Czech Science Foundation (GACR) under project 13-13142S and through SP2015/15, an SGS research project of VSB-TU Ostrava, and furthermore by the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296.

References

[1] Achdou, Y, and Pironneau, O.: *Computational Methods for Option Pricing*. Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[2] Hakala, J., and Wystup, U.: Making the most out of Multiple Currency Exposure: Protection with Basket Options. *The Euromoney Foreign Exchange and Treasury Management Handbook 2002*, Euromoney Institutional Investor PLC, 2002.

[3] Hecht, F.: New development in FreeFem++, *Journal of Numerical Mathematics* **20** (2012), no. 3-4, 251–265.

[4] Hozman, J.: Analysis of the discontinuous Galerkin method applied to the European option pricing problem. *AIP Conference Proceedings 1570*, 2013, 227–234.

[5] Koh, W. S., and Sulaiman, J., and Mail, R.: Numerical Solution for 2D European Option Pricing Using Quarter-Sweep Modified Gauss-Seidel Method, *Journal of Mathematics and Statistics* **8** (2012), no. 1, 129–135.

[6] Riviére, B.: *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. Society for Industrial and Applied Mathematics, Philadelphia, 2008.

Multivarietal Data Whitening of Main Trends in Economic Development

Radek Hrebik[1], Jaromir Kukal[2]

Abstract. The contribution deals with whitening and principal component analysis in field of main economic indicators. The aim is to show the possible way of using principal components in interpreting selected economic indicators and identify the main trends in the economic development. For the analysis can be selected more economic indicators. Principal component analysis reduce the dimensionality of origin data set. The input for presented research are statistic data describing the economic situation of more than thirty states during more than twenty years. The data come from statistics of European Commission and are published regularly. The contribution presents three basic ways of interpreting these data as input to principal component analysis. Firstly is the object represented by the state in a given year and all its indicators. Secondly is the object represented by the state and all its indicators. Thirdly all the indicators in a given year represents the object. Contribution shows and interprets the results.

The main aim is to show that principal component analysis can lead to another way of interpreting data and indicate main trends in economic development of some countries, to compare the trend of development between countries or group of countries and so on. The identification of some trends thanks to principal components can identify the development of European economics and also comparison with other world economics.

Keywords: Whitening, Principal component analysis, gross domestic product, trend, development.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The contribution is focused on principal component analysis (PCA). The aim is not to describe the principal component analysis itself in detail. The main idea of principal component analysis is reduction of dimensionality of some data set that consists of a large number of interrelated variables. The reduction retains as much as possible of the variation present in the data set. The aim is achieved by transforming to a new set of variables called the principal components. These principal components are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables. [2] In this research is the aim the reduction to two principal components (PC1 and PC2). This means that all selected indicators used as explanatory variables are reduced to two dimensional space to be easily interpreted from its graphical representation.

Paper deals with the basic economic data and shows the ways of possible interpretation to serve as input for principal component analysis. The aim is to search the main indicators, monitor the potential trend of concrete objects and find objects having something in common. This approach as itself is nothing new. The new is data input and expected conclusion. The authors expectation is to get the graphic representation of ecnomic indicators in time that will be easily readable and interpretable. One of the first published use of PCA for the economic time series data was presented in late forties by Stone

[1]Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, 115 19 Behov 7, Prague, Czech Republic, radek.hrebik@seznam.cz
[2]Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, 115 19 Behov 7, Prague, Czech Republic, jaromir.kukal@fjfi.cvut.cz

([2]). It also goes hand in hand for example with principal component analysis goal defined by Abdi and Williams – extracting the important information from the table to represent it as a set of new orthogonal variables called principal components and to display the pattern of similarity of the observations and of the variables as points in maps. [1]

The study dealing with principal component analysis to forecast a single time series with many predictors was presented by Stock and Watson. [5] The use of principal component analysis in connection with gross domestic product is discussed for example in [6]. Favero deals with comparison of two competing methods to estimate large-scale dynamic factor models based, respectively, on static and dynamic principal components. [7] Principal component analysis as alternative way to predict gross domestic product is presented in [8]. Research is based on publicly available economic indicators. Thanks to dimension reduction it gives a lot of possibilities which data to be used. This contribution presents three basic ways of using principal component analysis to interpret economic data. The way means interpreting the data set as objects. These objects represent different economic indicators through time and states. As the reason for doing such research can be also seen trying to predict the future development of some country and find the position of state if we know the basic economic prediction. The main is to capture some progress in time and easily read the development in time. The basic approach used as pattern is represented by objects as years. This means the number of explanatory variables can be so high. The aim of research is to present a big chance to show some new way of economic data interpretation. This is the main aim of this contribution. At this point the presentation of individual components is not the aim. Contribution deals not with the share of individual indicators in components.

2 Data Processing

Let $N, m \in \mathbb{N}$ be dimensionality of object description and length of time series. Let $\mathbf{y}_k \in \mathbb{R}^{\mathbb{N}}$ be k-th sample of original data series for $k = 1, 2, ..., m$. Optional preprocessing (if any) will transform the data to time series $\{\mathbf{x}_k\}_{k=1}^{M}$ where $M = m$ or $M = m - 1$ respectively. Omitted preprocessing as crucial case is represented by

$$\mathbf{x}_k = \mathbf{y}_k \tag{1}$$

for $k = 1, 2, 3, ..., m$. Therefore, the data are used as measured. This kind means no added value and take just the data published by the institutions.

In the case of positive descriptors i.e. $\mathbf{x}_{k,j} > \mathbf{0}$, which is typical for macroeconomic indicators, we can apply relative differences as

$$\mathbf{x}_{k,j} = (\mathbf{y}_{k+1,j} - \mathbf{y}_{k,j})/\mathbf{y}_{k,j} \tag{2}$$

rather use logarithmic differences

$$\mathbf{x}_{k,j} = \ln \frac{\mathbf{y}_{k+1,j}}{\mathbf{y}_{k+1,j}} \tag{3}$$

for $k = 1, 2, ..., m-1$, $j = 1, 2, ..., \mathbb{N}$.

The main processing and data visualisation operates only with time series $\{\mathbf{x}_k\}_{k=1}^{M}$.

3 PCA or Rather Whitening

Dimensionality reduction is based on Principal Component Analysis (PCA) [2]. We calculate mean value vector as

$$\mathbf{x}_0 = \frac{1}{M} \sum_{k=1}^{m} \mathbf{x}_k \tag{4}$$

and covariance matrix estimate as

$$\mathbb{C} = \frac{1}{M-1} \sum_{k=1}^{m} (\mathbf{x}_k - \mathbf{x}_0)(\mathbf{x}_k - \mathbf{x}_0)^{\mathrm{T}} \tag{5}$$

Eigen-Value Decomposition (EVD) is based on equation

$$(\mathbb{C} - \lambda\mathbb{I})\mathbf{v} = \mathbf{0} \tag{6}$$

with constrain

$$\|\mathbf{v}\| = 1 \tag{7}$$

where $\mathbb{I} \in \mathbb{R}^{N\times N}$ identifies matrix, $\lambda \geq 0$ is eigenvalue and $\mathbf{v} \in \mathbb{R}^N$ is corresponding eigenvector. Solutions of EVD can be ordered as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$ with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N$.

Traditional PCA of order $D \in \mathbb{N}$ is based on formula

$$\mathbf{p}_k = \mathbb{W}^{\mathrm{T}}(\mathbf{x}_k - \mathbf{x}_0) \in \mathbb{R}^D \tag{8}$$

where

$$\mathbb{W} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_D) \in \mathbb{R}^{N\times D} \tag{9}$$

Therefore, PCA method supposes user choice of dimension D ($D \leq 3$ for data visualization, $D = 2$ traditionally) and then direct application of (4), (5), (8), (9) to given data set. The PCA method generates $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m \in \mathbb{R}^D$ as data descriptors in D dimensional space. The main advantage of PCA is component decorrelation. It means the correlation matrix of $\mathbf{p}_k$'s is diagonal and their mean value is zero.

Data Whitening (DWH) [4] is a little bit improved process which guarantees unit covariance matrix of resulting vector

$$\mathbf{y}_k = \mathbb{L}^{-1/2}\mathbb{W}^{\mathrm{T}}(\mathbf{x}_k - \mathbf{x}_0) \in \mathbb{R}^D \tag{10}$$

where

$$\mathbb{L} = \mathrm{diag}(\lambda_1, \lambda_2, ...\lambda_{\mathrm{D}}) \in \mathbb{R}^{\mathrm{D}\times\mathrm{D}} \tag{11}$$

under supposition $\lambda_D > 0$.

Therefore whitening method (DWH) also begins with the same data set and given dimension D but apply (4), (5), (10), (11) to generate $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m \in \mathbb{R}^D$ as normalized descriptors in the same space as PCA. We prefer DWH for focusing on components with small variance. The main advantage of DWH is in normalization of decorrelated components. It means correlation matrix of $\mathbf{y}_k$'s is unit matrix, which guarantees unit variance of these components and zero mean value, of course.

4 Macroeconomical Descriptors of State Economics

In this research based in principal component analysis play the key role the input data set. It should be some economic time series. Used economic data has been selected from Statistical Annex of European Economy presented by European Commission in autumn 2014. [3] The selection of the indicators and period was the first task for the authors. There has been selected time series representing data of nine descriptors: Total population, Unemployment rate, Gross domestic product at current market prices, Private final consumption expenditure at current prices, Gross fixed capital formation at current prices, Domestic demand including stocks at current prices, Exports of goods and services at current prices, Imports of goods and services at current prices and Gross national saving. The selection of the descriptors can be made as of the author of analysis needs. This is the advantage to other model approaches. Thanks to dimensionality reduction there can be added more indicators and result of analysis will be still same readable.

As input to analysis serve the thirty five countries from the whole world, majority are the European countries. The observation take place in years 1993 to 2014. Selected indicators are the total population, unemployment rate, gross domestic product at current market prices, private final consumption expenditure at current prices, gross fixed capital formation at current prices, domestic demand including stocks, exports of goods and services, imports of goods and services and gross national saving. So totally nine indicators are monitored. As the time series go to year 2014, the 2014 is represented by prediction from autumn 2014.

As first possible interpretation of the data set is the object represented by a state in a given year. So the number of objects is relatively high. The total number of object is in this case seven hundred and eighty, it represents number of states multiplied by the number of observed years. As the number of object is high, the origin data set dimensionality is relatively small. In this case $N = 9$ representing each macroeconomic indicator and $m = 21 \times 35$ representing 21 years in each of 35 states.

In second case of possible use of principal component analysis the objects are represented by each state. So the properties are made of indicators in selected years. The number of object is thirty five. In comparison to first case of use the number of objects is dramatically fallen down. So the representation will be very simple and it will be clear which states are closed to each other. From graphic representation is expected to easily notice the groups of states. When one point represent one state there is very easily seen the groups of states with similar type of economy. In this case $N = 9 \times 21$ representing each macroeconomic indicator in each year and $m = 35$ representing each of 35 states.

The third kind of data interpretation is by objects representing calendar year. So there is only twenty one objects in this case so that $N = 21$. As the number of objects is very low, the number of descriptors is quite high, so the total number of indicators $m = 35 \times 9$, so each object is created by number of countries mal number of indicators. The number of properties is totally over three hundreds.

5 Data Analysis in Period 1993-2014

The data analysis is presented for all three approaches of representing the given economic data as objects. The input objects are represented by years, states and states in given years. In the first case of analysing the years as objects there is presented the role of preprocessing in this case.

5.1 Years as objects – Role of preprocessing

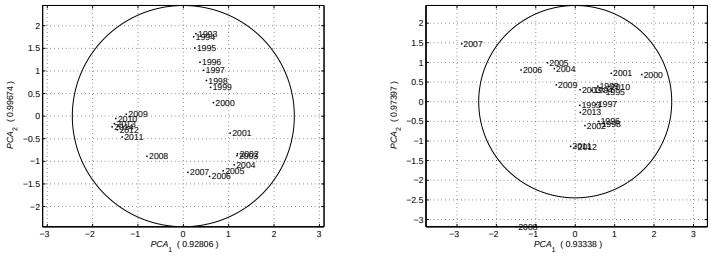
For the analysis there were selected last 21 years to show the main trends and try to identify the main milestones in this period. The next aim is to compare the development of Czech Republic with others. As it is seen from table 1 the best explanation of used data gives the raw data. The same conclusion can be seen in graphical representation. The research is based on searching the clear data interpretation. Comparing four possible ways it is clear that easily read can be using raw data. The term easily read means the clear identification of trend based on graphic representations of principal component analysis of objects represented as years.

Table 1: Principal Components

	PCA1	PCA2
PCA of raw data	0,9281	0,9967
Absolute differences	0,9334	0,9740
Relative differences	0,6965	0,8645
Logarithmic differences	0,6652	0,8533

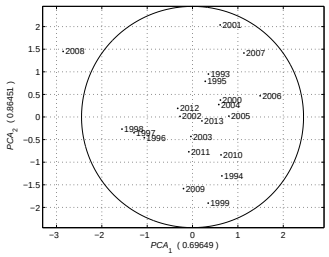
5.2 States and states in years as objects

The principal components are in this case counted from nearly two hundred indicators. So the reduction of dimensionality is quite high in this case. The indicators are created by the nine economy indicators

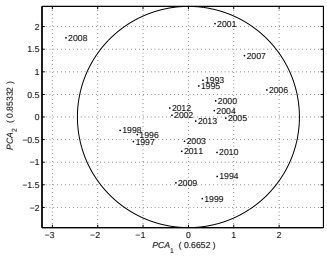


(a) Whitening

(b) Abs. diff. – Whitening



(c) Rel. diff. – Whitening



(d) Log. diff. Whitening

Figure 1: Year as Object of Macroeconomic Investigation

in twenty one years. As in the previous case of using principal component analysis also here are the biggest weights on gross domestic product and population. In case of first principal component is the population values included with bigger weight than in case of gross domestic product. Second principal component is preferring the values of gross domestic product in years. The values of first principal component are in most cases very close to zero, following the weights that implies that the population is without big changes having affect to component values. Second principal component is mostly counted from gross domestic product values. There also apparent the bigger range in values. The analysis ends with some outliers. Even the outliers where omitted in the new run of analysis with reduced number of indicators ended with the similar results. In the second run of the analysis the United States and Turkey were omitted. This led only to other group of outliers. The third attempt to attend the outliers ended with the nearly same result. So objects defined in this way gave no clear result as expected. This is the reason why this way is not to be recommended.

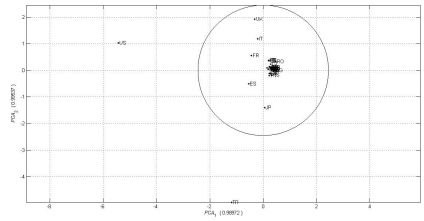
5.3 States in years as objects

The figure 2 shows that main points are concentrated by vertical axis. As representative state of vertical line can be selected for example Germany. As the state represented by movement also in horizontal line can be mentioned for example France. Because the number of objects is quite high, for better interpretation there are the objects for a given year grouped by the same colour.

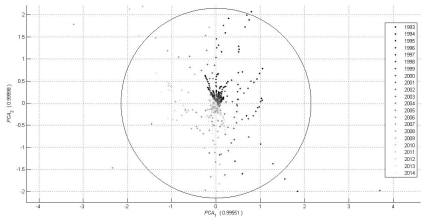
6 Summary

The aim of this contribution was to introduce one of the others, not commonly used, economic data analysis. This was principle component analysis with whitening. It was shown that in case of selected economic indicator the analysis can be done with the raw data without any preprocessing and the results are quite good. The aim was not only to show the analysis but to easily interpret the time series from its graphical representation. This aim was also fulfilled because there were presented the figures from which can be easily seen the movement in economic development. According to the expectation very easily is seen the changes in economy during the last financial crisis. There were presented three ways of interpreting selected economic indicators and two of them led to quite good results. The objects represented as states described by all in indicators in time gave no convincing result.

The two presented approaches to use principal component analysis leads to intuitive and clear interpretation of economic data. The main advantage is, that in two dimensional space is easily possible to



(a) PCA – States as objects



(b) PCA – State in Year as Object with legend

Figure 2: States and states in years as objects

catch more indicators. Thanks to two dimensional space the results are very easily presented in graphical form. The main aim of the research was presented in the figures. Based on principal components there is relatively easy way to present the data and search the directions and movements of economic indicators as a whole. The advantage of our approach is also in the qualification of new predictions and the new approach to evaluate the predictions. For example in case of our data the analysis can be made for real data or data include predictions to rate the position of some country and its future development. The next research will be based on comparison of the PCA results with other econometric models. The question is whether the PCA can predict or give warning sooner than other models. Nowadays it seems yes. Because the movements of years before crisis started are quite significant. This was not aim of this research but will be presented in next papers.

Acknowledgments.

The authors would like to acknowledge the support of the research grant SGS14/209/OHK4/3T/14.

References

[1] H. Abdi and L. J. Williams. *Principal Component Analysis*. (2010). [online]. [cited 2015-02-03]. http://www.utdallas.edu/~herve/abdi-awPCA2010.pdf.

[2] I. T. Jolliffe. *Principal Component Analysis – 2nd Ed.*. Springer (2002).

[3] European Comission. *Statistical Annex of European Economy: Autumn 2014*. Economic and Financial Affairs (2015). [online]. [cited 2015-01-15]. http://ec.europa.eu/economy_finance/publications/european_economy/2014/pdf/statistical_annex_autumn_2014_en.pdf.

[4] Eldar, Y., Oppenheim, A.,V. MMSE Whitening and Subspace Whitening. IEEE Transactions on Information Theory. (2003) 7, 1746-1851

[5] J. H. Stock and M. W. Watson. *Forecasting using principal components from a large number of predictors* Journal of the American Statistical Association, Vol. 97, No. 460, p. 1167-1179. (2002).

[6] C. Schumacher. *Forecasting German GDP using alternative factor models based on large datasets*. Bundesbank Discussion Paper 24/2005. (2005).

[7] C. Favero, M. Marcellino, F. Neglia. *Principal components at work: The empirical analysis of monetary policy with large datasets.* Journal of Applied Econometrics, 20, p. 603620. (2005).

[8] G. Chamberlin. *Forecasting GDP using external data sources.* Economic and Labour Market Review vol. 1(8), August 2007, p. 18-23. (2007).

Simulation of Production Derivatives with Maple: A Path to Corporate Performance Sustainability

Zuzana Chvatalova[1], Jiri Hrebicek[2]

Abstract. The paper deals with the simulation of the production possibilities frontier (PPF), which the advanced tools of the Maple system offer its users. The PPF illustrates selected basic economic principles including: scarcity, tradeoffs, efficiency, opportunity cost, and economic growth in the predefined special interactive Maple document, which is built into the subsection of Finance and Economics/Math Apps in Maple 2015, which has been released in March 2015 by Maplesoft, Inc. (Canada). The paper also points out its new efficient tools. In economics, the production function and its associated variables all belong to the fundamental concepts of solving management problems - their solution is often considered difficult, especially in the long term. Advanced information and communication technologies support understandable solutions, the visualization and the simulation of many of these issues from practice. Recently, some appropriated theories have been developed: for example, sustainable value (as a way of measuring sustainability performance) based on the notion of opportunity costs. Variables based on production theory are important; it is desirable to model them well.

Keywords: Maple, opportunity cost, production, production possibilities frontier, corporate performance sustainability.

JEL Classification: C61, C 63
AMS Classification: 91B38

1 Introduction

Production plays an important role in the economical sciences. It can be understood as a functional dependence of the output volume (production) on the inputs volumes (production factors) of the company in a certain time period. It is therefore tied in with society's technological advancements, and with work, capital, soil and natural resources (production factors), their use and their scarcity (from the perspective of usefulness and limitations). In [8]: Scarcity refers to the fact that society has limited resources and so it cannot produce all of the goods and services that people wish to have, and: "*…the economy faces a tradeoff between the production of Product A and the production of Product B - in order to produce more of A, a greater proportion of the fixed amount of resources must be dedicated to the production of A, thereby reducing the amount of resources available for the production of B*" [8].

Production itself has become an important phenomenon for instance in the creation of econometrics, which has established itself as a separate field of study in the 1930s and which strengthens the participation of mathematically-statistical methods in economics. The character of econometrics has been determined by matters of the production process (time series of adequate data for the United States of America have been processed for the establishing of the production function). The production functions are related to many economic derivatives (concepts and variables), whose knowledge can be essential in the company's management's decision-making in practice, but also for understanding essentials of economics in the process of education. The production possibilities frontier (PPF) is one of these characteristics. "*The PPF illustrates many basic economic principles, including: scarcity, tradeoffs, efficiency, opportunity cost, and economic growth*." [8], but also the marginal rate of substitution (MRS), the marginal rate of transformation (MRT), the profit-maximization problem, the isoprofit-lines, etc.

Today's advancements and the application of information and communication technologies (ICT) resources are supporting these trends. Models and related magnitudes need to "copy" the economic principles by means of using the linguistics, logic and symbolism of quantitative methods and their theories. The advantage of computer systems is their complexity and the fact that they are able to respond to practice stimuli. The Canadian company

[1] *Brno University of Technology, Faculty of Business and Management, Department of Informatics, Kolejni 4, 612 00 Brno, Czech Republic,* e-mail: chvatalova@fbm.vutbr.cz
[2] *Mendel University in Brno, Faculty of Business and Economics, Department of Informatics, Zemedelska 1, 61300 Brno, Czech Republic,* e-mail: hrebicek@mendelu.cz

Maplesoft Inc. has created the Maple system in a complex way. This development responds to the incentives of user groups (academic, professional, teacher, and students, those who apply it in practice, engineers, and others), disciplines (calculations, statistics, physics, economics, technology and many others), and different levels of education. It creates the background and supports a number of professional, technical and scientific forums and centers for communication, documentation and cooperation of their users and developers. From our point of view, it is beneficial: since the 17th release, there has been an expansion of the mathematical applications built into the system using predefined worksheets in the economics and finance area. The current release is Maple 2015, more in [5, 9]. Many instructions for solving practical problems are offered e.g. [4].

Quantification, modeling, and the need of production measurability and many related economical phenomena (that have been previously described mainly verbally) open a rich ambit of questions and problems for resolution. These often lead to optimization, monitoring of development, prediction, as well as to the identification and generalization of the patterns of economical phenomena and the related reciprocities also in the social and environmental area, and corporate governance (e.g., in view of human resources participation, cost-reduction, minimization of pollution or the non-pollution of the environment, not wasting of natural resources), more in [6].

Advanced criteria of company success are emerging. For instance, sustainable value based on the Figge-Hahn (FH) concept [1, 2], which currently plays a crucial role for monitoring the corporate performance sustainability and can be significant in the global development of society. This concept has been openly discussed. Its critics [7] point out: "*FH (2004) introduce the (relative) sustainable value (SV), defined as the difference of the economic output of the firm (e.g., gross or net value added, or some physical output measure) and the opportunity cost of its resource use*.", "*FH (2004, 2005) refer to the alternative activity as the benchmark."* They in [7] have created an identity analogous to that of the conceptual definition by FH. "*However, our definition of SV is more general in that it does not assume any particular functional form … but allows the resources to be interdependent*." Their identity determines SV: *"as a residual between the observed output and the production function*". We can see that the production function in the concept in [7] is a significant indicator of sustainable value identification.

The main goal of this paper is the production possibilities frontier simulation using integrated resources in Maple 2015 and associated analysis as a path to corporate performance sustainability.

2 The Production Possibilities Frontier in Maple

The production possibilities frontier (PPF) illustrates the various output combinations at a maximum given level of use of inputs and technologies. It is the boundary between the attainable and the unattainable production, thus it shows the economy's efficiency. PPFs are also related to opportunity costs. According to [10]: "*A necessary assumption to optimizing the production factors dislocation is to establish the category of the opportunity costs into economic thought. Here we differentiate: accounting profit = returns – accounting costs, economical profit = accounting profit – alternative costs, where alternative costs (opportunity costs) = value of another, the best version that has not been chosen or implemented*." Furthermore, [10]: "*When expressing the production possibilities of the economy, we need to take into account not only the physical limitations of resources and technologies, but also the limits of ecology, law, health, ethics and others; it is therefore suitable to distinguish physical and institutional PPFs*."

To analyze a simple PPF model, which can be visualized in two-dimensional space, we will assume an economy with two types of products as output: product A and product B, both of which need the same factors of production. We should note here: The situation can be generalized for any $n \in N$ using usual mathematical tools, but for $n > 3$ a situation cannot be visualized. The application of comparative statics can help sensibly resolve questions connected with the number of input magnitudes.

It can be seen that the possibility of the PPF simulation and of other interactive analyses provides a more reasonable base for practical decision-making.

2.1 Framework for the Production Possibilities Frontier

First, we briefly summarize the basic characteristics and definitions related to the production derivatives and the needed facts.

[3]: "***Pareto optimal*** *the term used to describe situations in which it is impossible to make one person better off without making at least some others worse off.* ***Pareto superior allocation*** *an allocation that at least one individual prefers and others like at least as well*." [11]: "***The production possibilities set*** *measures the set of outputs that are feasible given the technology and the amounts of inputs. The boundary of the production possibilities set is called* ***the production possibilities frontier***." The shape of this set is dependent on the technologies level. "*The slope of this production possibilities set measures* ***the marginal rate of transformation***" (MRT),

"*MRT measures the rate at which one good can be "transformed" into the other*". By [3]: "***Production possibilities frontier*** *the set of all possible output combinations that can be produced with a given endowment of factor inputs*". By [11] **the marginal rate of substitution** (MRS) is expressed as the slope of an indifference curve. "*The MRS measures the rate at which the consumer is just willing to substitute one good for the other*." From [3]: "***Edgeworth exchange box*** *a diagram used to analyze the general equilibrium of an exchange economy*". Using the **Edgeworth Box Diagram** and the **relevant contract curve,** we can create points on the PPF graph, more in [3]. We need to note here that the **Edgeworth box** is a scheme of two systems of indifference curves, the touching points forming the appropriate contract curve. For more details see, [3, 11]. By [11]: „*At each point on the production possibilities frontier, we can draw an Edgeworth box to illustrate the possible consumption allocations*", see Fig. 1. The fulfillment of conditions for the Pareto efficiency in an economy means that [11]: "*Each person's marginal rate of substitution between the goods must equal the marginal rate of transformation: the rate at which each person is just willing to substitute one good for the other must be the same as the rate at which it is technologically feasible to transform one good into the other*", see Fig. 1. Briefly according to [11] (in [11] this fact is derived in detail): $A = (a_1 + a_2)$, $B = (b_1 + b_2)$: the total amount of the two types of commodities produced and consumed by the two subjects 1, 2; for the transformation function T: the combination $(A, B) \in PPF <=> T(A, B) = 0$; the marginal rate of transformation *MRT:* $\frac{dB}{dA}$; the ordered pair (dA, dB): the small production change (feasible); the utility level for the subject 1, 2: u_1 , u_2 (let $\bar{u} = u_2(a_2, b_2)$). Thus:

$$\frac{\partial T}{\partial A} dA + \frac{\partial T}{\partial B} dB = 0, \quad i.e. \qquad \frac{dB}{dA} = -\frac{\partial T/\partial A}{\partial T/\partial B} \tag{1}$$

According to the Pareto efficient allocation, a maximization problem can be described as follows:

$$\max_{a_1, a_2, b_1 b_2} (u_1(a_1, b_1)),\ \bar{u} = u_2(a_2, b_2),\ T(A, B) = 0 \tag{2}$$

After designating the Lagrangian as $L = u_1(a_1, b_1) - \lambda(u_2(a_2, b_2) - \bar{u}) - (\mu(T(A, B) - 0))$, with the first-order conditions being $\frac{\partial L}{\partial a_1} = 0, \frac{\partial L}{\partial a_2} = 0, \frac{\partial L}{\partial b_1} = 0, \frac{\partial L}{\partial b_2} = 0$, and after the adjustment we get:

$$\frac{\partial u_1/\partial a_1}{\partial u_1/\partial b_1} = \frac{\partial T/\partial A}{\partial T/\partial B} \text{ and } \frac{\partial u_2/\partial a_2}{\partial u_2/\partial b_2} = \frac{\partial T/\partial A}{\partial T/\partial B} \tag{3}$$

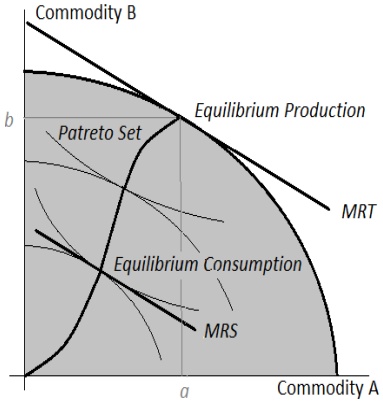


Figure 1 Production and the Edgeworth box. (Source: Authors´ elaboration according to [11])

Furthermore, we consider how to maximize of the company's profit *PR* (for two inputs *L* and *K* and the two outputs *A* and *B*) according to [11]:

$$\max_{A,B,L_A,L_B} (p_A A + p_B B - (w_A L_A + w_B L_B)) \tag{4}$$

where , $B =$ two outputs; p_A, p_B = the price per unit of *A*, price per unit of *B*; L_A, L_B= two inputs (two kinds of labour); w_A, w_B= the wage rates of two inputs. Let $L^* = w_A L_A^* + w_B L_B^*$ represent the labor costs of production (the company supposes, that in equilibrium the quantities of inputs L_A^* and L_B^* are optimal, L^* is constant). Then for the company's profit *PR* we get:

$$PR = p_A A + p_B B - L^* \tag{5}$$

The description of the company's isoprofit-lines (with a negative slope of $-\frac{p_A}{p_B}$ and an intercept of $\frac{L^* + PR}{p_B}$ with the vertical axis) is as follows, see Fig. 2:

$$B = \frac{L^* + PR}{p_B} - \frac{p_A}{p_B} A \tag{6}$$

Maximizing the company's profit means choosing the isoprofit-line which must represent a tangent to the production possibilities set - it has the highest possible intersection with the vertical axis and the slope of the production possibilities set is equal to the slope of the isoprofit-line, see Fig. 2:

$$MRT = -\frac{p_A}{p_B} \tag{7}$$

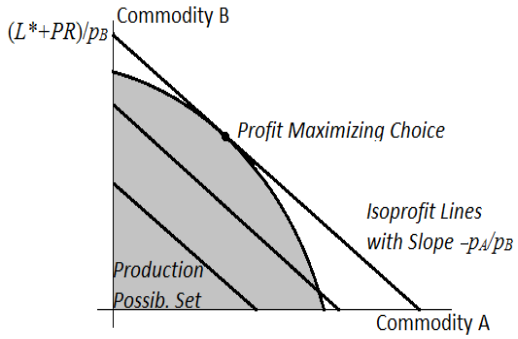


Figure 2 Profit maximization (Source: Authors´ elaboration according to [11])

These findings can be generalized for more companies.

2.2 The Production Possibilities Frontier Simulation by Maple 2015

Using Maplesoft activities we can solve economic and financial issues through the Maple system in various ways: create, or algorithmize our own proposals for solutions of given tasks in the Maple environment; inspire, or modify documents created by Maple users which are archived in the Maplesoft Maple application center in the categories *Economics* and *Financial Engineering* (currently there are more than 200 matching applications); to use the pre-built interactive Maple worksheets predefined in the Maple system in the *Math Apps* section from the menu on the top toolbar (the current release Maple 2015 contains almost 50 worksheets). Below, we will present an application from this area, which simulates the PPF issue.

The Production Possibilities Frontier Worksheet (PPFW) and its functionality are prescribed in *Startup Code* that uses modules, procedures, cycles, etc. The variables can be flexibly viewed, hidden and filtered. The PPFW can be easily modified, or edited using the toolbar in the top bar, see Fig. 3. The prescription for entering the PPF function can be determined, e.g., as a regression model using Maple Library Statistics. Functions coming directly from the built-in PPFW are presented in the examples below:

$$B = \sqrt{50000^2 - A^2} \quad \text{for} A \in < 0; 50000 > \tag{8}$$

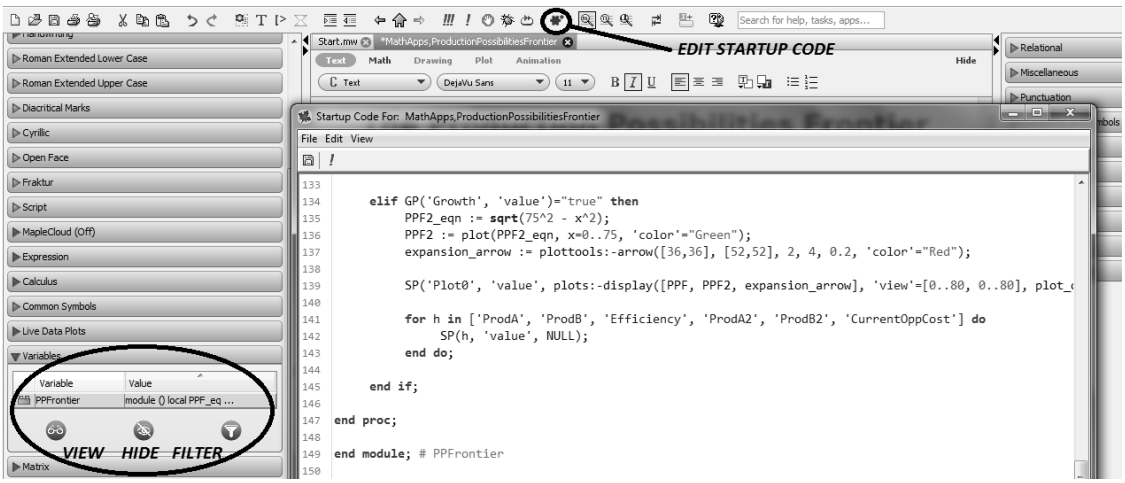


Figure 3 Part of the startup code - illustration (Source: According to [8])

The PPFW is structured into several areas. The main concept of the communication interface provides both a *theoretical background* on the issue of PPF (*Scarcity and Tradeoffs*, *Efficiency*, *Opportunity Cost*, *Economic Growth* through inserted sections) and *references* with links to specific pages. The *working part* includes components for essential calculations, visualization and simulation (*Production Efficiency*, *Opportunity Cost*, *Economic Growth*) Using the radio buttons we can choose which economic principle we want to investigate (by clicking and dragging the mouse on the plot, different production combinations can be viewed). The system automatically responds: it specifies the value of the selected products *A* and *B*, and evaluates the situation.

Production Efficiency: E.g., if we click on a combination *A* =22723 and *B*=29269, the system responds: *This is an attainable production combination with an inefficient use of resources*, see Fig. 4.

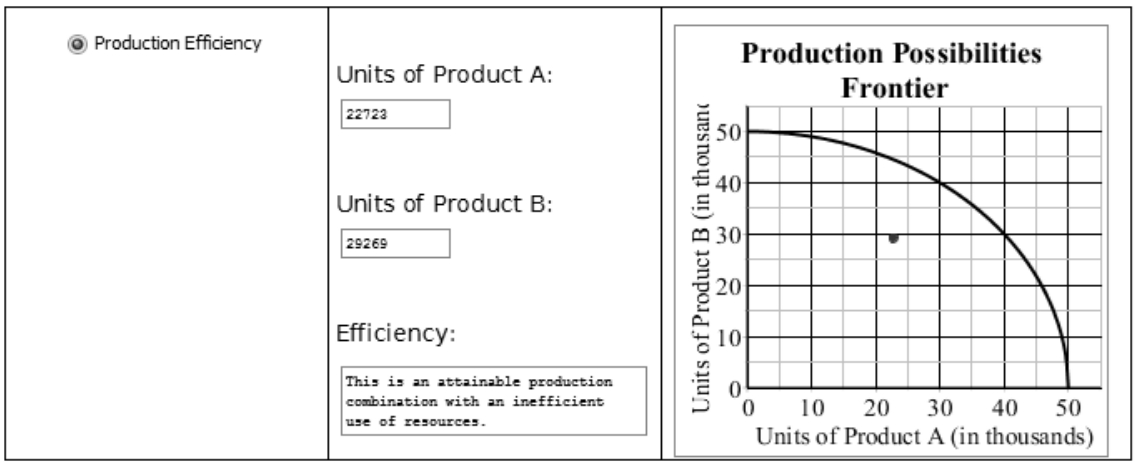


Figure 4 Production efficiency - illustration (Source: According to [8])

Note: The system responds in the case that the point lies on the PPF (e.g., *A*=27370, *B*=42134: *This is an attainable production combination with an efficient use of resources*; in the case that the point lies to the right on the PPF (e.g., *A*=40864, *B*=47325): *This is an unattainable production combination given the available resources*.

Opportunity Cost: After selecting a point on the PPF, the system will automatically again show the precise value and will draw a tangent line at this point and calculate its slope, i.e., it will compute the opportunity cost (what must be the sacrificed units of product *B* for the production of additional units of products with *A*). For instance, for products *A*=34696 and *B*=36002, the system responds, see Fig 5: *To produce one additional unit of A, the production of 0.964 units of B must be sacrificed*. (*Note*: The slope of tangents is automatically taken as negative but according to historical practices, we do not write a minus sign.) This slope can be expressed as a derivative of relationship (8) in the *A*. (According to below: $\frac{\partial B}{\partial A}(34696) = \frac{34696}{\sqrt{50000^2 - 34696^2}} = 0{,}9836822412.$)

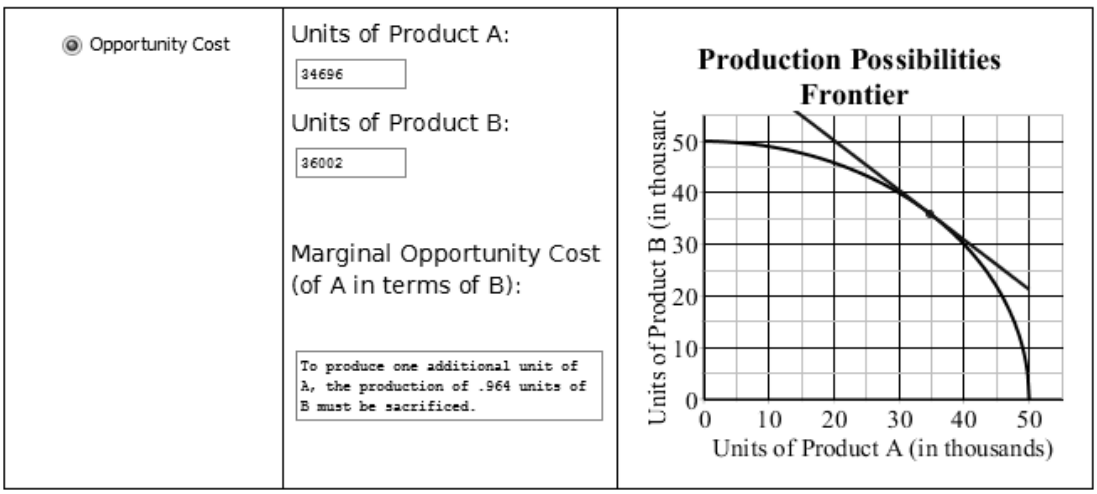


Figure 5 Opportunity cost - illustration (Source: According to [8])

Economic Growth: The system shows an economic growth, which can be also illustrated by moving the PPF curve towards the right or upwards (to the northeast), see Fig. 6.

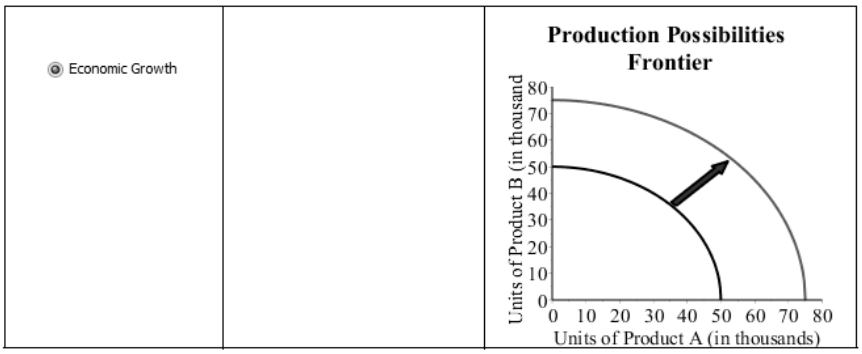


Figure 6 Economic growth (Source: According to [8])

3 Conclusion

It goes without saying, consequently, that the production functions theory closely relates not only to economical reciprocities but also to environmental politics, social potential and the level of every company's corporate government, because it impacts a whole range of activities and reciprocities. From the viewpoint of sustainable company performance as well as economic growth, production and magnitudes connected with production play an important role. Many advantages for practical decision-making, as well as for the process of education, are brought by quantification, algorithm development, production-related problem-solving, and applying the means provided by information and communication technologies. These methods also epitomize a challenge of searching for and applying modern scientific approaches in practice.

In the present contribution, we have shown the possibility of simulation, leading to achievable effective production. The reasoning is based on the Pareto efficient allocation, the link between MRS and MRT, and others. The simulations are presented using a predefined PPF worksheet in the Maple environment. We can see that the option of fast automatic calculations and especially visualizations can, in practice, lead to comfortable reasoning about limit values leading to profit, production achievability, non-wasting, but also about making use of capital goods. At the same time, getting good results definitely requires a workable relationship by which the PPF will be determined. Regression is the method of procuring this relationship obtained empirical values.

Acknowledgements

Supported by The Czech Science Foundation. Name of the Project: Measuring Corporate Sustainability in Selected Sectors. Reg. No. 14-23079S.

References

[1] Figge, F., and Hahn, T.: Sustainable value added: measuring corporate contributions to sustainability beyond eco-efficiency. *Ecological Economics* **48** (2004), 173-187.
[2] Figge, F., and Hahn, T.: The cost of sustainability capital and the creation of sustainable value by companies. Journal of Industrial Ecology **9** (2005), 47-58.
[3] Frank, R. H.: *Microeconomics and Behavior*. 7th ed. McGraw-Hill, Irwin, 2008.
[4] Gander, W., and Hřebíček, J.: *Solving Problems in Scientific Computing Using Maple and Matlab*. Springer, Heidelberg, Germany, 2004.
[5] Chvátalová, Z., Hřebíček, J., Koch, M., and Kalabis, P.: Corporate Performance Analysis Using the Maple system: Integration of Theory and Practice. (International conference *Perspectives of Business and Entrepreneurship Development*, Brno University in Technology, Faculty of Business and Management, May 2015) - in processing.
[6] Kocmanová, A., Hřebíček, J., Dočekalová, M., Hodinka, M., Hornungová, J., Chvátalová, Z., Kubálek, T., Popelka, O., Šimberová, I., Topolová, I., and Trenz, O.: *Měření podnikové výkonnosti.* Littera, Brno, 2013.
[7] Kuosmanen, T., and Kuosmanen, N.: How not to measure sustainable value (and how one might). *Ecological Economics*. vol. **69** (2009), 235-243.
[8] Maple (release Maple 2015): *The production Possibilities Frontier*. Tools/Math APPS/ Finance and Economics. Waterloo: Maplesoft, 2015.
[9] Maplesoft Inc.: *Company*. [online] [cit.2015-04-04]. 2015. Available from: http://www.maplesoft.com/
[10] Pavlačka, O.: *Ekonomie 1*. Basic lecture materials + further information. [online] [cit.2015-05-05]. Available from: http://aix-slx.upol.cz/~pavlacka
[11] Varian H. R.: *Intermediate Microeconomics*. *A modern approach*. 8th ed. W. W. Norton & Company, Inc., London, New York, 2010.

An Application of input-oriented CCR model and the Malmquist productivity index for Measuring the Development of the Bank Efficiency in the Visegrad Group

Lucie Chytilová[1], Jana Hančlová[2]

Abstract. The banking industry all around the world is very difficult nowadays. Especially the Central Europe is confronted not only with new rules from the European Commission or the 2008 crisis, but it also face problems resulting from recent historical changes. The aim of this article is to determine the efficiency of the banking industry in Visegrad Group (V4). We estimate the technical efficiency of banks in the V4 for years from 2008 to 2013. The input-oriented CCR model is used for the estimation. There are provided three types of models. Two models contain the input variable equity. In first case it is used as radial variable input and in second as so called quasi-fixed input. Then there is model without the variable equity, this model just includes the classical input and output variables for the banking industry. These models are examined and the development of calculated technical efficiency is analyzed in the cross section between the units and in the time. The Malmquist productivity index (MPI) is used to observe the changes in efficiency. The results show that the V4 is diverse group.

Keywords: Data envelopment analysis, input-oriented, CCR model, banks, technical efficiency.

JEL classification: C61, G21, D24
AMS classification: 90C05

1 Introduction

The last hundred years have been dramatic for most of the European countries. There have been World Wars, various political mindset, the European Union (EU) or different crisis. The European countries changed demographically, politically and economically. Since the World War II. countries as Czechoslovakia, Hungary and Poland were part of the so-called a social block. The social block was influenced by the Soviet Union. Economically it was characterized by a centrally planned economy with a focus on economic relations of countries within the block. The social block had fallen apart in 1989. This led some countries of social block to integrate and cooperate. In February 199 three countries - Czechoslovakia, Hungary and Poland have created the V4. Since 1993, the V4 consists four countries - the Czech Republic, Slovakia, Hungary and Poland. A market economy was created and opened to the rest of the world. The transformation of the banking system was an essential part of the transformation. The V4 countries are part of the EU since 2004. Slovakia joined the third stage of the European Monetary Union in 2009. Different monetary issues and the world globalization makes differences between countries everywhere and affect financial system. The V4 countries are geographically and historically close. Therefore the properties of their financial and banking systems should not show major differences. Identification of the of potential differences are one of the objectives of this paper.

The paper focuses on the identification of the efficiency of commercial banks in the V4 during time period from 2008 to 2013. We used a nonparametric approach - DEA (CCR input-oriented model) to

[1]VŠB-TU Ostrava, Faculty of Economics, Sokolská třída 33, Ostrava, Czech Republic, lucie.chytilova@vsb.cz
[2]VŠB-TU Ostrava, Faculty of Economics, Sokolská třída 33, Ostrava, Czech Republic, jana.hanclova@vsb.cz

compute the technical efficiency. This refers to the ability of the bank to produce existing level of outputs with the minimum of inputs. There have been two main approaches, based on the traditional microeconomic theory of the firm, for the bank efficiency evaluation as production unit. The production approach by Benston [1] and the intermediation approach by Sealy and Lindley [16]. We use the intermediation approach where deposits are considered as input. This approach identifies inputs and outputs precisely, but there is still question about the variable - equity. To clear this question, we created three models with different use of the variable equity. Models were calculated and compared for the banks in the V4 to see which one gives better picture of the situation. The MPI was used to distinguishes changes in technical efficiency to determine the exact source over time.

The remainder of this paper is organized as follows. Next section discuss the review of the literature. In Section 3 the methodology of DEA method and Malmquist approach is exposed. Section 4 introduces the used data while Section 5 presents empirical results. Section 6 provides final conclusions.

2 Review of the Literature

The empirical analysis of banking efficiency and other financial institutions are frequently discussed topics. Great attention is focused on the European banking industry. The situation there is difficult due to the different situation of each country in the EU. Fang et al. [6] identified the reasons for the interest. The main reasons according to them are three - the structural changes of banking regulation, financial market structure and competition environment in transition economies; the development of security markets and more banks are treated publicly in stock market; under the Basel Capital Accord, greater supervisory efforts have been focused on risk management and enhancing the capital ratio.

The efficiency for the banking industry was investigated for all Europe as well as for the V4. The work of Hančlová and Chytilová [11], Lyroudi and Angelidis [12] and other for the Europe may be mentioned as the inspiration. The investigations for the V4 from Řepková [15] or Stavárek [14] also exist.

There are known two main approaches for the bank efficiency evaluation as production unit. The main difference is the treatment of deposits. The production approach views deposits as output. Banks are producers of deposits, loans and other services. Inputs are define as physical variables. This approach was found by Benston [1]. Benston also found disadvantages - detailed database is required and it does not take into consideration the interest costs. The approach was used by Boďa and Zimková [2] for Slovak banks. The second approach was found by Sealy and Lindley [16] - intermediation approach. Banks are financial intermediaries between depositors and creditors. They collect deposits and other liabilities to apply them as interest-earning assets. Deposits are considered as input. In this case operating cost and interest cost are considered. It is the most common approach nowadays.

Even the approach is defined, there are question how to determine the exact type and number of variables for the model. Toloo and Tichý [17] documented that if the number of performance measures is high compared with the number of production units then a large percentage of the units is efficient. They extended both multiplier and envelopment forms of DEA models and propose two alternative approaches for selecting performance measures under variables returns to scale for banks in the Czech Republic.

There also exist two main techniques for determining the performance of production units - non-parametric and parametric approaches. DEA is one of the non-parametric approach. The optimal solution helps to find the efficient frontier. This approach is presented in work of Melecký and Staníčková [13] for example. The stochastic frontier analysis (SFA) is the parametric approach. The production system has only one output and stochastic frontier function is estimated by econometric methods. The paper by Hančlová [10] presents advantages and disadvantages of both approaches.

3 Methodology

3.1 DEA and CCR-I model

A new perspective on firm efficiency and productivity was introduced by work of Farrel [8], more precisely the calculation of the benchmark technology and efficiency measures. The basic premise was to leave the perfect input-output allocation and allow the inefficient operations. The inefficiency of firm is defined as the distance of a firm from a frontier production function accepted as the benchmark. The difference

between inefficient observed point and the production frontier measures how much is the firm inefficient - the production of firm lies below the frontier. The radial contraction is used for this measurement.

The former reflects the ability of production units to minimize input use as to produce a given amount of inputs (ie. input-oriented approach - I). Non-parametric approach may be used to calculate efficiency. Charnes et al. [4] introduced the calculation based on linear programming and DEA as CCR-I model under assumption of constant return-to-scale (CRS). Their work developed the piece-wise-linear convex hull approach to frontier estimation which were proposed by Farrel [8] in a input-oriented model. The CCR-I model is mathematically as it follows:

$$\begin{array}{ll} & \min_{\theta,\lambda}\theta_o, \\ \text{s.t.} & \sum_{j=1}^{n}\lambda_j x_{ij} \le \theta_o x_{io}, \quad i=1,2,...,m, \\ & \sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{ro}, \quad r=1,2,...,s, \\ & \lambda_j \ge 0, \quad j=1,2,...,n, \\ & \theta_o \in (-\infty,\infty), \end{array} \tag{1}$$

where $x=(x_1,...,x_m)$ and $y=(y_1,...,y_s)$ are vectors of inputs and outputs, respectively. λ_j is the weight for each DMU. It is dual-variable unit. The θ_o represents the efficiency score of DMU o-th. The DMU is efficient if and only if $\theta^*=1$. Note, the minimization problem is calculated for each bank of sample.

This type of model was chosen according to literature, see [9]. The assumption of CRS is better for the small sample size of DMUs and the comparison between small and large banks, according to [9].

3.2 CCR-I model with quasi variable

There exist many improvements of the classical CCR-I model. In our paper we used the model for the case where quasi fixed input variable is involved. This model is define below.

Suppose that there exist a set of n-banks ($j=1,...,n$), that produce a vector of s outputs $y=(y_1,...,y_s)$ using a vector of m classic/discretionary inputs $x^D=(x_1^D,...,x_m^D)$ and l quasi-fixed inputs $x^Q=(x_1^Q,...,x_l^Q)$. Then the technical efficiency for the o-th bank can be computed as the solution of the following linear programming model:

$$\begin{array}{ll} & \min_{\theta,\lambda}\theta_o, \\ \text{s.t.} & \sum_{j=1}^{n}\lambda_j x_{ij}^D \le \theta_o x_{io}^D, \quad i=1,...,m, \\ & \sum_{j=1}^{n}\lambda_j x_{kj}^Q \le x_{ko}^Q, \quad k=1,...,l, \\ & \sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{ro}, \quad r=1,2,...,s, \\ & \lambda_j \ge 0, \quad j=1,2,...,n, \\ & \theta_o \in (-\infty,\infty). \end{array} \tag{2}$$

The optimal value θ_o^* corresponds to the technical efficiency of o-the bank. It belongs to range $0<\theta_o\le 1$. If $\theta_o^*=1$ then the o-th bank is called Pareto-efficient. Linear model (2) measures the technical efficiency of o-the bank. It must be solved n times to obtain the technical efficiency of each individual bank.

3.3 Malmquist Index

The Malmquist index was introduced in work by Caves et al [3]. Fare et al. [7] generalized their work and defined an input-oriented productivity index, known as the geometric mean of two Malmquist indices. In producer analysis the Malmquist (Productivity) Index (MPI) may be used to construct indexes of input, output or productivity as ratios of input or output distance functions. The most famous method for measuring distance functions is the linear programming method. The MPI allows to measure the total productivity by means of distance-functions calculation. Of course, this can be estimated through the solution of mathematical programming problems of the DEA. This is done in this paper.

Based on the work of Zhu [18] the Malmquist productivity index $\text{MPI}_0^{t,t+1}$ for measuring the efficiency change of n-banks ($j=1,...,n$) where it is supposed the input vector $x_j^t=(x_{1j}^t,...,x_{mj}^t)$ and vector output $y_j^t=(y_{1j}^t,...,y_{sj}^t)$ at each time period t, $t=1,...,T$, is formulated in the following form:

$$M_o=\Big[\frac{\theta_o^t(x_o^t,y_o^t)}{\theta_o^t(x_o^{t+1},y_o^{t+1})}\frac{\theta_o^{t+1}(x_o^t,y_o^t)}{\theta_o^{t+1}(x_o^{t+1},y_o^{t+1})}\Big]^{1/2}, \tag{3}$$

where $\mathrm{MPI}_0^{t,t+1}$ measures the productivity changes between periods t and $t+1$. The productivity declines if $\mathrm{MPI}_0^{t,t+1} > 1$, remains unchange if $\mathrm{MPI}_0^{t,t+1} = 1$ and improves if $\mathrm{MPI}_0^{t,t+1} < 1$.

The following modification of M_o by Zhu [18] makes it possible to measure the change of technical efficiency and the movement of the frontier in terms of specific n^{th} bank:

$$M_o = \frac{\theta_o^t(x_o^t, y_o^t)}{\theta_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \cdot \left[\frac{\theta_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{\theta_o^t(x_o^{t+1}, y_o^{t+1})} \frac{\theta_o^{t+1}(x_o^t, y_o^t)}{\theta_o^t(x_o^t, y_o^t)}\right]^{1/2}, \tag{4}$$

where the first term on right side measures the magnitude of technical efficiency change (TEC) between periods t and $t+1$. The term on left side indicates if technical efficiency improves, remains or declines. The second terms measures the shift in the possibility frontier (FS), i.e. technology frontier shift, between periods t and $t+1$.

4 The Data and Measurement of Inputs and Outputs Variables

The most important task for the efficiency measurements is to identify the right and relevant variables for model. In Section 2 we introduced two main approaches for selecting variables for the banking efficiency measurements. The intermediation approach was chosen for this paper as the most used in this area. Solver in MS Excel, such as the DEA Frontier was used in this paper.

Even of the common use of the intermediation approach there is question about variables. Gulati and Kumar [9] establish two models for modelling of the bank efficiency. The first model included the non-interest income and the second one did not. They applied these models to Indian banking system and compared these models. Hančlová and Chytilová [11] presented the same models for the European banking industry and found that if the non-interest income is not included into output then it all leads to usual undervalued estimation of cost efficiency. Based on the facts, this paper uses the model where the non-interest income is included into outputs to represents the non-traditional activities. The rest of outputs are standard. Another question is how the variable *equity* should be used. To find out the answer we identify three models: model A - the variable equity is used as the discretionary input; model B - the variable equity is used as the quasi-fixed input; model C - the variable equity is not used at all.

	Variables	Description in the balance sheet	Units
Inputs	Physical capital (x_1 - FA)	Fixed assets	th Euro
	Labour (x_2 - LAB)	Number of employees	Number
	Loanable funds (x_3 - LF)	Deposits + Short term funding	th Euro
	Equity (x_4) - *Classical/Quasi-fixed*	Total equity	th Euro
Outputs	Advances (y_1 - ADL)	Loans + Advances to Banks	th Euro
	Investments (y_2 - INV)	Other Securities	th Euro
	Non-interest income (y_3 - NII)	Non-Earning Assets	th Euro

Table 1 Description of inputs and outputs

The required data set of inputs and outputs have been collected from the Bankscope[1]. In Table 1 is seen the description of inputs and outputs. This paper is based on period between 2008 to 2013. There have been many banks in the V4 in this period but we analyzed just 27 banks (8 - CZ, 8 - PL, 6 - SK, 5 - HU). This is due to the fact that some of them did not exist all the time or some variables were missing.

5 Empirical Analysis and Results

5.1 Comparison of three CCR-I models

The results showed that all three models give similar results. Table 2 shows the average efficiency score of the V4 banking industries for all models in the period. From Table 2 we can see that the relationship among efficiency scores of models is $av.EF\_ModelA \geq av.Ef\_ModelB \geq av.Ef\_ModelC$. This is in

[1] https://bankscope.bvdinfo.com/

consistent with theoretical result of [5] on page 111 (lemma 4.2). In Model B is equity as quasi fixed variable (non-radial). This efficiency (0.5852 - 0.7877) for the production model should not be higher than for the Model A (0.5880 - 0.7953). If the equity variable is excluded than the efficiency score of this product Model C have to even lower (0.5852 - 0.7842). This dependent on significant role of the equity variable in Model A and B. In terms of the differences of average efficiency scores we may see minimal differences between Models B and C. This suggest the possibility of omission the variable equity as input variable. Bigger differences are been between Models A and B. This shows the impact of inclusion the variable equity, ones as radial variable. We also can evaluate the results as development in time, where all Models A, B and C show similar trend in development - sharp rise of the average efficiency after 2008 until 2010 and relative oscillation around the level 77%. This is similar to result of [15].

	2008	2009	2010	2011	2012	2013	over all average by model
Model A	0.58799	0.68443	0.79285	0.76257	0.79529	0.75628	0.72593
Model B	0.58522	0.66860	0.78560	0.75850	0.78774	0.75531	0.71949
Model C	0.58522	0.66860	0.78417	0.75850	0.78421	0.75531	0.71873

Table 2 Average efficiency score of the V4 banking industry by each model

Table 3 provides information about the relative number of efficient banks in each year for the V4. We present this generally for the model because the number of the efficienc DMUs have been same across the all models after rounding to five digits.

number of efficient banks	2008	2009	2010	2011	2012	2013
all	14.81	22.22	33.33	29.63	25.93	18.52
in the Czech Republic	11.11	11.11	14.81	14.81	14.81	14.81
in Poland	0.00	3.70	3.70	3.70	0.00	0.00
in Hungary	0.00	0.00	7.41	7.41	3.70	0.00
in Slovakia	3.70	7.41	7.41	3.70	7.41	3.70

Table 3 The relative number of efficient banks in the V4

From Table 3 may be seen that relative number of efficient unit increased from 2008 from 14.81% to 33.33% in 2010. In the wake of financial and economic crisis there have been decrease to value 18.52% in 2013. In terms of trend comparison of V4 countries we can say that the number of efficient Czech banks have grown from 11.11% to 14.81% in 2010 and than it was stable. The number of Polish and Hungarian efficient banks have increased until 2010 to value 3.7% and 7.41%. In 2011 there have been decrease of efficient commercial banks again. Slovakian banks have oscillate around the share of 5% with small impact of the financial crises after 2010.

Over all we can say that most efficient in the V4 is the Czech Republic with a slightly growing trend. The banking industry is stable - same banks are identified as efficient in the period. As least efficient we can consider banking industry in Poland. This might be caused by the turbulent environment. Many banks went out of the market or changed the ownership. It is also important say that banks with missing data or change in the structure during the period have been excluded. This might cost some shortcomings and should be improvement in future analysis.

5.2 Productivity of banks and countries of the V4 by Mamlmquist Index

In this paper, the minimization-based Malmquist productivity index was used. This means that if Malmquist productivity index would be equal to 1 then there was no significant change in efficiency and performance. If the index would bigger than 1 then significant efficiency declines and performance is getting worse and significant efficiency is getting better and performance advancements if the index would be less than 1.

Malmquist index has been calculated for the Model A and C. The the Malmquist index values for both models are in Figure 1 for periods 2008-2010, 2011-2013 and all period 2008-2013. Previous results did not show many differences in models. This case is different. Same results for both models are seen just for DMUs - 8 (GE Money Bank a.s. - CZ), 10 (J&T Banka a.s. - CZ), 18 (Post Bank JSC-Postova

Banka a.s. - SK), 22 (Sberbank Slovensko a.s. - SK), 25 (Unicredit Bank Czech Republic and Slovakia a.s. - CZ) and 26 (UniCredit Bank Hungary Zrt - HU). The rest of results are different and in some cases there are also different over all results. These different results are for DMUs - 12 (mBank a.s. - PL), 13 (MKB Bank Zrt - HU), 21 (Raiffeisen stavebni sporitelna a.s. - CZ) and 23 (Stavebni Sporitelna Ceske Sporitelny a.s. - CZ). It is seen that the origin of the banks is probably not the reason.

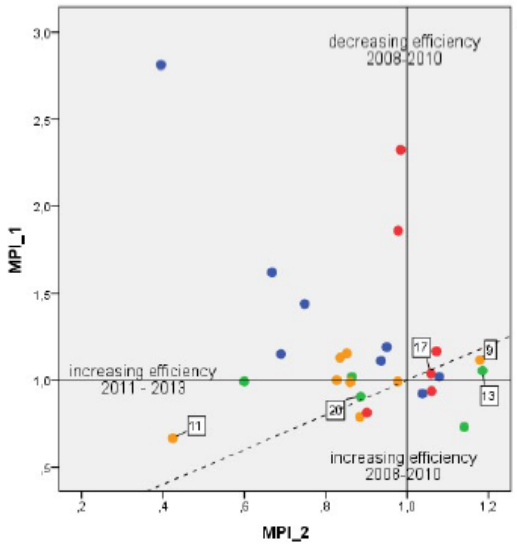


Figure 1 Malmquist index for the Model A and Model C

6 Conclusions

The main aim of the paper was to assess the possibility of influence to include the input variable equity into the non-parametric CCR input-oriented model to evaluate the efficiency of selected commercial banks from V4 countries and compare the trend of differences of the development by using MPI.

The results indicate that the input variable equity in the intermediation approach plays an important role in case of a radial variable. Results of the Model B and C demonstrate that it is possible to exclude this variable when the comparison with the fixed quasi input variable is made. The average technical efficiency of Model A, B and C were equal to 72.6%, 71.9% and 71.8%, respectively. This is in line with the theoretical concept of all individual models. The development of this efficiency in the case study pointed out a sharp rise since 2008 from a level of around 58% until the year 2010. Than the oscillation around a value of 77% followed. And therefore the financial crisis displayed at all banks of the V4.

The development of relative share of efficient banks in individual V4 countries is showing sharp rise until 2010. The number of efficient banks raised twice to value 33%, but it was followed by a steady decline to 18.5% of efficient banks in 2013. This trend is mainly made by Polish and Hungarian banks. The best share of efficient banks was found in the Czech Republic, the range was from 11.1 to 14.8%.

In the third part of the article the change of bank efficiency development has been evaluated for all banks in V4 countries for model involving the equity input variables and model without this variable, especially in regards to the breaking year 2010. These results classify banks into four basic groups according to the bank's growth or decline of efficiency in the two periods. Also the intensity of the change was explored. The first group was improving by the change of efficiency. The increasing efficiency across all periods are for example: 11 - Polish bank (mBank Hipoteczny a.s.) and 20 - Hungarian bank (Raiffeisen Bank Zrt). On the other hand, there are banks whose changes of efficiency in both periods are decreasing. The intensity of the decrease declines more after the year 2010 - Hungarian bank MKB Bank Zrt (13), Polish bank ING Bank Slaski a.s. (9) as well as Slovak bank OTP Banka Slovensko (17).

The results allow the comparison of bank efficiency in the V4, the comparison of their efficiency development in the period and classification of these banks according to the efficiency changes in the examined period. The closer analysis of the inefficiency causes depending on the size of banks and the prevailing domestic or foreign property could complemented the obtained empirical results in future.

Acknowledgements

This paper was created under SGS project (SP2015/117), Operational Programme Education for Competitiveness Project CZ.1.07/2.3.00/20.0296 and the grant no. 14-31593S of the Czech Science Foundation.

References

[1] Benston, G. J.: Branch Banking and Economies of Scale. *Journal of Finance* **20** (1965), 312-33.

[2] Boďa, M., and Zimková, E.: Service-oriented efficiency of Slovak banks. *The 7th International Days of Statistics and Economics* (2013), 164-172.

[3] Caves, D. W., Christensen L. R., and Diewert, W. E.: The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity'. *Econometrica* **50** (1982), 1393-1414.

[4] Charnes, A., Cooper, W. W., and Rhode, E.: Measuring the efficiency of decision making units, *European Journal of Operational Research* **2** (1978), 429-444.

[5] Cooper, W. W., Seiford, M. L., and Tone, K.: *Data Envelopment Analysis*. Springer, New York, 2007.

[6] Fang, Y., Hasan, I., Marton, K. and, Waisman, M.: Bank valuation in new EU member countries. *Economic Systems* **38** (2014), 55-72.

[7] Fare, R. et al.: Production Frontiers. Cambridge: University Press, 1994.

[8] Farrell, M. J.: The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society* **120** (1957), 253-290.

[9] Gulati, R., and Kumar, S.: Impact of non-traditional activities on the efficiency of Indian banks: an empirical investigation. *Macroeconomics and Finance in Emerging Market Economies* **4** (2011), 125-166.

[10] Hančlová, J.: An evaluation of a production performance across the selected EU regions. A stochastic frontier approach to Malmquist productivity index'. *Proceedings of International Conference on control, Decision and Information Technologies CoDIT 2014*, Metz (2014), 170-175.

[11] Hančlová, J., and Chytilová, L.: The impact of inclusion of non-traditional activities on efficiencies in European banking industry using the CCR-I model. *Proceedings of the 11th International Conference on Strategic Management and its Support by Information Systems 2015* (2015).

[12] Lyroudi, K., and Angelidis, D.: Measuring Banking Productivity of the Most Recent European Union Member Countries; A Non-Parametric Approach'. *Journal of Economics and Business* **9** (2006), 37-57.

[13] Melecký, L., and Staníčková, M.: National Efficiency Evaluation of Visegrad Countries in Comparison with Austria and Germany by Selected DEA Models. *Proceedings of 30th International Conference on Mathematical Methods in Economics*, Karviná, (2012), 575-580.

[14] Stavárek, D.: Banking efficiency in the context of European integration. *Eastern European Economics* **44** (2006), 5-31.

[15] Řepková, I.: Efficiency of the Slovak Commercial Banks Applying the DEA Window Analysis. *Economics and Management Engineering* **8** (2014), 1320-1325.

[16] Sealey, C. W. Jr., and Lindley, J. T.: Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions, *Journal of Finance* **32** (1977), 1251-1266.

[17] Toloo, M., and Tichý, T.: Two alternative approaches for selecting performance measures in data envelopment analysis. *Measurement* **65** (2015), 29-40.

[18] Zhu, J.: Manual DEA Frontier DEA Add-In for Microsoft Excel. Retrieved 10.3.2015 from http://www.deafrontier.net/.

Efficiency measurement in multi-period production systems

Josef Jablonský[1]

Abstract. The paper aims at efficiency measurement in multi-period production systems. A common approach how to analyze efficiency in multi-period systems in Malmquist index, window analysis, and the Park and Park model (PP model) formulated in [6] that attempts to measure the aggregative efficiency within multiple periods. The disadvantage of this model consists in its orientation to the "best" period of the decision making unit (DMU) under evaluation, i.e. the aggregative efficiency is given as the "best" efficiency score across all periods. The paper formulates original modifications of the PP model – the model that is oriented on the "worse" period of the DMU under evaluation, and the model that computes average efficiency across all periods. In addition, we propose a multi-period SBM and super-efficiency SBM models that measure inefficiencies using relative slacks, i.e. negative relative deviations in the input space and positive ones in the output space. The results of all models are illustrated and compared on the three-period example taken from [6]. The differences among the results of presented models are discussed.

Keywords: data envelopment analysis, efficiency, multi-period model, SBM model

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluation of relative efficiency of decision making units (DMUs) described by multiple inputs and outputs. This approach introduced in [2] is based on solving linear programming problems for each of the DMUs under evaluation. Conventional DEA models analyse relative technical efficiency of the set of n DMUs that are characterized by m inputs and r outputs in one period. The efficiency score θ_q of the DMU_q is defined as the weighted sum of outputs divided by the weighted sum of inputs:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}, \tag{1}$$

where u_k, $k = 1, 2, \ldots, r$ is the positive weight of the k-th output, v_i, $i = 1, 2, \ldots, m$ is the positive weight of the i-th input, and x_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ and y_{kj}, $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, n$ are non-negative values for the DMU_j of the i-th input and the k-th output respectively. Conventional DEA models maximize the efficiency score (1) under the assumption that the efficiency scores of all other DMUs do not exceed 1. The computational form of the conventional DEA models depends on various assumptions about input or output orientation of the model, about returns to scales, weights restrictions, etc. More information about them can be found e.g. in [9].

One of the streams in theory and practice of DEA models deals with evaluation of efficiency in multi-period production systems. In this case the models estimate the total efficiency in the context of time serial data. There were proposed various models how to deal with time serial data in DEA models in the past. Malmquist index [3] and dynamic analysis [7] are only two of them that are relatively widely applied. Park and Park [6] have formulated a two-stage DEA model that evaluates aggregative efficiency in multi-period case (further referenced as PP model). The paper aims at multi-period DEA models, discusses disadvantages of PP model, and formulates its modifications. In addition, an original slack based measure (SBM) based efficiency and super-efficiency model for multi-period systems is proposed. The results of all models are compared on an example presented in [6], and their differences are discussed.

The paper is organized as follows. Section 2 presents PP model and formulates its modifications that partly overcome its drawbacks. In addition, Section 2 contains formulation of an original SBM based multi-period model. Section 3 compares results of all models presented in previous parts of the paper on the data set taken from [6].

[1] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067 Praha 3, Czech Republic, jablon@vse.cz.

The results of numerical experiments are analyzed and discussed. The final section of the paper summarizes presented results and discusses possible future research in this field.

2 Multi-period DEA models

Conventional DEA models can be formulated either in their primal or dual form - see e.g. [9]. In this section we will use the dual form that has several advantages with respect to formulation of multi-period models. Let us suppose that the DMUs are described by the same set of inputs and outputs in T consecutive time periods $t = 1, 2, \ldots, T$, and assume that x^t_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and y^t_{kj}, $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, n$ are the values of the i-th input and the k-th output in the t-th period of the DMUj. The first phase of the output oriented PP model with constant (variable) returns to scale assumption that evaluates the aggregative efficiency of the DMU$_q$ is formulated as follows:

$$
\begin{aligned}
&\text{Maximize} && \theta_q && && \\
&\text{subject to} && \sum_{j=1}^{n} x^t_{ij}\lambda^t_j \le x^t_{iq}, && i = 1,2,\ldots,m, t = 1,2,\ldots,T\ , && (2)\\
& && \sum_{j=1}^{n} y^t_{ij}\lambda^t_j \ge \theta_q y^t_{iq}, && k = 1,2,\ldots,r, t = 1,2,\ldots,T\ , && \\
& && \left(\sum_{j=1}^{n} \lambda^t_j = 1, \right. && \left. t = 1,2,\ldots,T\right), && \\
& && \lambda^t_j \ge 0, && j = 1,2,\ldots,n, t = 1,2,\ldots,T, &&
\end{aligned}
$$

where λ^t_j $j = 1,2,\ldots,n$, $t = 1, 2,\ldots, T$ are the variables of the model – coefficients of linear (convex) combination of the DMUs in time period t, and θ_q is the total aggregative efficiency score of the DMU$_q$. Let θ^*_q be the optimal objective function value of model (2) – the value of aggregative efficiency score $\theta^*_q > 1$ indicates that the DMU$_q$ is not efficient in any period, $\theta^*_q = 1$ means full or weak efficiency in at least one of the periods. The units with their aggregative efficiency score $\theta^*_q = 1$ cannot be ranked using the PP model. In single period systems, this problem is solved by using of super-efficiency models or any other models that allow ranking of efficient DMUs. Information about the most important models of this class can be found e.g. in [5].

First super-efficiency model was introduced by Andersen and Petersen in [1]. Their model is based on the idea to remove the unit under evaluation from the set of units and then measure its distance from the new efficient frontier. A similar approach as in Andersen and Petersen model can be used for multi-period models too. For these purposes model (2) can be extended by additional constraints that ensure zero values for all weights of the DMU under evaluation, i.e.

$$\lambda^t_q = 0, \qquad t = 1,2,\ldots,T, \tag{3}$$

It is clear that the units with maximum aggregative efficiency scores in Model (2) have lower objective function values than 1 (for output oriented model) in model (2) with additional constraints (3). This allows their ranking quite easily.

Main disadvantage of PP model (2) consists in its orientation on the "best" period of the DMU under evaluation, i.e. the efficiency score given by this model is always its best efficiency score over all particular periods considered. That is why an application of this model can lead to inappropriate conclusions, e.g. the DMU that is rated as highly efficient in one of the periods only and highly inefficient in all other periods will be ranked higher than the unit that is efficient in all periods. That is why simple modifications of the model (2) where the aggregative efficiency is based on the "worse" period of the unit under evaluation and on average efficiency score over all periods. The aggregative efficiency score based on the "worse" period can be given using optimization model (4):

$$
\begin{aligned}
&\text{Maximize} && \left(\sum_{t=1}^{T} \theta^t_q\right) - \varepsilon\psi_q\,, && && \\
&\text{subject to} && \sum_{j=1}^{n} x^t_{ij}\lambda^t_j \le x^t_{iq}, && i = 1,2,\ldots,m, t = 1,2,\ldots,T && (4)
\end{aligned}
$$

$$\sum_{j=1}^{n} y_{ij}^t \lambda_j^t \geq \theta_q^t y_{iq}^t, \qquad k = 1,2,...,r, t = 1,2,...,T$$

$$\theta_q^t \leq \psi_q, \qquad t = 1,2,...,T$$

$$\lambda_j^t \geq 0, \lambda_q^t = 0, \qquad j = 1,2,...,n, t = 1,2,...,T,$$

where θ^t_q is the efficiency score in the period t and ψ_q is the aggregative efficiency score. A formulation of the model that maximizes average efficiency score over all periods need not be presented due to its simplicity.

SBM models are quite popular group of models that allow evaluation of efficiency independently on the orientation of the model. Probably the most often applied and discussed is the Tone's model in its standard and super-efficiency versions – see [8] – but this model is not suitable for its modification for multi-period analysis. We propose for this purpose a standard SBM model that measures inefficiencies using relative slacks, i.e. negative relative deviations in the input space and positive ones in the output space. This model can be easily written for multi-period system as follows:

Minimize $$\upsilon = \sum_{t=1}^{T} \left(\sum_{i=1}^{m} \left[s_{1i}^{t-} / x_{iq}^t \right] / m + \sum_{k=1}^{r} \left[s_{2k}^{t+} / y_{kq}^t \right] / r \right) / T,$$

subject to

$$\sum_{j=1}^{n} x_{ij}^t \lambda_j^t + s_{1i}^{t-} = x_{iq}^t, \qquad i = 1,2,...,m, t = 1,2,...,T, \tag{5}$$

$$\sum_{j=1}^{n} y_{kj}^t \lambda_j^t - s_{2i}^{t+} = y_{kq}^t, \qquad k = 1,2,...,r, t = 1,2,...,T,$$

$$\lambda_j^t \geq 0, \qquad j = 1,2,...,n,$$

$$s_{1i}^{t-} \geq 0, s_{2k}^{t+} \geq 0, \qquad i = 1,2,...,m, k = 1,2,...,r, t = 1,2,...,T,$$

where s_{1i}^{t-}, s_{2k}^{t+} are negative and positive slacks in the period t of the i-th input and the k-th output respectively. The other symbols have the same meaning as in model (4). The objective function of model (5) consists of two parts. The first part measures average inefficiency in input space and the second one the average inefficiency in the output space. The total objective function value is the average of both inefficiencies over all periods. The optimal objective function of model (5) $\upsilon^* = 0$ means that the DMU_q is efficient in all periods, otherwise the DMU_q is inefficient at least in one period. The units can be easily ranked according to the average efficiency score υ^*. The units that are efficient in all periods can be ranked according to their average super-efficiency measure that can be given using a modified version of the model proposed in [4]. Here, the super-efficiency is measured using undesirable slacks, i.e. using positive deviations in the input space and negative ones in the output space. The model is formulated as follows:

Minimize $$\rho = 1 + \sum_{t=1}^{T} \left(\sum_{i=1}^{m} \left[s_{1i}^{t+} / x_{iq}^t \right] / m + \sum_{k=1}^{r} \left[s_{2k}^{t-} / y_{kq}^t \right] / r \right) / T,$$

subject to

$$\sum_{j=1}^{n} x_{ij}^t \lambda_j^t + s_{1i}^{t-} - s_{1i}^{t+} = x_{iq}^t, \quad i = 1,2,...,m, t = 1,2,...,T, \tag{6}$$

$$\sum_{j=1}^{n} y_{kj}^t \lambda_j^t + s_{2i}^{t-} - s_{2i}^{t+} = y_{kq}^t, \; k = 1,2,...,r, t = 1,2,...,T,$$

$$\lambda_j^t \geq 0, \; j = 1,2,...,n, \; \lambda_q^t = 0, t = 1,2,...,T,$$

$$s_{1i}^{t-} \geq 0, s_{1i}^{t+} \geq 0, s_{2k}^{t-} \geq 0, s_{2k}^{t+} \geq 0, \qquad i = 1,2,...,m, k = 1,2,...,r, t = 1,2,...,T,$$

where s_{1i}^{t+}, s_{2k}^{t-} are positive and negative slacks in the period t of the i-th input and the k-th output respectively and other symbols correspond to model (5). It is possible to prove that the DMUs inefficient in all periods have the efficiency score given by model (6) $\rho^* = 1$ and the units that are efficient in all periods $\rho^* > 1$. That is why the inefficient DMUs can be ranked according to the efficiency scores given by model (5) and the units identified as efficient by this model can be discriminated according to the results of model (6).

3 An example

The results of all models presented in the previous section of the paper are illustrated on an example taken from [6]. The example consists in efficiency evaluation of 20 cable TV service operations units (SO) in South Korea over three years period from 1999 until 2001. There are defined two inputs – operating cost in 100 $ and the number of employees – and two outputs – revenues in 100 $ and the total number of subscribers (viewers). The data set for all three years is presented in Table 1.

DMU	**Operating cost** (100 $)			**Manpower** (number)			**Revenue** (100 $)			**Viewers** (100 numbers)		
	1999	2000	2001	1999	2000	2001	1999	2000	2001	1999	2000	2001
SO1	454	530	560	26	28	30	397	486	580	186	360	396
SO2	358	350	390	28	30	28	239	268	312	105	160	176
SO3	416	450	482	40	40	40	462	524	620	244	445	480
SO4	245	292	318	25	28	32	202	230	286	116	208	226
SO5	258	300	325	28	30	30	288	328	405	149	270	292
SO6	480	524	598	50	54	62	510	580	612	311	545	565
SO7	475	495	554	27	28	30	382	420	504	198	346	368
SO8	370	404	425	29	30	35	297	340	399	161	269	275
SO9	421	480	530	35	35	36	415	466	529	201	325	335
SO10	627	668	728	44	44	44	491	544	624	264	442	469
SO11	400	462	508	22	24	25	334	386	454	170	320	352
SO12	404	476	515	37	40	40	321	348	400	168	300	325
SO13	458	528	580	26	30	31	426	468	520	210	368	389
SO14	541	640	685	34	35	36	528	580	696	320	526	570
SO15	281	328	360	25	30	30	221	264	305	127	252	264
SO16	934	1018	1102	59	64	63	817	926	1098	361	615	665
SO17	863	1030	1100	49	50	50	884	1020	1104	424	828	900
SO18	571	674	738	31	35	34	486	568	660	258	482	520
SO19	339	386	414	28	30	30	385	462	554	196	365	400
SO20	454	530	560	26	28	30	189	215	254	90	180	188

Table 1 Data set

The results of the analysis are given in Table 2. We decided to apply models with constant returns to scale assumption in order to compare given results with those presented in original article [6]. Table 2 contains in its particular columns the following information:

(1) Identification of the DMUs
(2) Efficiency scores given by CCR output oriented model for the 1999 data set. The value greater than 1 indicates inefficiency (higher value less efficient unit), the values lower than 1 are super-efficiency measures of efficient units (lower values indicate higher ranking). Super-efficiency scores are computed using Andersen and Petersen model [1]. Columns (3) and (4) contain the same information for the next two years, i.e. 2000 and 2001.
(5) Aggregative efficiency (super-efficiency) scores derived from PP model (2). The results in Table 2 show that these values are really maximum efficiency scores over all particular periods.
(6) Aggregative efficiency scores based on simple average of three particular efficiency (super-efficiency) scores.
(7) Aggregative efficiency scores based on the "worse" period of the unit under evaluation. They are given as a result of the model (4) - ψ_q values.
(8) SBM aggregative efficiency given using model (5). The values lower greater than 0 indicate inefficiency in at least one period, 0 means efficiency in all periods considered. As evident only two units, SO17 and SO19, are efficient in all periods. In order to discriminate between them the multi-period SBM super-efficiency model (6) can be applied – super efficiency measure for SO17 is 1.0755 and for SO19 1.0396, i.e. the unit SO17 is ranked higher according to this model.

DMU	CCR-O 1999	CCR-O 2000	CCR-O 2001	PP model	Avg. agr.eff.	Worse agr.eff.	SBM model
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SO1	1.174	1.113	1.066	1.066	1.118	1.174	0.175
SO2	1.673	1.563	1.662	1.563	1.633	1.673	0.866
SO3	1.009	0.996	0.970	0.970	0.992	1.009	0.024
SO4	1.319	1.424	1.401	1.319	1.381	1.424	0.497
SO5	1.012	1.078	1.074	1.012	1.054	1.078	0.166
SO6	0.913	0.951	1.054	0.913	0.973	1.054	0.105
SO7	1.215	1.206	1.222	1.206	1.214	1.222	0.243
SO8	1.350	1.393	1.425	1.350	1.389	1.425	0.433
SO9	1.152	1.198	1.284	1.152	1.211	1.284	0.300
SO10	1.346	1.342	1.387	1.342	1.358	1.387	0.375
SO11	1.145	1.184	1.175	1.145	1.168	1.184	0.194
SO12	1.416	1.539	1.554	1.416	1.503	1.554	0.583
SO13	1.083	1.183	1.230	1.083	1.165	1.230	0.188
SO14	0.865	1.014	1.030	0.865	0.970	1.030	0.045
SO15	1.359	1.299	1.358	1.299	1.339	1.359	0.456
SO16	1.203	1.200	1.153	1.153	1.185	1.203	0.319
SO17	0.908	0.834	0.876	0.834	0.873	0.908	0.000
SO18	1.080	1.147	1.131	1.080	1.119	1.147	0.153
SO19	0.952	0.933	0.902	0.902	0.929	0.952	0.000
SO20	1.174	1.113	1.066	1.725	1.748	1.174	0.919

Table 2 Results (models with CRS assumptions)

The results presented in Table 2 show a very close similarity in rankings of DMUs given by all multi-period models that were applied. It is not surprising conclusion because all the DMUs have very close CCR efficiency scores in all three considered periods (see columns (2) to (4) in Table 2). One can easily imagine that it is possible to formulate an example with DMUs highly different in efficiencies in particular periods. In this case the results given by PP model that is oriented on the "best" period on one hand and the model oriented on the "worse" period and the SBM model on the other hand may be conflicting.

4 Conclusions

Analysis of efficiency of DMUs within multiple periods is an important task and many various models were proposed for these purposes in the past. The paper presents the PP model (2), which is one of the latest contributions in the area of multi-period DEA models. Except the conventional PP model several its modifications are formulated. The main motivation for formulation of these models consists in disadvantages of the PP model that were discussed mainly in Section 2 of the paper. A super-efficiency PP model and models that are, in the contrary to the PP model, based on the efficiency in the "worse" period and average efficiency over all periods of the unit under evaluation were formulated and tested on the real data set. In addition to the PP model and its modifications an original SBM and super-SBM multi-period model has been proposed as an alternative approach to conventional PP models.

The advantage of the proposed super-efficiency multi-period DEA model consists in its computational efficiency because it is necessary to solve only one optimization problem for each DMU in order to get its aggregative (super-) efficiency score. The disadvantage is given by the fact that the final results depend on the efficiencies of the DMUs in particular periods without any interconnections among periods. A future research in this field can be focused on analysis of multi-period production systems with interconnections among the inputs and/or outputs in particular periods, and in analysis of multi-period systems under the assumption of uncertainty in data set in all or at least some periods.

Acknowledgements

The research is supported by the Grant Agency of the Czech Republic, project no. P403/12/1387, and Internal Grant Agency of the University of Economics, Prague, project no. F4/62/2015.

References

[1] Andersen, P. and Petersen, N.C.: A procedure for ranking efficient units in data envelopment analysis. *Management Science*, **39** (1993), 1261-1264.

[2] Charnes, A., Cooper, W.W. and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2** (1978), 429–444.

[3] Fare, R. and Grosskopf, S.: *Intertemporal production frontiers with dynamic DEA*. Boston: Kluwer Academic Publishers, 1996.

[4] Jablonsky, J.: Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling*, **46** (2007), 1091–1098.

[5] Jablonsky, J.: Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal of Operations Research*, **20** (2012), 435-449.

[6] Park, K.S. and Park, K.: Measurement of multiperiod aggregative efficiency. *European Journal of Operational Research*, **193** (2009), 567-580.

[7] Sueyoshi, T. and Sekitani, K.: Returns to scale in dynamic DEA. *European Journal of Operational Research*, **161** (2005), 536-544.

[8] Tone, K.: A slack-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, **143** (2002), 32-41.

[9] Zhu, J.: *Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets and DEA Excel Solver*. Boston: Kluwer Academic Publishers, 2003.

Integration of Soft Factors in the Value-added Process-related Performance Analysis

Hendrik Jähn[1]

Abstract. Soft factors play an important role for the management and success within the economic sector. Networked production structures represent an important branch within most economies. In that context an innovative method for the inclusion of soft factors into a quantitative management procedure is introduced. Here the main focus lies on the analysis of the performance of an enterprise operating in a production network. That analysis is based both on hard and soft factors what means that the data of soft factors need to be quantified. This is realised by the application of the Repertory Grid-methodology. Additionally the value benefit analysis is used for the performance analysis in order to quantify an aggregated performance measure.

Keywords: Production Network, Soft factors, Performance Analysis.

JEL classification: M14
AMS classification: 91E45

1 Motivation

Successful networked production structures require efficient management structures. In that context among others both incentive and sanction mechanisms are applicable. The operationalisation of sanction mechanisms is primarily based on control and supervision systems. That fact is founded on the assumptions of the New Institutional Economics [9], e.g. asymmetrically distributed information among the actors, individual maximisation of utility, opportunistic behavior and bounded rationality of the decision makers. In order to support management structures a comprehensive approach for the value-added process-related measurement, assessment and evaluation of the service performance of an enterprise has been developed [5]. That process optimization approach is simply called "Performance Analysis Approach" (PAA). For the analysis both hard and soft factors are included.

In the following the basics of this PAA procedure are described in brief. That includes the modelling of a comprehensive algorithm. Here mainly a methodology for quantifying relevant soft factors and their evaluation are explained in a more detailed manner. In that context an example for considering the soft factor "cooperation quality" is given. In analogy the introduced procedure is also applicable for further soft factors such as confidence or enterprise culture.

2 Fundamentals and framework

In the following section some important fundamentals concerning the treatment of soft factors in a quantitative analysis are given. Subsequently the framework of the PAA and its procedure are introduced. The main pillar of the approach is formed by a combination of the Repertory Grid-methodology and the value benefit analysis.

[1]Chemnitz University of Technology, 09107 Chemnitz, Germany, hendrik.jaehn@wirtschaft.tu-chemnitz.de

2.1 Operationalisation

Every soft factor-specific performance parameter is represented by a key figure. The main challenge is that the characteristics and attributes of soft factors are not available in a quantitative form because they describe qualitative factors. They are normally formed by linguistic variables which are represented by linguistic expressions such as "high", "low" or "medium" or alternatively "good" and "bad". Hence, it is necessary to find a way to transfer their qualitative forming into a quantitative value.

Approaches for the consideration of soft factors can be found in different contexts in technical literature. Some useful approaches have been identified, see [8] and [1], which, however, cannot be used without modification, as their evaluation results are unusable in the present case. One further promising approach for the quantification of qualitative information is the Repertory Grid methodology [3]. There is a simple proceeding, linked with effective evaluation methods and a high level of acceptance. Additionally, a method is required, which helps to edit the data collected by the Repertory Grid methodology for the purpose of evolution. At this stage the value benefit analysis [10] has been chosen. It allows the determination of a weighted aggregated value, according to the information content of the different aspects to be considered.

2.2 Procedure of the analysis

As already mentioned the procedure of the analysis includes both the Repertory Grid-methodology and the value benefit analysis. At first, a Repertory Grid questionnaire, including all elements and constructs, which are relevant for the analysis, needs to be designed. Therefore elements and constructs of the Grid must be identified. Then, in order to guarantee a proper value benefit analysis the independence of the constructs must be verified. Additionally weightings need to be calculated. Subsequently the evaluation of the Grid can start after completing the data collection. The analysis will finally be realised within the scope of the value benefit analysis by converting the value benefit into an evaluation function for the performance parameter "cooperation climate". In Figure 1 the complete procedure is illustrated.

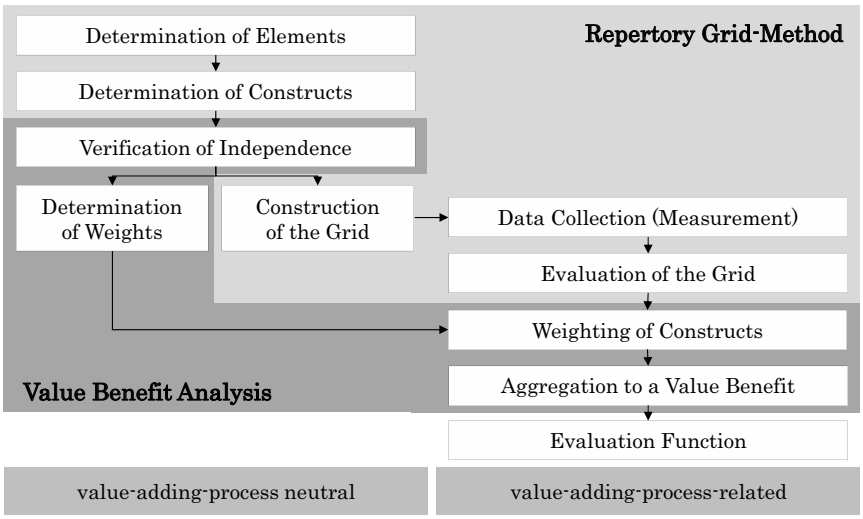


Figure 1 Procedure of soft factor integration

It is obvious that both applied methods appear in the value-added process-neutral and in the value-added process-related stream. That means some tasks need to be done before commencing the analysis (process-neutral) and some tasks are process-related. Some important contents of the single steps of the approach are explained consecutively.

2.3 Repertory Grid-methodology

The Repertory Grid-methodology is based on the Role Construct Theory which has been developed by Kelly [6]. The purpose is primarily to involve the subjective construction of reality with the help of the experiences of one or more persons. In the opinion of Kelly using that methodology allows the detection of the construct system of an individual. The essential components include elements and constructs. A collection of valuations (repertory, repertoire) is described by elements. They represent uniquely assignable events for the respondents and are used to transfer the meanings as for example roles, groups or situations of people. Here, the elements are characterised by the enterprises in which the company being valued is connected.

Constructs are not directly observable characteristics or beliefs referred to the elements and are shown in a Repertory Grid as polarised opinions. The use of constructs enables the structuring of the environment of individuals. A construct enables the realisation of behavior. The term grid defines the use of a data matrix in which the distinctive features of the properties are listed. In a subsequent step they are valued. The content of the Repertory Grid-methodology is for example determined by the use of qualitative questionnaires and the inclusion of quantitative data by means of the review by the numbers in the matrix. The most commonly used type of scale is the multi-level rating scale, where the rating scale here includes an assessment of -3 to +3. The individual occurrences are the following: (3): very pronounced, (2): moderately pronounced, (1): less pronounced and 0 neutral whereas values from 3 to 1 belong to the property pole and values from -1 to -3 belong to the contrary pole. The assessment of the resulting Repertory Grid can be done by cluster analysis or principal component analysis consecutively. The Repertory Grid-methodology offers the benefits of a systematic detection of a set of constructs.

2.4 Value Benefit Analysis

To perform the procedure correctly it is necessary to perform both differential and preferential independence test among the constructs. Detailed information can be found in the literature on decision theory, see [2]. In essence, only constructs may be included which are independent with regard to their meaning and content. That test secures that there are no redundancies.

By means of value-benefit analysis the various performance levels of the key figures of the relevant performance parameters (hard and soft factors) are aggregated to a total value. Therefore, it is necessary to establish a relationship between performance level and score for each performance parameter. This is possible best by modelling an evaluation function. An evaluation function can be modelled using a sufficiently large number of nodes by means of Lagrange interpolation. The performance levels are determined by quantification methods, e.g. the Repertory Grid-methodology. The appropriate evaluation function must be formed in advance. Subsequently, in the implementation of performance analysis easily the evaluation score can be determined. As a result ratings of all performance parameters are available. It is important to specify a range of values (interval) for the scores. Common is a scale from 0 to 10 credits in order to achieve a reasonable differentiation.

In the evaluation of soft factors, in many cases a soft factor consists of several sub-facts. In that case the performance levels of the sub-facts are determined according to the same pattern as described. Subsequently an aggregation of these values to a specific degree of fulfilment of the corresponding soft fact is required. In case of differing contents weightings can be introduced. For determination of weightings a large number of methods is available, cf. [2]. By the multiplication of weightings with the scores and the subsequent summing the products the aggregated score of one soft factor is determined.

The same procedure then is determined for the aggregation of the evaluation scores of all performance parameters. Again, specific weightings must be determined for the performance parameters according to their importance. Finally the aggregated total performance level of an enterprise as the most important evaluation result is available.

3 Exemplary implementation of the approach

Subsequently based on the soft factor "cooperation" the procedure of soft factor-integration is described. That includes the introduction of variables and some calculation schemes. Cooperation is one important feature of networked production structures. The success of a cooperation depends on the quality of a cooperation to a high degree.

3.1 Construction of a Grid and collection of data

The soft factor "cooperation" with its performance parameter "cooperation quality" is represented by the key figure "cooperation climate". Elements in a networked structure are all participating enterprises. In addition a network-coach can be considered, if relevant. For a comparison of values the consideration of an ideal actor is necessary. In a Repertory Grid elements are represented by columns. Constructs are represented by rows and include all relevant properties for the analysis. The relevant constructs for the soft factor cooperation are identified by the help of literature, e.g. [7]. After the verification of independence (reciprocal preference independence and difference independence) and the determination of weightings by an appropriate method according to [2] the construction of the Grid can commence. In case of a high number of constructs several construct categories can be formed. These categories are analysed separately. The results are aggregated in the next step. Weightings can be used again, if necessary.

Attribut Pole	Ideal Position	Enterprise 1	Enterprise 2	Enterprise 3	Enterprise 4	Netzwerk-Coach	Attribut Antipole
sensitive for problems							superficial
talkative							uncommunicative
consultative							denies consultation
selfless							selfish
open to compromises							focused on asserting
communicates success							reserved
convincing							repressive
solves conflicts							creates conflicts
oriented on cooperation							individualistic
adheres to rules							expands rules
able to accept criticism							unable to accept criticism
aspires harmony							rebellious
predictable							moody

Figure 2 Repertory Grid for cooperation quality

Figure 2 shows a typical grid for the soft factor cooperation. It consists of four construct categories with three to four constructs and six elements. In order to keep the effort of the analysis to a reasonable extend is recommended to keep the number of constructs relatively low.

The collection of data is not possible automatically for soft factors. Therefore a manual collection is required. After completion of a value added process the involved enterprises are supplied online with the Grid. It must be completed within a proper time slot by representatives of the enterprises. They will complete the grid with the evaluation figures for the contacted firms and for all constructs. Subsequently the results are forwarded to the evaluation instance.

3.2 Evaluation of a Grid

In this section an approach for the analysis and evaluation of questionnaires based on the Repertory Grid-methodology is described. That approach was specifically developed for considering soft factors in the performance analysis.

Evaluations are always interpreted depending on the desired, i.e. the ideal condition (construct "Ideal"). Hence first a relationship between the ratings a construct m of the element "Ideal" and the ratings of a construct m of a certain element (enterprise) needs to be established. Here it is assumed that this relationship is determined by the deviation of the actual state from the desired state. In this case it is the difference between the realised state (of an enterprise) and the ideal state.

For this purpose the actual rating e_i^m is subtracted from the ideal rating $e^{m,ideal}$. The absolute value of the result Δe_i^m represents the relevant deviation and serves as calculation and valuation basis. The following formula applies:

$$\Delta e_i^m = |e^{m,ideal} - e_i^m| \tag{1}$$

One problem occurs because the maximum possible deviations $e_i^{m,max}$ are not the same for all ideal ratings. According to evaluation metrics values from -3 to 3 are appiclable, a maximum deviation of 6 can result therefore. However, this is only possible in case the ideal rating $e^{m,ideal}$ is -3 and the actual rating results in 3 or vice versa. However was as ideal rating $e^{m,ideal}$ of 0 selected only a maximum deviation e_i^m of 3 is possible as a result of a rating of 3 or -3. Out of this situation, problems arise when aggregating because a priori no aggregate maximum deviation $E_i^{l,max}$ exists. However that value is required for a comparison of the actual and the target value for reasons of standardisation. An even more problematic situation occurs with an ideal value $e^{m,ideal}$ of 1 , 2 , -1 or -2. In that case there are positive and negative variations with varying frequency. Thus for example with an ideal value of 2 a deviation of 1 is possible twice (at a rating e_i^m by 3 or 1), a deviation of 0, 2, 3, 4, 5 but only once (at a rating e_i^m by 2, 0, -1, -2 and -3). In order to achieve simplification, it is assumed that only ideal values $e^{m,ideal}$ of 3, 0 and -3 are allowed, whereby the maximum deviation either is 6 or 3. Based on this idea, an aggregate maximum deviation $E_i^{l,max}$ is determined uniquely.

After determining the enterprise specific deviation of rating for each construct e_i^m the enterprise specific deviation for each construct category is to calculate. This is an aggregated measure. For this, the deviations of all to a construct category belonging constructs are aggregated. The values of the constructs Δe_i^m here can be both in weighted or unweighted form considered. Weightings have to be set in advance by an appropriate method. As a intermediate result, per Grid exists one value representing the deviation of ratings of an enterprise i with respect to the ideal value of a construct category l. That value is designated Δe_i^l.

By using that procedure procedure, for every enterprise i and every construct category l several values (from several grids) are available since according to their participation enterprises are multiple rated. These ratings come from various enterprises which participate in a cooperation. These values are to be aggregated subsequently. The easiest way is to calculate the average value of all deviations. However, a modification may consider that before aggregation of Δe_i^l outliers are eliminated in order to detect unrealistic reviews. These are not to be included in the evaluation finally. As a result, an aggregate value ΔE_i^l is calculated representing the deviation with respect to an enterprise i and a construct category l. It is obvious that a rating is better if is closer to the value 0.

In order to achieve a standardised classification of the aggregated deviation of a particular construct category and a particular enterprise the aggregated maximum deviation $\Delta E_i^{l,max}$ is to be caluculated. For this purpose, first for all constructs the maximum deviations $\Delta e_i^{m,max}$ need to be identified. With ideal values of $e^{m,ideal} = 3$ or -3 in both cases a maximum devation of 6 is relevant and with an ideal value of $e^{m,ideal} = 0$ as the maximum deviation only 3 is possible. Subsequently, these values are aggregated based on construct categories in analogy to Δe_i^l. There result the aggregated values $\Delta E_i^{l,max}$.

In order to allow a further aggregation the degree of fulfillment of the performance parameter quality cooperation is to determine. For that purpose, target and actual values for the deviation in the constructs and construct categories must be brought into a relationship. The actual values are represented by ΔE_i^l and the target value $\Delta E_i^{l,ideal}$ is 0 (no deviation). In order to quantify the deviation of the value for the maximum deviation $\Delta E_i^{l,max}$ is required. ΔE_i^l is between 0 and $\Delta E_i^{l,max}$ ("Target spectrum"). The

relation between ΔE_i^l and the target spectrum can be represented in various ways, for example as a linear relationship in the form of a function $f(\Delta E_i^l)$. As the result, there is a construct-specific or a construct category-specific scoring f_i^l, which is designated as characteristic value of a construct category.

The final aggregation to an enterprise-specific utility value can now take place. That value represents the performance of the enterprise considering a soft factor within a range of 0 to 10 credits. A further aggregation to an overall performance figure of an enterprise a possible by adding the utility values of all considered performance parameters. The application of weightings in that context is also possible.

4 Conclusions

Considering soft factors in a performance analysis of enterprises is essential. By using the performance parameter "cooperation quality", the procedure of soft factor-integration was introduced. It allows a quantitative collection, evaluation and analysis of the collected data. The result in the form of an aggregated measured value can be included in an integrated concept of the performance analysis for networked production structures [4] within the framewerk of the PAA-approach. Within that context in is possible to perform a comprehensive analysis including hard and soft factors. That results in a valuable extension of the operative controlling and management of production networks and participating enterprises.

References

[1] Burt, R.S.: *Structural Holes - The Social Structure of Competition*. Cambridge University Press, New York, 1992.

[2] Eisenführ, F., Weber, M., and Langer, T.: *Rational Decision Making*. Springer, Berlin, Heidelberg, 2010.

[3] Fransella, F., Bell, R., and Bannister, D.: *A Manual for Repertory Grid Technique*. 2nd ed., Wiley & Sons, Chichtester, 2010.

[4] Jähn, H., Fischer, M., and Burghardt, T.: Options for the Performance Analysis and Profit Distribution in networked Organizations based on industrial Enterprises. In: *Proceedings of the 10th International Conference on The Modern Information Technology in the Innovation Processes of the Industrial Enterprises (MITIP 2008)* (J. Hán, and P. Holejšovská, eds.), Prague, 2008, 268–273.

[5] Jähn, H.: Value-added process-related Performance Analysis of Enterprises acting in cooperative Production Structures, *Production Planning & Control* **20** (2009), 178–190.

[6] Kelly, G.A.: *The Psychology of Personal Constructs*. Vol. I, II. 2nd ed., Routhledge, New York, 1991.

[7] Schneider, H., and Knebel, H.: *Team und Teambeurteilung: Neue Trends in der Arbeitsorganisation*. Wirtschaftsverlag Bachem, Cologne, 1995.

[8] Scott, J.: *Social Network Analysis*. 3rd ed., SAGE, London, 2013.

[9] Williamson, O.E.: *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York, 1975.

[10] Zangemeister, C.: *Nutzwertanalyse in der Systemtechnik*. 4th ed., Wittemannsche Buchhandlung, Munich, 1976.

DETERMINANTS OF THE SHADOW ECONOMY IN THE EU COUNTRIES: EMPIRICAL VERIFICATION

Ondřej Jajkowicz[1]

Abstract. The main aim of this article is empirical verification on the influence, based on theory selected determinants of the shadow economy on this economic phenomenon size in the selected countries of the European Union (EU). Based on the research of the empirical literature a panel regression model for a selected group of countries is compiled and estimated. Due to the selection of a larger sample of countries examined in the empirical part of this article a panel regression method was applied to ensure exploration of the relationship between selected determinants of the shadow economy and the size of the shadow economy in the EU. Before equation estimation each time series were individually tested for the existence of a unit root in order to avoid the so-called spurious regression problem.

Keywords: Shadow Economy, Determinants of the Shadow Economy, Tax Burden, Regulatory Quality, Panel Regression

JEL Classification: E26, O17, K42
AMS Classification: 62J05, 62M10

1 Introduction

The main aim of this paper is to empirically verify the effect, based on theory selected determinants of the shadow economy on the shadow economy size in selected EU countries (Spain, Portugal, Greece and Italy) since 2000 to 2013. As determinants are meant determining parameters, factors, or factors that are critical for desirable or undesirable activity (moving activities into the shadow economy) of the man. Between the main determinants (in terms of causes) of the shadow economy belongs high tax burden of businesses, high levels of compulsory paid social contributions and excessive regulation of the formal economy (especially labor market) from the side of the state. The labor market is then connected to a large portion of the estimated size of the shadow economy called „Illicit work". Despite the importance of the shadow economy has not yet been deeper analysis of this phenomenon in these countries performed.

2 Determinants of shadow economy

According to Schneider (2011) for the shadow economy can be considered an economic activity and incomes, whose aim is to avoid government regulation, taxation or any capture. Already from this definition it is clear that the determinants that contribute to the existence or increase of the size of the shadow economy involves excessive tax burden on businesses and high level of regulation of the formal economy especially from the side of state. Besides these two major causes of existence and growth of the shadow economy the other equally important determinants can be included. With examination of specific determinants of the shadow economy in his article dealed Marinov (2008), which states that social and economic reasons which force businesses to move their activities into the shadow economy, are affected by government policy in the area of tax and regulatory measures. Based on the research of the existing literature it can be stated that the growth in the size of the shadow economy is caused by many different factors. Schneider (2011, 2012) among the most important determinants ranks the growth of the tax burden and the level of paid social security contributions, poor quality of government, a high level of regulation of the formal economy (especially the labor market), the complexity of the tax and legal system, reduction of weekly working hours, early retirement, drop the occasional honor and loyalty towards public institutions combined with a declining tax morale. The same causes of the existence of the shadow economy also included Schneider and Enste (2000). According to studies of Schneider (2006) or Startienė and Trimonis (2010), the tax burden has the strongest influence on the existence and growth of the size of the shadow economy.

3 Data and methods

The database for the estimation of the empirical model is composed of annual time series of selected variables from 2000 to 2013 for a group of countries surveyed. The response variable is the size of the shadow economy as a share of GDP. Time series of the size of the shadow economy are drawn from the works of Friedrich Schneider (2011,2013), a leading world expert in the field of exploration and estimates of the size of the shadow economy. Explanatory variables are based on conducted research of empirical literature. As determinants of the

[1] Vysoká škola báňská – Technická univerzita Ostrava, Ekonomická fakulta, ondrej.jajkowicz@vsb.cz

shadow economy representatives of the tax burden level of regulation and quality control were included. Furthermore, between the explanatory variables two control variables were incorporated. Specifically, the following explanatory variables were used:

• Tax burden:
  - The share of direct taxes in GDP (DT)
  - The share of indirect taxes in GDP (IT)
  - The share of social contributions in GDP (SC).
• The degree of regulation:
  - The share of government expenditure in GDP (GE).
• Quality of regulation:
  - Selected indicators of quality of governance so called "Worldwide Governance Indicators" (control of corruption (CC) and quality of regulation (QR)).
• Control variables:
  - Unemployment rate in % (U)
  - Urbanization rate, the share of urban population in total population in % (URB).

Tax burden is the ratio of direct and indirect taxes to GDP (%) and the share of compulsory paid social security contributions to GDP (%). Data on the share of direct taxes, indirect taxes and social contributions to GDP were drawn from the European Commission, especially from taxation trends in the EU time series (European Commission, 2014). The share of direct and indirect taxes to GDP represent suitable indicators to assess the tax burden on businesses. The ratio of social contributions to GDP was chosen for the reason that a large part of the underground economy consists of so-called. "illicit work". The main motive for "illicit work" offers the high mandatory levied social security contributions. Degree of regulation is the share of government expenditures to GDP (%). Shares of government expenditures to GDP for each country from Eurostat government statistics were drawn (Eurostat, 2014). The regulation is any government action that aims to influence the work of the private sector. As a regulation can therefore be all the economic policies of the government and all legislation. Because the capture the rate of regulation degree is very complicated, the share of government expenditures to GDP will be used for the aforementioned indicator. This method of capture rate regulation gives an overview of state involvement in the economy, but on the other hand, unfortunately, does not capture all regulatory measures. As proxy for the quality of regulation, indicator of regulatory quality [2]and control of corruption [3]indicator were chosen. Data were drawn from a database of World Bank indicators of the quality of public administration (ie. Worldwide Governance Indicators)(The World Bank, 2014). Among the control variables were included, the unemployment rate (in%) obtained from Eurostat labor market statistics (Eurostat, 2014) and the urbanization rate statistics from the World Bank, measrued as a percentage of urban population to total population (The World Bank, 2014). For the verification of the effect of the selected determinants of the shadow economy and the magnitude of this economic phenomenon in the selected countries panell regression method is applied. Exploration of panel data uses model way to solve selected problem, in which elements of time series analysis and elements of regression analysis appears. Panel analysis therefore essentially represents the next stage of modeling, which greatly appreciates usually hard-won information of a fact (Novák, 2007). The main advantages of using panel data is to obtain a large amount of observation and the ability to better analysis of hidden unobserved or unobservable random factors in the econometric structure of relations between units. The potential problem of heterogeneity, when the individual effects and time effects are not unimportant, addresses (based on Hausman test performance) the fixed effects model. Variable ν_i presents these unobserved fixed cross-section effects. For the purpose of empirical verification of the influence of selected determinants of the size of the shadow economy on this economic phenomenon, the estimated regression model of panel data in the following form was used:

$$SE_{it} = \alpha_1 + \alpha_2 QR_{it} + \alpha_3 CC_{it} + \alpha_4 U_{it} + \alpha_5 GE_{it} + \alpha_6 DT_{it} + \alpha_7 DT_{it-1} + \alpha_8 DT_{it-2} + \alpha_9 IT_{it} + \alpha_{10} IT_{it-1} + \alpha_{11} IT_{it-2} + \alpha_{12} SC_{it} + \alpha_{13} SC_{it-1} + \alpha_{14} SC_{it-2} + \alpha_{15} URB_{it} + \nu_i + \varepsilon_{it}$$

Changes in the variables representing the tax burden on businesses will not reflect in their behavior immediately. We can assume that in the real economy reactions of economic agents to change these variables

[2] An indicator of quality control that handles the World Bank, ranges from -2.5 (poor quality) to 2.5 (high quality) and includes the ability of the government to sensitize and implement appropriate policies and regulations to help develop the private sector.

[3] Control of corruption indicator, which also handles World Bank, captures perceptions of the extent to which public power is exercised for private gain. They take into account both small and large forms of corruption, as well as "capture" of the state by elites and private interests. Control of corruption indicator, ranges from -2.5 (perceived level of corruption is high) to 2.5 (perceived level of corruption is low).

comes with a lag. In order to know how late responding businesses to increase in the tax burden by moving their activities into the shadow economy, these variables would be delayed by several years. Another reason for the delay of these variables is the existence of a standard delay resulting from the formation of economic policy (for example, in approving the introduction of new taxes, changes in tax rates, etc.). Before equation estimation the individual test for each time series for the existence of a unit root is required. For this purpose for panel unit roots tests, e.g. Levin, Lin and Chu (2002); Im, Pesaran and Shin (2003), or ADF test and PP test according Maddalo and Wu (1999) were used. Unit root tests pointed to the non-stationarity of time series. Inserting these time series into the regression model could lead to the estimate so called „spurious regression". For this reason it has been necessary adjust the time series to stationary. The adjustment of time series in this case is made by conversion to the logarithm and the first differences:

$$DLOGSE_{it} = \alpha_1 + \alpha_2 DLOGQR_{it} + \alpha_3 DLOGCC_{it} + \alpha_4 DLOGU_{it} + \alpha_5 DLOGGE_{it} + \alpha_6 DLOGDT_{it} + \alpha_7 DLOGDT_{it-1} + \alpha_8 DLOGDT_{it-2} + \alpha_9 LOGIT_{it} + \alpha_{10} LOGIT_{it-1} + \alpha_{11} LOGIT_{it-2} + \alpha_{12} DLOGSC_{it} + \alpha_{13} DLOGSC_{it-1} + \alpha_{14} DLOGSC_{it-2} + \alpha_{15} DLOGURB_{it} + \nu_i + \varepsilon_{it}$$

To estimate the panel data model the data were transformed into the so-called "pooled panel data model". Among one of the advantages of this modification of the data is the fact that this method of input time series ignores the fact that we are working with data for various countries (Lee, 2002). The model was estimated through econometric program E-views version 7. Problem of autocorrelation and heteroskedasticity that are at the panel data fairly common, were resolved by model estimation using the "white period" estimation technique. In this case, the results of the standard deviations of parameter estimation and hypothesis tests are correct with respect to autocorrelation. This method ensures that the t-statistic and standard error are plausible, because they are corrected for the heteroscedasticity (Kotlánová, Kotlán, 2012). It is expected a positive relationship for the proportion of direct taxes, indirect taxes and social security contributions (% of GDP) and the size of the shadow economy (% of GDP). Growth of these variables (ie increases in the tax burden on businesses and the amount of social security contributions) will cause ceteris paribus the increase in size of the shadow economy. Furthermore, the positive impact of urbanization rate and the unemployment rate to the size of the shadow economy is expected. It is assumed that the growth rate of urbanization reinforces anonymity and therefore contributes to the development of the activities of the shadow economy. In the case of unemployment, people are replacing their official receipts by moving into the shadow economy activities which they lost as a result of job loss. In conclusion, the positive effect on the size of the shadow economy is also expected for indicators of degree of regulation, respectively the share of government spending to GDP. So if share of government spending in GDP is growing (growing degree of regulation of the formal economy) it leads to increase in the size of the shadow economy. Selected indicators of the quality of regulation can also be considered as indicators of the quality of the institutional environment. For these indicators is expected that they negatively affect the existence and size of the shadow economy. Assuming that if there is improvement in the indicator of control of corruption and regulatory quality (improvement in terms of positive growth of indicators towards the maximum of 2.5), the size of the shadow economy will be reduced.

4 Results

Estimated model results showing determinants effect on the size of the shadow economy in selected countries are presented in Table 1. The results indicate a fairly close relationship (with respect to the use of panel data), of the size of the shadow economy and selected determinants, which is represented by the coefficient of determination (R2 = 0.52). Almost 52% of the variation in the size of the shadow economy can be explained by the explanatory variables used. The findings don't confirm all theoretical predictions about influence of the selected determinants of the shadow economy on the size of the shadow economy in selected countries. Statistically significant is variable expressing the tax burden, concretely share of indirect taxes to GDP. So growing share of indirect taxes in GDP (IT) leads to increase of the tax burden, and ultimately to increase the size of the shadow economy (SE). Businesses are trying to avoid taxation and shifting part of their activities into the shadow economy. Another important variable determining the size of the shadow economy due to the results are the social security contributions (SP). Here it is necessary to realize again that the big part of the size of the shadow economy is made up of so-called. "illicit work". And in countries with high compulsory social contributions, "illicit work" can be considered as a major problem. In connection with this problem, we can mention for example the case of the Italy, where the amount of social security contributions and other employers paid cost in the long term moves above the EU average. If these contributions levied too high, employers lose interest to officially recruit employees. To avoid such payments employers employ workers "illegaly". The only exception in the context of fiscal variables is a variable representing the share of direct taxes in GDP (DT), which is in relation to the size of the shadow economy statistically insignificant. It was also found that the increase in unemployment (U) contributes to the growth of the size of the shadow economy. People who lose their jobs in the formal economy compensate their lost revenue by activities in the shadow economy. A statistically insignificant relationship

was demonstrated in the case of degree of regulation represented by a share of government expenditure in GDP (GE). Without the effect on the size of the shadow economy, according to the findings are variables representing the quality of regulation (QR), rate of urbanization (URB) and control of corruption (CC). However we can assume that at a certain time delay of these variables we would achieve a statistically significant result. This assumption must be taken into account for further research of this topic.

Independent Variable	**DLOG(SE)**
C	(0.816725) **0.818825**
DLOG(QR)	(-0.243247) **-0.021630**
DLOG(CC)	(-0.912040) **-0.010483**
DLOG(U)	(2.140574) **0.054370****
DLOG(GE)	(0.852732) **0.087196**
DLOG(DT)	(0.344802) **0.023882**
DLOG(DT)(-1)	(-0.751723) **-0.109254**
DLOG(DT)(-2)	(0.119742) **0.008228**
LOGIT	(-0.919034) **-0.112950**
LOGIT(-1)	(1.675829) **0.105049***
LOGIT(-2)	(1.166217) **0.127612**
DLOGSC	(1.180857) **0.245814****
DLOGSC(-1)	(1.198364) **0.191740****
DLOGSC(-2)	(0.364580) **0.037673**
DLOGURB	(-1.058121) **-0.217804**
Total observations	44
R^2	0.523886
Durbin-Watson stat.	1.959860
F-statistic	1.682868
Prob (F-statistic)	0.000112
The statistical significance on 1% (***), 5% (**) and 10% (*) level of significance	

Table 1

Source: Own estimate through the program E-views (version 7).
Note: In parentheses are the corresponding t-statistics.

5 Conclusion

Empirical research was conducted on a sample of four European countries (Portugal, Spain, Italy and Greece) since 2000 to 2013 using panel regression method. Findings partly confirm the theoretical predictions and provide new results in this little-studied issue. Specifically, the results confirm a statistically significant positive effect of indirect taxes and compulsory social security contributions (the tax burden), unemployment rate on the size of the shadow economy. Unemployment can be considered as the main cause of the current size of the shadow economy in the studied group of countries. The high unemployment rate especially in Spain and Greece led to the development of "illicit work", which forms a crucial part of the overall size of the shadow economy. Government should also focus on increasing the attractiveness of work in the official sector, for example by reducing non-wage costs. Besides improving the quality of regulation, tax system reform (simplification), reducing excessive tax burden, reduce the social security contributions, it is important to promote individual economic freedom for all actors in the economy, a functioning legal system (strict penalties for violations of law, law enforcement, etc. .). Finally, it should be noted that the issue will be further examined. The sample surveyed countries will be expanded as well as the amount of the determinants of the shadow economy, or is expected to use more sophisticated econometric methods.

Acknowledgements

This paper was financially supported within the VŠB - Technical University SGS grant project No. SP2015/115 "Institutional and Monetary Context of Economic Integration of European Countries Today".

References

[1] ENSTE, Dominik H. (2009). *Regulation and shadow economy: empirical evidence for 25 OECD-countries.* Constitutional Political Economy. 21: 231-248. ISSN: 1043-4062.

[2] EUROPEAN COMMISSION (2014). *Databases and indicators.* [online database]. Washington, DC, [cit. 12.12.2014]. Available at: <http://ec.europa.eu/economy_finance/db_indicators/index_en.htm>
[3] EUROSTAT (2014). *General and regional statistics.*[online database]. Washington, DC, [cit. 12.12.2014]. Available at: < http://ec.europa.eu/eurostat/data/database>
[4] IM, Kyung S.; PESARAN, M. Hashem and Yongcheol SHIN (2003). *Testing for unit roots in heterogeneous panels.* Journal of Econometrics, Vol. 115, Issue 1, p. 53-74. July 2003.
[5] JOHNSON, Simon, Daniel KAUFMANN and Andrei SHLEIFER (1997). *The Unofficial Economy in Transition*. Brookings Papers Econ. Act.
[6] KOTLÁNOVÁ, Eva a Igor KOTLÁN (2012). Vliv institucionálního prostředí na velikost korupce: Empirická analýza.Politická Ekonomie, No. 2, p. 167-186.
[7] LEVIN, Andrew; LIN, Chien-Fu; JAMES CHU, Chia-Shang (2002). *Unit root tests in panel data: asymptotic and finite-sample properties*. Journal of Econometrics, Vol. 108, Issue 1, May 2002, Pages 1-24.
[8] MADDALA, G. S. and Shaoven WU (1999). *A comparative study of unit root tests with panel data and a new simple test*. Oxford bulletin of economics and statistics, special issue. 0305-9049.
[9] MARINOV, Anton (2008). *Hidden economy in the rural regions of Bulgaria*. International Review on Public and Non profit Marketing. 5: 71-80. ISSN: 1865-1992
[10] NOVÁK, Petr (2007). *Analýza panelových dat*. Acta Oeconomica Pragensia, roč. 15, č. 1.
[11] SCHNEIDER, Friedrich and Dominik, H. ENSTE (2000). *Shadow Economies: Size, Causes, and Consequences.* Journal of economic literature. Vol. 38.
[12] SCHNEIDER, Friedrich (2006). *Shadow economies and corruption all over the world: what do we really know?* The Institute for the Study of Labor. Discussion Paper No. 2315.
[13] SCHNEIDER, Friedrich (2011). *The Shadow Economy in Europe, 2011*. Johannes Kepler University of Linz, *ATKEARNEY, VISA 2011*.
[14] SCHNEIDER, Friedrich (2011). *Size and Development of the Shadow Economy of 31 European and 5 other OECD Countries from 2003 to 2012: Some New Facts.* Johannes Kepler University of Linz [online]. [cit. 12.12.2014]. Available at: <http://www.economics.uni-linz.ac.at/members/Schneider/files/publications/2012/ShadEcEurope31.pdf>
[15] SCHNEIDER, Friedrich (2012). The Shadow Economy and Work in the Shadow: What Do We (Not) Know? IZA Discussion Paper. No. 6423.
[16] SCHNEIDER, Friedrich (2013). The Shadow Economy in Europe, 2013. *Johannes Kepler University of Linz. AT KEARNEY, VISA 2013.*
[17] STARTIENĖ, Gražina and Karolis TRIMONIS (2010). *Causes and consequences of non-observed Economy*. Economic and Management. (15), 275-80.
[18] WORLDBANK (2014). Worldwide Governance Indicators.[online database]. Washington, DC, [cit. 12.12.2014]. Available at: <http://data.worldbank.org/data-catalog/worldwide-governance-indicators>
[19] WORLDBANK (2014). *Indicators – Urban population.*[online database]. Washington, DC, [cit. 12.12.2014]. Available at: <http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>

Sequential radial approach to handle large *p*-median problems

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. In this paper, we introduce a sequential approximate radial approach to the very large *p*-median problems making use of a universal IP-solver. The approach overcomes the drawback of universal solvers, which often fail in solving large instances of the *p*-median problem due to big demands on memory and computational time. The radial approach uses an approximation of a common distance by some pre-determined distances given by so called dividing points. Covering formulation of the problem enables the implementation of the solving technique in the frame of commercial optimization software to obtain near optimal solution in a short term. Regardless of this approximate approach effectivity, huge instances of the *p*-median problem resist to attempts to solve them by the approach due to big size of the model, which depends on the cardinality of the possible service center location set. Suggested sequential method is based on successive reduction of the cardinality. At each step of the method, the radial formulation of the *p*-median problem is solved for such system of radii, which balances the cardinality of the possible service center location set.

Keywords: weighted *p*-median problem, approximate covering model, sequential problem size reduction

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

The number of possible center locations is the key size of the *p*-median problem and this size considerably influences solvability of the problem in acceptable time and accuracy [8]. Most of service system designers use the *p*-median problem to design a suitable deployment of the service centers and they also use commercial IP-solvers to obtain a good solution of the problem. For the designers, the solvability of huge problem instances is usually more substantial than obtaining the exact solution of the problem. The necessity to solve bigger instances of the *p*-median or the weighted *p*-median problems led to radial formulation of the problem [1], [2], [3], [9], which considerably outperforms the original location-allocation formulation, as a commercial IP-solver is used for the problem solving. To accelerate the computation, the homogenous radial formulation was developed [4], [5]. This approach represents a trade-off between time consumption and the result accuracy. The approach yields a near optimal solution with a controlled deviation from the optimal solution. The deviation depends on the number of radii, where each radius corresponds to one dividing point. The bigger number of radii is used, the smaller deviation from the optimal solution is obtained. The homogenous radial formulation uses a model, which size is proportional to product of the number of radii and the number of the possible center locations. It follows that bigger cardinality of the set of possible center locations can be handled, when small number of radii is considered. Within this paper, we suggest an approximate sequential method based on the radial approach. The method solves the *p*-median or weighted *p*-median problems at each step starting with full set of the possible center locations, where a given part of the center locations is chosen as the set of possible locations for the next step. The first step is performed with a minimal number of dividing points to keep the product of the number of possible locations and the number of dividing points constant for solvability of a partial problem. The remainder of the paper is organized as follows. The radial formulation of the *p*-median problem is briefly described and analyzed is Section 2. Section 3 is devoted to description of various schemes of the problem size reduction and Section 4 contains results of numerical experiments, which are used to choose the most effective way of the sequential reduction. The results are summarized in Section 5.

[1] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: jaroslav.janacek@fri.uniza.sk
[2] Ing. Marek Kvet, PhD., University of Žilina, Faculty of Security Engineering, 1.mája 32, 010 26 Žilina, Slovak Republic, marek.kvet@fbi.uniza.sk

2 Radial formulation of the *p*-median problem

The weighted *p*-median problem is a task of location of at most *p* centers so that the sum of individual distances from each user's location to the nearest located center multiplied by the number of users sharing the location is minimal. To describe the problem, we denote $I = \{1 \dots m\}$ a set of possible center locations and $J = \{1 \dots n\}$ the set of possible users' locations. The symbol b_j denotes the number of users located at the location *j*. The distance between a user at the location *j* and the possible center location *i* is denoted as d_{ij}.

The weighted *p*-median problem is broadly discussed in [1], [2], [3], [6] from the viewpoint of solving techniques suggested for fast solving of the huge instances. Within this paper, we focus on radial formulation used in the above papers and we deal with the approach based on the set of dividing points.

The strategic decision in the problem concerns location of centers at possible center locations from the set *I*. To model this decision at the particular location *i*, we introduce a zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1, if a center should be located at the location *i*, and it takes the value of 0 otherwise. To obtain upper or lower bounds of the original objective function, the range $[d_0, d_u]$ of all possible distance values $d_0 < d_1 < \dots < d_u$ from the matrix $\{d_{ij}\}$ is partitioned into $v+1$ zones. The zones are separated by a finite ascending sequence of so called dividing points $D_1, D_2 \dots D_v$ chosen from the sequence $d_0 < d_1 < \dots < d_u$, where $0 = d_0 = D_0 < D_1$ and also $D_v < D_{v+1} = d_u$. The zone *s* corresponds with the interval $(D_s, D_{s+1}]$. The length of the *s*-th interval is denoted by e_s for $s = 0 \dots v$. In addition, auxiliary zero-one variables x_{js} for $s = 0 \dots v$ are introduced. The variable x_{js} takes the value of 1, if the distance of the user at $j \in J$ from the nearest located center is greater than D_s and it takes the value of 0 otherwise. Then the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + \dots + e_vx_{jv}$ constitutes an upper approximation of the distance d_{j*} from user location *j* to the nearest located center. If the distance d_{j*} belongs to the interval $(D_s, D_{s+1}]$, then the value of D_{s+1} is the upper estimation of d_{j*} with the maximal possible deviation e_s. Let us introduce a zero-one constant a_{ij}^s for each triple $[i, j, s]$, where $i \in I$, $j \in J$ and $s \in \{0 \dots v\}$. The constant a_{ij}^s is equal to 1, if the distance d_{ij} between the user location *j* and the center location *i* is less than or equal to D_s, otherwise $a_{ij}^s = 0$. Then the radial-type weighted covering model can be formulated according to [6], [7] as follows:

$$\textit{Minimize} \qquad \sum_{j=1}^{n} b_j \sum_{s=0}^{v} e_s x_{js}, \tag{1}$$

$$\textit{Subject to:} \quad x_{js} + \sum_{i=1}^{m} a_{ij}^s y_i \geq 1 \qquad j = 1 \dots n, \quad s = 0 \dots v, \tag{2}$$

$$\sum_{i=1}^{m} y_i \leq p, \tag{3}$$

$$x_{js} \geq 0 \qquad j = 1 \dots n, \quad s = 0 \dots v, \tag{4}$$

$$y_i \in \{0, 1\} \quad i = 1 \dots m. \tag{5}$$

The objective function (1) gives the upper bound of the sum of weighted distance values. The constraints (2) ensure that the variables x_{js} are allowed to take the value of 0, if there is at least one center located in radius D_s from the user location *j*. The constraint (3) puts a limit *p* on the number of located centers.

Deployment of the dividing points influences the size of the covering model (1) – (5) and the accuracy of the result. The dividing points can be chosen only from the set of values $d_0 < d_1 < \dots < d_u$ of the distance matrix $\{d_{ij}\}$, based on frequency N_h of value d_h occurrence in the matrix $\{d_{ij}\}$. The possible approaches to the dividing point determination were broadly studied in [6], [7].

It can be easily found that the above presented model contains $n*(v+1)+m$ variables and $n*(v+1)+1$ structural constraints. These figures can be rounded to $n*v+m$ and $n*v$ respectively. As concerns the number of non-zero items on the left hand sides of the structural constraints, it can be estimated by the expression $n*v + n*v*m$, which overwhelming item is $n*v*m$.

3 Sequential reduction of the set of possible center locations

The suggested sequential reduction of the weighted *p*-median problem consists of successive application of the above described radial method denoted as *RMethod*(), which produces set $\underline{I}$ of the center locations, where the centers are to be located. The problem solved by the *RMethod*() is defined on a network by subset *I* of its nodes,

where centers can be located and by the numbers p and v. The number p limits the number of located centers and the figure v gives the number of dividing points. The other parts of input as distance matrix $\{d_{ij}\}$, users' location set J and the numbers b_j for $j \in J$ do not change during the sequential process, and that is why we give only I, p and v as arguments of the *RMethod*(I, p, v).

The suggested sequential reduction starts from huge instances of the weighted p-median problem characterized by I_0 and p, where $|I_0| = m_0$. The goal of the reduction is to reduce the original problem to a smaller one in w steps. The reduced problem is characterized by I_w and p, where $|I_w| = m_w << m_0$. We suppose that the reduced problem is solvable in admissible time and accuracy for given number of dividing points $v=v_w$. The reduction process is described by the following algorithm, which processes sequence of w problems.

1. Set $v_0=\lceil m_w * v_w / m_0 \rceil$.
2. For $i=0 \ldots w-1$ repeat $I_{i+1}=RMethod(I_i, m_{i+1}, v_i)$ and $v_{i+1}=\lceil m_0 * v_0 / m_{i+1} \rceil$.

Within this algorithm, we try to keep the product of $m_i * v_i$ constant at each step of the process with decreasing the cardinality m_i of I_i and increasing number v_i of dividing points. The sequence $m_0, m_1 \ldots m_w$ can be obtained either as arithmetic or geometric sequence for given entries w, m_0 and m_w.
We can use (6) for generating the sequence in the arithmetic case and (7) for the geometric one.

$$m_i = m_0 - i * \left\lceil \frac{m_0 - m_w}{w} \right\rceil \quad for\ i = 0, \ldots, w-1 \tag{6}$$

$$m_i = \left\lceil m_0 \left(\frac{m_w}{m_0} \right)^{\frac{i}{w}} \right\rceil \quad for\ i = 0, \ldots, w-1 \tag{7}$$

The following computational study can determine the way of decreasing and the number of steps, which keep the result accuracy and reach the reduction in acceptable computational time.

4 Numerical experiments

This section is devoted to searching for the most effective scheme of reduction of a huge p-median problem to the smaller instance from the point of the solution accuracy. Especially, we focus our effort on examination of the final reduced problem from the point of starting state of the final problem processing. We take into consideration that repeatedly used *RMethod* may burden the final optimization process by gradually reducing size of the problem. We organized the first series of experiments to find whether the fining of the reduction by way of a sequence of reduction steps may improve the accuracy of the final result. That is why, we solved mediate instances of the original problem by exact and approximate (radial) methods and compared their results to the solutions obtained by the suggested approach for various number of steps and schemes of the size reduction. Two schemes were studied here. The first one, denoted as "arithmetic", reduces the size of the problem by the same decrement at each step and the second, denoted as "geometric", reduces the size of the problem by a decreasing decrement according to (7).

The experiments were performed on the pool of benchmarks obtained from the real emergency health care system, which was originally designed for each self-governing region of Slovak Republic. The structure of the system is given by the deployment of ambulance vehicles at nodes of the associated road network. The instances are organized so that they correspond to the administrative organization of Slovakia. For each self-governing region (Banská Bystrica - BB, Košice - KE, Nitra - NR, Prešov - PO, Trenčín - TN, Trnava - TT and Žilina - ZA) all cities and villages with corresponding number of inhabitants b_j were taken into account. The coefficients b_j were rounded to hundreds. The number of possible service center locations $|I_0| = m_0$ is the same as the number of user locations $|J|$ in all solved instances. It means that each community (even the smallest) may represent a possible service center location. The value of parameter p corresponds to the original number of ambulance vehicles deployed in the given self-governing region. The final reduction of the set of possible locations was set at 60 % of the m_0 in this first series of experiments, which means $m_w=0.6*m_0$. The starting number v_0 of dividing points was computed according to the above mentioned algorithm, whereas the final number v_w of dividing points was set to 20 in all cases. The reduction process in the starting series of experiments is reported in the Table 1.

The table is organized into two sections, where the upper one contains brief description of the instance. The description consists of the number of possible service center locations $|I_0|$ and the number p of centers, which are to be located. The second section contains the deviations of the obtained solution objective function value from objective function value of the exact solution. The deviation is given in percentage, where the base is the objec-

tive function value of the exact solution. The rows are labeled by denotation "*n* steps – *x*", where *n* denotes the number of steps and *x* indicates the used scheme, which is either arithmetic (*x* = a) or geometric (*x* = g). The particular row denoted by "0 steps" corresponds to the results obtained by radial approach applied on the original set of possible service center locations.

Region	**BB**	**KE**	**NR**	**PO**	**TN**	**TT**	**ZA**	
$\|I_0\|=m_0$	515	460	350	664	276	249	315	
v_0	13	12	12	13	13	13	13	average deviations
p	52	46	35	67	28	25	32	
0 steps	0.02	1.00	0.00	0.00	1.13	0.00	0.11	0.32
1 step	0.16	0.26	0.48	0.44	1.21	0.26	0.10	0.42
2 steps - a	0.26	0.26	0.48	0.44	1.29	0.26	0.23	0.46
2 steps - g	0.26	0.26	0.48	0.44	1.29	0.26	0.23	0.46
4 steps - a	0.26	0.26	0.35	0.44	1.29	0.26	0.49	0.48
4 steps - g	0.26	0.26	0.48	0.44	1.29	0.26	0.23	0.46

Table 1 Deviations of the obtained solution objective function value from objective function value of the exact solution in percentage of the exact solution for individual benchmarks.

The second series of experiments with different final reduction were performed on the same pool of benchmarks to affirm validity of the results obtained in the first series. The final reduction was set at 70 % of the m_0 in this second series of experiments, which means $m_w=0.7*m_0$. The starting number v_0 of dividing points was computed according to the above mentioned algorithm. The reduction process in the second series of experiments is reported in the Table 2.

Region	**BB**	**KE**	**NR**	**PO**	**TN**	**TT**	**ZA**	
$\|I_0\|=m_0$	515	460	350	664	276	249	315	
v_0	15	14	14	15	15	15	15	average deviations
p	52	46	35	67	28	25	32	
0 steps	0.02	1.00	0.00	0.00	1.13	0.00	0.11	0.32
1 step	0.16	0.61	0.19	0.23	1.13	0.26	0.23	0.40
2 steps - a	0.26	0.61	0.19	0.23	1.13	0.26	0.23	0.42
2 steps - g	0.26	0.61	0.19	0.23	1.13	0.26	0.23	0.42
4 steps - a	0.26	0.61	0.19	0.23	1.13	0.26	0.23	0.42
4 steps - g	0.26	0.61	0.19	0.23	1.13	0.26	0.23	0.42

Table 2 Deviations of the obtained solution objective function value from objective function value of the exact solution in percentage of the exact solution for individual benchmarks.

All experiments were performed using the optimization software FICO Xpress 7.7 (64-bit, release 2014). The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with the parameters: 2.0 GHz and 8 GB RAM. The computational times are not reported here, but we note that the computational time for any instance solved by the radial approach does not exceed 2 seconds whereas the exact solution, where the location-allocation formulation was used, was completed at most in 1 minute.

The first and second series of experiments showed that the multi-step reduction does not bring any improvement of the solution accuracy compared to the one step reduction regardless of the used schemes. Nevertheless, all ways of the multi and one step reduction yield very accurate results, which differ from the exact solution at most by 0.5 percent.

To verify the suggested approach on larger instances of the problem, we organized the third series of experiments using the benchmark derived from the whole real emergency health care system of Slovak Republic. The cardinality of original set of possible service center locations $|I_0|$ is 2916. Due to this size, the attempts on solving the problem exactly failed, when the location-allocation model was used. That is why we compare the obtained result to the approximate solution by radial approach. The experiments were performed for the rates of reduction 60 and 70 percent of the original size and the results are reported in the Table 3.

Region:	**Slovakia**			
	Reduction 60 %		**Reduction 70 %**	
v_0	13		15	
	Deviation [%]	Computational time [s]	Deviation [%]	Computational time [s]
0 steps		33.80		33.80
1 step	0.09	12.15	0	13.00
2 steps - a	0.12	13.21	0	15.01
2 steps - g	0.12	13.62	0	14.48
4 steps - a	0.09	15.97	0	18.38
4 steps - g	0.12	16.29	0	19.08

Table 3 Deviations of the obtained solution objective function value from objective function value of the approximate radial solution in percentage of the approximate radial solution for given rates of reduction and computational times in seconds. The region has the following parameters $|I_0|$=2916 and p=292.

5 Conclusions

The paper is focused on mastering larger instances of the *p*-median problem using commercial IP-solver, which can fail in solving large *p*-median problem. The reduction method has been designed and examined to cope with the drawback. The suggested method is based on exploitation of multiple usage of the radial method of *p*-median problem solution with varying number of dividing points. Various schemes of the problem size reduction were studied for different number of reduction steps. It has been found that the studied ways of reduction yield very accurate results, which differ from the solution obtained directly, without any reduction, at most by 0.5 percent regardless of the number of steps or used scheme. Thus for the case of such instances, which are comparable to Slovakia, we can recommend to use one step reduction. This way of reduction demands minimal time in comparison with multi-step reduction and in addition, it pays for almost negligible loss of accuracy by considerable reduction of computational time comparing to direct solution. Thus we can conclude that we present a useful tool for large *p*-median problem instances. The tool can be easily implemented using common commercial optimization software. Further research in this field can be focused on applying and verifying the suggested approach on very large instances.

Acknowledgement

This work was supported by the research grants VEGA 1/0518/15 "Resilient rescue systems with uncertain accessibility of service", VEGA 1/0339/13 „Advanced microscopic modeling and complex data sources for designing spatially large public service systems" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks". This paper was also supported by the project: University Science Park of the University of Žilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the European Regional Development Fund. We would like to thank to „Centre of excellence for systems and services of intelligent transport" (ITMS 26220120050) for built up the infrastructure, which was used.

References

[1] Avella, P., Sassano, A., Vassil'ev, I..: *Computational study of large scale p-median problems*. In: Mathematical Programming 109, pp. 89-114, (2007).

[2] Current, J., Daskin, M., Schilling, D.: *Discrete network location models*. In: Drezner Z (ed) et al. Facility location. Applications and theory. Berlin: Springer, pp. 81-118, (2002).

[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, (2011).

[4] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: Proceedings of the 1st International Conference ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, (2008).

[5] Janáček, J., Kvet, M.: *Approximate solving of large p-median problems*. In: ORP3-Operational research peripatetic post-graduate programme: proceedings, Cádiz, Spain, September 13-17, 2011, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, 2011, ISBN 978-84-9828-348-8, pp. 221-225, (2011).

[6] Janáček, J., Kvet, M.: *Relevant Network Distances for Approximate Approach to Large p-Median Problems*. In: Operations research proceedings 2012: selected papers of the international conference on operations research: September 4-7, 2012, Hannover, Germany, ISSN 0721-5924, ISBN 978-3-319-00794-6, pp 123 – 128, (2012).
[7] Janáček, J., Kvet, M.: *Public service system design with disutility relevance estimation*. In: Proceedings of the 31st international conference:"Mathematical Methods in Economics", September 11-13, 2013, Jihlava, Czech Republic, ISBN 978-80-87035-76-4, pp 332 – 337, (2013).
[8] Janáčková, M., Szendreyová, A.: *An impact of set of centers of distribution system design problem on computational time of solving algorithm*. In: Mikulski J (ed). Advances in Transport Systems Telematics, Katowice, pp. 109-112, (2006).
[9] Marianov, V., Serra, D.: *Location problems in the public sector*. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp. 119-150, (2002).

Evaluation of absolute type in the Partial Goals Method

Věra Jandová[1], Jana Talašová[2]

Abstract. Methods of multiple criteria decision making usually work with evaluations of relative type. Such evaluations are dependent on the given set of alternatives and do not say too much about fulfillment of the decision-maker's goal. They are sufficient when the best alternative from a given set of alternatives is to be chosen. However, in the decision making about acceptance or rejection of alternatives entering one by one into a system (e.g. decision making about granting a credit in banking), an evaluation of absolute type with respect to the given goal is needed. The evaluations of absolute type presented in this paper are measurements on an absolute scale where one limit point expresses total fulfillment of the given goal while the second limit point means the total lack of its fulfillment. Such evaluations of absolute type are used in the Partial Goals Method which is implemented in its fuzzified version in the FuzzME software (Fuzzy Methods of Multiple Criteria Evaluation). In this paper, the evaluation process applied in the Partial Goals Method will be studied from the point of view of the measure theory.

Keywords: absolute evaluation, normalized measure, degree of fulfillment, weighted average, partial goals, MCDM

JEL classification: C44
AMS classification: 91B06

1 Evaluation of Absolute Type

The most frequently used methods of multiple criteria decision making and evaluation use *evaluation of relative type*, e.g. it is used in the utility theory (see [3]). The evaluations of relative type are dependent on the given set of compared alternatives. Nevertheless, there are situations where alternatives enter in the model one by one, and on the base of their evaluation, decision making about acceptance or rejection of them needs to be done (e.g. decision making about granting a credit in banking). In such case *evaluation of absolute type with respect to the given goal* is required. Evaluations of this type are independent on the set of compared alternatives and express degrees of fulfillment of the given goal, e.g. see [1] ,[2]. It means that evaluation is given on absolute evaluation scale where one limit point means the total fulfillment and the second one means the total lack of fulfillment of the given goal of evaluation.

The evaluation of absolute type with respect to the given goal will be illustrated on the following simple model. Let us assume that the goal G of an evaluation problem can be described by a countable set of attributes X. On account of this, also each alternative A can be associated with a set of attributes $x \subseteq X$ with respect to the goal G. Naturally, it holds:

- if $x = X$, then the alternative A absolutely fulfills the goal G;
- if $x = \emptyset$, then the alternative A does not fulfill the goal G at all;
- if $x \subset X$, $x \neq \emptyset$, then the alternative A fulfills the goal G at certain degree.

The *evaluation function* with respect to the given goal G will be defined by a normalized measure (a probability measure) μ (see [5]) on the set of attributes X. For an alternative A given by a set of attributes $x \subseteq X$, evaluation with respect to the goal G will be represented by $\mu(x)$. It means that

[1]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, vera.jandova@upol.cz
[2]Palacký University in Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, jana.talasova@upol.cz

evaluation will be given on the absolute scale $[0,1]$ with the following interpretation: if $\mu(x)=\alpha$, then the alternative A fulfills the goal G at $\alpha\%$. The evaluation function is defined formally as follows.

Definition 1.1. Let G be a goal of the evaluation problem and X a countable set of attributes which describes the goal G. Let $\mathcal{P}(X)$ be the power set of X. Let A be an alternative and $x \subseteq X$ the set of attributes which describes the alternative A with respect to the goal G. Let $\mu : \mathcal{P}(X) \to [0,1]$ be the normalized measure on X expressing fullfilments of the given goal for any alternative, i.e. μ fulfills

1. $\mu(\emptyset)=0, \quad \mu(X)=1;$
2. $\mu(\bigcup_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} \mu(X_i)$ for any countable sequence $\{X_i\}_{i=1}^{\infty}$ such that $X_i \subseteq X$ for each $i \in \mathbb{N}$, $X_i \cap X_j = \emptyset$ for $i \neq j$, $i,j \in \mathbb{N}$.

Then μ will be the evaluation function on X called the *measure of fulfillment* of the goal G. The value $\mu(x)$ is the evaluation of A with respect to G called the *measure (degree) of fulfillment* of the goal G by the alternative A.

For the finite case, the mentioned evaluation model will be given as follows: Let us assume the goal G of an evaluation problem is described by a finite set of attributes $X=\{a_1,a_2,\ldots,a_n\}$. Importances of the attributes $a_1,a_2,\ldots,a_n$ are expressed by normalized weights $\mu_1,\mu_2,\ldots,\mu_n$, $\sum_{i=1}^{n}\mu_i=1$, $\mu_i \geq 0$ for each $i=1,2,\ldots,n$. Then evaluation of an alternative A described by a set of attributes $x \subseteq X$ will be given by

$$\mu(x) = \sum_{i=1,2,\ldots,m:a_i \in x} \mu_i = \sum_{i=1}^{n} \mu_i \cdot \chi_x(a_i)$$

where $\chi_x : X \to \{0,1\}$ is the characteristic function of the set x. In this case $\mu : \mathcal{P}(X) \to [0,1]$, where $\mathcal{P}(X)$ is the power set of X, is the normalized measure generated by the normalized weights $\mu_1=\mu(a_1), \mu_2=\mu(a_2),\ldots,\mu_n=\mu(a_n)$.

2 The Partial Goals Method

The evaluation of absolute type with respect to the given goal is used in the Partial Goals Method which is described in detail in [3]. At first, the preliminaries of this method will be introduced. Then the method will be examined from the point of view of the measure theory. Finally, two simple models of multiplecriteria decision making based on the Partial Goals Method will be presented.

Preliminaries

Let us assume G is the overall goal of the evaluation. Let G be fully decomposed into m disjunctive partial goals $G_1,G_2,\ldots,G_m$. This requirement means that no interactions are considered among partial goals. Let us assume $w_1,w_2,\ldots,w_m$, $\sum_{i=1}^{m} w_i=1$, $w_i \geq 0$ for $i=1,2,\ldots,m$, are normalized weights of partial goals $G_1,G_2,\ldots,G_m$ which represent their shares in the overall goal G. Let partial evaluations of an alternative express the degrees of fulfillment of respective partial goals G_i, $i=1,2,\ldots,m$. It will be demonstrated that the overall evaluation of an alternative calculated by the weighted average of partial evaluations represents the degree of fulfillment of the overall goal.

Interpretation by the measure theory

The Partial Goals Method can be examined from the point of view of the measure theory. Let us assume that the overall goal is associated with a countable set of attributes X which describes it. Then also each partial goal G_i is described by a set of attributes X_i for each $i=1,2,\ldots,m$ such that

$$X = X_1 \cup X_2 \cup \cdots \cup X_m, \quad X_i \cap X_j = \emptyset \quad \text{for each} \quad i,j=1,2,\ldots,m,\ i \neq j.$$

The normalized weights $w_1,w_2,\ldots,w_m$ generate a normalized measure ν on the sets of partial goals $G_1,G_2,\ldots,G_m$. As partial goals are described by the sets of atributes $X_1,X_2,\ldots,X_m$, we can write $\nu : \mathcal{P}(\{X_1,X_2,\ldots,X_m\}) \to [0,1]$ where $w_1=\nu(X_1), w_2=\nu(X_2),\ldots,w_m=\nu(X_m)$. The partial evaluation functions $\mu_i : \mathcal{P}(X_i) \to [0,1]$ can be interpreted as normalized measures on the set of attributes which describe the partial goals G_i for each $i=1,2,\ldots,m$.

With regard to the fact that the overall goal is associated with the set of attributes, also any alternative can be considered as the set of attributes. Only attributes that relate to the overall goal are taken into account. Hence, each alternative A is described by the set of attributes $x \subseteq X$ and x can be decomposed into set of attributes $x_1 \subseteq X_1, x_2 \subseteq X_2, \ldots, x_m \subseteq X_m$ such that

$$x = x_1 \cup x_2 \cup \cdots \cup x_m, \quad x_i \cap x_j = \emptyset \quad \text{for each} \quad i, j = 1, 2, \ldots, m, \ i \neq j.$$

The overall evaluation of alternative A will be given by

$$\mu(x) = \sum_{i=1}^{m} w_i \mu_i(x_i) = \sum_{i=1}^{m} \nu(X_i)\mu_i(x_i).$$

The following theorem demonstrates that such defined function is a normalized measure $\mu : \mathcal{P}(X) \to [0, 1]$ on the set of attributes which describe the overall goal G.

Theorem 2.1. Let X be a countable set, let $X_1, X_2, \ldots, X_m$ be sets which form partition of X, i.e. $X = X_1 \cup X_2 \cup \cdots \cup X_m$ and $X_i \cap X_j = \emptyset$ for $i \neq j$, $i, j = 1, 2, \ldots, m$. Let $\mathcal{P}(\{X_1, X_2, \ldots X_m\})$ be the power set on $\{X_1, X_2, \ldots X_m\}$ and let $\mathcal{P}(X_i)$ be the power set of X_i, $i = 1, 2, \ldots, m$. Let $\nu : \mathcal{P}(\{X_1, X_2, \ldots X_m\}) \to [0, 1]$ and $\mu_i : \mathcal{P}(X_i) \to [0, 1]$, $i = 1, 2, \ldots, m$, be normalized measures. Let us define for each $x \subseteq X$, $x = x_1 \cup x_2 \cup \cdots \cup x_m$ where $x_i \subseteq X_i$, $i = 1, 2, \ldots, m$:

$$\mu(x) = \sum_{i=1}^{m} \nu(X_i)\mu_i(x_i).$$

Then the set function $\mu : \mathcal{P}(X) \to [0, 1]$, where $\mathcal{P}(X)$ is power set of X, is a normalized measure on X.

Proof: Let $X = \bigcup_{i=1}^{m} X_i$, $X_i \cap X_j = \emptyset$ for $i \neq j$, $i, j = 1, 2, \ldots, m$. Let $\nu : \mathcal{P}(\{X_1, X_2, \ldots X_m\}) \to [0, 1]$, $\mu_i : \mathcal{P}(X_i) \to [0, 1]$, $i = 1, 2, \ldots, m$, be normalized measures. Let us suppose $x = \bigcup_{i=1}^{m} x_i$, $x_i \subseteq X_i$.

1. From $\nu(X_i) \in [0, 1]$ and $\mu_i(x_i) \in [0, 1]$ for each $i = 1, 2, \ldots, m$, it holds $\mu(x) \in [0, 1]$ for each $x \subseteq X$:

$$0 \leq \nu(X_i)\mu_i(x_i) \leq \nu(X_i) \text{ for each } i = 1, 2, \ldots, m \quad \Rightarrow \quad 0 \leq \sum_{i=1}^{m} \nu(X_i)\mu_i(x_i) \leq \sum_{i=1}^{m} \nu(X_i) = 1.$$

2. If $x = \emptyset$, i.e. $x_i = \emptyset$ for each i, then $\mu(\emptyset) = \sum_{i=1}^{m} \nu(X_i)\mu_i(\emptyset) = \sum_{i=1}^{m} \nu(X_i) \cdot 0 = 0$.
3. If $x = X$, i.e. $x_i = X_i$ for each i, then $\mu(X) = \sum_{i=1}^{m} \nu(X_i)\mu_i(X_i) = \sum_{i=1}^{m} \nu(X_i) \cdot 1 = \nu(\bigcup_{i=1}^{m} X_i) = 1$.
4. Let us suppose $\{y_j\}_{j=1}^{\infty}$ is a countable set such that $y_j \subseteq X$ for each $j \in \mathbb{N}$, $y_j \cap y_k = \emptyset$ for each $j \neq k$, $j, k \in \mathbb{N}$. Let $y_j = \bigcup_{i=1}^{m} y_{ji}$, $y_{ji} \subseteq X_i$ for each $i = 1, 2, \ldots, m$, $j \in \mathbb{N}$. Then

$$\mu\left(\bigcup_{j=1}^{\infty} y_j\right) = \sum_{i=1}^{m} \nu(X_i)\mu_i\left(\bigcup_{j=1}^{\infty} y_{ji}\right) = \sum_{i=1}^{m} \nu(X_i) \sum_{j=1}^{\infty} \mu_i(y_{ji}) = \sum_{j=1}^{\infty} \sum_{i=1}^{m} \nu(X_i)\mu_i(y_{ji}) = \sum_{j=1}^{\infty} \mu(y_j). \qquad \square$$

2.1 Model I: Goal described by a finite set of atributes

The first model demonstrates use of the Partial Goals Method in such case where the overall goal is given by the finite set of attributes. It means that also each partial goal is given by a finite set of attributes.

Again, the overall goal G is fully decomposed into m disjunctive partial goals $G_1, G_2, \ldots, G_m$. The overall goal G is given by a finite set of attributes $X = \{a_1, a_2, \ldots, a_n\}$. Regarding to this, partial goals G_i, $i = 1, 2, \ldots, m$, are given by the set of attributes $X_i = \{a_j, j \in I_i\}$ where I_i is set of indexes determining which attributes are associated with the partial goal G_i, $i = 1, 2, \ldots, m$, i.e. it holds $\{1, 2, \ldots, n\} = I_1 \cup I_2 \cup \cdots \cup I_m$, $I_i \cap I_k = \emptyset$ for each $i \neq k$, $i, k = 1, 2, \ldots, m$. Normalized weights $w_1 = \nu(X_1), w_2 = \nu(X_2), \ldots, w_m = \nu(X_m)$, $\sum_{i=1}^{m} w_i = 1$, $w_i \geq 0$ for $i = 1, 2, \ldots, m$, of partial goals

$G_1, G_2, \ldots, G_m$ generate the normalized measure $\nu : \mathcal{P}(\{X_1, X_2, \ldots, X_m\}) \to [0,1]$. Evaluations of attributes which describe the partial goals G_i, $i = 1, 2, \ldots, m$, generate normalized measure on X_i, i.e. $\mu_{ij} = \mu_i(\{a_j\})$, where $j \in I_i$, generate for each $i = 1, 2, \ldots, m$ normalized measure

$$\mu_i : \mathcal{P}(\{a_j; j \in I_i\}) \to [0,1].$$

It holds $\mu_i(x_i) = \sum\limits_{j \in I_i, a_j \in x_i} \mu_{ij}$ for each $x_i = \bigcup\limits_{j \in I_i; a_j \in x_i} \{a_j\}$, $i = 1, 2, \ldots, m$.

The overall evaluation of alternative A described by $x = \bigcup\limits_{i=1}^{m} x_i$ where $x_i = \bigcup\limits_{j \in I_i; a_j \in x} \{a_j\}$, is given

$$\mu(x) = \sum_{i=1}^{m} \nu(X_i)\mu_i(x_i) = \sum_{i=1}^{m} \nu(X_i) \sum_{j \in I_i : a_j \in x} \mu_i(\{a_j\}) = \sum_{j=1,2,\ldots,n : a_j \in x} \mu(\{a_j\}).$$

Then the overall evaluations $\mu(\{a_1\}), \mu(\{a_2\}), \ldots, \mu(\{a_n\})$ of attributes, where $\mu(\{a_j\}) = \nu(X_i)\mu_i(\{a_j\})$, $j \in I_i$, $i = 1, 2, \ldots, m$, generate normalized measure $\mu : \mathcal{P}(X) \to [0,1]$.

2.2 Model II: An alternative approach

In decision making, it is not very common that partial goals are associated with the sets of attributes which determine the evaluation of an alternative. More common is the situation where fulfillment of each partial goal is directly measured on the absolute scale $[0,1]$ (e.g. it is given by the opinion of an expert or by the value of partial evaluation function defined for a quantitative characteristic of alternatives).

Again, the overall goal G is decomposed into partial goals $G_1, G_2, \ldots, G_m$ such that $G = \bigcup\limits_{i=1}^{m} G_i$, $G_i \cap G_j = \emptyset$ for $i \neq j$, $i, j = 1, 2, \ldots, m$. An alternative A satisfies certain proportion of each partial goal G_i; it will be denoted A_i, $i = 1, 2, \ldots, m$. A_i itself can be also considered as an alternative. For each partial goal, there is an alternative A_i^C which satisfies opposite proportion of G_i such that $G_i = A_i \cup A_i^C$, $A_i \cap A_i^C = \emptyset$ for each $i = 1, 2, \ldots, m$. The power set $\mathcal{P}(\{A_i, A_i^C\}) = \{A_i, A_i^C, \emptyset, G_i\}$ is σ-algebra on G_i, $i = 1, 2, \ldots, m$.

Partial evaluations $\mu_i(\{A_i\})$ generate for each $i = 1, 2, \ldots, m$ the normalized measure

$$\mu_i : \mathcal{P}(\{A_i, A_i^C\}) \to [0,1].$$

where $\mu_i(\{A_i\}) = \mu_i$ is given evaluation of A with respect to G_i, $\mu_i(\{A_i^C\}) = 1 - \mu_i$, $\mu_i(\emptyset) = 0$, $\mu_i(\{A_i, A_i^C\}) = \mu_i(G_i) = 1$.

Normalized weights $w_1 = \nu(G_1), w_2 = \nu(G_2), \ldots, w_m = \nu(G_m)$, $\sum\limits_{i=1}^{m} w_i = 1$, $w_i \geq 0$ for $i = 1, 2, \ldots, m$, of partial goals $G_1, G_2, \ldots, G_m$ generate normalized measure $\nu : \mathcal{P}(\{G_1, G_2, \ldots, G_m\}) \to [0,1]$.

Then the normalized measure

$$\mu : \mathcal{P}(\{A_1, A_1^C, A_2, A_2^C, \ldots, A_m, A_m^C\}) \to [0,1]$$

is defined for alternatives $B = B_1 \cup B_2 \cup \cdots \cup B_m$, where $B_i \in \mathcal{P}(\{A_i, A_i^C\})$, $i = 1, 2, \ldots, m$. The overall evaluation of alternative B is given

$$\mu(B) = \sum_{i=1}^{m} \nu(G_i)\mu_i(B_i).$$

The overall evaluation $\mu(A) = \sum\limits_{i=1}^{m} \nu(G_i)\mu_i(A_i)$ can be interpreted as the degree of fulfillment of the overall goal G by alternative A and $\mu(A^C) = \sum\limits_{i=1}^{m} \nu(G_i)\mu_i(A_i^C)$ as the degree of lack of fulfillment of the overall goal G by alternative A. It holds $\mu(A) + \mu(A^C) = 1$.

3 The Fuzzified Partial Goals Method

As well as in the crisp version, the Fuzzified Partial Goals Method (FPGM), which is implemented in FuzzME software (see [2]), can be examined from the point of view of fuzzy-number-valued (FNV) normalized measures. This method can be used again in two versions analogously to the crisp Partial Goals Method.

Let us assume the overall goal G is fully decomposed into m disjunctive partial goals $G_1, G_2, \ldots, G_m$. Normalized fuzzy weights $W_1, W_2, \ldots, W_m$ of the partial goals G_1,G_2,$\ldots$, G_m represent their fuzzy shares in the overall goal G. Normalized fuzzy weights are specified in definition 3.1

Definition 3.1. Let $W_1, W_2, \ldots, W_m$ be fuzzy numbers on $[0,1]$ and for any $\alpha \in (0,1]$ and each $w_i \in W_{i\alpha}$, $i \in \{1,2,\ldots,m\}$, there exist $w_j \in W_{j\alpha}$, $j = 1,2,\ldots,m$, $j \neq i$ such that

$$w_i + \sum_{j=1, j\neq i}^{m} w_j = 1.$$

Then $W_1, W_2, \ldots, W_m$ are called *normalized fuzzy weights.*

Note: Fuzzy number on $[0,1]$ is a fuzy number whose support is subset of $[0,1]$. The set of all fuzzy numbers on $[0,1]$ will be denoted $\mathcal{F}_N([0,1])$.

In FPGM, partial evaluations $U_1, U_2, \ldots, U_m$ of an alternative A express the fuzzy degrees of fulfillment of respective partial goals G_i, $i = 1,2,\ldots,m$. It will be demonstrated that the overall evaluation of an alternative calculated by the fuzzy weighted average of partial evaluations represents the fuzzy degree of fulfillment of the overall goal. The fuzzy weighted average is described in definition 3.2.

Definition 3.2. Let $W_1, W_2, \ldots, W_m$ be normalized fuzzy weights. Let $U_1, U_2, \ldots, U_m \in \mathcal{F}_N([0,1])$. Then the *fuzzy weighted average* of $U_1, U_2, \ldots, U_m$ with normalized fuzzy weights $W_1, W_2, \ldots, W_m$ is $U \in \mathcal{F}_N([0,1])$,

$$U = (\mathcal{F}) \sum_{i=1}^{m} W_i U_i,$$

with the membership function given for any $u \in [0,1]$ as follows

$$U(u) = \max\{\min\{W_1(w_1), W_2(w_2) \ldots, W_m(w_m), U_1(u_1), U_2(u_2), \ldots, U_m(u_m)\};$$
$$u = \sum_{i=1}^{m} w_i u_i, \sum_{i=1}^{m} w_i = 1, w_i \geq 0, u_i \in [0,1], i = 1,2,\ldots,m\}.$$

The Fuzzified Partial Goals Method can be interpreted by fuzzified measures. The fuzzy-number-valued measure on a finite set X is defined in definition 3.3.

Definition 3.3. Let X be a finite set, $X = \{X_1, X_2, \ldots, X_m\}$. Let $W_1, W_2, \ldots, W_m$ be normalized fuzzy weights expressing measures of $X_1, X_2, \ldots, X_m$. Then *fuzzy-number-valued (FNV) normalized measure* on X is the mapping $\mu^F : \mathcal{P}(\{X_1, X_2, \ldots, X_m\}) \to \mathcal{F}_N[0,1]$ which is defined as follows:

$$\mu^F(Y) = (\mathcal{F})(W_1 \cdot \chi_Y(X_1) + W_2 \cdot \chi_Y(X_2) + \cdots + W_m \cdot \chi_Y(X_m))$$

for each $Y \subseteq X$, where $\chi_Y : \{X_1, X_2, \ldots, X_m\} \to \{0,1\}$ is the characteristic function of the set Y.

FNV normalized measure given in definition 3.3 satisfies fuzzified axioms of normalized measure (see [4]):

- $\mu^F(\emptyset) = 0, \quad \mu^F(X) = 1;$
- for any pairwise disjoint $B_1, B_2, \ldots, B_k \subset X$, it holds

$$\mu^F\left(\bigcup_{i=1}^{k} B_i\right) = (\mathcal{F})(\mu^F(B_1) \cdot 1 + \mu^F(B_2) \cdot 1 + \cdots + \mu^F(B_k) \cdot 1 + \mu^F(X \setminus (B_1 \cup B_2 \cup \cdots \cup B_k)) \cdot 0).$$

Analogously as in the crisp version, it can be demonstrated by using FNV normalized measures that the overall evaluation of A described by $x \subseteq X$, $x = \bigcup_{i=1}^{m} x_i$, $x_i \subseteq X_i$ for each $i = 1, 2, \ldots, m$, given by

$$U = (\mathcal{F}) \sum_{i=1}^{m} W_i U_i,$$

where W_i, $i = 1, 2, \ldots, m$, represent normalized fuzzy weights of the partial goals, and U_i, $i = 1, 2, \ldots, m$, express the fuzzy degrees of fulfillment of partial goals by A, expresses the fuzzy degree of fulfillment of the overall goal G.

Let us suppose that the overall goal of evaluation G is described by a finite set of attributes X and the partial goals G_i by disjunctive sets of attributes $X_i \subseteq X$ for each $i = 1, 2, \ldots m$, $X = X_1 \cup X_2 \cup \cdots \cup X_m$. In the terms of measure, $\nu^F : \mathcal{P}(\{X_1, X_2, \ldots, X_m\}) \to \mathcal{F}_N([0, 1])$ represents FNV normalized measure expressing importances of partial goal sets. Functions $\mu_i^F : \mathcal{P}(X_i) \to \mathcal{F}_N([0, 1])$, $i = 1, 2, \ldots, m$, represent FNV normalized measures expressing partial evaluations of alternatives. Then FNV normalized measure $\mu^F : \mathcal{P}(X) \to \mathcal{F}_N([0, 1])$ given by

$$\mu^F(x) = (\mathcal{F}) \sum_{i=1}^{m} \nu^F(X_i) \mu_i^F(x_i)$$

represents the overall evaluation of alternative A described by $x \subseteq X$, $x = \bigcup_{i=1}^{m} x_i$ where $x_i \subseteq X_i$ for each $i = 1, 2, \ldots, m$.

Note: It holds $U = \mu^F(x), U_i = \mu_i^F(x_i), W_i = \nu^F(X_i)$, $i = 1, 2, \ldots, m$.

4 Conclusion

The evaluation of absolute type with respect to the given goal was defined by a normalized measure. It was explained that evaluation used in the Partial Goals Method can be interpreted this way. Moreover, it was demonstrated that if partial evaluations express degrees of fulfillment of partial goals and normalized weights represent shares of partial goals in the overall goal, then the overall evaluation calculated by the weighted average of partial evaluations expresses the degree of fulfillment of the overall goal. Two versions of this method were studied. The Fuzzified Partial Goals Method, which is implemented in the FuzzME software, was also examined from the point of view of the fuzzified normalized measures.

Acknowledgements

Supported by the grant IGA PrF 2015 013 Mathematical models of the Internal Grant Agency of Palacky University in Olomouc and by the grant GA14-02424S Methods of operations research for decision support under uncertainty of the Grant Agency of the Czech Republic.

References

[1] Bellman, R. E., Zadeh, L. A.: Decision-making in fuzzy environment. *Management Science* **17** (1970), 141–164.

[2] Holeček, P., Talašová, J., Müller, I.: Fuzzy Methods of Multiple-Criteria Evaluation and Their Software Implementation. In: *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies* (editors Mago V. K., Bhatia N.), IGI Global, 2012, 388–411.

[3] Talašová, J.: *Fuzzy methods of multicriteria evaluation and decision making.* (In Czech.) Publishing House of Palacky University, Olomouc, 2003.

[4] Talašová, J., Pavlačka, O.: Fuzzy probability spaces and their applications in decision making, *Austrian Journal of Statistics* **35** (2006), 347–356.

[5] Torra V., Narukawa Y.: *Modeling Decisions: Information Fusion and Aggregation Operators.* Springer-Verlag Berlin Heidelberg, 2007.

Geometry of Cobb-Douglas surfaces

Miloš Kaňka[1]

Abstract. The goal of this paper is to study the so-called Cobb-Douglas function $z(u,v) = u^\alpha \cdot v^\beta$, where u, v are positive real variables and α, β are positive real constants.
Due to Monge's idea, it is possible to study this function as a surface in $\mathbb{R}^3$ defined by the formula $x(u,v) = (u, v, u^\alpha \cdot v^\beta)$. We are interested in its geometrical properties, that is, Gaussian and mean curvature, and in some cases also the principal curvatures of this surface.
In the end, we give a relation between the Gaussian curvature of a Cobb-Douglas production function and the returns to scale.

Keywords: Tangent vectors, normal vector, Weingarten map, Gaussian and mean curvature, fundamental forms, shape operator, returns to scale.

JEL Classification: C65
AMS Classification: 53A05

1 Introduction

We can say, in a general sense, that the returns to scale express the relation between the changes of input and the change of output. Returns to scale do not, however, describe the whole long-term production function, as *they capture the connection between the proportional changes of input and the change of output induced by them*, cf. [5, p. 177].

As we are concerned in this work with special Cobb-Douglas surfaces, the following conclusions follow from the latter definition:

1. if $\alpha + \beta > 1$, the returns to scale are increasing,
2. if $\alpha + \beta = 1$, the returns to scale are constant,
3. if $\alpha + \beta < 1$, the returns to scale are decreasing.

The aim of this paper is to give some examples of Cobb-Douglas surfaces and some geometric characteristics of these surfaces. In case of increasing returns to scale, the Cobb-Douglas surfaces have the form

$$x(u,v) = (u, v, u^\alpha \cdot v^\beta), \quad \text{where} \quad u > 0, v > 0, \quad \alpha + \beta > 1, \alpha > 0, \beta > 0.$$

In case of decrease returns to scale, the Cobb-Douglas surfaces have the form

$$x(u,v) = (u, v, u^\alpha \cdot v^\beta), \quad \text{where} \quad u > 0, v > 0, \quad \alpha + \beta < 1, \alpha > 0, \beta > 0.$$

Analogously, in case of constant returns to scale, the Cobb-Douglas surfaces have the form

$$x(u,v) = (u, v, u^\alpha \cdot v^\beta), \quad \text{where} \quad u > 0, v > 0, \quad \alpha + \beta = 1, \alpha > 0, \beta > 0.$$

We are interested in the Gaussian curvature, mean curvature, and principal curvatures of these surfaces.

This paper does not give new and therefore unknown results in pure mathematics. The aim of this work is to offer certain kind of information to economists.

Let us remind the reader some basic definitions, see [4], [6].

A subset $S \subset \mathbb{R}^3$ is a *regular surface* if for every point $p \in S$ there exists an open set $U \subset \mathbb{R}^2$ and an open set $V \subset \mathbb{R}^3$ such that there is a regular map $x: U \to \mathbb{R}^3$ that is a homeomorphism of an open set $U \subset \mathbb{R}^2$ onto the open set $V \cap S$.

A surface map $x: U \to S \subset \mathbb{R}^3$, where U is an open set in $\mathbb{R}^2$, is called a *regular map* if it is smooth and the tangent vectors x_u and x_v are linearly independent at every point $(u,v) \in S$, and moreover the set of normal vectors $x_u \times x_v = N$ is a non-vanishing vector field on a regular surface S, everywhere perpendicular to S.

[1] University of Economics in Prague, Department of Mathematics, Ekonomická 957, 148 00 Prague, Czech Republic, kanka@vse.cz.

A regular surface is a surface $S \subset \mathbb{R}^3$ the atlas of which consists of regular maps.

In this paper the notion of a map will always mean a regular map, and we are to study smooth surfaces the atlas of which consists of regular maps. The basic tool for our study is the shape operator defined as follows, however, let us note that other techniques of differential geometry could in principle be successfully applied, such as Cartan's moving frames in [1] or [3], for example.

Definition 1. *Let $S \subset \mathbb{R}^3$ be a regular surface and let n be a surface unit normal to S defined in a neighbourhood of a point $p \in S$ ($p = x(u_0, v_0)$). For a tangent vector v_p to S we define*

$$\varphi(v_p) = -D_v n, \tag{1}$$

where $D_v n$ is the derivative of the vector field n in the direction v. Then φ is called a shape operator.

Lemma 1. *Let $U \subset \mathbb{R}^2$ and $S \subset \mathbb{R}^3$. Let $x: U \to S$ be a regular map. Then*

$$\varphi(x_u) = -n_u \quad and \quad \varphi(x_v) = -n_v.$$

Proof. If we fix $v = v_0$ and define a curve $\alpha(u) = x(u, v_0)$, we have

$$\varphi(x_u) = \varphi\big(x'(u, v_0)\big) = -D_{\alpha'(u)} n = -n_u,$$

see [2, Lemma 11.26].

Analogously, if we fix $u = u_0$ and define a curve $\beta(v) = x(u_0, v)$, we obtain

$$\varphi(x_v) = \varphi\big(x'(u_0, v)\big) = -D_{\beta'(v)} n = -n_v.$$

Lemma 2. *At each point p of a regular surface $S \subset \mathbb{R}^3$ the shape operator $\varphi: T_p(S) \to T_p(S)$, where $T_p(S)$ is the tangent space to the surface S at the point p, is a linear map.*

Proof. We use [2, Lemma 11.19]. Let Y, Z be differentiable vector fields on an open subset U of R^n and v_p, w_p tangent vectors to R^n at $p \in U$. Then

$$D_{av+bw} Y = a D_v Y + b D_w Y,$$
$$v_p[Y \cdot Z] = D_v Y \cdot Z(p) + Y(p) D_v Z.$$

From the first equation it follows that φ is linear. Using the second equation, if $Y = Z = n$, if we differentiate the equation $n \cdot n = 1$, we obtain

$$0 = v_p(n \cdot n) = (2D_v n) \cdot n_p = -2\varphi(v_p) \cdot n_p,$$

for any tangent vector v_p. From this equation it follows that $\varphi(v_p)$ is perpendicular to n_p, therefore $\varphi(v_p) \in S_p$.

Remark 1. Let x be a regular injective map. The equations $n \cdot x_u = 0$, $n \cdot x_v = 0$ give

$$0 = (n \cdot x_u)_u = n_u \cdot x_u + n x_{uu} \Rightarrow -n_u \cdot x_u = n x_{uu},$$
$$0 = (n \cdot x_u)_v = n_v \cdot x_u + n x_{uv} \Rightarrow -n_v \cdot x_u = n x_{uv} = n x_{vu},$$
$$0 = (n \cdot x_v)_v = n_v \cdot x_v + n x_{vv} \Rightarrow -n_v \cdot x_v = n x_{vv}.$$

Remark 2. Let $x: U \to \mathbb{R}^3$ be a regular injective map. Let us denote

$$l_{11} = -n_u \cdot x_u = n x_{uu},$$
$$l_{12} = -n_v \cdot x_u = n x_{uv} = -n_u x_v, \tag{2}$$
$$l_{22} = -n_v \cdot x_v = n x_{vv}.$$

The coefficients l_{11}, l_{12}, l_{22}, are called the coefficients of the second fundamental form F_{II} of x,

$$F_{\mathrm{II}} = l_{11} \mathrm{d}u^2 + 2 l_{12} \mathrm{d}u \mathrm{d}v + l_{22} \mathrm{d}v^2.$$

If we denote $g_{11} = \|x_u\|^2$, $g_{12} = x_u \cdot x_v$, $g_{22} = \|x_v\|^2$, the first fundamental form F_{I} of x can be written in the form

$$F_{\mathrm{I}} = g_{11} \mathrm{d}u^2 + 2 g_{12} \mathrm{d}u \mathrm{d}v + g_{22} \mathrm{d}v^2.$$

Theorem 1. *Let $x: U \to \mathbb{R}^3$ be a regular injective map. Then the shape operator φ is given with respect to the basis $x_u, x_v \in T_{(u,v)}(S)$ in the form*

$$\begin{aligned}\varphi(x_u) &= \frac{g_{22}l_{11} - g_{12}l_{12}}{g_{11}g_{22} - g_{12}^2}x_u + \frac{g_{11}l_{12} - g_{12}l_{11}}{g_{11}g_{22} - g_{12}^2}x_v,\\ \varphi(x_v) &= \frac{g_{22}l_{12} - g_{12}l_{22}}{g_{11}g_{22} - g_{12}^2}x_u + \frac{g_{11}l_{22} - g_{12}l_{12}}{g_{11}g_{22} - g_{12}^2}x_v,\end{aligned} \tag{3}$$

where $g_{11}g_{22} - g_{12}^2 > 0$.

Proof. As x is a regular injective map and x_u, x_v are linearly independent, we have

$$\begin{aligned}\varphi(x_u) &= \alpha_{11}x_u + \alpha_{21}x_v = -n_u,\\ \varphi(x_v) &= \alpha_{12}x_u + \alpha_{22}x_v = -n_v,\end{aligned} \tag{4}$$

for some functions α_{11}, α_{12}, α_{21}, α_{22} that need to be computed. From (2) and (4) we have

$$\begin{aligned}l_{11} &= -n_u \cdot x_u = g_{11}\alpha_{11} + g_{12}\alpha_{21},\\ l_{12} &= -n_u \cdot x_v = g_{12}\alpha_{11} + g_{22}\alpha_{21},\\ l_{12} &= -n_v \cdot x_u = g_{11}\alpha_{12} + g_{12}\alpha_{22},\\ l_{22} &= -n_v \cdot x_v = g_{12}\alpha_{12} + g_{22}\alpha_{22}.\end{aligned} \tag{5}$$

Equations (5) can be written in the form

$$\begin{pmatrix} l_{11} & l_{12} \\ l_{12} & l_{22} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix}\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix},$$

or

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix}^{-1}\begin{pmatrix} l_{11} & l_{12} \\ l_{12} & l_{22} \end{pmatrix}.$$

So, we have

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \frac{1}{g_{11}g_{22} - g_{12}^2}\begin{pmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{pmatrix}\begin{pmatrix} l_{11} & l_{12} \\ l_{12} & l_{22} \end{pmatrix}.$$

From which immediately follows (3).

Remark 3. The shape operator can be represented by the matrix

$$A(\varphi) = \begin{pmatrix} \dfrac{g_{22}l_{11} - g_{12}l_{12}}{g_{11}g_{22} - g_{12}^2} & \dfrac{g_{22}l_{12} - g_{12}l_{22}}{g_{11}g_{22} - g_{12}^2} \\ \dfrac{g_{11}l_{12} - g_{12}l_{11}}{g_{11}g_{22} - g_{12}^2} & \dfrac{-g_{12}l_{12} + g_{11}l_{22}}{g_{11}g_{22} - g_{12}^2} \end{pmatrix}.$$

As the Gaussian curvature $K = \det A(\varphi)$ and the mean curvature $H = \frac{1}{2}\operatorname{tr} A(\varphi)$, we have

$$K = \frac{l_{11}l_{22} - l_{12}^2}{g_{11}g_{22} - g_{12}^2} \quad \text{and} \quad H = \frac{l_{11}g_{22} - 2l_{12}g_{12} + l_{22}g_{11}}{2(g_{11}g_{22} - g_{12}^2)}.$$

Example 1. In this example we study the general case of the Cobb-Douglas surface $x(u,v) = (u, v, u^\alpha \cdot v^\beta)$, where $u, v > 0$ and $\alpha, \beta > 0$ are real constants.

Solution. The basis of the tangent space can be written in the form

$$x_u = \left(1, 0, \alpha \cdot u^{\alpha-1}v^\beta\right), \quad x_v = \left(0, 1, \beta \cdot v^{\beta-1}u^\alpha\right).$$

The functions g_{11}, g_{12}, g_{22} have the form

$$g_{11} = 1 + \alpha^2 u^{2\alpha-2} \cdot v^{2\beta}, \quad g_{12} = \alpha \cdot \beta \cdot u^{2\alpha-1} \cdot v^{2\beta-1}, \quad g_{22} = 1 + \beta^2 v^{2\beta-2}u^{2\alpha}.$$

Then the unit normal is given by

$$n = \frac{(-\alpha u^{\alpha-1} \cdot v^\beta, -\beta v^{\beta-1} \cdot u^\alpha, 1)}{\sqrt{\lambda}}, \tag{6}$$

where $\lambda = \alpha^2 u^{2\alpha-2} \cdot v^{2\beta} + \beta^2 v^{2\beta-2}u^{2\alpha} + 1$. Hence, we have

$$x_{uu} = \left(0,0,\alpha(\alpha-1)u^{\alpha-2}\cdot v^{\beta}\right),$$
$$x_{uv} = \left(0,0,\alpha\beta u^{\alpha-1}\cdot v^{\beta-1}\right), \tag{7}$$
$$x_{vv} = \left(0,0,\beta(\beta-1)v^{\beta-2}\cdot u^{\alpha}\right).$$

Further, (2), (6), (7) yield that the functions l_{11}, l_{12}, l_{22} can be written in the form

$$l_{11} = \frac{\alpha(\alpha-1)u^{\alpha-2}\cdot v^{\beta}}{\sqrt{\lambda}},$$
$$l_{12} = \frac{\alpha\beta u^{\alpha-1}\cdot v^{\beta-1}}{\sqrt{\lambda}},$$
$$l_{22} = \frac{\beta(\beta-1)v^{\beta-2}\cdot u^{\alpha}}{\sqrt{\lambda}}.$$

Finally, the Gaussian curvature is given by the formula

$$K = \frac{\alpha\beta\cdot[1-(\alpha+\beta)]\cdot u^{2\alpha-2}v^{2\beta-2}}{\left(\alpha^2u^{2\alpha-2}v^{2\beta}+\beta^2u^{2\alpha}v^{2\beta-2}+1\right)^2}. \tag{8}$$

2 Conclusion

From formula (8) it can be easily seen that the following implications are true:

- as $u,v>0$, then $\alpha+\beta>1 \Leftrightarrow K<0$, what means that the returns to scale are increasing and every point (u,v) of the Cobb-Douglas surface is hyperbolic and the principal curvatures k_1 and k_2 have opposite signs, which are the roots of the equation
$$k^2-2Hk+K=0,$$
that is, $k_{1,2}=H\pm\sqrt{H^2-K^2}$, $k_1k_2=K$, $\frac{k_1+k_2}{2}=H$;
- as $u,v>0$, then $\alpha+\beta=1 \Leftrightarrow K=0$, what means that the returns to scale are constant and every point (u,v) of the Cobb-Douglas surface is parabolic and at least one of the principal curvatures is zero;
- as $u,v>0$, then $\alpha+\beta<1 \Leftrightarrow K>0$,what means that the returns to scale are decreasing and every point (u,v) is elliptic and the principal curvatures k_1 and k_2 have the same sign.

It is a useful observation that for these special Cobb-Douglas surfaces the sign of their Gaussian curvature determines the monotonicity of the returns to scale and vice versa. Lastly, let us note that is interesting that a purely mathematical notion such as the Gaussian curvature is so closely related to a purely economic notion such as the returns to scale.

References

[1] Bureš, J., and Kaňka, M.: Some conditions for a surface in E^4 to be part of the sphere S^2. *Mathematica Bohemica* **4** (1994), 367—371.

[2] Gray, A.: *Modern differential geometry of curves and surfaces with Mathematica. Second Edition.* Chapman and Hall / CRC, New York, 1998.

[3] Kaňka, M.: Example of basic structure equations of Riemannian manifolds. *Mundus Symbolicus* **3** (1995), 57—62.

[4] Kobayashi, S., and Nomizu, K.: *Foundations of differential geometry*. Wiley-Interscience, New York, 1996.

[5] Soukopová, J., Hořejší, B., Macáková, L., and Soukup, J.: *Mikroekonomie, 3. doplněné vydání.* Management Press, Praha, 2003.

[6] Sternberg, S.: *Lectures on differential geometry*. AMS Chelsea Publishing, New York, 1999.

[7] Varian Hal, R., and Grega, L.: *Mikroekonomie, 1. vydání.* Victoria Publishing, Praha, 1995.

Scenario Generation via $\mathcal{L}_1$ Norm

Vlasta Kaňková[1]

Abstract. Optimization problems depending on a probability measure correspond to many economic and financial situations. It can be very complicated to solve these problems, especially when the "underlying" probability measure belongs to a continuous type. Consequently, the "underlying" continuous probability measure is often replaced by discrete one with finite number of atoms (scenario). The aim of the contribution is to deal with the above mentioned approximation in a special form of stochastic optimization problems with an operator of the mathematical expectation in the objective function.

The stability results determined by the help of the Wasserstein metric (based on the $\mathcal{L}_1$ norm) are employed to generate approximate distributions.

Keywords: One–stage stochastic programming problems, multistage stochastic problems, stability, Lipschitz property, $\mathcal{L}_1$ norm, Wasserstein metric, scenario generation, approximation error

JEL classification: C44
AMS classification: 90C15

1 Introduction

Let $(\Omega, \mathcal{S}, P)$ be a probability space, $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ an s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$, $F := F_\xi(z)$ the distribution function of ξ, P_F and Z_F the probability measure and the support corresponding to F, respectively. Let, moreover, $g_0 := g_0(x, z)$ be a real–valued function defined on $R^n \times R^s$, $X_F \subset X \subset R^n$ a nonempty set generally depending on F and $X \subset R^n$ a nonempty "deterministic" set. If E_F denotes the operator of mathematical expectation corresponding to F and if for an $x \in X$ there exists finite $\mathsf{E}_F g_0(x, \xi)$, then rather general one-stage (static) "classical" stochastic optimization problem can be introduced in the form:

$$\text{to find} \quad \varphi(F, X_F) = \inf\big\{\mathsf{E}_F g_0(x, \xi) : x \in X_F\big\}. \tag{1}$$

The objective function in Problem (1) depends linearly on the probability measure P_F. We shall try to include in our consideration also a little relax problems. In particular, we consider problems that can be covered by the following type:

$$\text{to find} \quad \bar{\varphi}(F, X_F) = \inf\big\{\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi)) : x \in X_F\big\}, \tag{2}$$

where $h := h(x, z)$ is an m_1–dimensional vector function defined on $R^n \times R^s$, $h = (h_1, \ldots, h_{m_1})$; $\bar{g}_0 := \bar{g}_0(x, z, y)$ is a real–valued function defined $\mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^{m_1}$.

Remark 1.
- The type of Problems (2) has begun recently to appear rather often in the literature (see, e. g., Ermoliev and Norkin [**?**]). Problem (2) covers Problem (1) with $\bar{g}_0(x, z, y) := g_0(x, z)$.
- Some problems from the class "Mean-Risk" can be covered by the type (2) (see, e. g., [**?**], [**?**], [**?**], [**?**]).

If $g_1 := g_1(x, z)$ is a real–valued function defined on $R^n \times R^s$, $g_2 := g_2(y, x, z)$ a real–valued function defined on $R^{n_1} \times R^n \times R^s$; $\mathcal{K}(x, z)$, for every $x \in X$, a measurable multifunction defined on $R^n \times R^s$, and

$$g_0(x, z) = g_1(x, z) + \inf\{g_2(y, x, z) : y \in \mathcal{K}(x, z)\}, \tag{3}$$

[1]Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: kankova@utia.cas.cz

then (1) is a two–stage stochastic programming problem (for more details see, e.g., [**?**] or [**?**]). However see also [**?**] to recognize that the well–known and often employed risk measure CVaR can be reformulated in the form of simple recourse problem and consequently in two–stage stochastic programming problem.

Two–stage stochastic programming problems correspond to applications in which it is necessary first to determine x on a base of the knowledge P_F only and, after the realization of the random element ξ, it is possible to correct the decision and to determine y.

The two–stage stochastic programming problems can be generalized to the multistage case. There are known a few types of different definitions of the multistage stochastic programming problems. We recall $(M+1)$–stage stochastic programming problem as the problem:

Find

$$\varphi_{\mathcal{F}}(M) = \inf\{\mathsf{E}_{F^{\xi^0}} g^0_{\mathcal{F}}(x^0,\, \xi^0)|\ x^0 \in \mathcal{K}^0\}, \tag{4}$$

where the function $g^0_{\mathcal{F}}(x^0,\, z^0)$ is defined recursively

$$\begin{aligned} g^k_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k) \quad &= \quad \inf\{\mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k}}\ g^{k+1}_{\mathcal{F}}(\bar{x}^{k+1},\, \bar{\xi}^{k+1})\,|x^{k+1} \in \mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k)\}, \\ k &= 0,\, 1,\, \ldots,\, M-1, \\ g^M_{\mathcal{F}}(\bar{x}^M,\, \bar{z}^M) \quad &:= \quad g^M_0(\bar{x}^M,\, \bar{z}^M), \quad \mathcal{K}_0 := X^0. \end{aligned} \tag{5}$$

$\xi^j := \xi^j(\omega)$, $j = 0,\, 1,\, \ldots,\, M$ denotes an s–dimensional random vector defined on a probability space $(\Omega,\, \mathcal{S},\, P)$; $F^{\xi^j}(z^j)$, $z^j \in R^s$, $j = 0,\, 1\, \ldots,\, M$ the distribution function of the ξ^j and $F^{\xi^k|\bar{\xi}^{k-1}}(z^k|\bar{z}^{k-1})$, $z^k \in R^s$, $\bar{z}^{k-1} \in R^{(k-1)s}$, $k = 1,\, \ldots,\, M$ the conditional distribution function (ξ^k conditioned by $\bar{\xi}^{k-1}$); $P_{F^{\xi^j}}$, $P_{F^{\xi^{k+1}|\bar{\xi}^k}}$, $j = 0,\, 1,\, \ldots,\, M$, $k = 0,\, 1,\, \ldots,\, M-1$ the corresponding probability measures; $Z^j := Z_{F^{\xi^j}} \subset R^s$, $j = 0,\, 1,\, \ldots,\, M$ the support of the probability measure $P_{F^{\xi^j}}$. Furthermore, the symbol $g^M_0 := g^M_0(\bar{x}^M,\, \bar{z}^M)$ denotes a continuous function defined on $R^{n(M+1)} \times R^{s(M+1)}$; $X^k \subset R^n$, $k = 0,\, 1,\, \ldots,\, M$ is a nonempty compact set; the symbol $\mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k) := \mathcal{K}^{k+1}_{F^{\xi^{k+1}|\bar{\xi}^k}}(\bar{x}^k,\, \bar{z}^k)$, $k = 0,\, 1,\, \ldots,\, M-1$ denotes a measurable multifunction defined on $R^{n(k+1)} \times R^{s(k+1)}$ with "values" subsets of R^n. $\bar{\xi}^k(:= \bar{\xi}^k(\omega)) = [\xi^0,\, \ldots,\, \xi^k]$; $\bar{z}^k = [z^0,\, \ldots,\, z^k]$, $z^j \in R^s$; $\bar{x}^k = [x^0,\, \ldots,\, x^k]$, $x^j \in R^n$; $\bar{X}^k = X^0 \times X^1 \ldots \times X^k$; $\bar{Z}^k := \bar{Z}^k_{\mathcal{F}} = Z_{F^{\xi^0}} \times Z_{F^{\xi^1}} \ldots \times Z_{F^{\xi^k}}$, $j = 0,\, 1,\, \ldots, k$, $k = 0,\, 1,\, \ldots,\, M$. Symbols $\mathsf{E}_{F^{\xi^0}}$, $\mathsf{E}_{F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k}}$, $k = 0,\, 1,\, \ldots,\, M-1$ denote the operators of mathematical expectation corresponding to F^{ξ^0}, $F^{\xi^{k+1}|\bar{\xi}^k=\bar{z}^k}$, $k = 0,\, \ldots,\, M-1$.

We have introduced three types of the stochastic optimization problems. The aim of the contribution is to suggest an approximate solution based on approximation of the continuous distributions by discrete one with finite number of atoms. To this end we employ the approach suggested in [**?**]. Furthermore, we generalize the former results in the case of "empirical" estimation approximation to the case of distributions with heavy tails. To this end we employ the stability results based on the Wasserstein metric with the "underlying" $\mathcal{L}_1$ norm.

2 Some Definitions and Some Assertions

First, we recall a few definitions and auxiliary assertions. To recall the first auxiliary assertion let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures on R^s and let the system $\mathcal{M}^1_1(R^s)$ be defined by the relation:

$$M^1_1(R^s) := \left\{\nu \in \mathcal{P}(R^s) : \int_{R^s} \|z\|_1 d\nu(z) < \infty\right\}, \quad \|\cdot\|_1 \quad \text{denotes} \quad \mathcal{L}_1 \quad \text{norm in} \quad R^s. \tag{6}$$

We introduce the system of the assumptions:

A.1
- $g_0(x, z)$ is either a uniformly continuous function on $X \times \mathbb{R}^s$, or X is a bounded convex set and there exists $\varepsilon > 0$ such that $g_0(x,\, z)$ is a convex on $X(\varepsilon)$ and bounded on $X(\varepsilon) \times Z_F$ ($X(\varepsilon)$ denotes the ε–neighborhood of the set$X$);
- $g_0(x, z)$ is for $x \in X$ a Lipschitz function of $z \in \mathbb{R}^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x;

B.1 $P_F, P_G \in \mathcal{M}_1^1(\mathbb{R}^s)$ and there exists $\varepsilon > 0$ such that

- $\bar{g}_0(x, z, y)$ is for $x \in X(\varepsilon), z \in R^s$ a Lipschitz function of $y \in Y(\varepsilon)$ with a Lipschitz constant L_y where $Y(\varepsilon) = \{y \in R^{m_1} : y = h(x, z) \quad \text{for some} \quad x \in X(\varepsilon),\ z \in \mathbb{R}^s\}$ and $\mathsf{E}_F h(x, \xi)$, $\mathsf{E}_G h(x, \xi) \in Y(\varepsilon)$;
- for every $x \in X(\varepsilon),\ y \in Y(\varepsilon)$ there exist finite mathematical expectations $\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi))$, $\mathsf{E}_F g_0^1(x, \xi, \mathsf{E}_G h(x, \xi))$, $\mathsf{E}_G g_0^1(x, \xi, \mathsf{E}_F h(x, \xi))$, and $\mathsf{E}_G g_0^1(x, \xi, \mathsf{E}_G h(x, \xi))$;
- $h_i(x, z), i = 1, \ldots, m_1$ are for every $x \in X(\varepsilon)$ Lipschitz functions of z with the Lipschitz constants L_h^i (corresponding to $\mathcal{L}_1$ norm),
- $\bar{g}_0(x, z, y)$ is for every $x \in X(\varepsilon), y \in Y(\varepsilon)$ a Lipschitz function of $z \in R^s$ with the Lipschitz constant $L_z(x,\, y)$ (corresponding to $\mathcal{L}_1$ norm),
- $\bar{g}_0(x,\, z,\, y)$ is for every $x \in X,\ z \in R^s$ a Lipschitz function of $y \in Y$ with the Lipschitz constant $L^y(x,\, z)$ corresponding to $\mathcal{L}_1$ norm;

B.2 $\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi))$, $\mathsf{E}_G \bar{g}_0(x, \xi, \mathsf{E}_G h(x, \xi))$ are continuous functions on X.

Proposition 1 ([**?**], [**?**]). *Let $P_F, P_G \in \mathcal{M}_1^1(R^s)$ and let X be a compact set. If*

1. *Assumption A.1 is fulfilled, then*

$$|\varphi(F,\, X) - \varphi(G,\, X)| \le L \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i, \tag{7}$$

2. *Assumptions B.1, B.2 are fulfilled, then there exist $\hat{C} > 0$ such that*

$$|\bar{\varphi}(F,\, X) - \bar{\varphi}(G,\, X)| \le \hat{C} \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{8}$$

The constant $\hat{C}$ can be estimated by the following relation

$$\hat{C} \le \mathsf{E}_F[L^y(x,\, \xi)] \sum_{i=1}^{s} L_h^i + L^z(x,\, \mathsf{E}_G(h,\, \xi)).$$

Proposition 1 reduces (from the mathematical point of view) an s–dimensional case to one–dimensional. We employ this fact to define atoms of the discrete approximate distribution functions. Of course a stochastic dependence between components of ξ is neglected by this approach. (The idea to reduce an s–dimensional case, $s > 1$ to one dimensional case is credited to G. Pflug [**?**] (see also Šmíd [**?**]).)

Evidently, if we approximate the continuous (w.r.t. Lebesque measure) probability measure P_F by a discrete one with the finite number of atoms, we obtain mostly (from the numerical point of view) more "pleasant" problem.

3 Approximation

3.1 Deterministic Case

To construct first discrete approach we introduce the following assumption:

A.2 $P_{F_i}, i = 1, \ldots, s$ are absolutely continuous w. r. t. the Lebesgue measure on R^1,
($F_i,\ i = 1,\ \ldots,\ s$ are one–dimensional marginal distribution functions corresponding to F.)

Evidently, if A.2 is fulfilled, then for given $M_i,\ \bar{M}_i > 0, i = 1, \ldots, s$ there exist natural numbers $m_i,\ \bar{m}_i,\ i = 1,\ \ldots,\ s$ and points $z_{i,\,j},\ \bar{z}_{i,\,k}, \in \bar{R}^1,\ j = 0,\ 1,\ \ldots,\ m_i,\ k = 1,\ \ldots,\ \bar{m}_i$ such that

$$-\infty = z_{i,0} < z_{i,1} < z_{i,2} < \ldots < z_{i,m_{i-1}} < z_{i,m_i} = \infty,$$

$$-\infty = \bar{z}_{i,0} < \bar{z}_{i,1} < \bar{z}_{i,2} < \ldots < \bar{z}_{i,\bar{m}_{i-1}} < \bar{z}_{i,\bar{m}_i} = \infty$$

and, simultaneously,

$$(L/s) \int_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)|dz_i \leq M_i,\ i = 1, \ldots, s,$$

$$(\hat{C}/s) \int_{-\infty}^{\infty} |F_i(z_i) - \bar{G}_i(z_i)|dz_i \leq \bar{M}_i,\ i = 1, \ldots, s,$$

where G_i, $\bar{G}_i$, $i = 1, \ldots, s$ are one dimensional discrete distribution functions with atoms in points $z_{i,j}, \bar{z}_{i,j},\ j = 1, \ldots, m_i,\ j = 1, \ldots, \bar{m}_i$, respectively.

Furthermore, it follows from the last relations that for every $M > 0$, $\bar{M} > 0$ there exist s–dimensional distribution functions G, $\bar{G}$ with marginals G_i, $\bar{G}_i$, $i = 1, \ldots, s$ such that

$$L \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)|dz_i \leq M, \tag{9}$$

$$\hat{C} \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - \bar{G}_i(z_i)|dz_i \leq \bar{M}. \tag{10}$$

We have proven the assertion

Proposition 2. *Let Assumption A.2 be fulfilled. Let moreover $M, \bar{M} > 0$. If*

1. *Assumption A.1 is fulfilled, then there exists a discrete distribution function G with discrete marginals G_i, $i = 1, \ldots, s$ such that*
$$|\varphi(F, X) - \varphi(G, X)| \leq M, \tag{11}$$
2. *Assumptions B.1, B.2 are fulfilled, then there exists a discrete distribution function $\bar{G}$ with discrete marginals $\bar{G}_i$, $i = 1, \ldots, s$ such that*
$$|\bar{\varphi}(F, X) - \bar{\varphi}(\bar{G}, X)| \leq \bar{M}. \tag{12}$$

A possibility to employ the above mentioned approach (in the case of Problem (1)) it is necessary to assume that $g_0(x, z)$ is a Lipschits function of z with the Lipschitz constant not depending on $x \in X$. It means, in the case of problem (3): if the function $g_1(x, z)$ fulfills this assumption, then a question arises if also

$$\inf\{g_2(y, x, z) : y \in \mathcal{K}(x, z)\} \tag{13}$$

fulfills this condition. To this end we consider two cases, separately. If (13) is a problem of linear programming, then the corresponding assertion can be found, e.g, in [**?**]. In the general nonlinear case we can find the corresponding assertion, e.g., in [**?**].

In the multistage case, we restrict to the case when the following assumption is fulfilled:

C.1 Random sequence $\{\xi^k\}_{k=-\infty}^{\infty}$ follows (generally) nonlinear autoregressive sequence

$$\xi^k = H(\xi^{k-1}\ \varepsilon^k),$$

where ξ^0, ε^k, $k = 1, 2, \ldots$ are stochastically independent s–dimensional random vectors defined on $(\Omega, \mathcal{S}, P)$ and, moreover, ε^k, $k = 1, \ldots$ identically distributed. $H = (H_1, \ldots, H_s)$ is a Lipschitz vector function defined on R^s. We denote the distribution function corresponding to $\varepsilon^1 = (\varepsilon_1^1, \ldots, \varepsilon_s^1)$ by the symbol F^{ε} and suppose the realization ξ^0 to be known.

Evidently, the multistage stochastic programming problem (4), (5) depends essentially on a system of (generally) conditional distribution functions

$$\mathcal{F} = \{F^{\xi^0}(z^0),\, F^{\xi^k|\bar{\xi}^{k-1}}(z^k|\bar{z}^{k-1}),\, k = 1,\, \ldots,\, M\}. \tag{14}$$

Consequently, if we replace $\mathcal{F}$ by another system $\mathcal{G}$

$$\mathcal{G} = \{G^{\xi^0}(z^0),\, G^{\xi^k|\bar{\xi}^{k-1}}(z^k|\bar{z}^{k-1}),\, k = 1,\, \ldots,\, M\}, \tag{15}$$

we obtain another multistage stochastic programming problem with the optimal value denoted $\varphi_{\mathcal{G}}(M)$.

Under Assumption C.1 the system $\mathcal{F}$ is determined by F^{ξ^0} and F^{ε}. Consequently, if we replace these two probability distribution functions by another G^{ξ^0} and G^{ε}, we obtain another system $\mathcal{G}$.

Considering, furthermore, the constraint sets $\mathcal{K}^{k+1}_{\mathcal{F}}(\bar{x}^k,\, \bar{z}^k)$, $k = 0,\, \ldots,\, M-1$ not depending on the probability measure, then the assumptions under which

$$|\varphi_{\mathcal{F}}(M) - \varphi_{\mathcal{G}}(M)| \leq \sum_{i=1}^{s} C^i_W \int\limits_{R^1} |F^{\varepsilon}_i(z_i) - G^{\varepsilon}_i(z_i)|dz_i$$

can be found in [**?**]. Consequently, if we define discrete distributions $G^{\xi_0}, G^{\varepsilon}_i$ $i = 1,\, \ldots,\, s$ determined by the approach of Proposition 2, then we have an approximating system $\mathcal{G}$ given by discrete mostly conditional distributional functions.

Furthermore, it follows from results of the above mentioned work that this approach can be generalized to the case when constraints sets are given by the individual probability constraints.

3.2 Empirical Estimates Case

We introduce the next assumptions:

A.3
- $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F,
- F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^{N}, N = 1, 2, \ldots$.

Proposition 3. *[**?**], [**?**] Let $P_F \in \mathcal{M}^1_1(R^s)$, X be a compact set. Let, moreover, Assumption A.3 be fulfilled. If*

1. *Assumption A.1 is fulfilled, then*

$$P\{\omega||\varphi(F^N, X) - \varphi(F, X)| \xrightarrow[N\to\infty]{} 0\} = 1, \tag{16}$$

1. *Assumptions B.1, B.2 are fulfilled, then*

$$P\{\omega||\bar{\varphi}(F^N, X) - \bar{\varphi}(F, X)| \xrightarrow[N\to\infty]{} 0\} = 1. \tag{17}$$

According to Proposition 3 we can see that $\varphi(F^N, X)$, $\bar{\varphi}(F^N, X)$ are, in the case of the "underlying" distributions F with finite first moments (under some additional assumptions) consistent estimates of $\varphi(F, X)$, $\bar{\varphi}(F, X)$. It means that these estimates are consistent also in the case of the heavy tailed distributions (including the stable distributions, for the definition of stable distribution see, e.g. [**?**]) if there exists first absolute moments.

Proposition 4. *[**?**] Let $P_F \in \mathcal{M}^1_1(R^s)$, $t > 0$, X be a compact set and for some $r > 2$ it holds that $\mathsf{E}_{F_i}|\xi_i|^r < +\infty, \quad i = 1, \ldots s$. Let, moreover, the constant γ fulfil the inequalities $0 < \gamma < 1/2 - 1/r$. If*

1. *Assumptions A.1, A.2, A.3 are fulfilled, then*

$$P\{\omega|\, N^{\gamma}|\varphi(F,\, X) - \varphi(F^N, X)| > t\} \xrightarrow[N\to\infty]{} 0. \tag{18}$$

1. *Assumptions B.1, B.2, A.2, A.3 are fulfilled, then*

$$P\{\omega|\, N^{\gamma}|\bar{\varphi}(F,\, X) - \bar{\varphi}(F^N, X)| > t\} \xrightarrow[N\to\infty]{} 0. \tag{19}$$

4 Conclusion

The paper deals with optimization problems depending on a probability measure. Especially, it is considered the case when the "underlying" probability measure is absolutely continuous with respect to the Lebesque measure and the aim of the paper is to suggest approximate problems with a discrete probability measure. Moreover, this approximation can be defined with respect to required error. The case of one–stage, two–stage and multistage cases are considered separately. However, problems with an "underlying" autoregressive random sequence are considered in the multistage case only.

A results for empirical estimates are recalled in the second part of the contribution.

Acknowledgment This work was supported by the Czech Science Foundation under grants 13-14445S, 15-10331S.

References

[1] Birge, J.R. and Louveaux, F.: *Introduction in Stochastic Programming.* Springer, Berlin 1999.

[2] Dupačová, J. and Kopa, M.: Robustness in stochastic programs with risk constraints. *Ann. Oper. Res.* **200** (2012), 55–74.

[3] Ermoliev, Y. M. and Norkin, V.: Sample everage approximation method for compound stochastic optimization problems. *SIAM J. Optim.* **23** (2013), 4, 2231–2263.

[4] M. Houda and V. Kaňková: Empirical estimates in economic and financial optimization problems. Bulletin of the Czech Econometric Society *19* (2012), 29, 50–69.

[5] Kall, P, and Mayer, J.: *Stochastic Linear Programming, Models, Theory and Computation.* Springer, New York 2005.

[6] Kaňková, V. and Šmíd, M.: On Approximation in Multistage Stochastic Programs; Markov Dependence. *Kybernetika* **40** (2004), 5, 625–638.

[7] Kaňková, V.: Convexity, Lipschitz property and differentiability in two–stage stochastic nonlinear programming problems. In: Proceedings of 3rd International Conference on Approximation and Optimization in the Caribean (B. Bank, J. Bustamante, J. Guddat, M. A. Jiménez, H. Th. Jongen and W. Romisch, eds.), Societed Matematica Maxicana 1998.

[8] Kaňková, V. and Houda, M.: Empirical estimates in stochastic programming. In: Proceedings of Prague Stochastics 2006 (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague 2006, 426–436.

[9] Kaňková, V.: Multistage stochastic programs via Autoregressive sequences and individual probability constraints. *Kybernetika* **44** (2008), 2, 151–170.

[10] Kaňková, V.: Empirical estimates in stochastic programming via distributions tails. *Kybernetika* **46** (2010), 3, 459–471.

[11] Kaňková, V.: Empirical estimates in optimization problems: survey with special regard to heavy tails and dependent samples. *Bulletin of the Czech Econometric Society* **19** (2012), 30, 92–111.

[12] Kaňková, V.: Risk measures in optimization problems via empirical estimates. *Czech Economic Review, Acta Universitatis Carolinea* **7** (2012), 3, 162–177.

[13] Kaňková, V.: A remark on multiobjective stochastic optimization in economic and financial problems. *Central European Journal of Operations Research* (submitted).

[14] Klebanov, L. B.: *Heavy Tailed Distributions.* MATFYZPRESS, Prague 2003.

[15] Mayer, J.: *Stochastic Linear Programming Algorithms: A Comparison Based on a Model Management System.* Gordon and Breach Science Publishers, Amsterdam 1998.

[16] Pflug, G. Ch.: Scenarion tree generation for multiperiod financial optimization by optimal discretization. *Math. Program. Ser. B.* **89** (2001), 251–271.

[17] Pflug, G. Ch. and Römisch, W.: *Modeling Measuring and Managing Risk.* World Scientific Publishing Co. Pte. Ltd, New Jersey 2007.

[18] Shapiro, A., Dentcheva, D. and Ruszczynski, A: *Lectures on Stochastic Programming (Modeling and Theory).* Published by Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia 2009.

[19] Šmíd, M.: The expected loss in the discretezation of multistage stochastic programming problems – estimation and convergence rate. *Ann. Oper. Res.* **165** (2009) 1, 29–45.

Bank Lending Determinants of EU Countries: A Bayesian model averaging evidence

Svatopluk Kapounek[1], Zuzana Kučerová[2]

Abstract.
The main objective of the paper is to identify the link between the European bank lending activities, lending standards, macroeconomic shocks and banking controls. The microeconomic data are provided by the Bankscope database, macroeconomic shocks are drawn from Eurostat and World Bank on-line databases. Our cross-sectional regression models consist up to 63 regressors which is too large to allow for direct estimation of each potential model, therefore we apply the Markov Chain Monte Carlo algorithm. The results confirm the significant impact of banking controls and market liquidity identified by total amount of central banks' assets. Surprisingly, there is no evidence of economic activity impact on lending behaviour.

Keywords: Banking industry, macroeconomic shocks, financial crisis, credit crunch.

JEL Classification: G2, C1
AMS Classification: 62C12

1 Introduction

It is generally agreed that lending activities tend to increase during the economic expansion and decrease during its contractions. The fall down of the lending activities was significantly observed during the financial crisis and consequences after 2007 which was caused by the decrease in investment demand and economic activities (Poměnková and Kapounek, [4]). However, the lending activities are not influenced only by demand side factors. Cuaresma, Fidrmuc and Hake [3] find that supply factors play a more significant role than the demand factors. Adams-Kane, Jia and Lim [1] contribute to the discussion and state that important bank lending determinants are changes in willingness to lend as a result of increased economic uncertainty, funding availability from interbank liquidity markets, and solvency from weakened balance sheets after the crisis.

Despite the wide discussions about the lending activity forces, the European economies are considered to be bank-based economies which are more reliant on bank credits and bank intermediation of savings, than the United States and the rest of the world. Chakraborty and Ray [2] study financial systems in a theoretical endogenous growth model. They state that that the level of investments and per capita GDP is higher and income inequality is lower under a bank-based system compared to market based system. It is mainly due to the fact that banks monitor the clients and it helps to solve agency problems and enables firms to borrow more. In market-based systems, the situation is slightly different and financial markets intermediate a lower amount of external finance to all firms.

The main objective of the paper is to identify the link between the European bank lending activities, lending standards, macroeconomic shocks and banking controls in the sample of EU countries. Thus, we expect significant impact of macroeconomics shocks (especially economic activity), market liquidity and banking controls (especially deposits) on the lending activities of the European banks. Moreover, we also focus on lending standard surveys (credit standards for approving loans and credit terms and conditions) and incorporate them into the analysis and discuss possible credit crunch across the European countries.

The structure of the paper is as follows. First section is introduction. Second section describes data and methods used in our paper. In the third section, results are presented. Fourth section brings conclusion.

[1] Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, Brno, Czech Republic, e-mail: kapounek@mendelu.cz.
[2] VSB-Technical University of Ostrava, Faculty of Economics, Dept. of Economic Policy, Sokolská tř. 33, Ostrava, Czech Republic, e-mail: zuzana.kucerova@vsb.cz.

2 Data and Methods

To identify determinants of the European banks' lending activities we used selected macroeconomic shocks, market specifics and various banking controls (the list of variables is presented in Table 1). The empirical analysis is based on the cross-sectional regression where net loans of bank $i = 1, \dots, N$ are regressed on an intercept α and number of explanatory variables selected from a set of k variables in a matrix X of dimension $N \times K$. Assume that rank $(\iota_N : X) = K + 1$, where ι_N is an N-dimensional vector of ones, and define β as the full k-dimensional vector of regression coefficients:

$$y = \alpha \iota_N + X_r \beta_r + \varepsilon, \tag{1}$$

where we assume $r = 1, \dots, R$ models, denoted by M_r and X_r is a $N \times k_r$ matrix containing (or all) columns of X. The N-vector of errors, ε, is assumed to be $N(0_N, h^{-1} I_T)$. Thus, $R = 2^K$ because there are 2^K possible subsets of X and 2^K possible choices for X_r (Koop, [8]).

We consider up to 63 regressors to be included in the model. That means 2^{63} different models to deal with which is far too many to evaluate. To solve this problem we applied Markov chain Monte Carlo techniques (MC^3) pioneered by Madigan and York [9]. The results are based on taking 1 100 000 draws and discarding the first 100 000 draws models as burn-in replications.

In the Bayesian framework we receive posterior model probabilities $p(M_r|y)$, for $r = 1, \dots, R$, where each model depends upon a vector of parameters θ_r and is characterized by prior $p(\theta_r|M_r)$ likelihood $p(y|\theta_r, M_r)$ and posterior $p(\theta_r|y, M_r)$. Assume vector of parameters ϕ which is function of θ_r for each of $r = 1, \dots, R$. Then we should obtain results for every model under consideration and average them where the weights in the averaging are the posterior model probabilities:

$$p(\phi \mid \mathrm{y}) = \sum_{r=1}^{R} p(\phi \mid \mathrm{y}, \mathrm{M_r}) p(\mathrm{M_r} \mid \mathrm{y}), \tag{2}$$

alternatively, if $g(\phi)$ is a function of ϕ, the rules of conditional expectation imply that

$$E[g(\phi) \mid \mathrm{y}] = \sum_{r=1}^{R} E[g(\phi) \mid \mathrm{y}, \mathrm{M_r}] p(\mathrm{M_r} \mid \mathrm{y}), \tag{3}$$

where $E[g(\phi) \mid \mathrm{y}, \mathrm{M_r}]$ and $p(\mathrm{M_r} \mid \mathrm{y})$ are calculated by posterior simulation (Koop, [8]).

To ensure that the noninformative prior for the intercept has the same implication for each model we followed approach applied by Fernandez et al. [7] who recommend the standardization of all explanatory variables by subtracting off their means. Due to the singularity of inversed matrix the generalized inverse, especially Moore-Penrose pseudoinverse concept was applied to compute regressions.

The dataset covers 4829 banks in 27 EU countries between the years 1998 and 2013. The microeconomic data are provided by the Bankscope database, macroeconomic data covering macroeconomic shocks are drawn from Eurostat on-line databases (Eurostat, [6]). We also use data from the ECB Bank Lending Survey (BLS) which are available since 2003 (ECB [5]). The dataset concerning bank lending standards works with the answers from this survey for selected Euro Area countries; banks in these countries provided quarterly information on credit standards for approving loans and credit terms and conditions. ECB (see ECB, [5]) uses the diffusion index based on the responses to questions concerning lending standards, i.e. the weighted difference between the share of banks reporting that lending standards have been tightened and the share of banks reporting that they have been eased. Positive values of the diffusion index indicate that a larger proportion of banks have tightened lending standards and *vice versa*.

3 Results

The main results of Bayes model averaging are presented in the Table 1. The first column provides information about the probability that the corresponding explanatory variable should be included in the model. The highlighted variables (using grey colour) are concerned being significant and important in the sample of all explanatory variables used in our sample and as such can be sorted and used for further analysis of the impact of these variables on the net loans of bank.

The results present the strong evidence of the impact of market liquidity represented by total assets of central banks and selected banking controls related particularly to loans, liquidity, equity and financial soundness indi-

cators. However, a few posterior means of balance sheet indicators are smaller than their standard deviations, namely capital funds/liabilities, liquid assets/funding, unreserved impaired loans/equity, equity/liabilities or capital funds/liabilities etc. Obviously, these indicators do not provide clear explanatory power.

	Explanatory Variable	BMA Post. Prob.	Post. Mean	Post. St. Dev.
1	Total Assets EUR (ln)	1.000	0.971	0.009
2	Loan Loss Reserve/Gross Loans (%)	0.058	0.000	0.001
3	Total Capital Ratio (%)	0.048	0.000	0.000
4	Capital Funds/Liabilities (%)	0.311	0.000	0.001
5	Net Interest Margin (%)	0.017	0.000	0.000
6	Return on Average Assets (ROA) (%)	0.998	0.032	0.007
7	Return on Average Equity (ROE) (%)	1.000	-0.003	0.001
8	Cost to Income Ratio (%)	0.037	0.000	0.000
9	Liquid Assets/Cust & ST Funding (%)	0.517	0.000	0.000
10	Loan Loss Prov/Net Int Rev (%)	0.016	0.000	0.000
11	Impaired Loans/Gross Loans (%)	0.999	0.011	0.002
12	Loan Loss Res/Impaired Loans (%)	0.045	0.000	0.000
13	Impaired Loans/Equity (%)	0.827	-0.001	0.001
14	Unreserved Impaired Loans/Equity (%)	0.283	0.000	0.001
15	NCO/Average Gross Loans (%)	0.029	0.000	0.000
16	NCO/Net Inc Bef Ln Lss Prov (%)	0.025	0.000	0.000
17	Tier 1 Ratio (%)	0.016	0.000	0.000
18	Equity/Tot Assets (%)	0.232	0.001	0.002
19	Equity/Net Loans (%)	1.000	-0.001	0.000
20	Equity/Cust & Short Term Funding (%)	0.991	-0.001	0.000
21	Equity/Liabilities (%)	0.284	0.000	0.001
22	Cap Funds/Tot Assets (%)	0.135	-0.001	0.002
23	Cap Funds/Net Loans (%)	0.999	-0.001	0.000
24	Cap Funds/Dep & ST Funding (%)	0.983	0.001	0.000
25	Cap Funds/Liabilities (%)	0.341	0.000	0.001
26	Subord Debt/Cap Funds (%)	0.055	0.000	0.000
27	Net Int Rev/Avg Assets (%)	0.208	0.002	0.004
28	Oth Op Inc/Avg Assets (%)	0.018	0.000	0.000
29	Non Int Exp/Avg Assets (%)	0.020	0.000	0.000
30	Pre-Tax Op Inc/Avg Assets (%)	0.191	-0.002	0.004
31	Non Op Items & Taxes/Avg Ast (%)	0.034	0.000	0.003
32	Dividend Pay-Out (%)	0.078	0.000	0.000
33	Inc Net Of Dist/Avg Equity (%)	0.016	0.000	0.000
34	Non Op Items/Net Income (%)	0.049	0.000	0.000
35	Recurring Earning Power (%)	0.9534	-0.0216	0.0073
36	Interbank Ratio (%)	1.0000	0.0004	0.0001
37	Net Loans/Tot Assets (%)	1.0000	0.0316	0.0006
38	Net Loans/Dep & ST Funding (%)	0.9997	0.0007	0.0002
39	Net Loans/Tot Dep & Bor (%)	0.0307	0.0000	0.0001
40	Liquid Assets/Dep & ST Funding (%)	0.4887	-0.0004	0.0004
41	Liquid Assets/Tot Dep & Bor (%)	0.0337	0.0000	0.0001
42	Interest Income/Average Earning Assets (%)	1.0000	-0.0036	0.0008

43	Reserves for Impaired Loans/Gross loans (%)	0.0450	0.0002	0.0010
44	Reserves for Imp Loans/Imp Loans (%)	0.0363	0.0000	0.0001
45	Customer Dep/Total Fund excl Derivatives (%)	0.9943	0.0017	0.0005
46	Net Interest Income EUR (ln)	0.0000	0.0000	0.0000
47	Number of Employees (ln)	1.0000	0.0492	0.0095
48	GDP per capita (ln)	0.0409	0.0014	0.0154
49	Final Consumption per capita (ln)	0.0316	0.0002	0.0130
50	Gross Capital Formation per capita (ln)	0.0409	-0.0019	0.0208
51	Export of Goods per capita (ln)	0.1299	0.0071	0.0230
52	Import of Goods per capita (ln)	0.0991	0.0050	0.0178
53	Unemployment (%)	0.2492	-0.0025	0.0050
54	HICP (%)	0.0204	0.0563	0.8995
55	GDP price changes (%)	0.0239	-0.0661	0.8903
56	NEER changes (%)	0.0296	0.0879	1.0020
57	REER changes (%)	0.0273	0.0604	0.7375
58	Money Market Interest Rate (3months) (%)	0.0225	-0.0001	0.0012
59	Marginal Lending Rate (%)	0.0282	-0.0002	0.0019
60	Lending Standards Enterprises (ln)	0.1803	0.0017	0.0042
61	Lending Standards Households, HouseLoans (ln)	0.1241	-0.0010	0.0032
62	Lending Standards Consumer Credit (ln)	0.0423	-0.0003	0.0019
63	Central Bank Financial Assets (ln)	0.9525	-0.0266	0.0105

Table 1 Bayesian Model Averaging Results

Table 2 shows posterior model probabilities calculated by MLM function and MC3 algorithm. It is clear that there are really small differences between these results. It confirms the existence of the convergence of the MC3 algorithm and the correctness of the results in Table 1.

The mean number of regressors involved into the model is 22.4816. Therefore, Table 3 presents the results of the most preferred model with 23 regressors with the highest posterior probabilities. Unfortunately, the posterior probability associated with the model is only 0.1 (see Table 2). However, the comparison of Table 1 and Table 3 confirms that posterior means of the explanatory variables are roughly the same. Thus, the similarity of the results in Table 1 and 3 supports the relevancy of the Bayesian Model Averaging approach.

	$p(M_r\|y)$	$p(M_r\|y)$
	Analytical	MC^3 estimate
1	0.1000	0.0660
2	0.1000	0.0954
3	0.1000	0.0832
4	0.1000	0.1808
5	0.1000	0.1209
6	0.1000	0.2282
7	0.1000	0.0298
8	0.1000	0.0149
9	0.1000	0.0675
10	0.1000	0.1133

Table 2 Posterior Model Probabilities for Top 10 Models

The estimated posterior probabilities for the most preferred model (in Table 3) depicts the results and presents 23 variables selected using the MC3 algorithm (for initial results for all 63 variables see Table 1). In this step, we produce new posterior results (for the preferred model) where posterior means may but do not have to show positive or negative impact. Therefore, we cannot conclude that there is strong evidence of negative effects

of liquidity on the bank lending. We only identify probability that there is an impact of the corresponding explanatory variable on the lending activities.

	Explanatory Variable	BMA Post. Prob.	Posterior Mean	Posterior St. Dev.
1	Total Assets EUR (ln)	1.0000	0.9705	0.0081
2	Capital Funds/Liabilities (%)	0.2915	0.0002	0.0004
3	Return on Average Assets (ROA) (%)	1.0000	0.0331	0.0056
4	Return on Average Equity (ROE) (%)	1.0000	-0.0032	0.0006
5	Liquid Assets/Cust & ST Funding (%)	0.5412	-0.0004	0.0004
6	Impaired Loans/Gross Loans (%)	1.0000	0.0116	0.0020
7	Impaired Loans/Equity (%)	0.8584	-0.0010	0.0005
8	Unreserved Impaired Loans/Equity (%)	0.2611	-0.0003	0.0005
9	Equity/Net Loans (%)	1.0000	-0.0011	0.0002
10	Equity/Cust & Short Term Funding (%)	0.9891	-0.0010	0.0002
11	Equity/Liabilities (%)	0.1396	-0.0001	0.0002
12	Cap Funds/Net Loans (%)	0.9996	-0.0008	0.0002
13	Cap Funds/Dep & ST Funding (%)	0.9828	0.0010	0.0003
14	Cap Funds/Liabilities (%)	0.2884	0.0002	0.0004
15	Recurring Earning Power (%)	1.0000	-0.0229	0.0047
16	Interbank Ratio (%)	1.0000	0.0004	0.0001
17	Net Loans/Tot Assets (%)	1.0000	0.0316	0.0006
18	Net Loans/Dep & ST Funding (%)	0.9999	0.0007	0.0002
19	Liquid Assets/Dep & ST Funding (%)	0.4658	-0.0004	0.0004
20	Interest Income/Average Earning Assets (%)	1.0000	-0.0034	0.0007
21	Customer Dep/Total Fund excl Derivatives (%)	0.9946	0.0017	0.0004
22	Number of Employees (ln)	1.0000	0.0482	0.0090
23	Central Bank Financial Assets (ln)	0.9957	-0.0263	0.0070

Table 3 Posterior Results for the Preferred Model

Finally, we estimate posterior model probabilities for the most preferred variables (see Table 4). The results point out uncertainty regarding the identity of the true model with several models receiving non-trivial posterior weight. This implies that there are two dominating models which account for more than 57% or the total posterior model probability.

	$p(M_r\|y)$ Analytical	$p(M_r\|y)$ MC^3 estimate
1	0.1000	0.3137
2	0.1000	0.0060
3	0.1000	0.0887
4	0.1000	0.2649
5	0.1000	0.0546
6	0.1000	0.0136
7	0.1000	0.0313
8	0.1000	0.0968
9	0.1000	0.1242
10	0.1000	0.0063

Table 4 Posterior Model Probabilities for Top 10 Preferred Models

4 Conclusion

The objective of the paper was to identify the link between the European bank lending activities, lending standards, macroeconomic shocks and banking controls included lending standards in the sample of EU countries. Therefore, we estimated the impact of macroeconomics shocks, market liquidity and banking controls on the lending activities of the European banks.

Using a Bayesian framework, we found that lending activity of the European banks is determined by specific banking controls, particularly ROA, ROE, share of the impaired loans, equity, the share of capital funds, net loans or share of interest income. All these indicators demonstrate the financial soundness of banking institutions supervised by either central banks or other supervisory institutions (national or supranational, e.g. Bank for International Settlements, International Monetary Fund). Banking institutions in the EU are thus focused on keeping the level of these indicators at optimum levels and this behaviour has an impact on their lending activities.

Our empirical analysis uncovered that there is no significant impact of economic activity and lending standards which is quite surprising. Thus, lending activities of banking institutions are not determined by aggregate demand or investment activities even in times of financial crisis. Equally, there is no evidence of credit crunch represented by changes in lending standards in the model. Thus, policy authorities should be aware of the fact that loose economic policy is not ready to support lending activities of the analysed banking institutions in the EU. However, we confirmed the evidence of strong impact of central banks financial assets which represent the macroeconomic indicator of market liquidity. From the policy point of view, the unconventional monetary policy (represented by this indicator in our paper) could be recommended, especially quantitative easing as the factor influencing lending activities of the banks across the EU27.

Acknowledgements

The results introduced in the paper have been funded with support from the Czech Science Foundation via grant No. P403/14-28848S “Financial Crisis, Depreciation and Credit Crunch in CEECs”.

References

[1] Adams-Kane, J., Jia, Y., Lim, J., J.: Global Transmision Channels for International Bank Lending in the 2007–09 Financial Crisist. Forthcoming in *Journal of International Money and Finance* (2015).
[2] Chakraborty, S., and Ray, T.: Bank-Based versus Market-Based Financial Systems: A Growth-Theoretic Analysis. *Journal of Monetary Economics* **53** (2006), 329–350.
[3] Cuaresma, J., C., Fidrmuc, J., Hake, M.: Demand and Supply Drivers of Foreign Currency Loans in CEECs: A Meta-analysis. *Economic Systems* **38** (2014), 26–42.
[4] Poměnková, J., Kapounek, S., 2013: Heterogeneous distribution of money supply across the euro area. In: *Proceedings of the 10th International Scientific Conference Economic Policy in the European Union member Countries: Selected Papers* (Tvrdoň, M., and Majerová, I., eds.). Silesian University in Opava, Opava, 2013, 268–279.
[5] European Central Bank: *Statistical data warehouse: Bank lending survey*. [online database]. (2015), [cit. 2015-02-10]. Available at: https://sdw.ecb.europa.eu/browse.do?node=9484572.
[6] Eurostat: *European Statistics* [online database]. (2015), [cit. 2015-05-02]. Available from Internet: http://ec.europa.eu/eurostat/data/database.
[7] Fernandez, C., Ley, E., Steel, M.: Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* **16** (2001), 563–576.
[8] Koop, G.: *Bayesian Econometrics*. John Wiley and Sons, New York, 2003.
[9] Madigan, D., York, J.: Bayesian Graphical Models for Discrete Data. *International Statistical Review* **63** (1995), 942–963.

Mathematical model for sustainability assessment of agriculture farms with biogas plants

Edward Kasem, Oldřich Trenz, Jiří Hřebíček, Oldřich Faldik[1]

Abstract. Nowadays, the awareness of sustainability impacts of the agriculture farms with biogas plants rises. Current trends of sustainability assessment depend on more comprehensive sustainability dimensions; environmental, social, economic and governance performance which demonstrate the link between the organization's strategy and commitment to a sustainable global economy.
In this paper, we tried to provide an overview about the set of key performance indicators (KPIs) based on two international standard frameworks: the Global Reporting Initiative (GRI) and Sustainability Assessment of Food and Agriculture (SAFA). We also introduce the Data Envelopment Analysis (DEA) which is a very powerful tool for mathematical modeling of corporate performance evaluation. Our main goal is evaluating the level of efficiency which represents the level of corporate performance of farms with biogas plants. In this study, the input indicators are described as environmental, social and corporate governance indicators. In addition to the efficiency score, the weights of inputs are also provided. These weights are used to find the contribution of particular criteria to the achieved score of farms.

Keywords: agriculture, biogas plants, corporate performance, DEA, GRI, SAFA, sustainability assessment.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Biogas plant technology in agriculture sector is one of the promising environment-friendly technologies driven by purely energy-sustainable processes that don't contribute to the climate change [26]. Because of that the interest in renewable biomass and organic biogas is rising and it is also supported by the strategy of the European Union aiming at reduction of greenhouse gas (GHG) emissions. It can be used for heating and electricity generation and improves the security of energy supply. Organic biogas production combines renewable energy production and smart farming [8]. Both are important concepts regarding sustainable development [7]. There are many reasons for considering biogas production in smart farming [9], for example: renewable energy production, climate protection, closed nutrient cycles, optimizing crop rotation and cropping system, increased crop yields and quality, alternative sources of income and independent energy supply.

In our research, we considered a methodology which depends on optimization algorithms and used it in sustainability assessment. It consists of many different models like eco-efficiency models, multi-attribute, multi-criteria decision-making models [20] and Data Envelopment Analysis (DEA) models. DEA is very a powerful mathematical model of farm sustainability performance evaluation. It can be used as an appropriate method of decision-making in agriculture biogas farms with biogas plants. The main purpose of applying this analysis is evaluating the level of efficiency of chosen farms [27]. The efficiency in our case is represented as a share of output in a weighted sum of inputs. In other words, it depicts a certain degree to which desirable outputs can offset economic, environmental, social and corporate governance indicators. DEA analysis can also be described as a non-parametric methodology aimed at evaluating the relative efficiencies of comparable decision-making units (DMUs) by means of a variety of mathematical programming models [4, 22].

The main goal of this paper is corporate performance evaluation for farms with biogas plants in the South-Moravian region of the Czech Republic. According to key performance indicators (KPIs) which are chosen from different global frameworks and the DEA, the farms with biogas plants can be divided into efficient farms and non-efficient ones. We have to consider the Global Reporting Initiative (GRI) methodology [12, 13] and the Sustainability Assessment of Food and Agriculture (SAFA) [24], which are briefly described in Section 2. We also use one of the DEA models for performance assessment and implement this in Maple [23]. Section 3 depicts the relation between the farms score and simple environmental, social and governance (ESG) factors to develop

[1] Mendel University/ Faculty of Business and Economics, Department of Informatics, Zemědělská 1, 61300 Brno, Czech Republic, {edward.kasem, oldrich.trenz, jiri.hrebicek, oldrich.faldik}@mendelu.cz.

and compute pressure-specific composite indicators of corporate performance. Finally, Section 4 concludes the paper.

2 Materials and Methods

2.1 Framework for Sustainability Assessment

Global Reporting Initiative: GRI is the most common non-profit organization that focuses on developing a comprehensive, credible and transparent sustainability reporting framework regardless of the size, sector or location of the studied organization. The last version of this framework is the GRI G4 Guidelines [12, 13]. The G3.1 Guidelines [11] were the starting point for the G4 Guidelines, where some important changes were made to generalize sustainable reporting. The G4 Guidelines are presented in two parts: the Reporting Principles and the Standard Disclosures [12] "what must be reported", and the Implementation Manual [13] "how organizations can report against G4 Guidelines criteria". Their vision is a sustainable global economy where organizations manage their ESG and economic performance impacts responsibly, and report transparently. The *Standard Disclosures* guideline part [13] consists of *General Standard Disclosures* with reporting topics G4-I, I = 1, …, 58 and *Specific standard disclosures* which consists of the *Disclosures on Management Approach* (DMA) with G4-DMA topic and *Indicators and Aspects* specific DMAs, which are sorted into three categories: *Economic* with indicators G4-ECI, I = 1,…,9; *Environmental* with indicators G4-ENI, I = 1,…,34 and *Social category*, which is further divided into four sub-categories *Labor Practices and Decent Work* with indicators G4-LAI, I = 1,…,16; *Human Rights* with indicators G4-HRI, I = 1,…,12; *Society* with indicators G4-SOI, I = 1,…,11 and *Product Responsibility* with indicators G4-PRI, I = 1,…,9. The General Standard Disclosure introduced reporting topics: Strategy and Analysis (G4-1, G4-2); Organizational Profile (G4-9, …, G4-16); Identified Material Aspects and Boundaries (G4-17, …, G4-23); Stakeholder Engagement (G4-24, …, G4-27); Report profile (G4-28, …, G4-33); Governance (G4-34, …, G4-55); Ethics and Integrity (G4-56, G4-57, G4-58).

Sustainability assessment of food and agriculture system framework: The SAFA is a holistic global framework of the Food and Agriculture Organization of the United Nations (FAO) [8] designated for the assessment of sustainability along food and agriculture value chains [16, 24]. This framework can be applied to SMB/SMEs and large-scale companies and organizations. The SAFA framework seeks to harmonize the long-term objective of sustainable transformation of corporate agriculture and food systems, providing a transparent and aggregated framework designed for assessing corporate sustainability. Its structure is built upon a combination of various standards and frameworks like: ISO 14040:2006, the ISEAL Code of Good Practice, the Reference Tools of the Global Social Compliance Programme [10], the GRI G4 Sustainability Reporting Guidelines and its Food Sector Supplement [14]. The guiding vision of the SAFA framework is that food and agriculture systems worldwide are characterized by four dimensions of sustainability: good governance, environmental integrity, economic resilience and social well-being [15, 24]. The second level of the SAFA framework contains a set of 21 core sustainability issues or universal "themes". At this level, policy-makers and national governments can work towards establishment of a holistic scope of sustainability without defining the specific pathways. On the third level of the SAFA framework, each of the 21 sustainability themes is divided into sub-themes, or individual issues within SAFA themes. This level is composed of 58 sub-themes, is relevant for the supply chain actors doing an analysis, which identifies hot spot areas, as well as gaps in existing sustainability efforts.

2.2 Data Envelopment Analysis Model

In this section, we will highlight the modelling approach used in our study - Data Envelopment Analysis (DEA) which is a relatively new "data oriented" approach for evaluating the efficiency of a number of producers. In DEA the producers are usually referred to as the Decision Making Units (DMUs) which convert multiple inputs into multiple outputs. This approach refers to assessment of sustainability performance methodology [6] for farms with biogas plants which is based on the optimization algorithm. Relative efficiency is defined as a ratio of the total weighted output to the total weighted input. DEA can be used as a very powerful service management and benchmarking technique to evaluate nonprofit and public sector organizations [3, 6]. This analysis helps us in the performance assessment of different DMUs (farms with biogas plants) taking into consideration the observed quantities of marketable inputs and outputs. The advantage of this approach is that it provides the information about the sustainable management and efficiency of agricultural systems under different conditions at specific biogas stations. It also allows us to use a complex set of indicators for all the different sustainability dimensions [19, 20].

In this study the dual-role factors DEA model [5] is adopted. This model maximizes the dual-role efficiency of biogas plants according to the chosen output and dual core factors. The efficiency of this model can be described using Eq. 1.

$$Max\,\theta_k = \sum_{r=1}^{s} u_r y_{kr} + \gamma w_k - \beta w_k \tag{1}$$

Subject to:

$$\sum_{i=1}^{m} \upsilon_i x_{ik} = 1 \tag{2}$$

$$\sum_{r=1}^{s} u_r y_{kr} + \gamma w_k - \beta w_k - \sum_{i=1}^{m} \upsilon_i x_{ij} \le 0, j = 1,2,\ldots n \tag{3}$$

$u_r, \upsilon_i, \gamma, \beta \ge 0, \ \forall r, i$

where n is the number of evaluated $DMUs_j$ (j = 1, …, n). Each DMU consists of m inputs and s outputs with x_{ij} (i = 1, ..., m) and y_{rj} (r = 1, ..., s) values, respectively. w_j is a dual factor used in our assessment. u_r, v_i are the weights given to the output r and the input i, respectively. γ and β are the weights given to the dual-role factor. DMU_k consumes x_{ik} (i = 1, …, m), the amount of input i, to produce y_{rk}(r = 1, ..., s), the amount of output r. Let us consider the cross-efficiency evaluation, each DMU determines a set of input and output weights individually, leading to n sets of weights for n DMUs. The n sets of weights are used to assess the efficiencies of the n DMUs, resulting in n efficiency values for every DMU. After that, the efficiency values for each DMU are finally averaged as an overall efficiency value of the DMU [1]. u_r^*(r = 1, ..., s) and v_i^*(i = 1, ...,m) are the optimal solutions of Eq. 1. Then the dual-role efficiency of DMU_k is computed according to E.4.

$$\theta_k^* = \sum_{r=1}^{s} u_r^* y_{kr} + \gamma^* w_k - \beta^* w_k \tag{4}$$

This value is the best relative efficiency that DMU_k can achieve. Therefore a cross-efficiency value of DMU_j which reflects the peer evaluation of DMU_k to DMU_j (j = 1, ..., n; $j \neq k$) is calculated according to Eq. 5.

$$\theta_j = \frac{\sum_{r=1}^{s} u_r^* y_{rj}}{\sum_{i=1}^{m} \upsilon_i^* x_{ij}} \tag{5}$$

Where DMU_k is target DMU as presented in [27]. The above mentioned model is solved n times for each target DMU (given farm) using the Maple program [23]. As a result, there are n sets of input and output weights for n DMUs and each DMU will have one dual-role efficiency value and ($n-1$) cross-efficiency values. The n efficiency values are then averaged as the overall performance of the DMU. Based on their average cross-efficiency values, DMUs then can be compared or ranked.

3 Results

In order to determine the efficiency of farms with biogas plants depending on the sustainability key performance indicators (KPIs), we tried to cover all sustainability dimensions by relating them to the available GRI and SAFA frameworks. However, the limitation in data availability made us consider only a few performance indicators in each dimension. Without the loss of the overall picture, in order to make a performance assessment, a few KPIs are chosen and summarized in Table 1. The main output indicator EVA is calculated from Eq. 6, using the annual economic farm reports [2].

$$EVA = (ROE - r_e).E \tag{6}$$

where E describes the market value of the farm's equity, Return On Equity (ROE)=Net Income/E and r_e the alternative equity costs.

Our study involved 17 (A, B, C, …, Q) South Moravian farms with biogas facilities. To calculate their efficiency, we considered the KPIs: the economic indicators, EC1and EC2, as organizations' outputs and others, ENV1, ENV2, SO1, GOV1, as organizations' inputs [3], because there is a gradual shift towards the performance business assessment through the economic value added (EVA) [18] and cash flow indicators [25]. In addition, SO2 is considered a dual-role factor. These core indicators are relating to measuring ESG and economic activities. For the input and output dataset, mainly environmental, financial and sustainability reports of farms were used.

Indicator	Description	Unit
EC1	Economic Value Added (EVA)	CZK
EC2	Cash flow	CZK
ENV1	The amount of hazardous waste	tons
ENV2	The amount of other waste	tons
SO1	The number of employees.	num
SO2	The average employees salary and bonus	CZK
GOV1	Percentage of women in management of the company	%

Table 1 Core indicators

We applied the model described above and computed the efficiency score of selected organizations (DMUs) using the linear programing in Maple [17], where the Optimization [LPSolve] package for solving linear programming problems was applied. We used the LPSolve (obj, constr, bd, opts) procedure [23] for finding the optimum of various types of multivariate functions with various types of constraints. It solves a linear programming problem, which involves computing the minimum (or maximum) of a linear objective function subject to linear constraints.

Table 2 identifies the set of farms as A, C, …, Q with their efficiency scores. Looking at this table, we can distinguish between efficient and non-efficient farms with biogas plants. Figure 1 depicts two types of curves. The red one describes the threshold of sustainability performance for these farms, whereas the blue curve presents the efficiency score of each studied farm in the South Moravian region of the Czech Republic.

Company	Score	v_{ENV2}	v_{SO1}	v_{GOV1}	v_{ENV1}
A	0.71	0.03	0.04	0	0
B	0.95	0	0.01	0	0
C	1	0	0.01	0	0
D	1	0	0.01	0	0
E	0.51	0.01	0.01	0	0
F	0.57	0	0.01	0	0
G	0.65	0	0.01	0	0
H	1	0	0.14	0	0
I	0.69	0	0.03	0	0
J	0.89	0	0.01	0	0
K	1	0.02	0.03	0	0
L	1	0	0.14	0	0
M	0.87	0	0.01	0	1.3
N	0.93	0.01	0.03	0	0
O	0.93	0	0.01	0	0.63
P	0.6	0.01	0.13	0	0
Q	0.9	0	0.01	0	0

Table 2 Efficiency score and input weights

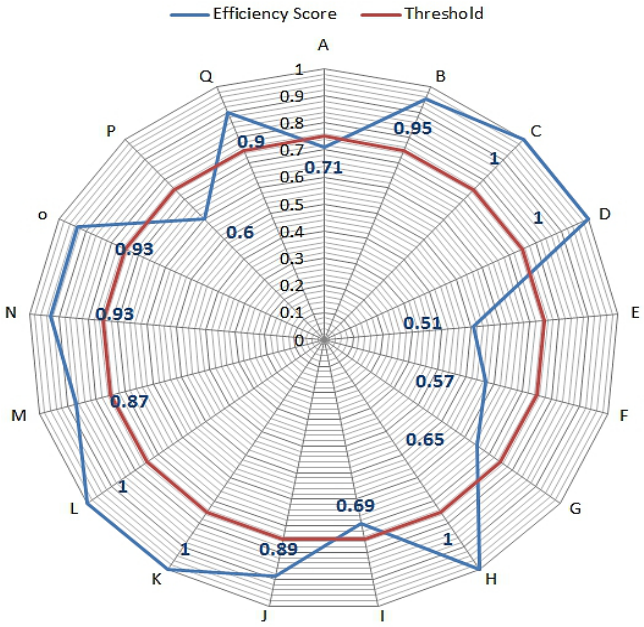


Figure 1 The level of efficiency of the studied biogas farms

These results were achieved during sustainability performance evaluation using the DEA model. The set of organizations C, D, H, K and L are the most efficient farms with relative efficiency scores approximately equal to 1. However, the farms A, E, F, G, I and P with a score less than 0.75 are considered inefficient [21]. Depending on our detailed results, we can conclude that company H is the most efficient organization with the best sustainability performance, its efficiency is equal 1. All other parameters that appear in Table 2 can help us in determining the weight of each input KPI which affects the sustainability performance. These weights can be used to formulate an equation for sustainability assessment. As presented in Table 2 and depending on the Table 1, the number of employees (SO1) has the largest effect on the biogas farm efficiency, whereas the governance indicator (GOV1), which depicts the percentage of women in company management, doesn't have any effect on the farms performance. Our final results show different weights for each KPI depending on the given farm. For our future work these weights can be combined with using a suitable method to produce a unique weight of each indicator which will help all farms with biogas plants in the South Moravian region of the Czech Republic to assess their sustainability performance. These results will be a part of the our developed Web Environmental Benchmarking and Reporting System (WEBRIS)[2] which consists of two modules; one *Survey* collects questionnaires and their management and the second one *Reporting* provides evaluation of the forms, and display of results as a diagram in a comparison with the average for the sector.

4 Conclusion

One of the main factors which are intertwined with the management of the farms with biogas plants is sustainability performance assessment. This process should be comprehensive and transparent to achieve best performance. For this reason, two main global frameworks – Global Reporting Initiative G4 and Sustainability assessment of Food and Agriculture system framework – are presented. These frameworks form a backbone of sustainability performance assessments by determining the main key performance indicators (KPIs) which should be reported. Since our research concentrates on biogas farms in the South Moravian region of the Czech Republic which is a part of the agriculture sector, previously mentioned frameworks are the most suitable reporting tools. After that we present the DEA model for measuring farm with biogas plants sustainability performance considering a dual-role factor and cross-efficiency technique simultaneously. This model can provide us with an indication of corporate performance level and determine the weight of each KPI. Finally, this implemented model can be integrated into information and communication tools of farms and used by management of them as a powerful technology for monitoring their efficiency scores, determining their weak points and proving them.

Acknowledgements

This paper is supported by the Czech Science Foundation. Project name: Measuring corporate sustainability performance in selected sectors [Nr. 14-23079S].

[2] http://gacr.pefka.mendelu.cz/gacr403/en/

References

[1] Adler, N., Friedman, L., and Sinuany-Stern, Z.: Review of ranking methods in the data envelopment analysis context. *European Journal of Operations Research*, **140** (2002), 249–265.

[2] Amadeus Database. [Online]. Available on: http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/Amadeus.aspx.

[3] Callens, I. and Daniel T.: Towards indicators of sustainable development for firms: a productive efficiency perspective. Ecological Economics 28(1999): 41–53.

[4] Charnes, A., Cooper, W. W., and Rhdes, E.: Measuring the efficiency of decision making units. *European-Journal of Operational Research* **2** (1978), 429–444.

[5] Cisneros, J.M., Grau, J.B., Anton, J.M., de Prada, J.D., and Degioanni, A.J.: Assessing multi-criteria approaches with environmental, economic and social attributes weights and procedures: a case study in the Pampas, Argentina. *Agricultural Water Management*, **98** (2011), 1545–1556.

[6] Dong, F., Mitchell, P. D., and Colquhoun, J.: Measuring farm sustainability using data envelope analysis with principal components: The case of Wisconsin cranberry. *Journal of environmental management*, **147** (2015), 175–183.

[7] Environmental Indicators for Agriculture. Volume 3. Methods and Results. OECD Publishing, Paris, 2001.

[8] FAO, "Climate-Smart' Agriculture – Policies, Practices and Financing for Food Security, Adaptation and Mitigation. Food and Agriculture Organization of the United Nations, Rome, 2010.

[9] Gerlach, F., Grieb, B., and Zerger U.: *Sustainable biogas production: a handbook for organic farmers.* Published by FiBL Projekte GmbH, Germany, 2013.

[10] GSCP (2010): Reference tools of the global social compliance programme. [Online]. Available at: http://www.gscp net.com/working-plan.html.

[11] G3.1 Guidelines (2011). [Online]. Available at: http://www.globalreporting.org/ReportingFramework/G31Guidelines/.

[12] G4 Guidelines (2013a): Reporting principles and standard disclosures. [Online]. Available at: https://www.glob- alreporting.org/reporting/g4.

[13] G4 Guidelines (2013b): Implementation manual. [Online]. Available at: https://www.gl -obalreporting.org/reporting/g4/.

[14] G4-Food-Processing-Sector-Disclosures (2014). [Online]. Available at https://www.globalreporting.org/resourcelibrary/GRI-G4-Food-Processing-Sector-Disclosures.pdf.

[15] Hřebíček, J., Soukopová, J., Štencl, M., and Trenz, O.: Integration of Economic, Environmental, Social and Corporate Governance Performance and Reporting in Enterprises. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **59** (2011). 157–177.

[16] Hřebíček, J., Trenz, O., and Vernerová, E.: Optimal set of agri-environmental indicators for the agricultural sector of Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, **6** (2013), 2171–2181.

[17] Hřebíček, J., Trenz, O., Chvátalová, Z., and Soukopová, J.: Optimization of corporate performance using data envelopment analysis with Maple. *In Aurelio Araujo et al. Engineering Optimization 2014*. London: CRC Press, (2014). 763–767.

[18] Ilic, M.: Economic value added as a modern performance indicator. Perspectives of Innovations, Economics and Business, PIEB, 3 (2010), 94–97.

[19] Jablonský, J., and Dlouhý, M.: *Models of evaluating the effectiveness of production units*. Professional Publishing, Praha, 2004.

[20] Jablonský, J.: *Operations research: Quantitative models for economic decisions (in Czech)*, Praha: Professional Publishing. 2007.

[21] Křen J.: *Metodika hodnocení trvalé udržitelnosti systémů rostlinné produkce pro podmínky ČR*. Mendelova univerzita v Brně, 2011.

[22] Lee K.H., and Saen, R.F.: Measuring corporate sustainability management: A data envelopment analysis approach. *International Journal of Production Economics* **140** (2012), 219–226.

[23] Maple Online Help. 2014. Optimization [LPSolve] – solve a linear program. [Online]. Available at: http://www.maplesoft.com/support/help/Maple/view.aspx?path=Opti mization/LP Solve.

[24] SAFA Guidelines (2013): Sustainability Assessment of Food and Agriculture systems. SAFA Guidelines version 2.0. [Online]. Available: http://www.fao.org/nr/ sustainability/sustainability-assessments-safa/en/.

[25] Rappaport, A: Creating Shareholder Value: The New Standard for Business Performance, Free Press, New York, 1986.

[26] Torquati, B., Venanzi, S., Ciani, A., Diotallevi, F., and Tamburi, V.: Environmental Sustainability and Economic Benefits of Dairy Farm Biogas Energy Production: A Case Study in Umbria. *Sustainability*, **6** (2014), 6696–6713.

[27] Wang, Y.M., and Chin, K.S.: Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics*, **128** (2010), 332–338.

Calculation and evaluation of new AOQL single sampling plans for inspection by variables

Nikola Kaspříková[1]

Abstract. The average outgoing quality limit (AOQL) sampling plans which minimize the mean inspection cost per lot of process average quality when the remainder of rejected lots is inspected were originally designed by Dodge and Romig for the inspection by attributes. Then the single sampling plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of the rejected lots is inspected by attributes) were proposed and it has been shown that these plans are in many situations more economical than the corresponding Dodge-Romig single sampling plans for the attributes inspection. This paper recalls some of the properties of the AOQL single sampling plans when the remainder of rejected lots is inspected and then the new AOQL sampling plans for the inspection by variables are proposed, using the exponentially weighted moving average statistic in the decision procedure. The plans are calculated using the operating characteristic for the known standard deviation case and the economic characteristics of the plans are discussed.

Keywords: EWMA, acceptance sampling, inspection cost, optimization.

JEL classification: C44
AMS classification: 62D99

1 Introduction

The AOQL sampling plans minimizing the mean inspection cost per lot of process average quality when the remainder of rejected lots is inspected were originally designed by Dodge and Romig (see e.g. [2]) for the inspection by attributes. Plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of rejected lots is inspected by attributes) were then proposed and it was shown that these plans are in many situations more economical than the corresponding Dodge-Romig attribute sampling plans. The AOQL plans for inspection by variables and attributes have been introduced in [8], using approximate calculation of the plans. Exact plans, using non-central t distribution in calculation of the operating characteristic, have been reported in [9] and implemented in R package LTPDvar [5] The operating characteristics used for these plans are discussed by Jennett and Welch in [3] and by Johnson and Welch in [4]. It has been shown that these plans are in many situations superior to the original attribute sampling plans and similar results have been obtained for the LTPD plans, the analysis is provided in [6] and in [7]. Recent development of acceptance sampling plans includes the work by Aslam et al. in [1] where the exponentially weighted moving average (EWMA) statistic is used for a design of the (p_1, p_2) sampling plans, i.e. sampling plans which satisfy the requirement to control the producer's risk and the consumer's risk. Using the EWMA statistic enables some savings in the cost of inspection as it allows using information on the quality in the previous lots. The new AOQL plans for the inspection by variables and attributes, which are designed using the EWMA statistics, are proposed in the present paper and a discussion of the economic performance of such plans is provided. The structure of the paper is as follows: the AOQL plans for the inspection by attributes are recalled first, then the AOQL plans for the inspection by variables and attributes using EWMA statistic are introduced and finally the analysis of the economic performance of the new plans is provided.

[1]University of Economics in Prague,
Nám. W. Churchilla 4, Praha, Czech Republic, nb33@tulipany.cz

2 Attributes inspection sampling plans

Under the assumption that each inspected item is classified as either good or defective (acceptance sampling by attributes), Dodge and Romig (see e.g. [2]) introduced sampling plans (n, c) which minimize the mean number of items inspected per lot of the process average quality, assuming that the remainder of the rejected lots is inspected

$$I_s = N - (N - n) \cdot L(\bar{p}; n, c) \tag{1}$$

under the condition

$$\max_{0<p<1} AOQ(p) = p_L. \tag{2}$$

The notation in equations (1) and (2) is as follows:
N is the number of items in the lot (the given parameter),
$\bar{p}$ is the process average fraction defective (the given parameter),
p_L is the average outgoing quality limit (the given parameter, denoted AOQL),
n is the number of items in the sample ($n < N$),
c is the acceptance number (the lot is rejected when the number of defective items in the sample is greater than c),
$L(p)$ is the operating characteristic (the probability of accepting a submitted lot with the fraction defective p).

The function AOQ is the *average outgoing quality, AOQ*(p) is the mean fraction defective after inspection when the fraction defective before inspection was p. The average outgoing quality (where all defective items found are replaced by good ones) is approximately

$$AOQ(p) = \left(1 - \frac{n}{N}\right) \cdot p \cdot L(p; n, c). \tag{3}$$

Therefore the condition (2) can be rewritten as

$$\max_{0<p<1} \left(1 - \frac{n}{N}\right) \cdot p \cdot L(p; n, c) = p_L. \tag{4}$$

The condition (2) protects the consumer against having an average outgoing quality higher than p_L (the chosen value), regardless of what the fraction defective p is before inspection.

3 AOQL plans for inspection by variables and attributes

The new AOQL plans for the inspection by variables and attributes are designed under the following assumptions: the measurements of a single quality characteristic X are independent, identically distributed normal random variables with unknown parameter μ and known parameter σ^2. For the quality characteristic X there is given either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit L (the item is defective if its measurement is smaller than L).

The inspection procedure is as follows:

Draw a random sample of n items from the lot and compute the sample mean $\bar{x}$ and the statistic T at time t as $T_t = \lambda\bar{x} + (1 - \lambda)T_{t-1}$, where λ is a smoothing constant between 0 and 1. The values of the smoothing constant over 0.5 give more weight to the sample in the current lot. Accept the lot if

$$\frac{U - T_t}{\sigma} \geq k \quad \text{or} \quad \frac{T_t - L}{\sigma} \geq k. \tag{5}$$

Suppose that c_s^* is the cost of inspection of one item by attributes and c_m^* is the cost of inspection of one item by variables and that the sample is inspected by variables. Then the inspection cost per lot with proportion defective p, assuming that the remainder of rejected lots is inspected by attributes (the inspection by variables and attributes), is $n \cdot c_m^*$ with probability $L(p, n, k)$ and $n \cdot c_m^* + (N - n) \cdot c_s^*$ with probability $1 - L(p, n, k)$. The mean inspection cost per lot of process average quality $\bar{p}$ is therefore

$$C_{ms} = n \cdot c_m^* + (N - n) \cdot c_s^* \cdot (1 - L(\bar{p}; n; k)) \tag{6}$$

Dividing (6) by c_s^* gives the objective function

$$I_{ms} = n \cdot c_m + (N - n) \cdot (1 - L(\bar{p}; n; k)), \tag{7}$$

where $c_m = c_m^*/c_s^*$ is the ratio of the cost of inspection of one item by variables to the cost of inspection of this item by attributes (this parameter has to be estimated in each real situation, it is usually $c_m > 1$). Note that both the function $I_{ms} = C_{ms}/c_s^*$ and the function C_{ms} have a minimum for the same acceptance plan (n, k). Therefore, we shall look for the acceptance plan (n, k) minimizing (7), instead of (6), under the condition (4).

Setting the value of c_m to 1 can be used in situations, when both the sample and the remainder of rejected lots are inspected by variables. Acceptance sampling by variables can thus be considered just as a special case of acceptance sampling by variables and attributes. Then instead of I_{ms} we may use notation I_m and setting $c_m = 1$ in (7) we obtain

$$I_m = N - (N - n) \cdot L(\bar{p}; n; k), \tag{8}$$

i. e. the mean number of items inspected per lot of process average quality, assuming that both the sample and the remainder of rejected lots is inspected by variables.

The task to be solved is to determine plan (n, k) minimizing (7) under the condition (4) for given values of input parameters N, c_m, p_L and $\bar{p}$. The operating characteristic is (see e.g. [1])

$$L(p, n, k) = \Phi((u_{1-p} - k)A), \tag{9}$$

where

$$A = \sqrt{\frac{n(2 - \lambda)}{\lambda}}. \tag{10}$$

The function Φ in (9) is a standard normal distribution function and u_{1-p} is a quantile of order $1 - p$ (the unique root of the equation $\Phi(u) = 1 - p$).

4 Calculation and economic performance of the new AOQL plans

We shall calculate the AOQL acceptance sampling plan for sampling inspection by variables when the remainder of rejected lots is inspected by attributes below. The task will be solved using the operating characteristic given by (9) and using the EWMA statistic with smoothing constant $\lambda = 0.9$. The resulting sampling plan will be compared with the corresponding Dodge-Romig plan available in [2] and evaluated with regard to the economic characteristics. Let us consider the following situation. There is a lot with $N = 3500$ items considered in the acceptance procedure. The average outgoing quality limit is given to be $p_L = 0.015$. It is known that the average process quality is $\bar{p} = 0.01$. A cost of inspecting an item by variables is 70% higher than the cost of inspecting the item by attributes, so the parameter c_m equals 1.7. We are to find the optimal AOQL acceptance sampling plan for sampling inspection by variables when the remainder of rejected lots is inspected by attributes.

The sampling plan can be calculated using a modified version of the code available in LTPDvar package [5] for R software [10]. The resulting plan is $n = 34$, $k = 1.9461$.

For the values of the input parameters given in our problem, there is plan (165, 4) for the acceptance sampling by attributes in [2]. Let us compare the plans ($n = 34$, $k = 1.9461$) and ($n = 165$, $c = 4$) with regard to the economic efficiency. We shall use the parameter e defined as

$$e = \frac{I_{ms}}{I_s} \cdot 100. \tag{11}$$

The expression $(1 - e)$ then represents the percentage of savings in mean inspection cost per lot of proces average quality when the sampling plan for the inspection by variables and attributes is used instead of the corresponding plan for the inspection by attributes. Let us denote the plan for the inspection by variables and attributes as (n_1, k) and the corresponding plan for the inspection by attributes as (n_2, c). Then the parameter e is defined as

$$e = \frac{n_1 \cdot c_m + (N - n_1) \cdot (1 - L(\bar{p}, n_1, k))}{N - (N - n_2) \cdot L(\bar{p}, n_2, c)} \cdot 100. \tag{12}$$

Since we get

$$e = 34.32,$$

it can be expected that approximately **66% savings in the inspection cost** can be made using the AOQL plan for inspection by variables and attributes (34, 1.9461), in place of the corresponding Dodge-Romig plan $(165, 4)$.

The operating characteristic values of the resulting variables inspection plan and of the corresponding sampling plan for inspection by attributes for the set of input parameters values considered in the problem solved are shown in Figure 1. It may be observed that the operating characteristic curve of the resulting variables inspection plan seems more favourable than the operating characteristic curve of the attributes inspection plan.

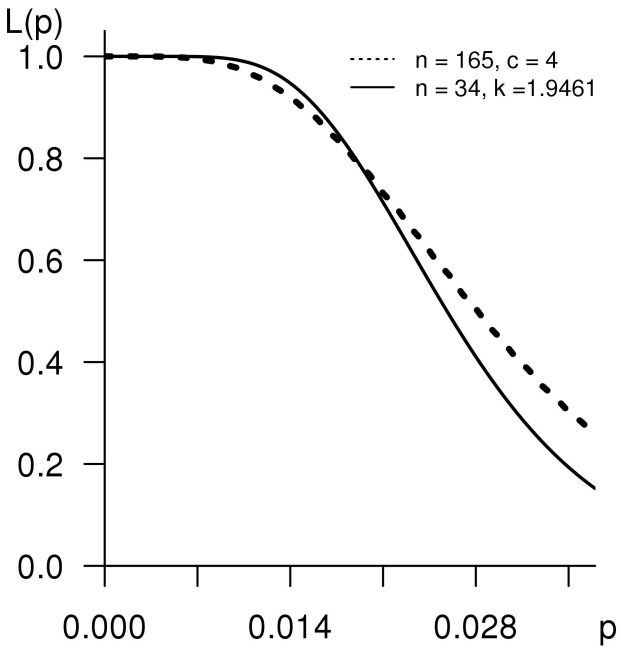


Figure 1 Operating characteristics curves of sampling plans

5 Conclusion

The new AOQL plans for the inspection by variables and attributes minimizing the mean inspection cost per lot of process average quality, which are designed to use the EWMA statistic in the decision procedure, have been introduced and it has been shown that these plans seem to be promising with respect to the economic characteristics.

Acknowledgements

This paper has been produced with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Aslam, M., Azam, M., and Jun, C.: A new lot inspection procedure based on exponentially weighted moving average, *International Journal of Systems Science* **46** (2015), 1392–1400.

[2] Dodge, H., and F., Romig, H. G.: *Sampling Inspection Tables: Single and Double Sampling.* John Wiley, New York, 1998.

[3] Jennett, W. J., and Welch, B. L.: The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale, *Supplement to the Journal of the Royal Statistical Society* **6** (1939), 80–88.

[4] Johnson, N. L., and Welch, B. L..: Applications of the Non-central t distribution, *Biometrika* **38** (1940), 362–389.

[5] Kasprikova, N.: *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.1. http://CRAN.R-project.org/package=LTPDvar. 2012.

[6] Kasprikova, N., and Klufa, J.: AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes, *Quality technology and quantitative management* **6** (2015), 133–142.

[7] Kasprikova, N., and Klufa, J.: Calculation of LTPD single sampling plans for inspection by variables and its software implementation. In: *International Days of Statistics and Economics*, University of Economics, Prague, 2011.

[8] Klufa, J.: Dodge-Romig AOQL single sampling plans for inspection by variables, *Statistical Papers* **38** (1997), 111–119.

[9] Klufa, J.: Dodge-Roming AOQL plans for inspection by variables from numerical point of view, *Statistical Papers* **49** (2008), 1 – 13.

[10] R Core Team: *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.R-project.org

Two-dimensional Bin Packing Problem in batch scheduling.

František Koblasa[1], Miroslav Vavroušek[2],František Manlig[3]

Abstract. Recently, there is great focus on production planning scheduling as a tool how to meet customer demand. This is based on technical attributes e.g. constraints like processing, setup transportation time etc., but also on specific requirements as maximal dimensions, temperature, material. That is why Advanced Planning and Scheduling systems become more and more important. However, it heavily depends on kind of used constrains and it is usual to focus only on time availability of machines.
This article is following recent trends in constrained scheduling considering special technologies as heat treatment, blackening, nitration etc. There is necessary to take in account manufacturing batch which is usually containing parts of several preceding batches so products with different dimensions which have to be processed in work space of given technology. This article is proposing early research done by reviewing two-dimensional Bin Packing Problem and by developing and testing algorithm inspired by Maximal Rectangle approach with goal to discuss further use in batch scheduling of heat treatment technologies.

Keywords: Bin Packing Problem, Scheduling, Heat treatment, Maximal rectangle.

JEL Classification: C61
AMS Classification:90C27

1 Introduction

Production planning and scheduling is solved by several approaches. The simplest one is considering material availability (Material Requirements Planning) followed by most used which is machine availability in the terms of utilization (Manufacturing Resource Planning). Most advanced approach (Advanced Planning and Scheduling) is using scheduling approaches where a machine can perform only one job at a time as in classical scheduling models as Open Shop, Flow Shop or Job Shop Scheduling Problems. However, there are technologies whose are able to process several jobs at time, so they are able to manufacture jobs in batches. Batch processing machines occurs in many different types. Base on the type, it is necessary to include various additional constraints to correctly determine processing time of whole batch.

There are two basic types of setting batch processing time [1]:

- S-batch – batch processing time is equal to sum of each task in batch.
- P-batch – batch processing time is equal to the maximum processing time among all tasks.

This article is focusing on the p-batch type of processing which has great representation in scheduling in various fields of manufacturing and services as:

- Mechanical engineering – used for operations which change internal and external properties as in heat treating furnaces [2], during blasting [3]or burnishing [4].
- Electrical engineering – during chip test operation called burn-in [5], which test to find hidden defects by long term heating or bake-out [6] operation which is used to manufacture ceramic semiconductors, another application of p-batch processing is during wafer fabrication [7]. However it is mostly used in semiconductor industry [8]
- Medical processing – classical example of p-batch scheduling is during the operation of sterilization as shown in [9,10]

P-batch scheduling is often connected with the problem of placing incompatible job families with product different dimensions in to the working space which is classical case of Bin Packing Problem. Following chapter deals with two-dimensional Bin Packing Problem. There are reviewed basic classification and solution methods

[1]Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering,
Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: frantisek.koblasa@tul.cz.
[2]Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering,
Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: miroslav.vavrousek@tul.cz.
[3] Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering,
Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: frantisek.manlig@tul.cz..

of mentioned problem together and by preliminary testing of Maximal Rectangle algorithm. Batch scheduling problem of heating furnaces and suitability of tested algorithm is discussed.

2 Two-dimensional Bin Packing Problem

Cutting and packing problems appear in the practice a lot by different names as problems wary by its use in particular type of industry [11, 12]. There are problems as:

- Cutting stock and trim loss problems [13] were is usual focus to minimize leftovers during cutting material.
- Bin packing, dual bin packing, strip packing and vector packing, where there is focus on minimizing number of bins necessary to pack items [14].
- Knapsack problems, where we are trying to maximize value of items [15].
- Vehicle loading, pallet loading, container loading and car loading problems etc. [16].

Our problem which contains batch scheduling of various types of product focus on placing object in heating treatment devices with goal to minimize or overall process time so number of batches. This problem can be modeled or described as bin packing problem. We are focusing on bin packing problem, where complex shape of inserted object in to the bin is simplified to two-dimensional object - rectangle.

Type of this problem is usually classified as 2BP [17] and described as there is a set of rectangles R with their width (w) and height (h) ($R_1;R_2;\ldots;R_n$);$R_j = (w_j; h_j)$ and the task is to pack items into a minimum number of bins of size (W;H).This combinatorial problem is considered as NP-hard and exists in two basic rule variations given by constraints and ways to solve:

- Orientation: the items may have either a fixed orientation or 90 degrees rotated.
- Type of cuts: whether or not the cutting patterns should meet the requirement of guillotine cuts.

Those rules are dividing 2BP in to for following classes[17]:

- 2BPOG: Items are oriented (O) and guillotine cuts are required (G).
- 2BPRG: Items may be rotated by 90 degrees (R) and guillotine cuts are required (G).
- 2BPOF: Items are oriented (O) and cutting is free (F).
- 2BPRF: Items may be rotated by 90 degrees (R) and cutting is free (F).

2BPOG and 2BPRG are usually used in the case of scheduling cutting technologies and are usually solved in the terms of constructive algorithms by shelf or guillotine algorithms using rules like Shorter/Longer Axis, Shorter/Longer Leftover or Max/Min Area. 2BPOF and 2BPRF are more general and can be solved by e.g. Maximal Rectangle or Skyline algorithms using rule of Bottom-Left, Best Area Fit, Best Short/Long Side Fit e.g.

This study focuses on further use in heat treating processes, where guillotine cuts are not required and parts are already oriented i.e. models and algorithms from class 2BPOF are used. When choosing type of algorithm to design, it was selected Maximal Rectangle approach as it is considered as best in general white skyline algorithms excels in particular cases, but in general have worse results.

2.1 Maximum Rectangle Algorithm

Minimal free space – Maximal rectangle scheme is inspired by Maximal rectangle algorithms presented in [18].The set of placed rectangles is arranged according to the size from the lowest to the highest. In the beginning, the first bin of given proportions and a rectangle of free space covering the entire space of the bin are created. This is a preparation phase of the algorithm.

The placed rectangles are consecutively chosen from the arranged set. In case the first rectangle is bigger than the space of the bin the task has no solution. In case the task has a solution the first rectangle is placed in the corner of the free space. The free space is divided into two new free spaces which cover the space of the cut out rectangle. A function was created to calculate the remaining free space, which takes over two rectangles and returns the difference of the first rectangle without a common free space with the second rectangle. The resulting difference is described using four rectangles of maximal proportions at the most. An identifier of the bin is set for the placed rectangle and the coordinates of the center are calculated. The free spaces are arranged according to size (see Fig. 1).

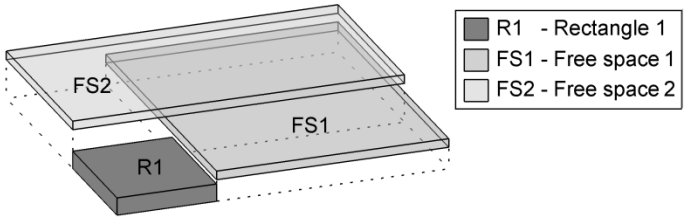


Figure 1 Location of the first rectangle in the bin

Next, the second rectangle is chosen from the set. The free spaces are looked through consecutively and they are arranged according to size from the smallest to the biggest. If the free space is big enough, the rectangle is placed. After that the calculation of the remaining free space is carried out. The calculation has to be done for all the free spaces to which the rectangle reaches. The free spaces are arranged according to size. The set of free spaces is optimized. The optimization lies in deleting free spaces which are only subspaces of other spaces (see Fig.2).

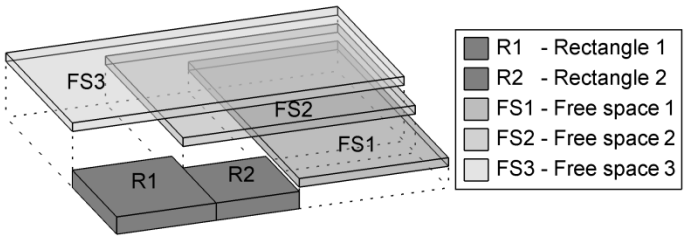


Figure 2 Location of the second rectangle in the bin

If no free space in any bin is big enough, a second bin is created into which the rectangle is placed. The procedure repeats for all rectangles from the given set. By placing the biggest rectangle into the smallest free space the space is used effectively. By going through all bins, starting with the first, the free spaces are filled by placing bigger rectangles.

Data structure

Bins contain set of Bin
Bin contain set of FreeSpace
FreeSpace = {w; h; centerX; centerY}
Rectangles contain set of Rectangle
Rectangle = {w; h; centerX; centerY}

Initialize:

```
01 create new Bin B in the Bins
02 add new FreeSpace F = (W; H) to B
```

Pack:

```
03 sort Rectangle R = (w; h) in the Rectangles by size of surface from the biggest
04 for each Rectangle R in the Rectangles
05 |   Bool placed = False
06 |
07 |   for each Bin B in the bins
08 |   |   for each FreeSpaceF= (w; h) in the Bin B
09 |   |   |   if F.w>=R.w and F.h>=R.h
10 |   |   |   |   placed = True
11 |   |   |   |   compute F - R and subdivide the result into at most four new FreeSpaces G1.;..;G4
12 |   |   |   |   add G1;...;G4 to B
13 |   |   |   |   break
14 |   |   if placed = False
15 |   |   |   break
16 |   if placed = True
17 |   |   create new Bin B in the Bins = (W; H)
18 |   |   add new FreeSpace F = (W; H) to B
19 |   |   compute F - R and subdivide the result into at most four new FreeSpaces G1;...;G4
20 |   |   add G1;...;G4 to B
21 |   optimize FreeSpaces in the Bin B (Erase FreeSpace, which is only subspace)
22 |   sort FreeSpaces in the Bin B by size of the surface from the smallest
```

2.2 Test problems and Results

Minimal free space – Maximal rectangle is tested in this section on two common known datasets. The first set was generated by Berkey and Wang [19] and consists of 6 classes (Class-I to VI):

- Class I : w_j and h_j uniformly random in [1,10], $W = H = 10$;
- Class II : w_j and h_j uniformly random in [1,10], $W = H = 30$;
- Class III :w_j and h_j uniformly random in [1,35], $W = H = 40$;
- Class IV :w_j and h_j uniformly random in [1,35], $W = H = 100$;
- Class V : w_j and h_j uniformly random in [1,100], $W = H = 100$;
- Class VI :w_j and h_j uniformly random in [1,100], $W = H = 300$.

The second set was generated by Martello and Vigo [20] and consists of 4 classes (Class VII to X). There are four different item types:

- w_j uniformly in [1,1/2W] and h_j uniformly in [2/3H,H].
- w_j uniformly in [2/3L,W] and h_j uniformly in [1,1/2H].
- w_j uniformly in [1/2L,W] and h_j uniformly in [1/2H,H].
- w_j uniformly in [1,1/2W] and h_j uniformly in [1,1/2H].

The items in classes VII–X are generated as following and the bin sizes are $W = H = 100$:

- Class VII 70% of type 1 items and 10% each for types 2, 3 and 4.
- Class VIII 70% of type 2 items and 10% each for types 1, 3 and 4.
- Class IX 70% of type 3 items and 10% each for types 1, 2 and 4.
- Class X 70% of type 4 items and 10% each for types 1, 2 and 3.

Table 1 presents results of those 500 problems (each class contains 10 models per problem) comparing designed algorithm (MinfMaxr) with known optimal solution [21] by dividing our solution with the optimal one. Table also show results of well know Alternate Directions (ADOF) constructive algorithm [22].

Class	Problem	20	40	60	80	100	Average
I	ADOF	1,120	1,090	1,070	1,060	1,050	1,078
	MinfMaxr	1,000	1,015	1,030	1,011	1,019	1,015
II	ADOF	1,000	1,100	1,100	1,070	1,030	1,060
	MinfMaxr	1,000	1,053	1,160	1,129	1,077	1,084
III	ADOF	1,200	1,150	1,130	1,100	1,090	1,134
	MinfMaxr	1,059	1,076	1,066	1,048	1,054	1,061
IV	ADOF	1,000	1,000	1,150	1,100	1,030	1,056
	MinfMaxr	1,000	1,053	1,130	1,133	1,081	1,079
V	ADOF	1,140	1,110	1,100	1,090	1,090	1,106
	MinfMaxr	1,015	1,034	1,039	1,050	1,043	1,036
VI	ADOF	1,000	1,400	1,050	1,000	1,070	1,104
	MinfMaxr	1,000	1,267	1,048	1,000	1,125	1,088
VII	ADOF	1,100	1,100	1,070	1,060	1,040	1,074
	MinfMaxr	1,091	1,073	1,038	1,049	1,026	1,055
VIII	ADOF	1,130	1,080	1,060	1,060	1,060	1,078
	MinfMaxr	1,052	1,045	1,057	1,022	1,029	1,041
IX	ADOF	1,010	1,020	1,020	1,020	1,010	1,016
	MinfMaxr	1,007	1,007	1,005	1,002	1,001	1,004
X	ADOF	1,100	1,090	1,110	1,100	1,100	1,100
	MinfMaxr	1,048	1,041	1,071	1,073	1,065	1,060

Table 1 2BP constructive algorithm results comparison

Designed approach provides promising results despite the fact it is only constructive algorithm. MinfMaxr was able to find several optimal solutions (class II, IV and VI in problem 20). Overall results were equal or

slightly better in average comparing with ADOF algorithm, with only two exceptions in the Class II and IV. However, MinfMaxr can be used in heat treating operations only if meets conditions of given technology.

3 Batch scheduling of Heat treatment

Heat treatment furnace operations are used usually to improve the structural properties of steel castings which are subjected to stressful mechanical and thermal loads. Although there are many types of heat treatment operations based on the type of castings, in general, those kind manufacturing process takes a significant part of the total processing time comparing to the classical milling operations. That makes heat treatment workshops bottlenecks. Scheduling of those bottlenecks is complex problem where technology constraints often interfere with classical batch scheduling. This problem is often threated by decomposition to several stages and for its multi constrained property is usually solved by simulation software [23, 24].

There are only a few studies which focus on addressing the problem of scheduling heat treatment furnaces [25, 26].However there are known additional constraints as:

- Scheduling heterogeneous batch.
- Processors with incompatible job families.
- Non-identical job sizes.
- Dynamic job arrivals to maximize the utilization of the batch processors to minimize the total weighted tardiness.

These constraints are usually decomposed to those steps:

1. Loading available batches (including part properties – to indicate job families of heating process).
2. Sorting parts by families.
3. Design manufacturing batches from families (solved by bin packing problem) and create "space decomposition schedule" for operative planning.
4. Assign respective bins to a suitable machine.
5. Determine the completion time of each task (bin) by selecting longest completion of all parts from bin.
6. Start to produce batches.

Product families from step 4.and also batch processing time in step 5. is usually influenced by:

- Heating technology (in the case of soft annealing there are shorter processing times than in normalized)
- Material (steel, cast-iron)
- Mass (depends on thicker dimension - bigger mass, longer heating and cooling)
- Surface (bigger surface – heat can go faster in material)
- Distance between parts (influences surface taking heating and cooling process)

The greatest problem in the matter of scheduling is the last step of proceeding batches i.e. decision, if it is suitable to release batch to processing. Maximal utilization of processor space could prolong processing time (more load cause longer heating and cooling) so it is necessary to determine suitable multi-objective function. That is usually called threshold policy [24] - specifies the minimum number of lots that should be present in the queue to initiate processing (considering dynamic job arrivals), which is less than or equal to the maximum number of lots that the furnace can process at a time.

4 Conclusion an further research

Designed MinfMaxr algorithm shoved promising result and could be used operation connected to heat treating technologies as a tool to design manufacturing batches. However, it is necessary to make in-deep analysis of particular technology process at defined machine to schedule whole system in the holistic point of view. Further research will focus on improving designed algorithm by allowing part rotation, its use in 3D space, implementing special conditions of part placement in the workspace of the specific machine and on definition of the multi-objective function, which could cover most of the enterprise requirements as machine utilization, meeting due dates etc.

Acknowledgements

Supported by the grant No. SGS-21010 of the Grant Agency TUL-SGS.

References

[1] Brucker, P. Scheduling Algorithms. 5th ed. Berlin: Springer, 2007, 371 s. ISBN 978-3-540-69515-8.

[2] Ramasubramanian, M,Mathirajan, M and Ramachandran V. Minimizing makespan on a Single Heat-Treatment Furnace in Steel Casting Industry. In: *International Journal of Services and Operations Management* (Inderscience Publication), Vol. 7, (2010) pp. 112-142..
[3] Kumar, V.Kumar, S.Chan, F.T.S.Tiwari, M.K.: *Auction-based approach to resolve the scheduling problem in the steel making process.* International Journal of Production Research, 44 (8). pp. 1503-1522. (2006) ISSN 1366-588,
[4] Przybylski, W., &Ściborski, B.: Studies on the influence of technological variants of finishing machining on flow of parts in flexible manufacturing. In: *Advances in Manufacturing Science and Technology*, 34(1), (2010), 21-32.
[5] Koh, S.G., Koo, P.H., Kim, D.C., Hur, W.S. Scheduling a single batch processing machine with arbitrary job sizes and incompatible job families. In:*International Journal of Production Economics*.Vol. 98, pp 81-96. (2005), ISSN 0925-5273.
[6] Koh, S.-G., Koo, P.-H., Ha, J.-W., Lee, W.-S. Scheduling parallel batch processing machines with arbitrary job sizes and incompatible job families. In:*International Journal of Production Research.Vol.42*, No. 19, pp 4091-4107.(2004), ISSN 0020-7543,.
[7] Bang, J. Y., Kim, Y. D., & Choi, S. W. Multiproduct Lot Merging–Splitting Algorithms for Semiconductor Wafer Fabrication. In: *Semiconductor Manufacturing*, IEEE Transactions on, 25(2), (2012), 200-210.
[8] Mathirajan, M. and A.I. Sivakumar,: A literature review, classification and simple meta-analysis on scheduling of batch processors in semiconductor. In*: J. Adv. Manuf. Technol*., 29:(2006), 990-1001..
[9] Ozturk, O.Espinouse, M-L.Mascolo,M.D,Gouin.A.:Makespan minimisation on parallel batch processing machines with non-identical job sizes and release dates. In:*International Journal of Production Research*, Taylor & Francis: STM, Behavioural Science and Public Health Titles, 50 (20), (2012), pp.6022-6035,
[10] Rossi, A., Puppato, A., Lanzetta, M.: Heuristics for scheduling a two-stage hybrid flow shop with parallel batching machines: application at a hospital sterilisation plant. In: *International Journal of Production Research*, 51(8), (2013), 2363-2376,
[11] Dyckhoff,A.: A typology of cutting and packing problems, In: *European Journal of Operational Research*, Volume 44, Issue 2, , Pages 145-159, 1990 http://dx.doi.org/10.1016/0377-2217(90)90350-K.
[12] Wäscher, G., Haußner, H., & Schumann, H.: An improved typology of cutting and packing problems. In: *European Journal of Operational Research*,183(3), (2007). 1109-1130.
[13] Cherri, A. C., Arenales, M. N., Yanasse, H. H., Poldi, K. C., &Vianna, A. C. G.: The one-dimensional cutting stock problem with usable leftovers–A survey. In: *European Journal of Operational Research*, 236(2), 2014. 395-402.
[14] Delorme, M., Iori, M., & Martello, S.: Bin Packing and Cutting Stock Problems: Mathematical Models and Exact Algorithms. In *DECISION MODELS for SMARTER CITIES.*(2014).
[15] Lust, T., &Teghem, J. The multiobjective multidimensional knapsack problem: a survey and a new approach. In: *International Transactions in Operational Research*, 19(4), (2012). 495-520.
[16] Morabito, R., & Morales, S.: A simple and effective recursive procedure for the manufacturer's pallet loading problem. In: *Journal of the Operational Research Society*, (1998). 819-828.
[17] Cui, Y. P., Cui, Y., & Tang, T.: Sequential heuristic for the two-dimensional bin-packing problem. In: *European Journal of Operational Research*,240(1), (2015). 43-53.
[18] Jylänki, J. A thousand ways to pack the bin-a practical approach to two-dimensional rectangle bin packing. [online]. [2015-05-06] available: http://clb. demon. fi/files/RectangleBinPack. pdf (2010).
[19] Berkey, J. O. and Wang, P. Y.: Two-Dimensional Finite Bin-Packing Algorithms. In: *The Journal of the Operational Research Society*, 38(5), (1987). 423-429.
[20] Martello, S.Vigo, D.: Exact Solution of the Two-Dimensional Finite Bin Packing Problem. In: *Management Science*, 44(3), (1998),388-399.
[21] DEIS - Operations Research Group Library of Instances: Best Known Solution for 2BP|O|F [online]. [2015-05-06]. available:http://www.or.deis.unibo.it/research_pages/ORinstances/soluz_2dbp.htm
[22] Lodi, A., Martello, S., & Vigo, D. Heuristic and metaheuristic approaches for a class of two-dimensional bin packing problems.In: *INFORMS Journal on Computing*, 11(4), (1999). 345-357.
[23] Ram, B. Patel, G.: Modeling furnace operations using simulation and heuristics. In*: Proceedings of the 30th conference on Winter simulation*. IEEE Computer Society Press, (1998) p. 957-964.
[24] Akcali, E, et al. Alternative loading and dispatching policies for furnace operations in semiconductor manufacturing: a comparison by simulation. In: *Proceedings of the 32nd conference on Winter simulation. Society for Computer Simulation International*, (2000). 1428-1435.
[25] Mathirajan, M., Sivakumar, A. I. and Chandru, V.: Scheduling algorithms for heterogeneous batch processors with incompatible job-families, In: *Journal Of Intelligent Manufacturing* 15(6), (2004)787-803.
[26] Mathirajan, M. and Sivakumar, A. I.. Minimizing total weighted tardiness on heterogeneous batch processing machines with incompatible job families, In: *International Journal Of Advanced Manufacturing Technology* 28(9), (2006) 1038-1047.

Robust formulation for allocation problem

Maria Kobzareva[1], Jan Pelikán[2]

Abstract. Robust optimization is a modeling tool, which is designed to solve problems with uncertain data set. This kind of problem is frequently solved problem in many fields and it is more appropriate for real-life applications, because in practice data are hardly deterministic, but tend to change unexpectedly. The paper introduces robust optimization methodology for allocation problem and suggests a possibility of solving allocation problem with uncertain data set using robust optimization model. To solve the problem we use robust linear programming. The problem takes into account the fact that demand in certain nodes could not be the same as predicted, but real demand could be greater due to unexpected events. In this paper we present the method, which transforms this non-linear problem into sequence of linear problems. The algorithm is programmed in VBA for MS Excel, which is connected with Lingo and uses it to solve the problem. The application is described on a case study and provided with mathematical models and its detailed description. The paper also presents computation experiments and its results.

Keywords: robust optimization, data uncertainty, robustness, linear programming, allocation problem

JEL Classification: C44, C6, D81

AMS Classification: 90C05, 90-08, 90B06, 62G35

1 Introduction

In classic optimization problems we consider data to be deterministic, in other words we assume input data to be known precisely at the moment we solve the problem. Unfortunately, in reality it is rarely the case, as uncertain elements are the part of every problem, are we talking about vehicle routing problem, flow optimization problems, allocation problems and others. For example, if we are planning optimal route for a vehicle we have to take into account the fact that the route could be extended due to traffic jam, traffic light malfunction, traffic accident, road reconstruction or bad weather conditions. Similar situation could occur at the securities market, where predicted future profits of shares can be affected by a number of factors, which are impossible to predict, or when locating a material storage in allocation problem, where the demand at different centers can change unexpectedly. If we assume the input data are exact and will not change under any circumstances and in reality the changes in given inputs will occur, we can find ourselves in the situation when the generated optimal solution violates critical constraints, can become non-optimal and present non-efficient results from the objective function point of view. That's why in the changing world it is important to take into account all possible factors that can affect the generated solution when optimizing. That is the aspect this paper is going to work with – searching for an optimal solution under uncertainty with the help of robust optimization technique.

Robust optimization is a modern approach to uncertainty modelling, where we assume different options of data changing and consider parameter Γ, which is called protection level. This parameter assumes, that negative changes of input values will occur in maximum Γ cases; the method will search all the subsets with the size of Γ elements from an original set and select the subset, with the maximum increase of input values (for minimization problem). This problem is called Γ –robust, which means that the optimal solution of a given minimization problem expects an increase in original values for maximum amount of Γ variables. This problem can be solved with mathematical model, which transforms the problem into the sequence of linear problems. When solving greater complicated problems heuristic methods could be used. Thus robust optimization works with uncertain data and optimizes the value of objective function due to a certain amount of stated cases, which will meet the changes in data. Robust optimization is a modern modelling approach under uncertainty, where coincidence is not modeled in stochastic way, but in deterministic set-based way.

[1]University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: mariakobzareva@gmail.com

[2] University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: pelikan@vse.cz

The robust optimization technique will be studied on allocation problem, where we have to decide where to locate central storage from where the particular centers will be supplied, while the demand in different centers is not known at the moment of solving – it is the quantity that can change due to unexpected circumstances. Thus we're talking about allocation problem with uncertain input data, which is very popular in dynamic environment of today's world.

2 Allocation problem

Let us introduce allocation problem, where we have to decide where to locate a material storage for supplying different centers, respecting the condition that only a certain number of storages will be built and the center will only be supplied from the particular node if a storage is built in that node. The aim is to minimize transportation costs. The problem that the demand of certain centers is not strictly set at the moment we solve the problem – it can change under the impact of unexpected circumstances. Let us first introduce classis allocation problem with given certain demand in different nodes.

2.1 Mathematical description of allocation problem

Let n be the number of centers which need to be supplied from the storages, located in some of the existed centers. Let us introduce binary variable $x_i \in \{0,1\}$ i =1,.., n, where $x_i = 1$ in case the central storage is located in center i, $x_i = 0$ otherwise. Next we consider c_{ij} distance between center i and center j for i, j =1,.., n, d_i is the demand of center i for i =1,.., n and K the number of storages to be built. Variable z_{ij} is binary for i, j =1,.., n; if $z_{ij} = 1$, then center j gets material from storage located in center i, $z_{ij} = 0$ otherwise.

The mathematical model can then be written as follows:

$$min \sum_{i}^{n} \sum_{j}^{n} c_{ij} d_i z_{ij} \tag{1}$$

$$\sum_{i=1}^{n} x_i = K, \tag{2}$$

$$\sum_{i=1}^{n} z_{ij} = 1, \qquad j = 1,..,\mathrm{n}, \tag{3}$$

$$\sum_{j=1}^{n} z_{ij} \leq n x_i, \qquad i = 1,..,\mathrm{n}, \tag{4}$$

where total transportation costs in ton kilometers calculated as (1) are minimized under the condition (2) which ensures that only K centers will be selected where the storage will be built. To ensure supply of center j the equation (3) must be equal. Condition (4) assures that if the storage is not built at center i, that it is not possible to supply other centers from this center.

3 Robust formulation of the allocation problem

We will now focus on the allocation problem with uncertain demand.

Let is introduce binary variable $y_i \in \{0,1\}$ i =1,.., n, where $y_i = 1$ when unexpected event will appear at center i, associated with greater demand d_i', $y_i = 0$ otherwise. Let Γ is maximum number of such unexpected events, Γ is given.

Then robust formulation of the allocation problem is

$$\min_{z} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i z_{ij} + \max_{y} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i' y_i z_{ij} \right), \tag{5}$$

$$\sum_{\mathrm{i}=1}^{12} \mathrm{y_i} \leq \Gamma, \tag{6}$$

$$\sum_{i=1}^{n} z_{ij} = 1, j = 1,2, \dots, n, \tag{7}$$

$$\sum_{i=1}^{n} x_i = K, \tag{8}$$

$$\sum_{j=1}^{n} z_{ij} \leq n x_i, \qquad i = 1, .., n, \tag{9}$$

$$x_i \in \{0,1\}\ i = 1,2, \dots, n, z_{ij} \in \{0,1\}\ i = 1,2, \dots, n, j = 1,2, \dots, n, \tag{10}$$

where (5) assures, total transportation costs in ton kilometers associated with unexpected demand d_i' are minimized under the condition (6) which assures that the unexpected material demand in center *i* depended on variable y_i could appear only in at most Γ cases. Condition (7) then again ensures that only *K* centers will be selected where the storage will be built, equation (8) ensures supply of center *j*, condition (9) assures that if the storage is not built at center *i*, that it is not possible to supply other centers from this subdivision.

Comment 1. The robust problem is nonlinear.

The presented model is two step optimization problem, to solve this problem we can use the method presented in [1].

4 Method for solving robust optimization problem

The method considers basic robust problem objective function presented as

$$F = \min_{z \in P} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i z_{ij} + \max_{y} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i' y_i z_{ij} \right), \tag{11}$$

where $P = \{x_i, z_{ij}$ fulfilling (6), (7), (8), (9), (10)$\}$.

The procedure of finding optimal solution for (11) is based on the following proposition 1.

Proposition 1 [1]**.**

The assumption is

$$c_{i1j1} d_{i1}' \geq c_{i2j2} d_{i2}' \geq \cdots \geq c_{injn} d_{in}' \geq 0. \tag{12}$$

The optimal solution is

$$F = \min_{l = 1,2, \dots, n+1} G_{iljl,} \tag{13}$$

where

$$G_{iljl} = \Gamma c_{iljl} d_{il}' + \min_{z \in P} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i z_{ij} + \sum_{k=1}^{l} \left(c_{ikjk} d_{ik}' - c_{iljl} d_{il}' \right) z_{ikjk} \right). \tag{14}$$

We can formulate algorithm that solves robust optimization problem (11) according the proposition 1.

Algorithm 1[1]**:**

Step 1. Solve the problem and calculate values of $G_{iljl}, l = 1,2, \dots, n$ as follows:

$$G_{iljl} = \Gamma c_{iljl} d_{il}' + \min_{z \in P} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} d_i z_{ij} + \sum_{k=1}^{l} \left(c_{ikjk} d_{ik}' - c_{iljl} d_{il}' \right) z_{ikjk} \right), \tag{15}$$

Step 2. Calculate the value $F = min\{G_{i1j1}, G_{i2j2}, \dots, G_{injn}\} = G_{i*j*}$. The optimal robust solution is z_{ij}^*, x_i^*, which is optimal solution for the problem G_{i*j*}.

Comment 2.

The problem (11), (13), (14) are linear optimization problems with binary variables.

Comment 3.

The value Γ close to zero indicates an optimistic forecast of the future value of objective function and conversely high value of Γ means the pessimistic estimation of the value of objective function.

5 Case study – robust solution formulation

Power supply network-maintenance company with 12 service centers solves the problem of allocating two material storages for maintaining electricity net in north and middle Czech. Every center has workers, which eliminate malfunction on electricity net in specific region of Czech Republic. The average year consumption of material calculated from historic data is defined as d_i, d_i' is additional consumption of center i, where we also take into account that in some years under the unexpected circumstances (flooding, snow calamity, hurricanes, enc.) the consumption can be greater than usual, the real consumption will be $d_i + d_i'$. The aim is to allocate central material storage in two centers, from which the other centers will be supplied with material with minimum transportation costs and with possibility to cover the material demand in case Γ unexpected events in the region will occur.

The input data for described case study are presented in table 1, where places from 1 to 12 are the following: Česká Lípa, Děčín, Chomutov, Jablonec, Liberec, Litoměřice, Louny, Most, Praha, Teplice, Ústí, Varnsdorf. Table 1 contains distance matrix between nodes of graph, the nodes demand d_i and unexpected demand d_i'.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**d_i**	**$d`_i$**
1	0	33	113	68	55	42	85	90	84	69	60	30	1566	1132
2	33	0	86	89	76	53	73	64	115	36	27	51	2915	1344
3	113	86	0	175	162	71	34	23	103	51	69	137	2148	998
4	68	89	175	0	13	110	145	153	106	125	116	68	1430	770
5	55	76	162	13	0	97	140	140	111	112	103	55	3071	2036
6	42	53	71	110	97	0	43	48	63	37	26	72	1214	321
7	85	73	34	145	140	43	0	27	69	37	54	113	2518	1993
8	90	64	23	153	140	48	27	0	96	28	47	115	879	244
9	84	115	103	106	111	63	69	96	0	94	88	114	1005	167
10	69	36	51	125	112	37	37	28	94	0	19	87	1476	1082
11	60	27	69	116	103	26	54	47	88	19	0	78	935	657
12	30	51	137	68	55	72	113	115	114	87	78	0	840	708

Table 1: List of input data – travel matrix, demand

5.1 Robust solution for formulated allocation problem

The problem was solved using the proposed method form chapter 3. Parameter Γ was set to 4, which means that we assume maximum number of 4 nodes to deal with the greater, unexpected demand. In the presented study we were looking for two nodes, where the material storage could be located.

The results for the presented problem, calculated according to (13) are shown in the table 2. In the table you can find the objective values of the robust model for particular values of l and the location of material storages.

l	G_{iljl}	x_i											
1-44	1959667-1047349	0	0	0	0	1	0	0	0	0	1	0	0
45-68	1043810-949434	0	0	0	1	0	0	0	0	0	1	0	0
69-76	945216-934470	0	0	0	0	1	0	0	0	0	0	1	0
77-94	925880-868064	0	0	0	0	1	1	0	0	0	0	0	0
95	868064	1	0	0	0	0	0	0	1	0	0	0	0
96-104	856788-837477	0	0	0	0	1	1	0	0	0	0	0	0
105-113	837142	0	0	0	0	1	1	0	0	0	0	0	0
114-144	838123-917071	0	0	0	0	1	1	0	0	0	0	0	0

Table 2: List of interactions calculating robust optimal solution

From the table above we can see how the value of objective function varies for different *l*. For example for l = 1-44 the optimal storage location is in node 5 and node 10 and the objective function value is between 1959667 and 1047349. The optimal objective function value for robust optimization problem can then be calculated as $F = min\{G_{i1j1}, G_{i2j2}, \dots, G_{i144j144}\} = G_{i105j105} = G_{i113j113}$. The optimal solution is similar for l = 105-113, where the objective function is at its minimum. In the optimal solution the storages are located in Liberec (node 5), from where nodes 4 and 5 only are supplied and in Litoměřice (node 6), from where all other nodes except for 4 and 5 are supplied. Minimum objective function value is 837 142.

We could also change the value of parameter Γ to see how the number of potential unexpected increase in demand can affect the solution.

Γ	F	x_i											
0	640339	0	0	0	0	1	0	0	0	0	1	0	0
1	752319	0	0	0	0	1	0	0	0	0	1	0	0
2	794128	0	0	0	0	1	1	0	0	0	0	0	0
3	816919	0	0	0	0	1	1	0	0	0	0	0	0
4	837142	0	0	0	0	1	1	0	0	0	0	0	0
5	854155	0	0	0	0	1	1	0	0	0	0	0	0

Table 3: List of robust optimal solution for Γ=0-5

From the table above we can see how the optimal value of objective function varies for different Γ. Generally, the greater is Γ the greater is objective function value. It is obvious that for Γ=0 and Γ=1 the storages are located in node 5 (Liberec) and node 10 (Teplice), for Γ greater than one the location of storages remains at nodes 5 (Liberec) and 6 (Litoměřice).

6 Conclusion

This paper deals with method which solves optimization problems under changing circumstances and reviews one of the modern approaches to modeling under uncertainty. The method is illustrated on allocation problem with uncertain demand, where we have to locate certain number of storages, considering the possibility of increased demand in several centers. The problem takes into account the fact that demand of certain centers could not be the same as predicted, but an increase in the form of unexpected demand can occur. The described potential increase in demand is considered only in a few centers, their number is given by parameter Γ. When considering parameter Γ we take into account maximum possible increase in total transportation costs associated with unexpected demand for given centers. In this paper we review the method, which transforms this non-linear problem into sequence of linear problems. The presented method is described on a case study, providing with computation experiments and its results.

Acknowledgements

This research was supported by the Czech Science Foundation, grant No. P403/12/1947 and by the project F4/11/2013 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Bertsimas, D., and Weismantel, R.: *Optimization over integers*, Dynamic Ideas, Belmont, 2005.

[2] Bertsimas, D., and Brown, D. B., and Caramanis, C.: *Theory and applications of Robust Optimization*, SIAM Rev., 2007, 53, (3), 464–501.

[3] Kouvelis, P., and Yu, G.: *Robust discrete optimization and its applications*, Kluwer academic publishers, Dordrecht, 1997.

[4] Sim, M.: *Robust Optimization by Melvyn Sim*, Massachusetts Institute of Technology, Massachusetts , 2004.

The estimation of an extended Kaldor model parameters with Mathematica

Jan Kodera[1], Quang Van Tran[1], Miloslav Vošvrda[2]

Abstract. In this article we try to analyse inflation dynamics using a four equation continuous dynamic model. The model is a traditional two-equation Kaldor model, extended by two additional equations: one describes an adaptive inflation expectations, the other the continuous dynamics of the money market. We expect instability of velocity of money, which in our case is dependent on the expected inflation. We use the quarterly data of the Czech economy from 2000 to 2014 to estimate the parameters of this model using non-linear estimation methods since the model is non-linear in its parameters. In our opinion, if the model is properly estimated, the result gives us a nonlinear model which is able to generate the complex dynamics as an alternative to the one of a traditional DSGE approach.

Keywords: four equation Kaldorian model, Nonlinear least squares method, Mathematica, Czech economy data

JEL classification: C51, E32
AMS classification: 62P20, 91B62

1 Introduction

Theoretical non-linear dynamic models are often inspired to be alternatives of their discrete counterparts in the literature. However, the parameters of these models are rarely estimated as is it not an easy task due to their continuous dynamics. This problem is often solved by discretization. Another problem connected with their estimation lies in the fact that these models are not only non-linear in variables, but often also in the parameters, which does not make its estimation a coveted target for researchers. Despite their estimation difficulty, if properly implemented, the result will give us a type of non-linear model that creates complex dynamics of economic variables resembling their actual behaviour including of aperiodic oscillations. To fill this gap in the literature, in the paper we conduct the estimation of the parameters of a continuous dynamic model consisting of four differential equations, some of which contain non-linearities is implemented using non-linear estimation technique.

The theory of continuous non-linear dynamical models is described in the famous book Guckenheimer-Holmes [**?**]. We can also meet a variety of modern works further developing present knowledge, like Kuznetsov [**?**], Perko [**?**]. The theory of non-linear dynamical systems is particularly interesting because the traditional relations as investment and consumer functions and many others are in principle of non-linear shape. Such non-linear functions are able to generate more complex dynamics, for example in [**?**], [**?**], [**?**] and others. We will work with a model that is based on the work Chiarella [**?**]. Chiarella's model is a two equations model with variables: expected inflation and inflation. In our paper we extend this model by adding two equations of the famous Kaldor model which can be found in [**?**].

The estimation of parameters of this four equation model is performed on quarterly data of the gross product, capital stock, investment and the inflation of the Czech economy from period 2000 - 2014. We solve it as an econometric problem, in which one has to deal with a model usually nonlinear not only in variables, but also in its parameters. The shape of the likelihood function can make some more difficulties, therefore the solution of the whole problem requires a proper selection of non-linear estimation methods

[1]University of Economics in Prague, Department of Statistics and Probability, nám. W. Churchilla 4, Praha 3 - 130 67, {kodera,tran}@vse.cz
[2]Institute of Information Theory and Automation, Department of Econometrics, Pod Vodárenskou věží 4, Praha 8, 182 08, vosvrda@utia.cas.cz

as well as the choice of the most suitable program environment to implement the estimation procedure. Both issues will be discussed in the paper.

2 Model Formulation

Model, which appears in Chiarella's book [**?**] is formed by two differential equations. The first equation is the equation for the dynamics of the money market, where the imbalance between supply and demand affect price developments. For example, if the money supply is greater than demand for money, the excess money generated on current accounts are used to buy commodities, which will raise prices and thus the demand for money increases and the initial excess money supply over demand is reduced. The reverse process occurs when the money supply is less than demand for money.

Fisher equation is the basic equation for modelling the money market in neoclassical economics. It determines the correct amount of money in circulation. This amount depends on the velocity of circulation of money, the price level and real output.

$$M^d(t)V(t) = P(t)Y(t). \tag{1}$$

Symbol M^d denotes the demand for money, V denotes velocity of circulation of money and P the price level and Y gross domestic product. All of these variables are function of time. Velocity of circulation of money in the neoclassical approach is understood as a given technological conditions. The change can only occur with a change in these conditions, so it is relatively slow. Faster a larger changes can occur in relation to the revolutionary changes in technology of payment.

The instability of money circulation velocity is often assumed in deterministic dynamical models. The velocity has its limits, so it can vary within a certain range between the lower and upper limits given technological parameters of circulation. Let K be its infimum and K_1 its supemum. The velocity of circulation of money is assumed to be an increasing function of expected inflation. Let us denote the expected inflation at time t by symbol $\pi(t)$. Then the velocity of money circulation can be written as a composite function:

$$V(t) = K\mathrm{e}^{\theta(\pi(t))}.$$

where θ is an inside function whose infimum is 0 and supremum is $\log K_1 - \log K$. Substituting it into equation (**??**) and after taking logarithm it gets the following form:

$$m^d(t) + k + \theta(\pi(t)) = y(t) + p(t), \tag{2}$$

where lower case letters stand for logarithms of upper case letter variable. Price dynamics will result from imbalances in the money market. We denote money supply by $M(t)$ and its logarithm by $m(t)$. To express the demand for money, we use equation (**??**), so we get

$$\dot{p}(t) = \gamma[m(t) + k + \theta(\pi(t)) - y(t) - p(t)]. \tag{3}$$

In our model the adaptive expectation of inflation is assumed, which is expressed by the following equation

$$\dot{\pi}(t) = \omega[\dot{p}(t) - \pi(t)]; \quad \omega \in (0, 1). \tag{4}$$

If actual inflation $\dot{p}(t)$ is higher than expectation of inflation $\pi(t)$, then the expectation increases. In the case of lower actual inflation, than the expectation inflation expectation decreases. Let us rearrange equation (**??**) by substituting $\dot{p}(t)$ in (**??**) into (**??**) and we get:

$$\dot{\pi}(t) = \omega\left\{\gamma\left[m(t) + k + \theta(\pi(t)) - y(t) - p(t)\right] - \pi(t)\right\}. \tag{5}$$

Differential equations (**??**) and (**??**) describe the dynamics of prices $p(t)$ and expected inflation $\pi(t)$ using the imbalances in the money market (equation (**??**)) and the adaptive expectations of inflation (equation (**??**)). Variables $m(t)$ and $y(t)$ are exogenous variables. Let the quantity $m(t)$ be a constant and the variable $y(t)$ becomes endogenous variable extending the model by two differential equations of so called Kaldor model of the investment cycle. The first differential equation describes the dynamics of the commodity market. The original version of this equation is formulated as

$$\dot{Y}(t) = \alpha[I(Y(t), K(t)) - S(Y(t))], \tag{6}$$

where $K(t)$ is the capital stock and $I(Y(t), K(t))$ investment as an increasing function of production $Y(t)$ and decreasing function of capital $K(t)$, Savings are a function of production (income) $Y(t)$. Investment function I is modified as a product of the level of the propensity to investment $j(.)$ and production $Y(t)$. Function $j(.)$ is increasing function of the productivity of capital $Y(t)/K(t)$.

$$I(Y(t), K(t)) = j\left(\frac{Y(t)}{K(t)}\right) Y(t) = j\left(e^{y(t)-k(t)}\right) Y(t) = i(y(t) - k(t))Y(t),$$

where $i = j \circ \exp$ and $k(t) = \log K(t)$, $y(t) = \log Y(t)$. Savings function fulfils the condition that from a certain level of income is increasing and concave what corresponds to economic experience.

$$S(Y(t)) = (s_0 + s_1 \log Y(t))Y(t) = (s_0 + s_1 y(t))Y(t).$$

Substituting from the above two equations into equation (**??**) and dividing both sides of equation $Y(t)$ we get:

$$\dot{y}(t) = \alpha[i(y(t) - k(t)) - (s_0 + s_1 y(t))]. \tag{7}$$

The second equation of Kaldor's model describes capital formation.

$$\dot{K}(t) = i(y(t) - k(t))Y(t) - \delta_0 K^{1+\delta_1}(t),$$

where $\delta_0 K^{1+\delta_1}$ is a function describing depreciation. Dividing both sides of the equation by $K(t)$ we get in logarithms

$$\dot{k}(t) = i(y(t) - k(t))e^{y(t)-k(t)} - \delta_0 e^{\delta_1 k}. \tag{8}$$

Equations (**??**) and (**??**) form modified Kaldor model.

3 Specification of non-linearities in the model

In the model formed by equations (**??**), (**??**), (**??**) and (**??**) there are two non-linear functions. They are function $\theta(\pi(t))$ which is the logarithm of the velocity of circulation of money (up to a constant) and the propensity to invest function $i(y(t) - k(t))$. The velocity of circulation of money is an increasing function of expected inflation defined on the interval $(-\infty, \infty)$ and its range is restricted on interval $[K, K_1]$. The inside function θ is assumed to be a logistic function proposed as:

$$\theta(\pi(t)) = \frac{\theta_0}{1 + \theta_1 e^{-\theta_2 \pi(t)}}; \quad \theta_\mu > 0,\ \mu = 0, 1, 2.$$

The infimum of the above function is equal 0, its supremum is equal $\theta_0/(1 + \theta_1) = \log K_1 - \log K$.

Propensity to invest function is also logistic in our model.

$$i(y(t) - k(t)) = \frac{i_0}{1 + i_1 e^{-i_2(y(t)-k(t)}}; \quad i_\mu > 0.\ \mu = 1, 2, 3$$

The infimum of propensity to invest function is 0 and its supremum is $\frac{i_0}{1+i_1}$. Substituting specified non-linearities for θ and i in equations (**??**), (**??**), (**??**) and (**??**) we get

$$\dot{p}(t) = \gamma\left[m(t) + k + \frac{\theta_0}{1 + \theta_1 e^{-\theta_2 \pi(t)}} - y(t) - p(t)\right] \tag{9}$$

$$\dot{\pi}(t) = \omega\left\{\gamma\left[m(t) + k + \frac{\theta_0}{1 + \theta_1 e^{-\theta_2 \pi(t)}} - y(t) - p(t)\right] - \pi(t)\right\} \tag{10}$$

$$\dot{y}(t) = \alpha\left[\frac{i_0}{1 + i_1 e^{-i_2(y(t)-k(t)}} - (s_0 + s_1 y(t))\right] \tag{11}$$

$$\dot{k}(t) = \frac{i_0}{1 + i_1 e^{-i_2(y(t)-k(t)}} e^{y(t)-k(t)} - \delta_0 e^{\delta_1 k(t)} \tag{12}$$

The parameters of the model formed by equations (**??**), (**??**), (**??**) and (**??**) will be estimated using data of Czech economy in the next section.

4 Parameter estimation

The parameters of the four equation Kaldor model described above are estimated from the Czech economic data. We use public available quarterly data on real GDP, money aggregate M2, investment and quarter to quarter inflation rates from period 2000 - 2014. Regarding capital stocks data, as only yearly data for capital stocks are available, we use quarterly data on investment and depreciation to calculate the missing values at the end of the first three quarters in a year to get quarterly data on capital stocks. The descriptive statistics of data used for parameters estimation are shown in Table 1.

	GDP	Aggreggate M2	Cap. stock	Investment	Inflation
Mean	803899.5	244550.3	23244631	2079574	100.0263
Median	828681.0	242469.0	23216414	2040597	100.1000
Maximum	936276.0	322648.0	26310589	3126444	101.0000
Minimum	591380.0	178900.0	19776757	1099047	99.20000
Std. devi	105719.9	34861.19	2110535	654439	0.342536
Skewness	-0.401773	0.110802	-0.039520	0.029389	-0.086410
Kurtosis	1.747132	2.146762	1.634164	1.471155	3.452682
Jarque-Berra	5.261488	1.845671	4.445421	5.559453	0.557621
Probability	0.072025	0.397391	0.108315	0.062055	0.756683
Observations	57	57	57	57	57

Table 1 Data descriptive statistics

4.1 Parameter estimation of capital formation equation

We use data for capital and production to estimate five parameters $i_0, i_1, i_3, \delta_0, \delta_1$ in equation (**??**) in Mathematica. For minimization of sum of non-linear least squares which is in our case a non-convex function of parameters we use Mathematica procedure ArgMin. For their corresponding confidence intervals, a non-parametric bootstrapping approach methods is applied. For this purpose procedures EmpiricalDistribution and Quantile are used. The estimation results and two-sided 95% confidence intervals are displayed in Table **??**.

Parameter	Estimate	0.025-Quantile	0.975-Quantile
i_0	0.201644	0.184421	0.276908
i_1	-3.53261	-3.57472	-3.45759
i_2	15.2786	6.21251	27.6901
δ_0	1.57668	-6.90258	8.61736
δ_1	-1.31202	-8.16069	-1.24718

Table 2 Estimation and significance of parameters for capital formation equation

Table **??** shows that the estimates of i_0, i_1, i_2 are statistically significant while the estimates of δ_0, δ_1 are not. In our opinion statistical insignificance of parameters δ_0 and δ_1 is probably caused by nature of capital stock data, which are always less reliable than other data used for estimations purpose.

4.2 Parameters estimate for production dynamics equation

Production dynamics is defined by equation (**??**) which contains an expression for propensity to invest

$$\frac{i_0}{1 + i_1 e^{-i_2(y(t)-k(t)}} \tag{13}$$

whose parameters are estimated in the subsection **??**. These values are further used for the numerical calculation of that expression. The calculated value of propensity to invest expressed by (**??**) together

Parameter	Estimate	Standard Error	t-Statistic	P-Value
α	2.21465	0.372078	5.95211	$2.15765 * 10^-7$
αs_0	-2.09031	0.871809	-2.39767	0.0200549
αs_1	0.179532	0.0664417	2.7021	0.00923525

Table 3 Estimation and significance of parameters for production dynamic equation

with logarithm of production y are substituted into equation (**??**) which takes a linear form in the parameters $\alpha, \alpha s_0, \alpha s_1$. These parameters are estimated by Mathematica procedures LinearFit. The estimation results of equation (**??**) parameters are shown in table **??**. The results of tests are statistically significant. They can be used for further work.

4.3 Parameters estimate of the equation for adaptive inflation expectation

Adaptive expectation of inflation is displayed by equation (**??**). Let approximate $\dot{\pi}(t)$ by $\pi_{t+1} - \pi_t$ and $\dot{p}(t)$ by $\Delta p_t = p_t - p_{t-1}$. The equation turns

$$\pi_{t+1} - \pi_t = \lambda[\Delta p_t - \pi_t]$$

or

$$\pi_{t+1} = \lambda \Delta p_t + (1 - \lambda)\pi_t],$$

where $\lambda \in (0.1)$. Solving the equation recurrently with respect to π_{t+1} we get

$$\pi_{t+1} = \lambda \sum_{k=0}^{\infty} (1 - \omega)^k \Delta p_{t-k}.$$

For our experiment we use approximation for $k = 0, 1$ of the above equation For the purpose of estimation we substitute Δp_t for π_{t+1}, thus

$$\Delta p_{t+1} = \lambda \Delta p_t + \lambda(1 - \lambda)\Delta p_{t-1}. \tag{14}$$

To estimate λ, we use series of price indices for Czech economy and the bootstrap method is used to obtain its statistical significance. The estimation results are shown in Table **??**. λ is statistically significant, so it can be used into equation (**??**).

Parameter	Estimate	0.025-Quantile	0.975-Quantile
λ	0.264801	0.12941	0.485373

Table 4 Estimation results for adaptive inflation expectation

4.4 Parameters estimate of the equation for price dynamics

First we experiment with estimation of equation (**??**). This experiment brings unsatisfactory results. The estimates are statistically insignificant and economically uninterpretable. Therefore we have decided to estimate equation (**??**) in a linear form as follows

$$\dot{p}(t) = c_0 + c_1 y(t) + c_2 p(t),$$

which is slightly different from equation (**??**). The estimation results are displayed in Table **??**. We notice a problem with parameter c_2 where its p-value is higher than 0.05, namely it is $0.071756 < 0.1$, which can be also accepted with looser tolerance.

Parameter	Estimate	Standard Error	t-Statistic	P-Value
c_0	0.369006	0.161683	2.28228	0.0265141
c_1	-0.040378	0.0176408	-2.28889	0.0260989
c_2	0.0408325	0.0222227	1.83743	0.071756

Table 5 Estimation and significance of parameters for price dynamics

5 Conclusion

The objective of this paper is to investigate the feasibility of estimation of the parameters of a non-linear dynamics model. Such models are non-convex in the parameters, therefore methods for global minimization have to be used. There are several professional packages offering such methods which do not always give precise results. We have tested some of them and decided to use package Mathematica and its procedure ArgMin which, in our opinion, provides quite reliable results. We also apply the linear procedures of Mathematica, whenever we estimate parameters of equations which are linear in parameters. The results we have obtained at this stage are not always satisfactory, especially with regard to the statistical significance of the estimates. For the time being these preliminary results of our experiment can be accepted as a first step of the development of non-linear dynamical inflation modelling. Despite many difficulties of estimation process, it can be a promising alternative to the DSGE approach. It can give room to many future research topics whose objective is to improve formulation of these models and to find more progressive non-linear methods for their estimation.

Acknowledgements

The support from the Czech Science Foundation under Grant P402/12/G097 is gratefully acknowledged.

References

[1] Flaschel, P., Franke, R., Semmler, W.: *Dynamic Macroeconomics*. MIT Press, Cambridge, Massachusetts, 1997.

[2] Gabisch , G., Lorenz, H. W.: *Business Cycle Theory*. Springer Verlag, Berlin, 1989.

[3] Guckenheimer, J., Holmes, P.: *Non-Linear Oscillation, Dynamical Systems and Bifurcations of Vector Fields*. Springer Verlag, New York, 1986.

[4] Chiarella, C.: *The Elements of a Nonlinear Theory of Economic Dynamics*. Springer Verlag, Berlin, 1990.

[5] Kaldor, N.: A Model of the Trade Cycle, *Economic Journal* **50** (1940), 78–92.

[6] Kodera, J., Tran, Q. V., Vošvrda, M.: Complex Price Dynamics in the Modified Kaldorian Model, *Prague Economic Papers* **2013/3**, 358–384.

[7] Kuznetsov, Y. A.: *Elements of Applied Bifurcation Theory*. Springer Verlag, Berlin, 1998.

[8] Lorenz, H. W.: *Nonlinear Dynamical Economics and Chaotic Motion*. Springer Verlag, Berlin, 1994.

[9] Martinez, W. L., Martinez, A. R.: *Computational Statistics Handbook with MATLAB*, Chapman & Hall, Bo Raton, 2007.

[10] Perko, L. : *Differential Equations and Dynamical Systems*. Springer Verlag, Berlin, 2001.

Modeling and handling of uncertain demand in counting zones tariff system

Michal Kohání[1]

Abstract. One of various approaches to the tariff system in public transport is the zone tariff, where the whole region is divided into the smaller sub-regions - tariff zones. The price of travelling in such system depends on the number of travelled zones. Designing of counting zone tariff system is connected with the decisions about optimal price for travelling and optimal zone partitioning of the region. One of the fundamental input data in the zone tariff planning is origin-destination (OD) matrix as the demand matrix. However, it is often not known exactly, but it can have a major influence on the system design. In this paper we propose the integer programming approach to solve the problem with uncertain demand, where we use the theory of fuzzy sets to model the OD matrix. We formulate mathematical model of the problem and to test behavior of proposed approach we make a computational study using a universal IP-solver on the test data from selected region and we evaluate the influence of uncertain demand on system design.

Keywords: tariff planning, tariff zones design, IP solver, fuzzy approach.

JEL Classification: C44, D81, R42
AMS Classification: 90C10, 90C035, 90C70

1 Introduction

When transport authorities plan the regional public transportation, one of the problems they deal with is the problem of the tariff and the ticket prices. As was mentioned in [2] and [6], one of possible tariff system is *zone tariff*. In this system the region is divided into smaller sub-regions - tariff zones and the price for travelling depends on the origin and the destination zone and usually on the number of travelled zones on the trip. Zone tariff systems can be divided into two groups. One of them is *the counting zone tariff system*, where the price of trip is calculated according to the origin zone, the destination zone and the number of travelled zones. For all trips hold that the price for trips passing the same number of zones must be equal. Example of the counting zone tariff system in Bratislava region in Slovakia is in the Figure 1.

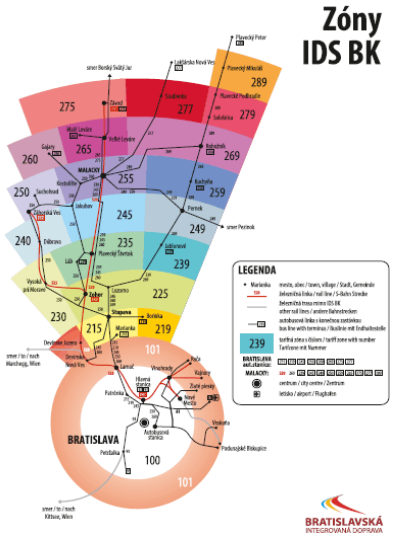


Figure 1: *Example of the counting zones tariff system in Bratislava region (www.bid.sk)*

Another important task in the tariff planning is designing of tariff zones and new fares. Hamacher and Schöbel in [2], Schöbel in [6] proposed approaches for the zone design problem with arbitrary prices. Hamacher and Schöbel in [2] and Schöbel in [6] mentioned solving of the counting zones tariff system where the goal is to

[1] University of Žilina, Department of Mathematical Methods and Operations Research, Univerzitná 8215/1, 01026 Žilina, Slovakia, michal.kohani@fri.uniza.sk.

design the zones so that the new and the old price for most of the trips are as close as possible. A note on fair fare tariff was mentioned also by Paluch in [5].

One of the fundamental input data in the zone tariff planning is origin-destination (OD) matrix. However, this matrix is often not exactly known, but it can have a major influence on the system design. We can use various approaches to model this uncertain OD matrix. One of them is to use the means of fuzzy set theory [7] [8]. Uncertain values can be modeled as triangle fuzzy numbers and the objective function can be transformed to the fuzzy inequality with the level of satisfaction *h*. Then the model can be solved using Tanaka-Asai's approach to maximize the level of satisfaction and find appropriate solution of the counting zone tariff system.

This paper will be organized as follows. In the section 2, we present the model of the tariff zones design problem with the average deviation criterion for counting zones tariff system and the way of linearization of the model. In the chapter 3 we describe the transformation of the problem to the form which models uncertain demand in OD matrix using the theory of fuzzy sets. In the chapter 4 we perform computational study using universal IP-solver on the test data from selected region and we evaluate the influence of uncertain demand on system design.

2 Mathematical model of the tariff zones design problem

Let all stops/stations in the public transport network constitute the set of nodes *I*. The stop *i* and *j* from the set *I* are connected by the edge $(i,j) \in E$, if there is direct connection by public transport line between these two stops. Symbol *E* denotes the set of edges. The distance between stops *i* and *j* is denoted as d_{ij}. For each pair of stops *i* and *j* is c_{ij} the current price of travelling between these two stops. We assume that for each pair of stops *i* and *j* $c_{ij}=c_{ji}$. The number of passengers between stops *i* and *j* is b_{ij} (OD matrix).

If we want to calculate new price of the trip between nodes *i* and *j*, we need to calculate the number of zones crossed on this trip. This calculation can be easily replaced by the calculation of crossed zone borders as was used in [2] and [4]. We assume that the stop can be assigned only to one zone and then the border between zones is on the edge. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in E$, which is equal to 1 if stops *r* and *s* are in different zones and is equal to 0 otherwise. For calculation of the number of crossed borders we need to determine the used path for travelling between stops *i* and *j*. We introduce parameter a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge *(r,s)* is used for travelling from stop *i* to stop *j* and 0 otherwise.

When we want to set a new price for travelling in proposed system, there are more possibilities how to do it. In this paper we use two different unit prices, as was mentioned in [3] – price f_1 for travelling in the first zone and unit price f_2 for travelling in each additional zone. The final new price will be calculated as the sum of price for the first zone and number of other travelled zones multiplied by the unit price for additional zones. New price n_{ij}, determined by the number of crossed zones will be calculated according to the definition (1) as follows:

$$n_{ij} = f_1 + \sum_{(r,s)\in E} f_2 a_{ij}^{rs} w_{rs} \tag{1}$$

Construction of the zone tariff model was inspired by the model of the p-median problem. We introduce binary variables y_i, which represent the "fictional" centre of the zone. Variable y_i is equal to 1 if there is a centre of the zone in node *i* and 0 otherwise. For each pair of stations *i* and *j* we introduce variable z_{ij}. Variable z_{ij} is equal to 1 if the station *j* is assigned to the zone with centre in the node *i* and 0 otherwise. We expect to create at most *p* tariff zones. In our model we will use the average deviation between current and new price for all passengers, as was recommended by experts in [6]. The current (or fair) price between stations *i* and *j* will be denoted by c_{ij}. The mathematical model of zone tariff problem with fixed prices and number of zones can be written in following form (2) – (9):

$$Minimize\ dev_{avg} = \frac{\sum_{i\in I}\sum_{j\in I}\left|c_{ij}-n_{ij}\right|b_{ij}}{\sum_{i\in I}\sum_{j\in I} b_{ij}} \tag{2}$$

$$subject\ to \sum_{i\in I} z_{ij} = 1,\ for\ j \in I \tag{3}$$

$$z_{ij} \le y_i, for i, j \in I \tag{4}$$

$$z_{ij} - z_{ik} \le w_{jk}, for i \in I, (j,k) \in E \tag{5}$$

$$\sum_{i \in I} y_i \le p \tag{6}$$

$$z_{ij} \in \{0,1\}, for\ i,\ j \in I \tag{7}$$

$$y_i \in \{0,1\}, for\ i \in I \tag{8}$$

$$w_{ij} \in \{0,1\}, for\ (i,j) \in E \tag{9}$$

Conditions (3) ensure that each stop will be assigned to only one zone exactly. Conditions (4) ensure that the stop *j* will be assigned only to existing centre of the zone. Conditions (5) are coupling constraints between variables for allocation of the stop to zone and variables for determining the zone border on the edge *(j,k).* Condition (6) ensures that we will create at most *p* tariff zones.

Because the objective function (2) in this model is not a linear function, we need to modify this objective function to the linear form. We introduce new variables u_{ij} and $v_{ij.}$ Variables u_{ij} represent calculated prices for travelling in case that the new price is lower than current one and variables v_{ij} represent the calculated prices for the opposite case. Then we can reformulate mathematical model (2) – (9) to the linear form:

$$Minimize\ F = \frac{\sum_{i \in I}\sum_{j \in I} u_{ij} b_{ij} + \sum_{i \in I}\sum_{j \in I} v_{ij} b_{ij}}{\sum_{i \in I}\sum_{j \in I} b_{ij}} \tag{10}$$

$$subject\ to\ (3) - (9)$$

$$c_{ij} - n_{ij} = u_{ij} - v_{ij}, for i, j \in I \tag{11}$$

$$u_{ij} \ge 0, for i, j \in I \tag{12}$$

$$v_{ij} \ge 0, for i, j \in I \tag{13}$$

3 Modeling of uncertain demand in tariff zones design problem

We can use various approaches to model uncertain demand in our problem. One of them is to use the means of the fuzzy set theory [7] [8]. Uncertain demand can be modeled as a triangle fuzzy number $\underline{b_{ij}} = <b_{ij}^1, b_{ij}^2, b_{ij}^3>$ with its associated normalized membership function $\mu(b_{ij}) \in <0,1>$.

Based on [8] the minimization process of optimized objective function can be replaced by the intention that the objective function value of the resulting solution should belong to the fuzzy set of sufficiently small objective function values at as high level of satisfaction *h* as possible. To construct the fuzzy set of sufficiently small objective function values, we determine two real crisp values F_1 and F_2 of the original objective function, where each value lower than F_1 can be considered as certainly big and each value higher than F_2 can be considered as certainly small. Values F_1 and F_2 can be determined by crisp optimization for the best (with the usage of b_{ij}^1 in the objective function) and worst (with the usage of b_{ij}^3 in the objective function) cases of the problem. Situation is described by the Figure 2.

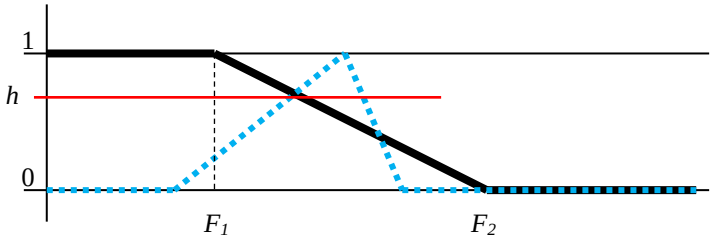


Figure 2: *Fuzzy inequalities and objective function transformation*

Result of the objective function transformation to the form of fuzzy inequality is described by (15) and this transformation is based on rules described in [8]. Mathematical model of finding the highest level of satisfaction h can be described as follows:

$$Maximize\ h \tag{14}$$

$$subject\ to\ (1-h)\left(\frac{\sum_{i\in I}\sum_{j\in I}u_{ij}b_{ij}^1+\sum_{i\in I}\sum_{j\in I}v_{ij}b_{ij}^1}{\sum_{i\in I}\sum_{j\in I}b_{ij}^1}\right)+h\left(\frac{\sum_{i\in I}\sum_{j\in I}u_{ij}b_{ij}^2+\sum_{i\in I}\sum_{j\in I}v_{ij}b_{ij}^2}{\sum_{i\in I}\sum_{j\in I}b_{ij}^2}\right)\le hF_1+(1-h)F_2 \tag{15}$$

$$subject\ to\ (3)-(9),(11)-(13),h\ge 0$$

This model can be solved using Tanaka-Asai's approach which maximizes the level of satisfaction and finds appropriate solution of the counting zone tariff system. We gradually increase the value of h and for given value of h we solve linear problem (3) – (9), (11) – (13), (15) with original decision variables and original objective function (10). The process terminates, when no feasible solution exists. Optimal solution obtained for the last (highest) value of h is the resulting solution of the problem.

4 Computational study

The goal of the computational study is to search for optimal solutions of zone tariff problem under uncertain demand represented by OD matrix. Numerical experiments were performed on the test data set created from the real public transport network in the Zvolen Region in Slovak Republic. The stops in the networks are represented by the municipalities or part of municipalities. Test networks have 25 and 51 stops/municipalities and are shown in the Figure 3. Black circles represent stops in the network, the size of the circle represents approximate number of inhabitants and links represent existing direct connections between stops (regional buses).

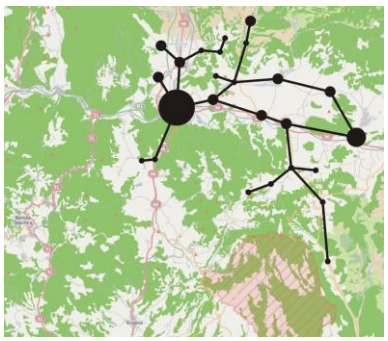
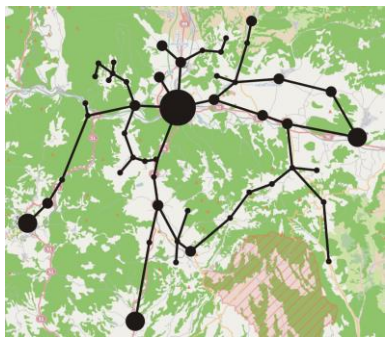
(a)  (b) 

Figure 3: *Test networks with 25 stops (a) and 51 stops (b), Map source:* openstreetmap.org

Current prices were calculated according to real prices which depend on the distance for travelling by regional buses. The original OD matrix was estimated using the gravity model as in [1], where the number of passengers between nodes i and j is calculated as follows:

$$\frac{b_i b_j}{d_{ij}}$$

where the parameter b_i represents the number of inhabitants in the node i. The experiments were performed in Xpress [9] on a computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM.

In the computational study, values of parameters f_1, f_2 and p are optimal values of these parameters which were obtained in previous research [3] [4]. To study behavior of fuzzy approach to solve this problem with uncertain demand, we use various triangle fuzzy numbers $<b_{ij}^1, b_{ij}^2, b_{ij}^3>$ to describe OD matrix. First setting *Fuz1* is described by triangle fuzzy number $<0.8*b_{ij}, b_{ij}, 1.2*b_{ij}>$, where $0.8*b_{ij}$ means 80% from original values in OD matrix etc. Second setting, denoted as *Fuz2*, is described by triangle fuzzy number $<0.9*b_{ij}, b_{ij}, 1.2*b_{ij}>$ and setting *Fuz3* is described by triangle fuzzy number $<0.8*b_{ij}, b_{ij}, 1.1*b_{ij}>$. In settings *Fuz4* and *Fuz5* we changed the main value of the triangle fuzzy number to $0.95*b_{ij}$ (*Fuz4*) or $1.05*b_{ij}$ (*Fuz5*) respectively, left and right values are equal to left and right values of triangle fuzzy numbers in *Fuz1*.

Table 1 represents results of numerical experiments for network with 25 stops and Table 2 represents solutions of numerical experiments for network with 51 stops. Values of parameter p and optimal prices f_1^*, f_2^* were obtained in [4]. Rows denoted as F^* represent the best average deviation calculated according to the (10). Rows denotes as *Time [s]* represent computational time and rows denoted as h represent best obtained values of the level of satisfaction h. Columns denoted as *Fuz1 – Fuz5* represent mentioned settings of triangle fuzzy numbers.

p_max	***f_1****	***f_2****		***Fuz1***	***Fuz2***	***Fuz3***	***Fuz4***	***Fuz5***
4	0.8	0.6	***F****	4940.00	4940.00	4940.00	4792.37	6964.00
			Time [s]	82.56	88.38	83.96	84.79	80.27
			h	0.87	1.00	0.85	0.96	0.87
10	0.7	0.4	***F****	3345.00	3345.00	3345.00	3177.00	3512.25
			Time [s]	36.74	38.95	34.40	36.85	33.60
			h	0.70	0.75	0.65	0.77	0.65
16	0.6	0.2	***F****	4489.10	4489.10	4489.10	4264.64	4713.55
			Time [s]	25.39	26.91	26.40	26.38	24.85
			h	0.75	0.89	0.70	0.82	0.70
20	0.5	0.1	***F****	6949.70	6949.70	6949.70	6602.22	7297.18
			Time [s]	21.46	22.31	21.01	22.13	21.00
			h	0.67	0.76	0.60	0.73	0.62
22	0.4	0.1	***F****	9093.50	9093.50	9093.50	8638.82	9548.18
			Time [s]	23.83	24.57	23.59	24.21	22.92
			h	0.93	1.00	0.91	1.00	0.83

Table 1 *Results for the test network with 25 stops*

p_max	***f_1****	***f_2****		***Fuz1***	***Fuz2***	***Fuz3***
10	0.8	0.5	***F****	9505.82	9702.08	9458.77
			Time [s]	419.38	481.31	401.10
			h	0.81	1.00	0.77
16	0.8	0.4	***F****	9511.50	9414.00	9654.80
			Time [s]	405.69	524.85	395.39
			h	0.69	1.00	0.62
20	0.7	0.4	***F****	7735.71	7735.71	7735.71
			Time [s]	197.40	214.44	204.41
			h	0.68	0.85	0.73
25	0.7	0.3	***F****	8356.17	8356.17	8356.17
			Time [s]	222.12	247.80	226.57
			h	0.66	0.76	0.60
30	0.7	0.2	***F****	8704.85	8704.85	8704.85
			Time [s]	144.89	150.19	141.03
			h	0.91	0.76	0.89

Table 2 *Results for the test network with 51 stops*

5 Conclusion

In this paper we proposed integer programming approach to solve the tariff zone design problem with uncertain demand. We formulated mathematical model of the problem and appropriate linearization of the non-linear objective function. To handle uncertainty of OD matrix in the model we used the theory of fuzzy sets with triangle fuzzy numbers and Tanaka-Asai's approach. To test behavior of proposed approach we made a computational study using a universal IP-solver on the test data from selected region.

In the Table 1 we can find results of numerical experiments for the test network with 25 stops and in the Table 2 are results for the test network with 51 stops. In both networks we obtained highest values of h with the triangle fuzzy number setting *Fuz2*. This is caused by the small difference between left value b_{ij}^1 and main value b_{ij}^2 of the triangle fuzzy number.

As we can see from results, the results are consistent with inputs. The more we decrease left value b_{ij}^1 of the triangle fuzzy number, the level of satisfaction h decreases, as we can see from the comparison of values h in settings *Fuz1* and *Fuz2* in both tables. The same effect on the value h has the change of the main value b_{ij}^2 as we can see from the comparison of columns *Fuz4* and *Fuz5* in the Table 1. From the point of view of the value of objective function, in most cases we obtained equitable solutions, as we can see from the values of F^*.

Computational time of all mentioned experiments is in acceptable range, we can see, that smaller value of p can increase the computational time several times in comparing with obtained computational times in the cases with bigger value p.

Acknowledgements

This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks". We would like to thank to "*Centre of excellence for systems and services of intelligent transport*" (ITMS 26220120050) for built-up infrastructure, which was used.

References

[1] Cenek. P., Janacek. J., Janosikova. L., 2002, Location of Transportation Districts at Modeling of Tranportation Processes in a Region. *Communications - Scientific Letters of the University of Zilina* **4**, No. 1-2, 5-9.

[2] Hamacher. H. W., Schöbel. A.: Design of Zone Tariff Systems in Public Transportation. *Operations Research* **52** (2004), 897-908.

[3] Kohání, M.: Designing of zone tariff in integrated transport systems. In: *Communications: scientific letters of the University of Žilina* **15** (2013), 29-33.

[4] Kohání. M.: Zone partitioning problem with given prices and number of zones in counting zones tariff system. *In: SOR ´13: proceedings of the 12th International Symposium on Operational Research: Dolenjske Toplice, Slovenia* (2013), 75-80.

[5] Palúch. S.: On a fair fare rating on a bus line. In: *Communications: scientific letters of the University of Žilina* **15** (2013), 25-28.

[6] Schöbel. A.: *Optimization in Public Transportation: Stop Location. Delay Management and Tariff Zone Design in a Public Transportation Network*, Springer, 2006.

[7] Teodorović. D., Vukadinović. K.: *Traffic Control and Transport Planning: A Fuzzy Sets and Neural Networks Approach*. Kluver Academic Publishers, Boston, Dordrecht, London, 1998.

[8] Ramík. J., Vlach. M.: *Generalized Concativity in Fuzzy Optimization and Decision analysis*. Kluwer Academic Publishers, Boston, 2002.

[9] *XPRESS-Mosel User guide*, Fair Isaac Corporation, Birmingham, 2012.

Out-of-sample SSD efficiency of mean-CVaR efficient portfolios

Miloš Kopa[1]

Abstract. The paper deals with out-of-sample efficiency of particular portfolios with respect to the second order stochastic dominance (SSD). Firstly, using Conditional Value at Risk (CVaR) as the measure of risk, we compute three mean-risk efficient portfolios from the mean-CVaR in-sample efficiency frontier. These portfolios are identified as the optimal solutions of the CVaR minimizing problem under the condition on minimal required portfolio mean return. Three levels of minimal required portfolio mean return are considered. Then we test out-of-sample efficiency of all these portfolios with respect to the second order stochastic dominance. We apply this procedure to US stock market daily data 1927-2014 such that we use one year moving window period for both in-sample and out-of-sample analysis. Hence we identify 87 in-sample sets of three mean-CVaR efficient portfolios and we test SSD efficiency of all these 261 portfolios. Moreover, we analyze the out-of-sample efficiency (measure of SSD inefficiency) for each minimal required level of portfolio mean return separately.

Keywords: Mean-CVaR model, Second order stochastic dominance, portfolio efficiency, out-of-sample analysis

JEL classification: D81, G11
AMS classification: 91B16, 91B30

1 Introduction

When solving portfolio selection problem several approaches can be used: mean-risk models, maximizing expected utility problems, stochastic dominance criteria, etc. In this paper we employ the mean-CVaR model as a special case of the general mean-risk model and we combine it with the efficiency with respect to the second order stochastic dominance. One may adopt stochastic dominance approach to test an efficiency of a given portfolio with respect to considered set of utility functions if the information about the risk attitude of a decision maker is not perfectly known. The second order stochastic dominance is the most common stochastic dominance relation because of its risk aversion interpretation and its relation to CVaR, see e.g. Ogryczak and Ruszczynski (2002), Levy (2006), Kopa and Chovanec (2008), Dupačová and Kopa (2012) or Kopa and Post (2015).

In the context of portfolio selection problem, Post (2003) and Kuosmanen (2004) firstly develop linear programming tests for testing if a given portfolio is SSD efficient relative to all possible portfolios formed from a set of assets. Applying first-order optimality condition for portfolio optimization with concave utility functions, Post (2003) derives a computationally efficient LP test. Using the formulation in terms of cumulative returns, Kuosmanen (2004) derives a test that identifies another, SSD efficient portfolio that, moreover, dominates the evaluated portfolio (if the latter is inefficient). Later on, Kopa and Chovanec (2008) propose a new linear programming test in terms of conditional values of risk. Recently, Dupačová and Kopa (2012) or Kopa and Post (2015) introduce modifications allowing for a general discrete probability distribution and general weighting scheme. These SSD efficiency tests can also be seen as a special type of DEA models, see Branda and Kopa (2012, 2014) for more details. Finally, Lizyayev (2012) and Grechuk (2014) present the additional SSD efficiency testing ideas and results.

In this paper, we analyze the out-of-sample performance of particular in-sample mean-CVaR efficient portfolios. For three different levels of minimal required portfolio mean return, we first compute in-sample

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, kopa@karlin.mff.cuni.cz

mean-CVaR efficient portfolios for 95% level of CVaR. It would be easy and straightforward to analyze out-of-sample performance in the sense of mean-CVaR, however, it is unlikely to happen that in-sample mean-CVaR efficient portfolio remains out-of-sample efficient at the same level of CVaR. Therefore we test out-of-sample SSD efficiency of these three portfolios, what can be seen as a modification of mean-CVaR efficiency when all possible levels of CVaR are considered. We slightly modify Kopa and Post (2015) test to guarantee that a SSD inefficient portfolio is mean-CVaR efficient for no level of CVaR. Hence, only SSD efficient portfolios may be mean-CVaR efficient for some level of CVaR.

We use one-year moving window for both in-sample and out-of-sample analysis. We consider US industry benchmark portfolios from the Kenneth French library and their daily returns from 1927 to 2014. Hence we consider 87 in-sample periods and 87 one year-ahead out-of-sample periods. Firstly, we identify out-of-sample SSD efficient portfolios. Secondly, we compare the SSD inefficiency measures of other portfolios for the particular years between each other. Finally, we present a comparison of out-of-sample SSD efficiency behavior for three different levels of minimal required portfolio mean return.

The remainder of this paper is structured as follows. Section 2 presents notation, basic definitions of conditional value at risk as well as of the second order stochastic dominance. It is followed by a summary of Kopa and Post (2015) SSD portfolio efficiency test and its modification in Section 3. Section 4 shows the results of the empirical part. The paper is summarized and concluded in Section 5.

2 Preliminaries

Consider a random vector $\mathbf{r} = (r_1, r_2, ..., r_N)'$ of returns of N assets in T equiprobable scenarios. The returns of the assets for the various scenarios are given by

$$X = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^T \end{pmatrix}$$

where $\mathbf{x}^t = (x_1^t, x_2^t, \ldots, x_N^t)$ is the t-th row of matrix X. We will use $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)'$ for a vector of portfolio weights and the portfolio possibilities are given by

$$\Lambda = \{\boldsymbol{\lambda} \in R^N | \mathbf{1}'\boldsymbol{\lambda} = 1, \ \lambda_n \geq 0, \ n = 1, 2, \ldots, N\}.$$

We follow Pflug (2000) and Rockafellar Uryasev (2000, 2002) in defining *conditional value-at-risk* (CVaR) for portfolio losses $(-\mathbf{r}'\boldsymbol{\lambda})$.

Definition 1. Let $\alpha \in \langle 0, 1\rangle$. Conditional value-at-risk of portfolio $\boldsymbol{\lambda} \in \Lambda$ at level α is the optimal value of objective function of the following optimization problem:

$$\mathrm{CVaR}_\alpha(\boldsymbol{\lambda}) = \min_{a \in \mathbb{R}} \{a + \frac{1}{1-\alpha}\mathbb{E}[-\mathbf{r}'\boldsymbol{\lambda} - a]^+\}$$

where $[x]^+ = \max(x, 0)$.

Since we use scenario approach, it can be rewritten as a linear programming problem:

$$\begin{aligned} \mathrm{CVaR}_\alpha(\boldsymbol{\lambda}) &= \min_{a, w_t} \ a + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T} w_t \\ s.t. \quad w_t &\geq -\mathbf{x}^t\boldsymbol{\lambda} - a \\ w_t &\geq 0 \end{aligned} \tag{1}$$

and adding condition on the mean return of the portfolio we get the following formulation of mean-CVaR

model:

$$\begin{aligned} \min_{a,w_t,\boldsymbol{\lambda}\in\Lambda} \quad & a + \frac{1}{(1-\alpha)T}\sum_{t=1}^{T} w_t \\ \text{s.t.} \quad w_t &\geq -\mathbf{x}^t\boldsymbol{\lambda} - a \\ \sum_{t=1}^{T} T^{-1}\mathbf{x}^t\boldsymbol{\lambda} &\geq \mu \\ w_t &\geq 0. \end{aligned} \tag{2}$$

Let $F_{\mathbf{r}'\boldsymbol{\lambda}}(x)$ denote the cumulative probability distribution function of returns of portfolio $\boldsymbol{\lambda}$. The *twice cumulative probability distribution function* of returns of portfolio $\boldsymbol{\lambda}$ is given by:

$$F^{(2)}_{\mathbf{r}'\boldsymbol{\lambda}}(t) = \int_{-\infty}^{t} F_{\mathbf{r}'\boldsymbol{\lambda}}(x)dx \tag{3}$$

Definition 2. Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the second order stochastic dominance ($\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau}$) if

$$F^{(2)}_{\mathbf{r}'\boldsymbol{\lambda}}(t) \leq F^{(2)}_{\mathbf{r}'\boldsymbol{\tau}}(t) \qquad \forall t \in \mathbb{R}$$

with at least one strict inequality.

The equivalent definition, presented in e.g. Hanoch and Levy (1969), Levy (2006), Kopa and Chovanec (2008) or Kopa and Post (2015) is based on comparison of expected utility of portfolio returns:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau} \iff \mathrm{E}u(\mathbf{r}'\boldsymbol{\lambda}) \geq \mathrm{E}u(\mathbf{r}'\boldsymbol{\tau})$$

for all concave utility functions u with strict inequality for at least some concave utility function.

3 SSD efficiency tests

In this session, we present an application of the second order stochastic dominance in portfolio efficiency and we present a SSD portfolio efficiency test derived in Kopa and Chovanec (2008) and more general formulation suggested in Kopa and Post (2015).

Definition 3. A given portfolio $\boldsymbol{\tau} \in \Lambda$ is *SSD inefficient* if and only if there exists portfolio $\boldsymbol{\lambda} \in \Lambda$ such that $\mathbf{r}'\boldsymbol{\lambda} \succ_{SSD} \mathbf{r}'\boldsymbol{\tau}$. Otherwise, portfolio $\boldsymbol{\tau}$ is *SSD efficient*.

Applying the duality property between CVaR and the second order stochastic dominance derived in Ogryczak and Ruszczynski (2002) and inspired by multicriteria approach of Roman, Darby-Dowman and Mitra (2006), Kopa and Chovanec (2008) proved that portfolio $\boldsymbol{\tau}$ is SSD efficient if and only if the optimal objective value of the following linear programming problem:

$$\max_{D_k,\lambda_n,b_k,w_k^t} \sum_{k=1}^{T} D_k \tag{4}$$

$$\begin{aligned} \text{s.t.} \quad \mathrm{CVaR}_{\frac{k-1}{T}}(\boldsymbol{\tau}) - b_k - \tfrac{1}{(1-\frac{k-1}{T})T}\textstyle\sum_{t=1}^{T} w_k^t &\geq D_k, && k = 1,2,\ldots,T \\ w_k^t &\geq -\mathbf{x}^t\boldsymbol{\lambda} - b_k, && t,k = 1,2,\ldots,T \\ w_k^t &\geq 0, && t,k = 1,2,\ldots,T \\ D_k &\geq 0, && k = 1,2,\ldots,T \\ \boldsymbol{\lambda} &\in \Lambda. \end{aligned}$$

is equal to zero. The idea is simple, the problem (4) tries to find a portfolio having smaller CVaRs than the tested portfolio at all considered levels $\alpha = 0, \frac{1}{T}, ..., \frac{T-1}{T}$. If such portfolio exists, then the portfolio SSD dominates the tested one and, moreover, it is SSD efficient. More general formulation, allowing for general weighting scheme was suggested in Kopa and Post (2015):

$$\xi(\boldsymbol{\tau}) = \max_{d_s,v_{t,s},\boldsymbol{\lambda}} \sum_{s=1}^{T} w_s d_s \tag{5}$$

$$\begin{aligned} \text{s.t.} \quad -T^{-1}\mathbf{x}^t\boldsymbol{\lambda} + \tfrac{1}{s}d_s - v_{t,s} + \tfrac{1}{s}\textstyle\sum_{k=1}^{T} v_{k,s} &\leq -\tfrac{1}{Ts}\textstyle\sum_{k=1}^{s}\mathbf{x}^k\boldsymbol{\tau} \\ & \quad t,s = 1,2,...,T \\ v_{t,s} &\geq 0, \quad t,s = 1,2,...,T \\ d_s &\geq 0, \quad s = 1,2,...,T. \\ \boldsymbol{\lambda} &\in \Lambda \end{aligned}$$

where weights $w_s > 0, s = 1,...,T$ can be chosen arbitrarily. In this case, the problem (5) tries to identify a portfolio which outperforms the tested portfolio $\boldsymbol{\tau}$ in terms of the second quantile function (or cumulative returns). If $\xi(\boldsymbol{\tau}) > 0$ then a portfolio $\boldsymbol{\lambda}^*$ which SSD dominates portfolio $\boldsymbol{\tau}$ exists, hence $\boldsymbol{\tau}$ is SSD inefficient. Moreover, such portfolio $\boldsymbol{\lambda}^*$ is already SSD efficient. If $\xi(\boldsymbol{\tau}) = 0$ then such dominating portfolio does not exist, and hence, portfolio $\boldsymbol{\tau}$ is SSD efficient. See Kopa and Post (2015) for more details. If we slightly modify problem (5), then SSD inefficient portfolio $\boldsymbol{\tau}$ (i.e. satisfying $\xi_\epsilon(\boldsymbol{\tau}) > 0$ where $\xi_\epsilon(\boldsymbol{\tau})$ is defined by problem (6)) would be mean-CVaR efficient for no level of CVaR. The modification makes use of additional arbitrary small $\epsilon > 0$:

$$\xi_\epsilon(\boldsymbol{\tau}) = \max_{d_s, v_{t,s}, \boldsymbol{\lambda}} \sum_{s=1}^{T} w_s d_s \tag{6}$$

$$\begin{aligned} \text{s.t.} \quad -T^{-1}\mathbf{x}^t\boldsymbol{\lambda} + \tfrac{1}{s}d_s - v_{t,s} + \tfrac{1}{s}\textstyle\sum_{k=1}^{T} v_{k,s} &\leq -\tfrac{1}{Ts}\textstyle\sum_{k=1}^{s}\mathbf{x}^k\boldsymbol{\tau} \\ & \quad t,s = 1,2,...,T \\ v_{t,s} &\geq 0, \quad t,s = 1,2,...,T \\ d_s &\geq \epsilon, \quad s = 1,2,...,T. \\ \boldsymbol{\lambda} &\in \Lambda. \end{aligned}$$

4 Empirical application

We consider daily returns of 49 US representative industry portfolios as the base assets. Data are taken from the Kenneth French data library from January 1st, 1927 till December 31st, 2014. For each year (from January 1st to January 1st) we firstly compute the mean return of the market portfolio proxied by CRSP, the minimal and the maximal mean return over all 49 assets. We define m_2 as the mean return of the market portfolio in the considered year. Similarly, let m_1 (m_3) be the average of m_2 and the minimal (maximal) mean return over all 49 assets.

Secondly, using (2), we identify the mean-CVaR portfolios on 95% level of CVaR and considering $\mu = m_i, i = 1,2,3$. Finally, applying (6), each of these portfolios is tested whether it is SSD efficient or not and the measures of inefficiency $\xi_\epsilon(\boldsymbol{\tau})$ are depicted on Figure 1 for all 87 periods and for all three levels of μ. The weights were chosen equally, i.e. $w_s = 1, s = 1,...,T$ and $\epsilon = 10^{-8}$.

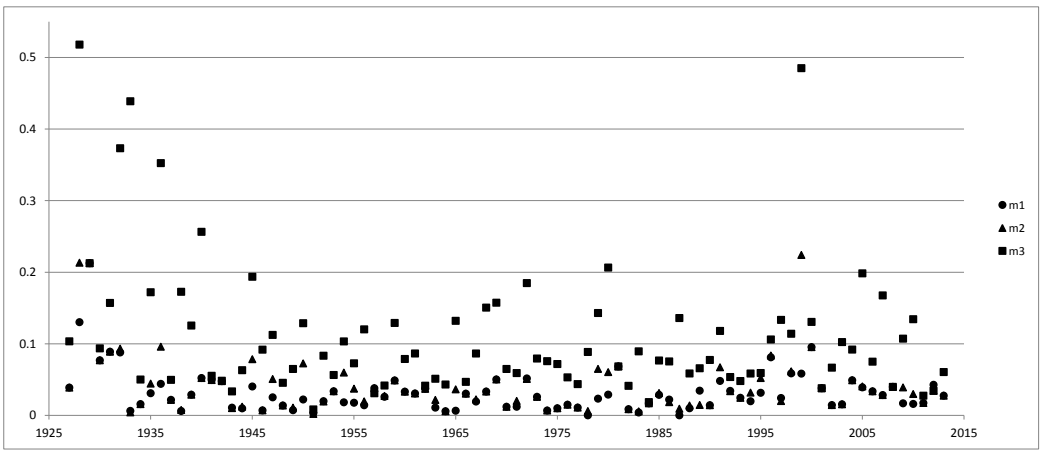


Figure 1 The figure captures out-of-sample SSD inefficiency measure of 261 in-sample mean-CVaR portfolios, considering 87 different time periods (years) and three different levels of minimal required mean return $m_i, i = 1,2,3$.

5 Conclusions

The paper deals with out-of-sample testing of in-sample mean-CVaR efficient portfolios in terms of the second order stochastic dominance efficiency. Modifying the Kopa and Post (2015) test we were able to prove that if the portfolio is out-of-sample SSD inefficient then it is out-of-sample mean-CVaR inefficient for any level of CVaR. We apply this methodology on the US representative industry portfolios which serve as the base assets. Figure 1 shows that we can find only two (year 1978 and 1987) of 261 portfolios that are out-of-sample SSD efficient (i.e. measure of SSD inefficiency is equal to zero), in both cases $\mu = m_1$ was considered. Hence, almost all of the considered portfolios are out-of-sample SSD inefficient, and therefore, they cannot be out-of-sample mean-CVaR efficient for any level of CVaR. Comparing the measure of SSD inefficiency over the time periods, we can see significantly higher values around year 1930, what can be explained by the big financial crises. Perhaps surprisingly, high SSD inefficiency measure was observed for in-sample mean-CVaR portfolios in year 1999, too. Finally, we can easily see, that higher in-sample required minimal mean return implies nearly always higher (or the same) out-of-sample SSD inefficiency measure. Therefore, when focusing on the out-of-sample performance in terms of SSD efficiency, we suggest to choose the in-sample portfolio with globally minimal CVaR or with as small mean return as possible if the globally minimal CVaR portfolio is not feasible.

We applied mean-CVaR approach because of the linear programming formulations. However, one can easily repeat the methodology using another measure of risk. Moreover, this kind of analysis could be similarly done also with the first order stochastic dominance (Kopa and Post (2009)), higher order stochastic dominance (Post and Kopa 2013) or decreasing absolute risk aversion stochastic dominance (Post, Fang and Kopa (2015)). Finally, one can apply some robust versions of the first or second order stochastic dominance and efficiency tests see e.g. Dentcheva and Ruszczynski (2010), Kopa (2010) and Dupačová and Kopa (2012, 2014). Unfortunately, all these modifications would either significantly increase the computational burden or lead to non-tractable formulations.

Acknowledgements

The research was partially supported by the Czech Science Foundation under grant 15-02938S.

References

[1] Branda, M. and Kopa, M. (2012): DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* 62, No. 2, 106–124.

[2] Branda, M. and Kopa, M. (2014): On relations between DEA-risk models and stochastic dominance efficiency tests, *Central European Journal of Operations Research* 22, No. 1, 13–35.

[3] Dentcheva, D., and Ruszczyński, A. (2010): Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming, Series B* 123, (2010), 85–100.

[4] Dupačová, J. and Kopa, M. (2012): Robustness in stochastic programs with risk constraints. *Annals of Operations Research* 200, No. 1, 55–74.

[5] Dupačová, J. and Kopa, M. (2014): Robustness of optimal portfolios under risk and stochastic dominance con-straints. *European Journal of Operational Research* 234, No. 2, 434–441.

[6] Grechuk, B. (2014): A simple SSD-efficiency test, *Optimization Letters* 8, No. 7, 2135–2143.

[7] Hanoch, G. and Levy, H. (1969): The efficient analysis of choices involving risk. *Review of Economic Studies*, 36, No. 3, 335–346.

[8] Kopa, M. (2010): Measuring of second-order stochastic dominance portfolio efficiency, *Kybernetika* 46, No. 3, 488–500.

[9] Kopa, M. and Chovanec, P. (2008): A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* 44, No. 2, 243–258.

[10] Kopa, M. and Post, T. (2009): A portfolio efficiency test based on FSD optimality. *Journal of Financial and Quantitative Analysis* 44, No. 5, 1103–1124.

[11] Kopa, M. and Post, T. (2015): A general test for SSD portfolio efficiency. *OR Spectrum* 37, No. 3, 703–734.

[12] Kuosmannen, T. (2004): Efficient diversification according to stochastic dominance criteria. *Management Science* 50, No. 10, 1390–1406.

[13] Levy, H. (2006): *Stochastic dominance: Investment decision making under uncertainty.* Second edition, Springer, New York.

[14] Lizyayev, A. (2012): Stochastic dominance efficiency analysis of diversified portfolios: classification, comparison and refinements. *Annals of Operations Research* 196, No. 1 , 391–410.

[15] Ogryczak W. and Ruszczyński A.: Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization* 13, 60–78.

[16] Pflug, G. (2000). Some Remarks on the value-at-risk and the Conditional value-at-risk. In: *Probabilistic Constrained Optimization* (S.P. Uryasev, ed.), Kluwer, Dordrecht, 272–281.

[17] Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance* 58, 1905–1932.

[18] Post, T., Fang Y. and M. Kopa (2015): Linear Tests for DARA Stochastic Dominance. *Management Science* 61, No. 7, 1615–1629.

[19] Post, T. and Kopa M. (2013): General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research*, 230, No. 2, 321–332.

[20] Roman, D., Darby-Dowman, K. and Mitra, G. (2006): Portfolio construction based on stochastic dominance and target return distributions. *Mathematical Programming* 108, 541–569.

[21] Rockafellar, R.T. and Uryasev, S. (2000): Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2, 21–41.

[22] Rockafellar, R.T. and Uryasev, S. (2002): Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26, 1443–1471.

Automatic Detection of Buying or Selling Signal Using Head and Shoulders Chart Pattern with Neckline

Kot'átková Stránská, Pavla, Navrátil, Martin and Heckenbergerová, Jana [1]

Abstract. Head and Shoulders chart pattern (H&SP) is commonly used for detection of buying or selling signal. Straight line connecting the breakpoints between the head and shoulders, known as neckline, play an important role in such detection. When after the occurrence of the H&SP formation the stock price drops below (rises above) the neckline support (resistance) by more than five percent, we say that the H&SP top (bottom) is confirmed and it gives the right signal to sell (buy) the shares. Prices are then expected to move in the same direction to approximately the difference between head and neckline.

The aim of this contribution is presentation of algorithm for automatic detection and confirmation graphical formation Head and Shoulders Pattern with neckline. Methodology for buying-selling signal detection will be presented as well. Whole procedure is tested on selected time sections of Dow Jones Industrial Averages. Result of the analysis will be applied to quantify the change in trend of stock market prices after H&SP confirmation. Sensitivity of H&SP detection will be analyzed as dependence on selected coefficient of determination and size the window.

Keywords: Stock market price, Chart pattern detection, Neckline, Graphical formation, Nonlinear regression model, Coefficient of Determination, Buying and selling signal, Sensitivity analysis

JEL classification: C14, G17
AMS classification: 62J02, 62J05, 62B15

1 Algorithm for Detection of Head and Shoulders Chart Pattern with Neckline

1.1 Head and Shoulders Chart Pattern

Head and Shoulders Chart Pattern (H&SP) also known as camel's hump is specific sequence of local maxima and minima usually comes after a long period of growth or fall and signalizes a fundamental change in the trend of stock market prices [9], [10] and [3]. Shape of this chart pattern corresponds to left shoulder, head and right shoulder. An important characteristic of this formation is the neckline, straight line connecting the breakpoints between the head and shoulders. There are two types of H&SP in terms of head position, H&SP top and H&SP bottom.

The H&SP formation can be detected for different lengths of time window and for different levels of index of determination. It is mostly searched only on the basis of subjective visual detection in graph of security market prices; nevertheless the automatic algorithm for the detection of Head and Shoulders formation, based on an approximation of nonlinear functions and a linearized regression model to estimate the unknown parameters, was presented in [4], [5]. Detected H&SP top (bottom) is confirmed when the stock price drops below (rises above) neckline support (resistance) by more than five percent. This

[1]Faculty of Electrical Engineering and Informatics, Department of Mathematics and Physics, University of Pardubice, Studentska 95, 532 10 Pardubice, Czech Republic, pavla.kotatkovastranska@upce.cz; martin.navratil1@student.upce.cz; jana.heckenbergerova@upce.cz

breakpoint gives us the right signal to sell (buy) the shares. After chart pattern confirmation, the price is expected to move in the same direction to approximately the difference between head and neckline [9], [1].

1.2 Regression Model and Parameters Estimation

Let us assume that each H&SP can be described by nonlinear function:

$$f(x) = \beta_1 + \beta_2 x + \beta_3 \frac{\sin(\kappa x)}{x} \tag{1}$$

with four characteristic parameters $\beta_1, \beta_2, \beta_3$ and κ, where x represents the time (centered around zero). First two parameters describe linear trend of chart pattern, parameter β_3 represents height of the head, and the last parameter κ can be derived from H&SP time duration.

Neckline equation of this chart pattern can be written in form:

$$p(x) = \beta_1 + \beta_2 x - 0.2172\kappa\beta_3. \tag{2}$$

Chart pattern parameters β_1, β_2 and β_3 are dependent on stock market prices and vector parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ has to be estimated using nonlinear regression model and methodology fully described in [5]. Let us assume that linearized model regular model of incomplete measurement without constrains and it is in form

$$\mathbf{Y} \sim N_n\left(\boldsymbol{\Phi}_0 + \mathbf{F}\delta\boldsymbol{\beta}, \mathbf{I}_n\right), \text{where } \mathbf{F}_i = \frac{\partial f(x_i, \boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}'} = \left(1, x_i, \frac{sin(\kappa x_i)}{x_i}\right), \quad i = 1, \ldots, n, \tag{3}$$

$\boldsymbol{\beta}^0$ is an initial approximation given by extremal data and $\delta\boldsymbol{\beta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$. The estimator of unknown parameter $\boldsymbol{\beta}$ can be therefore evaluated as $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^0 + \delta\widehat{\boldsymbol{\beta}}$, where $\delta\widehat{\boldsymbol{\beta}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'(\mathbf{Y} - \boldsymbol{\Phi}_0)$. When $\widehat{\boldsymbol{\beta}}_3$ is positive, the Head and Shoulders Pattern - Top is expected signalizing end of growing market; for $\widehat{\boldsymbol{\beta}}_3 < 0$ H&SP - Bottom could be recognized. If $|\widehat{\boldsymbol{\beta}}_3| < \beta_3^0$, then last parameter is not significant, straight line approximation can be expected and H&SP can not be detected in selected time frame.

Parameter κ describes the length of graphical formation and therefore it can be evaluated directly from series of selected trading times [4]. Let n be duration of expected chart pattern, then $\kappa = \frac{6\pi}{n}$. Table 1 summarizes values of parameter κ for selected lengths of pattern frame.

n	11	13	15	17	19	21	23	25	27
κ	1.7136	1.4500	1.2566	1.1088	0.9921	0.8976	0.8195	0.7540	0.6981

Table 1 Parameter κ for several values of n

1.3 Detection of H&SP Formation

Coefficient of determination is used as a detection criterion:

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Y_i} - \widehat{\mathrm{Y}}_\mathrm{i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Y_i} - \overline{\mathrm{Y}}_\mathrm{i})^2}, \tag{4}$$

where Y_i are real prices, $\widehat{Y}_i$ are modelled prices based on selected chart pattern and $\overline{Y}$ is average stock market price for given time frame. Head and Shoulders Pattern is recognized when R^2 is greater than given threshold (e.g. 0.85).

Figure 1 illustrates numerical results of H&SP approximation. As coefficient of determination R^2 is greater than preset threshold $\mathrm{R}^2_{\min} = 0.85$, we can say that H&SP was DETECTED in given dataset. Negative value of last estimated parameter $\widehat{\boldsymbol{\beta}}_3$ characterizes bottom type of chart pattern. One can see that DJIA values break the neckline and rise high above it immediately after pattern detection. This is the CONFIRMATION of H&SP detection and therefore business day No.22 can be marked as BUYING SIGNAL.

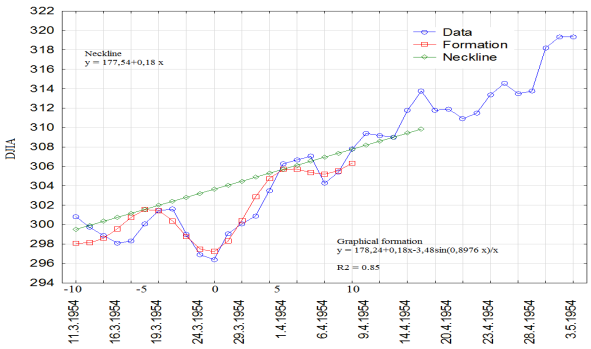


Figure 1 Head and Shoulders Pattern in DJIA dataset from March 11, 1954 till April 9, 1954

2 Confirmation of H&SP Detection and Buying–Selling Signal Determination

Chart pattern neckline [Eq.(2)] has to be evaluated for confirmation of H&SP detection and for correct selection of buying(selling) time. The neckline breakout signalize significant change in price movement. The detected H&SP top formation can be confirmed if and only if there exists calibrated time index $x_{i_s} \in \langle p - \frac{n}{8}, p + \frac{n}{4} \rangle$ such as $Y_{i_s} \leq p(x_{i_s}) - 0.05\beta_3$. This time index x_{i_s} then gives us the selling signal. Confirmation of bottom chart pattern detection works analogically. Whole confirmation process is summarized in following methodology:

1. preparation of dataset (time series of share prices or market indexes)
2. presetting of fixed parameters – frame length n (expected duration of pattern) and minimal coefficient of determination $\mathrm{R}^2_{\min}$ and evaluation of κ parameter and centering index p
3. select data from given floating window
4. evaluation of initial parameters $\boldsymbol{\beta}^0$, chart pattern parameters β_1, β_2 and β_3 and corresponding determination index R^2
5. chart pattern detection:
   (a) if $\mathrm{R}^2 \leq \mathrm{R}^2_{\min}$ then H&SP is not recognized and go to 3
   (b) if $\mathrm{R}^2 > \mathrm{R}^2_{\min}$ then H&SP is detected
6. H&SP type recognition:
   (a) if $|\beta_3| < \beta_3^0$ then detected H&SP is not significant and go to 3
   (b) if $\beta_3 < 0$ then detected formation is H&SP bottom and go to 7
   (c) if $\beta_3 > 0$ then detected formation is H&SP top and go to 8
7. H&SP bottom confirmation:
   (a) evaluation of neckline resistance $p(x)$
   (b) if $\forall x_i \in \langle p - \frac{n}{8}, p + \frac{n}{4} \rangle$: $Y_i < p(x_i) - 0.05\beta_3$ then pattern is not confirmed and go to 3
   (c) if contrary $\exists x_{i_b} \in \langle p - \frac{n}{8}, p + \frac{n}{4} \rangle$: $Y_{i_b} \geq p(x_{i_b}) - 0.05\beta_3$ then H&SP bottom is confirmed in selected time frame
   (d) x_{i_b} is the buying signal
8. H&SP top confirmation:
   (a) evaluation of neckline support $p(x)$
   (b) if $\forall x_i \in \langle p - \frac{n}{8}, p + \frac{n}{4} \rangle$: $Y_i > p(x_i) - 0.05\beta_3$ then pattern is not confirmed and go to 3
   (c) if contrary $\exists x_{i_s} \in \langle p - \frac{n}{8}, p + \frac{n}{4} \rangle$: $Y_{i_s} \leq p(x_{i_s}) - 0.05\beta_3$ then H&SP top is confirmed
   (d) x_{i_s} is the selling signal
9. use some market strategy for buying and selling and go to 3

3 Case study

Large data series of Dow Jones Industrial Averages (DJIA) is used in following section. Indexing of DJIA started in 18th February 1895 with initial price of 28.9563 and it is running till now, whole time series can be downloaded with one-day resolution from [2].

In previous studies [4], [5] performance of H&SP detection and confirmation algorithm was described for pre-selected time frames and fixed input parameters. Following section presents the results of sensitivity analyses when whole DJIA dataset is searched, frame length is fluent and coefficient of determination is increasing. Moreover usability of H&SP detection and confirmation algorithm is verified by quantification of the change in trend of stock market prices after H&SP confirmation. To be able to search through the whole DJIA database we decided to utilize H&SP detection and confirmation algorithm into the simple Java script.

3.1 Sensitivity of H&SP detection algorithm on selected determination index threshold and length of floating window

Table summarizing the numbers of detected and confirmed chart pattern when frame length is changing from 11 to 51 and determination index threshold is set to 50%, 85%, 90%, 93% and 97% respectively can be found in [6]. Dependence of confirmed pattern number on frame length and coefficient of determination is illustrated in Figure 2.

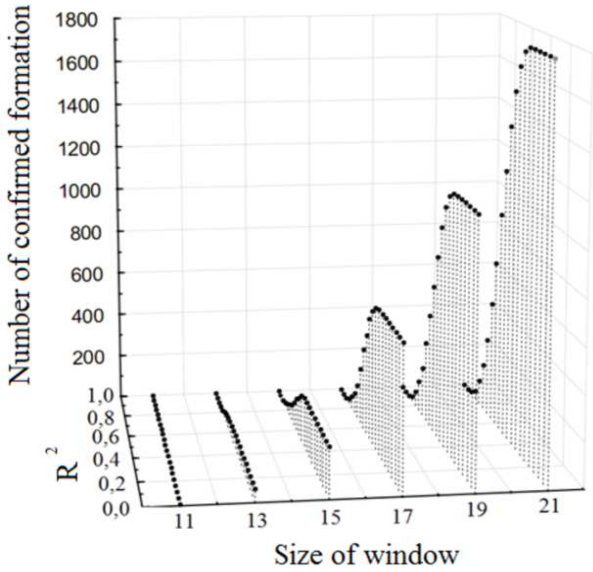


Figure 2 Dependence of confirmed pattern number on frame length and coefficient of determination

It is obvious that not all detected H&SP are confirmed, the confirmation percentage is around 75% in average. One can also see that number of detected (confirmed) chart patterns is highly dependent on selected determination index and size of floating window. This is the reason why we decided to test their independence and find the power of dependence.

Pearson χ^2 test of independence [7, 8] is based on marginal frequencies comparison. P-value of this test statistic is equal to zero with significance level $\alpha = 0.05$. Therefore null hypothesis about independence of determination coefficient and frame length has to be rejected. Power of dependence can by quantified by Cramer coefficient that is ranging from 0 to 1. Higher values of Cramer coefficient signalize more tight dependence. In our case the Cramer coefficient value ($V = 0.023418$) shows only weak dependence between frame length and determination index.

3.2 Analysis of DJIA absolute difference after H&SP detection

For this analysis only few detected H&SP formations were selected, they are summarize on Table 2. Detected chart pattern was chosen only for $R^2 > 0.9$ and buying(selling) signal following the detection. Absolute differences represent the change in trend of stock market prices after H&SP confirmation and they are evaluated from comparison of final and buying(selling) value. Evolution of absolute error values for H&SP TOP is visualized on Figure 3. Figure 4 captures absolute error values for selected H&SP BOTTOM formations.

Everyone can see the great results of this analysis which are proving that after detected top formation DJIA values are falling and bottom formation signalize following grow of market. This means that our algorithm for detection and confirmation of H&SP is fully functional and can be used for market speculations. Nevertheless, we are still unable to predict the end of growing and falling of market. More detailed analyses have to be done to understand this behavior.

Tag	Type	Duration	Length	R^2	Buying–selling signal
$B93_{21}01$	B	3/1973–4/1973	21	93	22
$B93_{19}01$	B	10/1954–10/1954	19	93	21
$B90_{21}01$	B	2/1909–2/1909	21	90	24
$B90_{21}02$	B	3/1973–4/1973	21	90	24
$B90_{19}03$	B	10/1954–10/1954	19	90	22
$T90_{21}01$	T	8/1903–8/1903	21	90	23
$T90_{21}02$	T	11/1941-12/1941	21	90	26
$T90_{21}03$	T	9/1943–10/1943	21	90	26
$T90_{21}04$	T	5/1961–1/1962	21	90	23
$T90_{21}05$	T	6/1969–7/1969	21	90	26
$T90_{21}06$	T	7/1977–8/1977	21	90	23
$T90_{21}07$	T	5/2008–6/2008	21	90	27
$T90_{21}08$	T	10/2012–11/2012	21	90	23
$T90_{19}01$	T	10/2012–11/2012	19	90	22
$T90_{17}01$	T	3/1990–3/1990	17	90	17
$T90_{17}02$	T	7/1994–8/1994	17	90	21

Table 2 Summary of analyzed H&SP formations

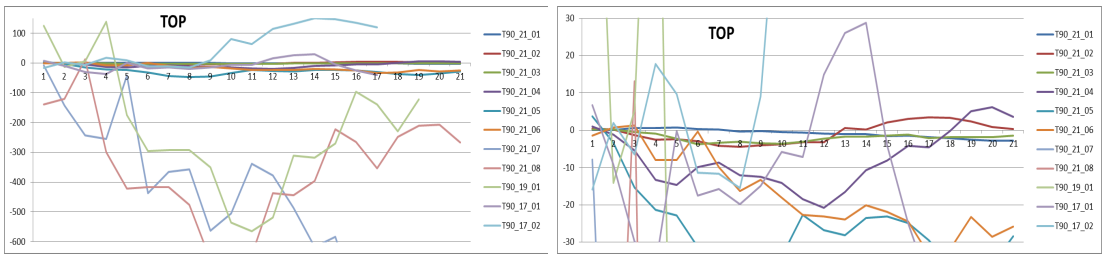


Figure 3 Absolute increments of formation H&SP TOP

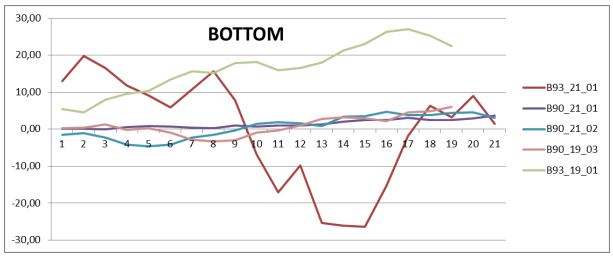


Figure 4 Absolute increments of formation H&SP BOTTOM

4 Conclusion

The aim of this contribution was the presentation of algorithm for automatic detection and confirmation graphical formation Head and Shoulders Pattern with neckline and the introduction of methodology for buying-selling signal detection. Results of presented analyses, performed on selected time sections of Dow Jones Industrial Averages, show weak dependence of determination index on window size. Moreover they prove that after detection of top formation DJIA values are falling and bottom formation signalize following grow of market. This means that our algorithm for detection and confirmation of H&SP is fully functional and can be used for market speculations.

Acknowledgements

This contribution has been supported by the Ministry of Education, Youth and Sports of the Czech Republic project No. CZ.1.07/ 2.3.00/30.0058, and partly supported by SGS project and institutional support of the University of Pardubice, Czech Republic.

References

[1] Bender, J. C., Osler, C. L, Simon, D.: *Noise Trading and Illusory Correlations in US Equity Markets.* Review of Finance, 2012. DOI:10.1093/rof/rfr037

[2] Dow Jones Index. [online], http://measuringworth.com/m/datasets/DJA/

[3] Friesen, G. C., Weller, P. A. and Dunham, L. M.: Price trends and patterns in technical analysis: A theoretical and empirical examination. *Journal of Banking and Finance* **33** (2009), 1089-1100. ISSN 0378-4266. DOI: 10.1016/j.jbankfin.2008.12.010

[4] Heckenbergerová, J., Marek, J. and Mejznar, J.: *Automatic head and shoulders pattern recognition using nonlinear statistical modelling.* Proc. of 12th Conference on Applied Mathematics, APLIMAT2013, Bratislava Feb. 2013, ISBN 978-80-227-3865-1.

[5] Heckenbergerová, J., Marek, J. and Kot'átková Stránská, P. *Algorithm for detection and confirmation of Head and Shoulders Pattern with Neckline.* Proc. of 14th Conference on Applied Mathematics, APLIMAT 2015, Bratislava Feb. 2015, pp. 357-369, ISBN 978-80-227-4314-3.

[6] Kot'átková Stránská, P., Navrátil, M. and Heckenbergerová, J.: *Automatic Detection of Buying or Selling Signal Using Head and Shoulders Chart Pattern with Neckline.* [online] appendix available at http://www.upce.cz/fei/kmf/download/mme2015.pdf or http://www.upce.cz/fei/kmf/download/mm2015.pdf

[7] Kubaček, L., Kubačková, L. and Volaufova, J.: *Statistical models with linear structures.* Veda, Publishing House of the Slovak Academy of Sciences, Bratislava, 1995.

[8] Kubaček, L., Kubačková, L.: *Statistics and Metrology* (in Czech). Publishing House of Palacký University, Olomouc, 2000.

[9] Neely, Ch., Weller, P. and Dittmar, R.: Is Technical Analysis in the Foreign Exchange Market Profitable?: A Genetic Programming Approach,*Journal of Financial and Quantitative Analysis* (1997).

[10] Wang, Z., Zeng, Y. and Ping L. A Nonparametric Kernel Regression Method for the Recognition of Visual Technical Patterns in China's Stock Market. *In Third International Conference Business Intelligence and Financial Engineering (BIFE)*, 2010, 296-300. ISBN 978-1-4244-7575-9. DOI: 10.1109/BIFE.2010.76

On the estimation of the state price density

Noureddine kouaissah[1], Sergio Ortobelli Lozza[2]

Abstract. In this paper we compare alternative methods to estimate the state price density under the classical hypothesis of the Black and Scholes model. In particular, we examine different methodologies to evaluate the conditional expectation and its relationship with the state price density under the hypothesis of the Black Scholes model. Thus, using this relationship we are able to evaluate the state price density of several assets. Finally, we propose a comparison among two estimators of the state price density.

Keywords: state price density, conditional expectation estimators, empirical analysis.

JEL Classification: C14, G13
AMS Classification: 28A25

1 Introduction

The fundamental work of Black and Scholes [3] (hereinafter BS) is the most used methodology among derivatives pricing analysis and has a great importance for improving research on the option pricing techniques. Unfortunately, widespread empirical analyses point out that a set of assumptions, particularly normally distributed returns and constant volatility, result in poor pricing and hedging performance. Indeed, the presence of skewness and kurtosis in the market complicates the situation significantly**.** Nowadays, practitioners still quote options based on the volatility implied by the BS model. A common market practice is to use the market price as an exogenous variable to be put into the BS formula. However, different generalizations of the BS model have been proposed: jump diffusion model proposed by Merton [10], stochastic volatility model by Heston [11] or a combination of the last approaches proposed by Bates [5]. Generally, most models that have been proposed so far usually relax some assumptions of Black Scholes, and then trying to be justified via general fundamental theorem of asset pricing-FTAP, Harrison and Kreps [9].

State Price Density (hereinafter SPD) is a fundamental entity asset pricing theory. Among no-arbitrage models, the SPD is frequently called risk-neutral density, which is the density of the equivalent martingale measure with respect to the Lebesgue measure. The existence of the equivalent martingale measure follows from the absence of arbitrage opportunities, while its uniqueness demands complete markets. In order to estimate SPD, two main approaches have been presented the so called parametric and nonparametric approaches. The former assumes that the dynamics of the underlying asset belong to a certain parametric family. The parameters are then estimated by minimizing the distance between observed prices and theoretical no-arbitrage prices. The main example is the BS model, under which the underlying asset follows a geometric Brownian motion and the SPD is lognormal. However, widespread evidence on the implied volatility smile suggests that a lognormal SPD is too restrictive assumption. In the nonparametric approach normally no specific model is assumed for the SPD or the asset return dynamics. In this technique the call pricing function is estimated via nonparametric regression of the observed call prices from multiple variables as strike price, stock price and time-to-maturity. The SPD is then derived as the second derivative of call pricing function with respect to the strike price, using Breeden and Litzenberger result [4].

The paper follows the above mentioned arguments and presents a new method to estimate SPD. In particular, we estimate the real mean return function using local polynomial smoothing technique. Then, according the hypothesis of the Black and Scholes model, we are able to estimate the SPD. To this end we follow two distinguished approaches to recover SPD, the first one based on nonparametric estimation techniques "kernel" which are natural candidates (see [1], [2] among others for an application to options), then a new method based on conditional expectation estimator proposed by [12].The kernel non-parametric regression method allows estimating the regression function, which is a realization of the conditional expectation $\boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{X})$, while the second approach estimate the conditional expectation (intended as a random variable), based on an appropriate approximation of the σ-algebra generated by X. The two approaches estimate the SPD under the hypothesis of the Black and Scholes model. Thus, we are able to evaluate the SPD of several assets using different conditional expectation estimators. The main goal of this paper is to discuss and compare the conditional expectation method and

[1] University of Bergamo, Italy - VSB-TU of Ostrava, Czech Republic, noureddine.kouaissah@unibg.it
[2] University of Bergamo, Italy - VSB-TU of Ostrava, Czech Republic, sol@unibg.it

the nonparametric technique, which can be useful tool for different aspects such as the presence of arbitrage opportunities in the market or even pricing some complex derivatives.

The rest of the paper is organized as follows. Section 2 describes some theoretical properties of two approaches. Section 3 concludes an empirical analysis. In particular we present an application to the S&P 500 index using daily data from the sample period 04/28/2013 to 04/28/2015. Concluding remarks are contained in Section 4.

2 An alternative methodology to estimate the SPD

Fisher Black and Myron Scholes [3] achieved a major breakthrough in European option pricing. In this model we assume that the price process follows a standard geometric Brownian motion defined on filtered probability space $(\Omega, \Im, \mathrm{P}, \{\Im_t\}_{t\geq 0})$, where $\{\Im_t\}_{t\geq 0}$ is the natural filtration of the process completed by the null sets. Under these assumptions we know that $E(\mathrm{S}_T \mid \Im_t) = \mathrm{E}(\mathrm{S}_T \mid \mathrm{S}_t)$ as consequence of Markovian property. The model of stock price behavior used is defined as:

$$dS = \mu S dt + \sigma S dB, \tag{2.1}$$

where, μ is the expected rate of return, σ is the volatility of stock return and B denotes a standard Brownian motion. Under this hypothesis we know that the log price is normally distributed:

$$\ln S_T \sim \phi\left[\ln S_0 + \left(\mu - 0.5\sigma^2\right)T, \sigma^2 T\right], \tag{2.2}$$

where, S_T is the stock price at future time T, S_0 is the stock price at time 0 and ϕ denotes a normal distribution. Please note that μ in equation (2.1) represents the expected rate of return in real world, while in BS model (risk neutral world) becomes risk-free rate r.[3]

For a stake of clarity, denotes S^{RW} for a real world price and the risk neutral price as S^{RN}. Under the hypothesis of the BS model it is straightforward to write:

$$S_T^{RN} = S_T^{RW} e^{-(\mu - r)T}, \tag{2.3}$$

Since $S_t = e^{-r(\mathrm{T}-\mathrm{t})} E(S_T^{RN} \mid \Im_t)$, we can write $S_t = e^{-r(\mathrm{T}-\mathrm{t})} E(S_T^{RW} e^{-(\mu - r)T} \mid \Im_t)$ from which we obtain:

$$E(S_T^{RN} \mid \Im_t) = e^{-\mu T + rt} E(S^{RW} \mid \Im_t), \tag{2.4}$$

If we assume μ changes over time in model (2.1), then equation (2.4) becomes

$$e^{-\int_0^T (\mu(\tau) - r) d\tau} E(S_T \mid \Im_t) = \tilde{E}(S_T \mid \Im_t), \tag{2.5}$$

where, $\tilde{E}(S_T \mid \Im_t)$ denotes expectation under risk neutral world and $E(\mathrm{S}_T \setminus \Im_t)$ the conditional expected price under real world. Moreover, (2.5) is equivalent to:

$$e^{-\int_0^T \mu(\tau) d\tau} \int_0^\infty s q_{RW}(\mathrm{s}) ds = e^{-rT} \int_0^\infty s q_{RN}(\mathrm{s})\, \mathrm{ds}, \tag{2.6}$$

where, $\mathrm{q}_{RW}(\mathrm{s})$ and $q_{RN}(s)$ denotes SPDs under real and risk neutral world respectively.

Please note that under the BS hypothesis S_{T-t} has the same distribution of $S_T e^{-\mu t}$.

The first step in this paper is to propose a direct method of estimating the real mean return function. Therefore, we use a local estimator that automatically provides an estimate of the real mean function and its derivatives. The input data are daily prices. Denoting the intrinsic value by $\tilde{\mu}_i$ and the true function by $\mu(\mathrm{t}_i)$, $i = 1, ..., n$, we assume the following regression model:

$$\tilde{\mu}_i = \mu(\mathrm{t}_i) + \varepsilon_i, \tag{2.7}$$

where, ε_i models the noise, n denotes the number of data considered. The local quadratic estimator $\hat{u}(\mathrm{t})$ of the regression function $u(\mathrm{t})$ in the point t is defined by the solution of the following local least squares criterion:

[3] For more details about BS assumptions we refer to Hull (2012)

$$\min_{\alpha_0,\alpha_1,\alpha_2} \sum_{i=1}^{n} \left\{ \tilde{\mu}_i - \alpha_0 - \alpha_1(\mathrm{t}_i - t) - \alpha_2(\mathrm{t}_i - t)^2 \right\} k_h(\mathrm{t} - t_i), \tag{2.8}$$

where, $k_h(\mathrm{t}-t_i) = \frac{1}{h} k\left(\frac{t-t_i}{h}\right)$ is *kernel* function, see Fan and Gijbels for more details. Comparing last equation with the Taylor expansion of u yields:

$$\alpha_0 = \hat{\mu}(\mathrm{t}_i), \alpha_1 = \hat{\mu}'(\mathrm{t}_i), \; 2\alpha_2 = \hat{\mu}''(\mathrm{t}_i), \tag{2.9}$$

which make the estimation of the regression function and its two derivatives possible. The second step towards estimating state price density is to use two methodologies, namely OLP estimator and kernel estimator, to estimate the quantity $E(\mathrm{S}_T / \mathrm{S}_t)$.

2.1 Nonparametric conditional expectation estimators

Regression analysis is surely one of the most suitable and widely used statistical techniques. In general, it explores the dependency of the so-called dependent variable on one (or more) explanatory or independent variables.

$$Y = E(Y / X = x) + \varepsilon = g(x) + \varepsilon. \tag{2.10}$$

It is well known that, if we know the form of the function $g(x) = E(Y / X = x)$, (e.g. polynomial, exponential, etc.), then we can estimate the unknown parameters of $g(x)$ with several methods (e.g. least squares). In particular, if we do not know the general form of $g(x)$, except that it is a continuous and smooth function, then we can approximate it with a non-parametric method, as proposed by [6] and [8]. The aim of non-parametric technique is to relax assumptions on the form of regression function, and allows data search for an appropriate function that represents well the available data, without assuming any specific form of the function. Thus, $g(x)$ can be estimated by:

$$\hat{g}_n(x) = \frac{\sum_{i=1}^{n} y_i k\left(\frac{x - x_i}{h(n)}\right)}{\sum_{i=1}^{n} k\left(\frac{x - x_i}{h(n)}\right)}, \tag{2.12}$$

where, $k(\bullet)$ is a density function such that: i) $k(x) < \mathrm{C} < \infty$, ii) $\lim_{x \to \pm\infty} |xk(x)| = 0$, iii) $h(n) \to 0$ when $n \to \infty$. h is a bandwidth, also called a smoothing parameter, which controls the size of the local averaging. The function $k(x)$ is denoted by *kernel*, observe that kernel functions are generally used for estimating probability densities non-parametrically (see [13]). An overview of nonparametric regression or smoothing techniques may be found, e.g. among others [7].

An alternative non-parametric approach for approximating the conditional expectation denoted by "*OLP*" has been given in [12]. Define by $\Im_X$ the σ-algebra generated by X (that is $\Im_X = \sigma(X) = X^{-1}(\mathcal{B}) = \{X^{-1}(B): B \in \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on$\mathbb{R}$). Observe that the regression function is just a "pointwise" realization of the random variable $E(Y|\Im_X)$, which can equivalently be denoted by $E(Y|X)$. The following methodology is aimed at estimating $E(Y|X)$ rather than $g(x)$. For this reason, we propose the following consistent estimator of the random variable$E(Y|X)$.

Let $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ be integrable random variables in the probability space $(\Omega, \Im, P)$ and define by $\Im_X$ the σ-algebra generated by X (that is, $\Im_X = \sigma(X) = X^{-1}(\mathcal{B}) = \{X^{-1}(B): B \in \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$). Notice that: $E(Y|X)$ is equivalent to $E(Y|\Im_X)$. We can approximate $\Im_X$ with a σ-algebra generated by a suitable partition of Ω. In particular, for any $k \in \mathbb{N}$, we consider the partition $\{A_j\}_{J=1}^{b^k} = \{A_1, \ldots, A_{b^k}\}$ of Ω in b^k subsets, where b is an integer number greater than 1 and:

- $A_1 = \{\omega: X(\omega) \leq F_X^{-1}\left(\frac{1}{b^k}\right)\}$,
- $A_h = \{\omega: F_X^{-1}\left(\frac{h-1}{b^k}\right) < X(\omega) \leq F_X^{-1}\left(\frac{h}{b^k}\right)\}$, for h=2,…,b^k–1,
- $A_{b^k} = \Omega - \cup_{J=1}^{b^k-1} A_j = \{\omega: X(\omega) > F_X^{-1}\left(\frac{b^k-1}{b^k}\right)\}$.

Thus, starting with the trivial sigma algebra $\Im_0 = \{\emptyset, \Omega\}$, we can generate a sequence of sigma algebras generated by these partitions obtained by varying k (k=1,…,m,…). Thus, $\Im_1 = \sigma\{\emptyset, \Omega, A_1, \dots, A_b\}$ is the sigma algebra generated by $A_1 = \{\omega: X(\omega) \leq F_X^{-1}(1/b)\}$, $A_s = \left\{\omega: F_X^{-1}\left(\frac{s-1}{b}\right) < X(\omega) \leq F_X^{-1}\left(\frac{s}{b}\right)\right\}$, s=1,...,b-1 and $A_b = \{\omega: X(\omega) > F_X^{-1}((b-1)/b)\}$, moreover:

$$\Im_k = \sigma\left(\{A_j\}_{J=1}^{b^k}\right), k \in \mathbb{N} \tag{2.13}$$

Under these hypotheses [12] proved that:

$$E(Y|X) = lim_{k\to\infty} E(Y|\Im_k) \text{ a.s.} \tag{2.14}$$

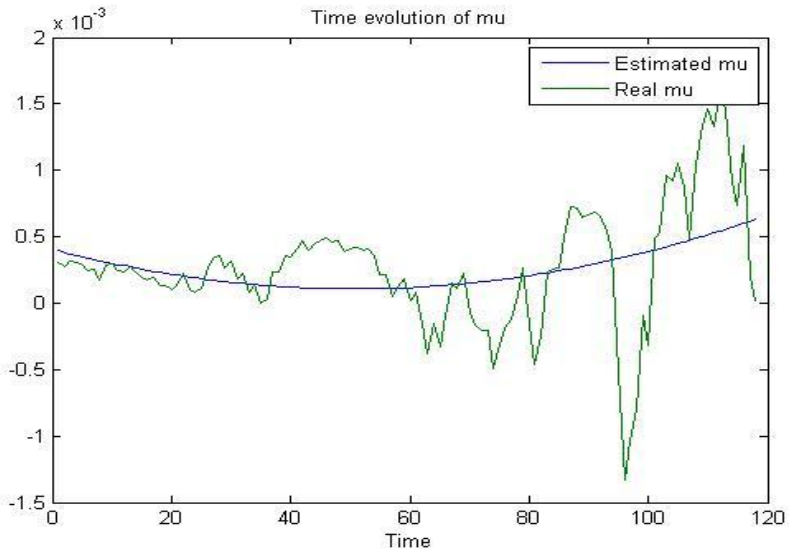
where, $E(Y|\Im_k)(\omega) = \sum_{j=1}^{b^k} E(Y|A_j)1_{A_j}(\omega)$ a.s. and $1_{A_j}(\omega) = \begin{cases} 1 & \omega \in A_j \\ 0 & \omega \notin A_j \end{cases}$.

When b is large enough, even $E(Y|\Im_1)$ can be a good approximation of the conditional expected value $E(Y|X)$. On the one side, given N i.i.d. observations of Y, we get that $\frac{1}{n_{A_j}}\sum_{y\in A_j} y$ (where n_{A_j} is the number of elements of A_j) is a consistent estimator of $E(Y|A_j)$. On the other side, if we know that the probability p_i is the probability of the i-th outcome y_i of random variable Y, we get $E(Y|A_j) = \sum_{y_i\in A_j} y_i p_i / P(A_j)$, otherwise, we can give uniform weight to each observation, which yields the following consistent estimator of $E(Y|A_j)$: $\frac{1}{n_{A_j}}\sum_{y_i\in A_j} y_i$, where n_{A_j} is the number of elements of A_j. Therefore, we are able to mate $E(Y|\Im_k)$, that is a consistent estimator of the conditional expected value $E(Y|X)$ as a consequence of [12] Proposition 1.

3 Empirical analysis

In this section, we report some numerical experiments obtained with the techniques introduced to estimate the SPD, namely the nonparametric estimator based on kernel estimator and OLP estimator. To assess the empirical relevance of these estimators and the corresponding SPD estimate, we present an application to the S&P 500 index using daily data for the sample period April 28, 2013 to April 28, 2015. Firstly, we estimate the real mean return function using local polynomial smoothing technique (2.8). Thus, we evaluate the conditional expected price using formula (2.12) and (2.14). The results of this analysis are reported in Figure 1

Figure 1: Real mean return function estimation over time

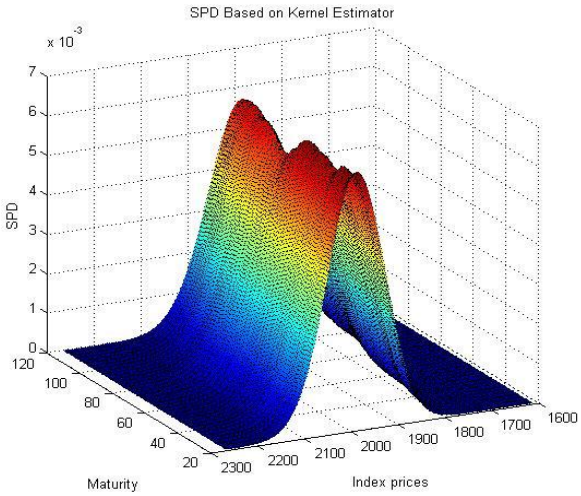


Secondly we use both estimators, namely kernel estimator and OLP, to estimate $E(S_T \mid S_t)$ as described above. Then, we use the relationship (2.5) in order to recover the SPD. In this context, we use Treasury Bond 3 months as a riskless interest rate for a period matching our selecting data. Figure 2 shows the results of estimated SPD for different maturities under both estimators.

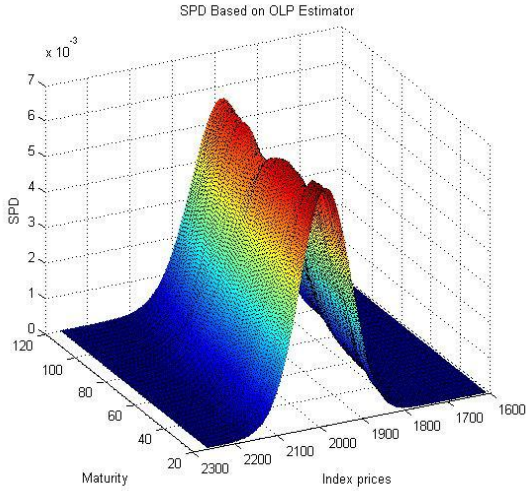
From Figure 2 we observe a slight difference in the result obtained from both methodologies. This result can be explained by the nature of the two estimators. In particular, OLP method proposed by [12] yields consistent estimators of random variables $E(X \mid Y)$ which can be used to evaluate the SPD.

Figure 2: State Price Densities obtained with Kernel estimator (A) and OLP estimator (B)

A: kernel estimator



B: OLP estimator



4 Concluding remarks

In this paper, we propose two different approaches to estimate state price density. In particular, the classical nonparametric estimator based on *kernel* estimator and a recent alternative the so called OLP estimator which uses a different approach to evaluate the conditional expectation consistently. Differently from previous studies we estimate SPD directly from the underlying asset under the hypothesis of the Black and Scholes model. To do so, firstly we examine the so called real mean return function using local polynomial smoothing technique. Then, we estimate the conditional expectation under real probability density. According the hypothesis of the Black and Scholes model, we are able to derive a closed formula for approximating the conditional expectation under risk neutral probability. This simple analysis is the starting point for extrapolating arbitrage opportunities and relevant information from different markets (futures and options) consistently with the analysis of the underlying.

Acknowledgements

This paper has been supported by the Italian funds ex MURST 60% 2014 and 2015 and MIUR PRIN MISURA Project, 2013–2015. The research was also supported through the Czech Science Foundation (GACR) under project 15-236998 and through SP2015/15, an SGS research project of VSB-TU Ostrava, and furthermore by access to the supercomputing capacity, and the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296

References:

[1] Ait-Sahalia, Y., and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* **53**(2), 499–547

[2] BENKO, M., FENGLER, M., HARDLE, W., KOPA, M (2007). On Extracting Information Implied in Options. *Computational statistics*, **22**, 543–553

[3] BLACK, F., SCHOLES, M (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637-659
[4] Breeden, D. T., and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business* **51**(4), 621–51
[5] D. S. Bates (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. Review of Financial Studies, **9**, 69–107
[6] E. A. Nadaraya, (1964). "On estimating regression," *Theory of Probability and its Applications*, **9**(1), 141-142
[7] Fan J, Gijbels I (1996). Local polynomial modelling and its applications. Chapman and Hall, London
[8] G. S. Watson (1964), "Smooth regression analysis," *Sankhya, Series A*, **26**(4), 359-372
[9] Harrison J, Kreps D.M (1979) Martingales and arbitrage in multiperiod securities markets. *Journal Economic Theory* **20**,381–408
[10] R. C. Merton (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, **3**, 125–144
[11] S. Heston (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343
[12] S. Ortobelli, F. Petronio, T. Lando, (2015) "A portfolio return definition coherent with the investors preferences," to appear in IMA-*Journal of Management Mathematics*.
[13] V. A. Epanechnikov, (1965). "Non-parametric estimation of a multivariate probability density," *Theory of Probability and its Applications*, **14**(1), 153-158

The use of linear programming to solve routing tasks in practice

Petr Kozel[1], Šárka Michalcová[2]

Abstract. Routing tasks are tasks in which the attention is focused on the proposal of the optimal route of service vehicle with a view to minimize selected criteria (costs, distance, etc.). To design the route, it is possible use linear programming, which allows us to solve these tasks. In addition to basic tasks (such as the operation of all vertices and edges of a network) it must also be taken into account other conditions which correspond to demands of practice. These may include the consideration of one-way roads or requests for partial operation of a network. This paper deals with the consideration of some specifics resulting from practice. The contribution presents data experiments based on actual practice.

Keywords: linear programming, routing tasks, optimization problems, municipal waste collection

JEL classification: C61
AMS classification: 90C05

1 Introduction

Routing tasks are tasks in which the attention is focused on the proposal of the optimal route of service vehicle with a view to minimize some selected criteria (costs, distance, etc.). To design the route, it is possible use linear programming, which allows us to solve these tasks. Basic tasks can be categorized according to two aspects: either the subject is a service of the vertex (or of the edge) or a network on which the operation is performed. Approach used to solve tasks there are focused on an edge service or an undirected network can be found in [6]. Procedure, that takes into account the existence of one-way roads, is solved in [5].

All basic approaches of solving routing tasks which are focused on an edge operation (finding the optimal route), operate with a network in which it is necessary to serve all edges at least once. These are mainly tasks of municipal waste collection, road maintenance etc. - thus the existence of an isolated network, in which the optimal route is searched. This optimal route contains each edge of the network at least once. In practice, however, we do not encounter with the existence of completely isolated network. Most defined of networks for the service are a part (subnet) of a larger network (abovenet). The network can be a specific city transport infrastructure and the subnet may be a specific set of roads (streets) which is necessary to be served. It is still true that in the subnet which is necessary to serve, every edge must be operated at least once. However, in terms of minimization, it may be used an edge of abovenet. The use of this edge will have a positive impact on the optimization criterion namely its minimizing.

In this paper, the attention is focused on the edge operation of a specific subnet which must be operated with the possibility of using edges of abovenet. Full approach is presented for mixed networks that contain directed and undirected edges representing one-way roads.

[1]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, 701 21 Ostrava, petr.kozel@vsb.cz
[2]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, 701 21 Ostrava, sarka.michalcova@vsb.cz

2 Formulation of problem

The network is specified as $N(V, E, l)$, which defines a mixed graph. This network has three attributes: $V = 1, ..., m$ represents a set of vertices, $E = 1, ..., n$ is a set of edges (directed and undirected edges) and l_{ij} represents evaluation of the edge ij, where $ij \in E$ is a road length (in kilometers). This network can be divided into two subnets. The first subnet $N_1(V_1, E_1, l)$ represents a set of edges that is necessary to be served. The second subnet $N_2(V_2, E_2, l)$ represents a set of edges that can be used for operation. It is true that $N = N_1 \cup N_2$. The aim of the task is to find a route that passes through every edge of the subnet N_1 at least once with possibility of using edges from the subnet N_2, so the total length of traveled distance is minimal and the route begins and ends at the depot. At the same time it has to be fulfilled that directed edges located both in the subnet N_1 and the subnet N_2 can only pass through in the same direction with the edge orientation. This requirement corresponds to respecting one-way roads in practice.

The basic problem has been first formulated by *Mei-Ko-Kuan* and is known as "*The Chinese Postman Problem*" [8]. Solution of this problem can be found using the Eulerian path. Assuming conditions for existence of the Eulerian path are fulfilled, solution can found using the *Edmonds' algorithm* [1]. If conditions for existence of the Eulerian path are not fulfilled, it is necessary to choose an alternative procedure that removes this shortage. In practice, it is unproductive passing through some sections. Approaches leading to find the Eulerian path in undirected and mixed graphs, and which do not fulfilled conditions for the Eulerian path existence, are described in detail in [6] and in [5].

It should be noted that the network (respectively graph) in which the task of the optimal route finding is solved, must be strongly connected. Thus for every two vertices i, j a directed path from the vertex i to j and a directed path from j to i can be found. Situation model is shown in Figure 1. The network N consists of 8 nodes and of 16 edges. The subnet N_1 is indicated by solid lines and represents edges that must be operated. The subnet N_2 is shown by dashed lines and represents edges that can be used for operation.

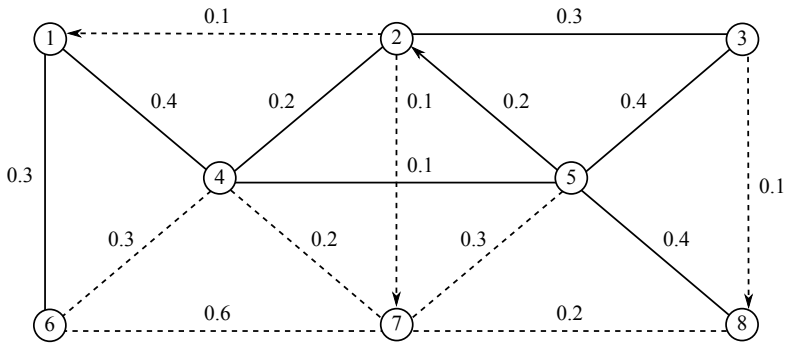


Figure 1 The network N with subnets N_1, N_2

3 Mathematical model

To find the optimal route, it can be used a mathematical model that determines how many times will be used the particular edge ij on route. Mathematical model (1) - (6) is based on a model that is designed for a mixed network and which has been published in [5]. Input data of mathematical model correspond to both subnets N_1, N_2 and form together the network N. Evaluation of existing edges in the subnet N_1 is marked with the constant d_{ij}, evaluation of edges in the subnet N_2 as e_{ij} and the following relationships are valid: $\{d_{ij}\} \subset \{l_{ij}\}$ and $\{e_{ij}\} \subset \{l_{ij}\}$. This way of constants defining ensures to work only with existing edges and corresponds to the spare input data matrices. The same situation is dealt with dynamic declaration in computational environment.

In this paper, the practice requirement has been taken into account: respecting of one-way roads. Possibility of run through a directed edge against the direction of the orientation is penalized by introduction of a prohibitive constant T, in detail in [5].

Values shown in the Table 1 correspond to the given example.

Input data																
E_1	(14)	(16)	(23)	(24)	(25)	(32)	(35)	(41)	(42)	(45)	(52)	(53)	(54)	(58)	(61)	(85)
d_{ij}	0.4	0.3	0.3	0.2	T	0.3	0.4	0.4	0.2	0.1	0.2	0.4	0.1	0.4	0.3	0.4
E_2	(12)	(21)	(27)	(38)	(46)	(47)	(57)	(64)	(67)	(72)	(74)	(75)	(76)	(78)	(83)	
e_{ij}	T	0.1	0.1	0.1	0.3	0.2	0.3	0.3	0.6	T	0.2	0.3	0.6	0.2	T	

Table 1 Input data given example

$$\text{Min} \sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij} \cdot z_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{m} e_{ij} \cdot y_{ij} \tag{1}$$
$$\text{exists } d_{ij} \qquad\qquad \text{exists } e_{ij}$$

$$z_{ij} + z_{ji} \geq 1, \text{ for } i,j = 1,...,m, \text{ where exists } d_{ij} \tag{2}$$

$$\sum_{\substack{j=1 \\ \text{exists } d_{ij}}}^{m} z_{ij} + \sum_{\substack{j=1 \\ \text{exists } e_{ij}}}^{m} y_{ij} = \sum_{\substack{j=1 \\ \text{exists } d_{ij}}}^{m} z_{ji} + \sum_{\substack{j=1 \\ \text{exists } e_{ij}}}^{m} y_{ji}, \text{ for } i = 1,...,m \tag{3}$$

$$z_{ij} \in \mathbb{Z}^+, \text{ for } i,j = 1,...,m, \text{ where exists } d_{ij} \tag{4}$$

$$y_{ij} \in \mathbb{Z}^+, \text{ for } i,j = 1,...,m, \text{ where exists } e_{ij} \tag{5}$$

The objective function (1) represents the total distance you need to travel during the operation of route. The first part of the expression is the distance traveled during the operation of edges of the subnet N_1("necessary" edges), and the second one represents the distance traveled using edges from the subnet N_2("possible" edges). The variable y_{ij} is an integer variable that expresses the number of edge runs in the subnet N_1. The variable y_{ij} is an integer which determines the number of runs through edges of the subnet N_2. The condition (2) ensures that each existing edge in the subnet N_1 will be included in the move at least once. Prohibitive constant $P \in \mathbb{Z}^+$ restricts the number of use the specific edge ij in the subnet N_2. The condition (3) provides that every visited vertex will be consequently leaved. Obligatory conditions (4) and (5) ensures that variables z_{ij} and y_{ij} can take only non-negative integers.

4 Numerical experiments

In this chapter will be presented experiments that have been realized with the mathematical model (1) - (6) in computational environment *Xpress-IVE* with real data. Before the presentation of results, the procedure of solution is introduced on a given example. Let us suppose the network which is shown in the Figure 1. At first, let us focus the attention on the solution that may be obtained during the route design, assuming it is not possible to use other edges than edges of the subnet N_1. The optimal route (1-4-5-8-5-2-3-5-2-4-1-6-1) corresponding to this solution is shown on Figure 2 and its length is **3.6** km.

The optimal solution is obtained using the mathematical model (1) - (6) which also allows us to use edges of the subnet N_2. This optimal route (1-6-4-5-3-2-7-8-5-2-4-1) is shown in the Figure 3. The route length is **2.9** km.

Now we focus our attention on solving real problems. The feasibility of the proposed procedure and of the mathematical model was validated in terms of municipal waste collection in *Rychnov nad Kněžnou*. The network of this town consists of 761 vertices (intersection and the end of streets) and of 917 edges (streets). In the current state, the operation of the area is divided to three parts corresponding to three days (*Monday, Tuesday, Wednesday*). There were solved three tasks on basis of the initial situation. These tasks correspond to the defined waste collection routes.

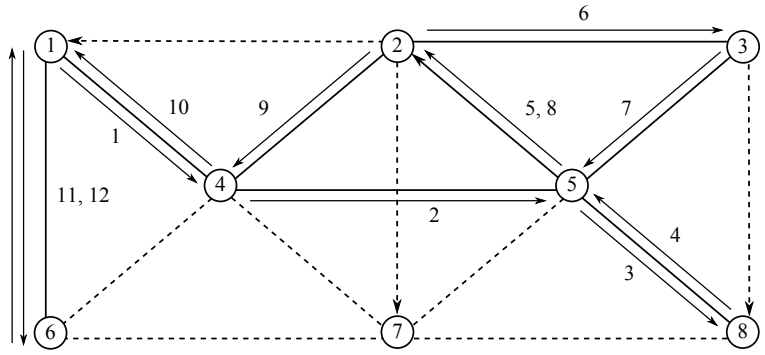


Figure 2 Solution corresponding to the exclusive use of edges of the subnet N_1

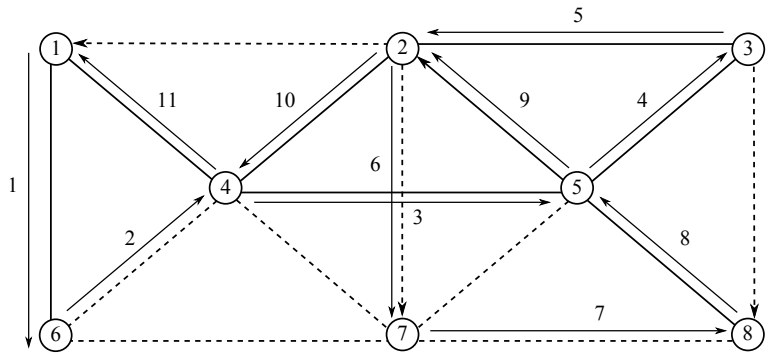


Figure 3 Solution corresponding to the use of edges in the subnets N_1 and N_2

For each task (*Monday, Tuesday, Wednesday*) was at first solved a variant (*default network*) when the optimal route is searched in the subnet N_1, which contains only edges that are necessary to be served. Then, it was subsequently solved a variant (*extended network*), when it is possible to use edges of the subnet N_2 for the operation of the default network.

Results of experiments are shown in the Tables 2 - 4.

Problem	Problem size [vertices/edges/one-way]	Total distance traveled [km]	Computational time [sec]
Default network	263/299/11	56.22	0.375
Extended network	498/618/27	54.42	4.329

Table 2 Results of computational experiments - *Monday*

Table 2 summarizes results that are related to the first task: the optimal route designing for *Monday*. The subnet N_1 (*default network*), which must be operated, is formed by 263 vertices and 299 edges, 11 of which are directed (one-way roads). The subnet N_2 (*extended network*), which may be used for the operation in the subnet N_1, consists of 498 vertices and of 618 edges, 27 of which are directed. Achieved savings of the total distance traveled are **1.8** km and there were used **15** edges from the extended network. Thanks to the effective runs, the objective function has been degreased from the original value 56.22 km to 54.42 km.

Table 3 summarizes results that are related to the second task: the optimal route designing for *Tuesday*. The subnet N_1 (*default network*), which must be operated, is formed by 211 vertices and 237 edges, 25 of which are directed (one-way roads). The subnet N_2 (*extended network*), which may be used for the operation in the subnet N_1, consists of 550 vertices and of 680 edges, 14 of which are directed. Achieved savings of the total distance traveled are **4.08** km and there were used **28** edges from the extended network. Thanks to the effective runs, the objective function has been degreased from the original value 53.1 km to 49.02 km.

Table 4 summarizes results that are related to the third task: the optimal route designing for *Wednes-*

Problem	Problem size [vertices/edges/one-way]	Total distance traveled [km]	Computational time [sec]
Default network	211/237/25	53.1	0.25
Extended network	550/680/14	49.02	1.1

Table 3 Results of computational experiments - *Tuesday*

Problem	Problem size [vertices/edges/one-way]	Total distance traveled [km]	Computational time [sec]
dDefault network	339/387/2	127.78	8.234
Extended network	422/530/35	115.07	140.304

Table 4 Results of computational experiments - *Wednesday*

day. The subnet N_1 (*default network*), which must be operated, is formed by 339 vertices and 387 edges, 2 of which are directed (one-way roads). The subnet N_2 (*extended network*), which may be used for the operation in the subnet N_1, consists of 422 vertices and of 530 edges, 35 of which are directed. Achieved savings of the total distance traveled are **12.71** km and there were used **33** edges from the extended network. Thanks to the effective runs, the objective function has been degreased from the original value 127.78 km to 115.07 km.

5 Conclusion

In this paper, the attention was focused on design of optimal service routes of vehicles using linear programming. There was introduced a mathematical model, which builds on a previous research [5], and takes into account practical requirements. This is respecting one-way roads, but mainly respecting the case, when for the service of the subnet N_1 (*default network*) it is also possible to use other edges that are related to the solved network. This fact offers both a possibility of effective runs and also a possibility of reducing value of the objective function as the total distance traveled.

Introduced mathematical model was applied on practical tasks. The task of municipal waste collection in *Rychnov nad Kněžnou* has been solved in the framework of computational experiments. When evaluating results, the total distance traveled was compared in two variants of calculation:

- the operation of the subnet N_1 (*default network*), which includes only edges that must be served,
- the operation of the subnet N_1, when can also be used edges from subnet N_2 (*extended network*).

As seen on the Tables 2 - 4, in all tasks, there were obtained in the operation of subnet N_1 with using the edges from the subnet N_2 obtained better solutions than in the operation of the subnet N_1 without this possibility.

Experiments were realized in the environment *Xpress-IVE* and the computational time was maximally in order of tens seconds in all cases. Due to the fact, that there are not some tasks of the operational management, time-consuming calculations are negligible.

Acknowledgements

This paper was supported within Operational Programme Education for Competitiveness - Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Bondy, A., Murty, U.: *Graph Theory*. Springer, San Francisco, 2008.

[2] Friedrich, V.: *Matematika na počítači pro nematematiky*. Vydavatelství VŠB - TU Ostrava, Ostrava, 2013.

[3] Guéret, Ch., Prins, Ch., et al.: *Applications of optimization with Xpress-MP*. Blisworth: Dash Associates, United Kingdom, 2002.

[4] Janáček, J., et al.: *Modelovanie and Optimalizácia*. Žilina, Žilinská univerzita, 2013.

[5] Kozel, P., Michalcová, Š.: The use of linear programming in solving service tasks of mixed transport networks. In: *Proceedings of 11th International Conference Strategic Management and its Support by Information Systems 2015*, Uherské Hradiště, 2015.

[6] Kozel, P., Michalcová, Š., and Friedrich, V.: The using of linear programming for solving the municipal waste collection problem. In: *Proceedings of 32nd International Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 483–488.

[7] Rybička, J.: *LaTeXpro začátečníky*. Brno, Konvoj, 2003.

[8] Skiena, S., and Pemmaraju, S.: *Computational Discrete Mathematics*. Cambridge University Press, New York, 2003.

[9] Wolfram, S. *The Mathematica Book*. Cambridge University Press, New York, 1999.

[10] *XPRESS-Mosel "User guide"*. Blisworth: Dash Associates, United Kingdom, 2005.

Measuring the Age of Machinery and Equipment in Czech Republic

Igor Krejčí[1], Petr Mazouch[2]

Abstract. Balances of fixed capital are constructed on the basis of international standards. Because the indicators of fixed capital are required to be in market prices and based on the real service lives of assets, the business bookkeeping data are insufficient. Moreover, the stocktaking for the whole economy is impossible. Therefore, processes of retirement, decay and obsolescence of fixed capital are subject of the modelling. Stocks and consumption of fixed capital are estimated by a Perpetual Inventory Method.
In the paper, the transformation of the Czech model of the Perpetual Inventory Method is applied to estimate the indicator, which is not in the scope of annual accounts. On the basis of Markov chain approach to maintenance models we estimate the age structure and the average age of chosen assets in the Czech economy. The analysis uses also the basic demographic indicators to describe the ageing situation in Czech industries. Due to the different service lives in various industries, the average age of machinery and equipment is compared on the basis of life expectancy and quantile analysis. Industries with extreme age of machinery and equipment are presented in the paper.

Keywords: Average age, average service life, fixed capital, Markov chain.

JEL Classification: E22**,** C44
AMS Classification: 91B25

1 Introduction

Balances of fixed capital for the statistics of the national economy should be constructed on the basis of international standards [8]. Because the indicators of fixed capital are required to be in market prices and based on the real service lives of assets the business bookkeeping data are insufficient [6, 10]. Moreover, the stocktaking for the whole economy is impossible. Therefore, processes of retirement, decay and obsolescence of fixed capital are subject of modelling. Consequently, stocks and consumption of fixed capital are estimated by a Perpetual Inventory Method (PIM) [17].

Despite the goal of national accounts aims on measuring the value of assets, it is possible to use same inputs used for the PIM to estimate the average age of assets as weighted average of surviving investments [9, 19]. Such indicator can describe the development in the industries and could be used as a quantitative indicator of modernization for S.M.A.R.T defined goals [7, 13] as the modernisation is a common goal for public policy in regulated industries like agriculture [14] or education [15, 16].

The aim of the paper is to estimate the average age and the age structure of the machinery and equipment in the Czech Republic. On the basis of Markov chain approach to maintenance models, we estimate the age structure and the average age of chosen assets in the Czech economy [11]. The analysis uses also the basic demographic indicators to describe the ageing situation in the Czech industries. Due to the different service lives in various industries, the average age of the machinery and equipment is compared on the basis of life expectancy and quantile analysis. Industries with extreme age of machinery and equipment are presented in the paper.

In the first part of the paper we describe the used methodology. Second part contains the application on official data sources of the Czech statistical office. Finally, we discuss the limitation of proposed approach and topics for future research.

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague 6, 16521, krejcii@pef.czu.cz.
[2] University of Economics, Prague, Department of Economic Statistics, nám. W. Churchilla 4, Prague 3, 13067, mazouchp@vse.cz.

2 Material and methods

2.1 Average age estimation

Despite there are other ways to estimate the age, we apply the transformation to Markov chains, which we have already supported by software implementation [12]. Although there are other ways to estimate the average age [1, 20], the applied model must estimate (and also record) the consumption and retirement of fixed assets separately for each yearly investment.

The net capital stock value is decreasing due to the wear and tear, obsolescence etc. Therefore, our estimation of age characteristics is based on the gross stock value. We consider using net capital stock for such estimation similar to use labour productivity for age groups to weight the average age of human population.

For the estimation we use same inputs as the Czech Statistical Office for construction of balances of fixed capital [3]. Most important inputs are time series of investment, price indices and other changes. For model parameters we use the average service lives used for the official PIM, which vary from 4.0 to 20.5 for transport equipment and 6.9 to 20.0 for other machinery and equipment, retirement function is lognormal [2].

The aging/retirement process in the PIM depends only on a present state and transition probabilities; the retirement process has the Markov property (see e.g. [18]):

$$P\left(X_n = i_n \middle| X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\right) = P\left(X_n = i_n \middle| X_{n-1} = i_{n-1}\right). \tag{1}$$

We transform the lognormal retirement function used in the standard PIM into transition matrix $\boldsymbol{P}$ with m+1 rows and columns, where m states the maximum age and m+1th state aggregates the retired assets. The conditional probability that the asset will age to period n+1 is $p_{n,n}$+1. Conditional probability of retirement in nth period is $p_{n,m}$+1. The probability of asset retirement in nth period is a_n; we take this probability from official retirement function. Survival probability is r_n. The elements of $\boldsymbol{P}$ are from equations (2)-(4):

$$r_{i-1} - r_i = a_i, i = 0, \ldots, m-1, r_0 = 1, \tag{2}$$

$$p_{n,m+1} = \frac{a_n}{r_{n-1}}, \tag{3}$$

$$p_{n,n+1} = \frac{r_n}{r_{n-1}}. \tag{4}$$

Then we calculate aging and retirement of fixed capital in the year t by (5):

$$\boldsymbol{g}_t^T = \boldsymbol{g}_{t-1}^T \boldsymbol{P}. \tag{5}$$

Each element of vector $\boldsymbol{g}_t$ from (5) contains surviving investments from previous years, i.e. first element contains current investment, second element of the vector contains one year old investment etc.

Information about the distribution of age of fixed capital can be described not only by average but also by other quantiles as lower quartile (25% quantile), median (50% quantile), upper quartile (75% quantile) or deciles (especially the first and the ninth is used). Together with mean it provides more complex information about distribution with respect to extreme values (by which mean is influenced). Results are described by box-plot graphs.

3 Result and discussion

Figure 1 shows the average age of the machinery and equipment in period 2000-2013 in the Czech Republic. The chosen industries represent the industries with highest and lowest average age in 2000 and in 2013 and the industries with biggest difference between the min and max average age in the period 200-2013. Industries are from classification CZ-NACE [4].

As the age is dependent on the average service life, we measure also the change of ratio age to service life. In the given period, the industries with extreme behaviour are same for the absolute and relative change. Activities of membership organisations represents the industry with lowest average age in 2000 but also the industry the highest increase of average age in the period 2000-2013.

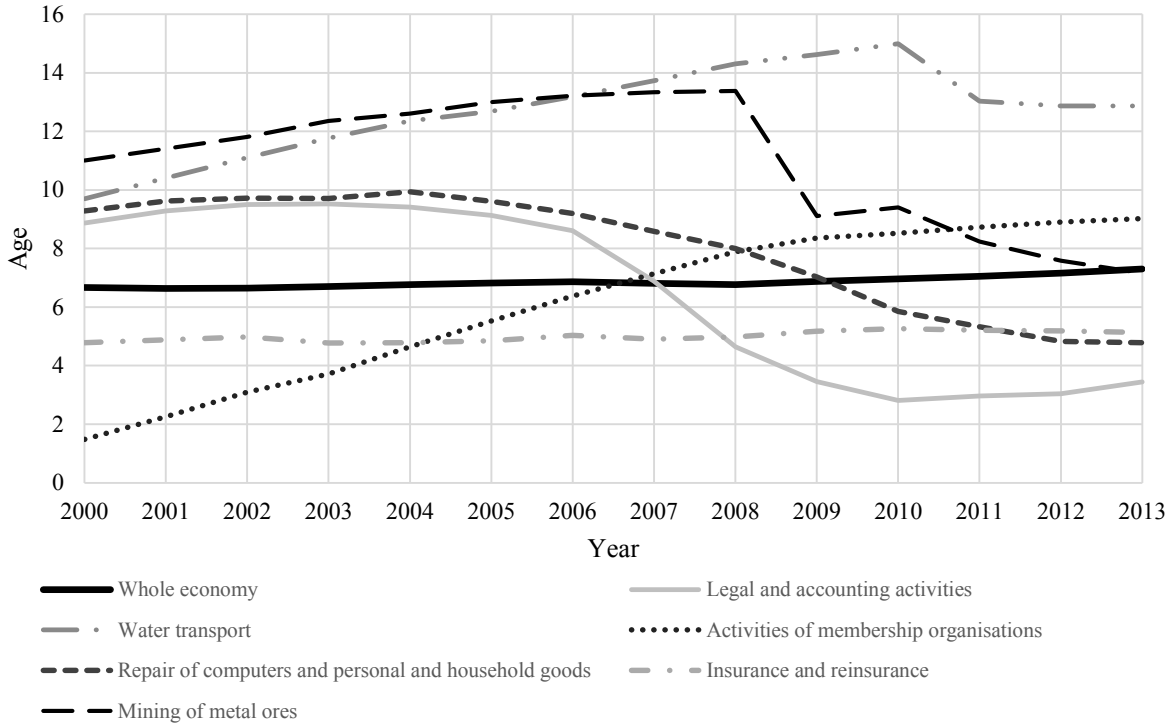


Figure 1 Age of the machinery and equipment in the Czech Republic in 2000-2013

The average in the whole Czech economy has stable characteristics; in period 2000-2013 average age is 6.86 and the standard deviation 0.19. The average age has increased by 0.63 years in that period. Slight change in average is caused by change of whole distribution as we can see from Figure 2. First decile (lower whisker in the plot) increased from 0.9 to 1.3 years between 2000 and 2013, lower quartile (lower edge of the box) increased from 2.6 to 3.1, median (middle line in the box) increased from 5.3 to 6.2, upper quartile (upper edge of the box) increased from 9.5 to 10.2 and the ninth decile (upper whisker) increased from 14.3 to 14.8 years.

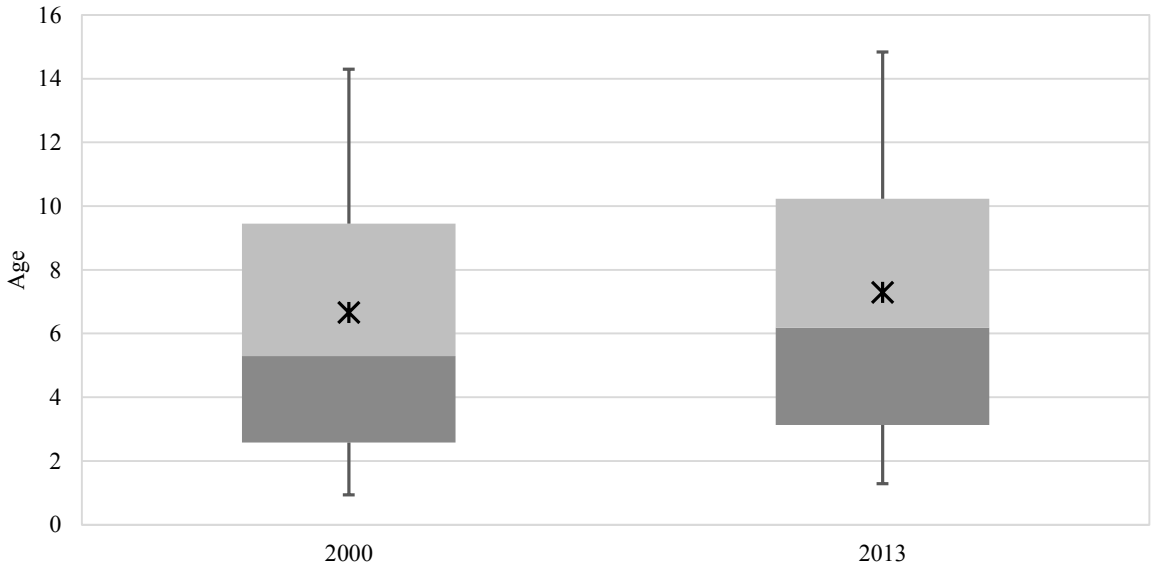


Figure 2 Age distribution of the machinery and equipment in the Czech Republic in 2000 and 2013

As the characteristics for whole economy are stable (or almost stable) there are differences among industries in the economy (and also we can find differences between transport equipment and the other machinery and equipment. In general, the industries with higher service life tends to have the higher average age too, therefore, the more important is the development in time.

Figure 3 compares the average service life and average age of transport equipment in 2000 and 2013. We can identify that transport equipment aged in majority of industries. The biggest change is in the industry of Activities of membership organisations (the change is 8.5 years between 2000 and 2013) and in the industry of Transport via pipeline (change 8.05 years). The average age of the industry of Transport via pipeline is also very close to its average service life, all other industries are below the line (average age of the transport equipment is lower than average service life).

For all industries with the highest average service life as Water transport, Manufactories of basic metals or Mining of hard coal between 2000 and 2013 average age increased significantly – about five years for all those industries.

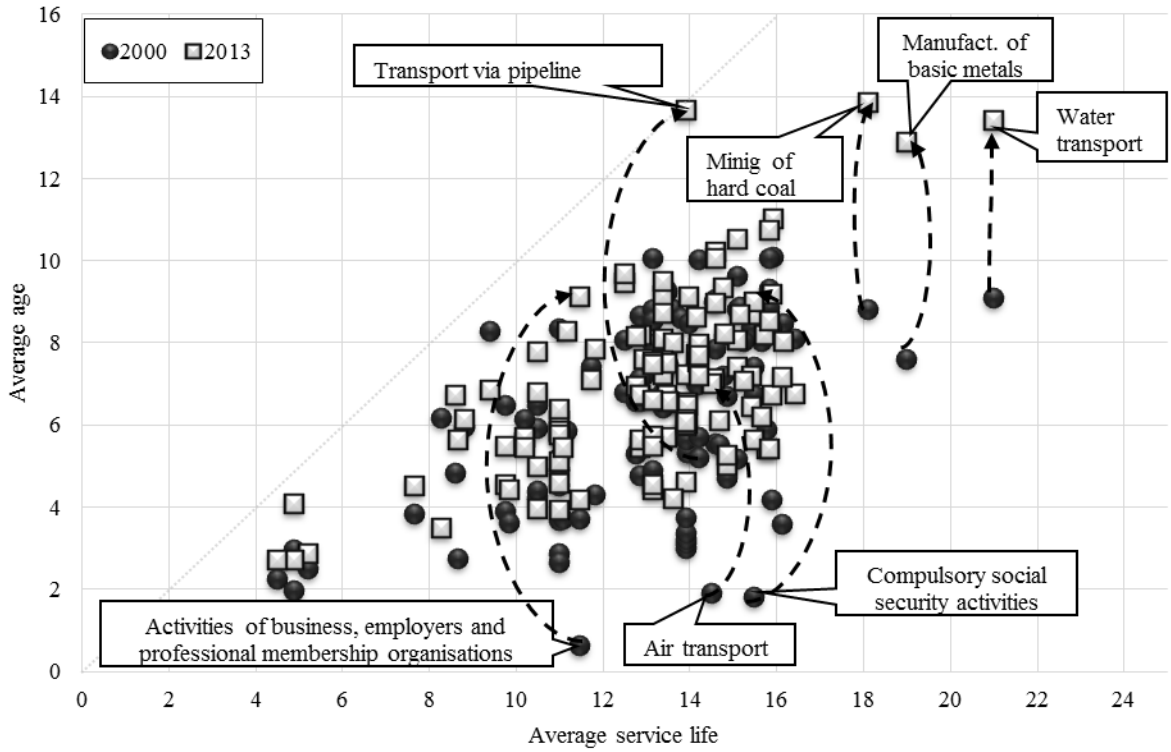


Figure 3 Comparison of Average service life and Average age of Transport Equipment in industries, 2000 and 2013

For other machinery and equipment the situation is slightly different. We can see that industries in right part of the graph (industries with the highest average service life) aged, i.e. all points moved to higher position. In 2000 there was one industry (Repair of computers and personals and household goods) above the dotted line in graph, which represents equity between average service life and average age. Investments in this industry caused that in 2013 it is below the line and average age is lower than average service life (average age has dropped about 5 years).

The biggest changes in the average age were in industry of Rail transport (increase by 7.4 years between 2000 and 2013), Manufacture and distribution of gas (increase of 6.3 years) and Veterinary activities (increase by 6.4 years). In opposite way, the age in industries of Repair of computers and personal and household goods has dropped (decrease by 5.0 years between 2000 and 2013).

As for transport equipment, other machinery and equipment industries with the highest average service life aged significantly - Manufacture of paper and paper products (2 years between 2000 and 2013), Manufacture basic metals (4.7 years) and Manufacture of wearing apparel (3 years).

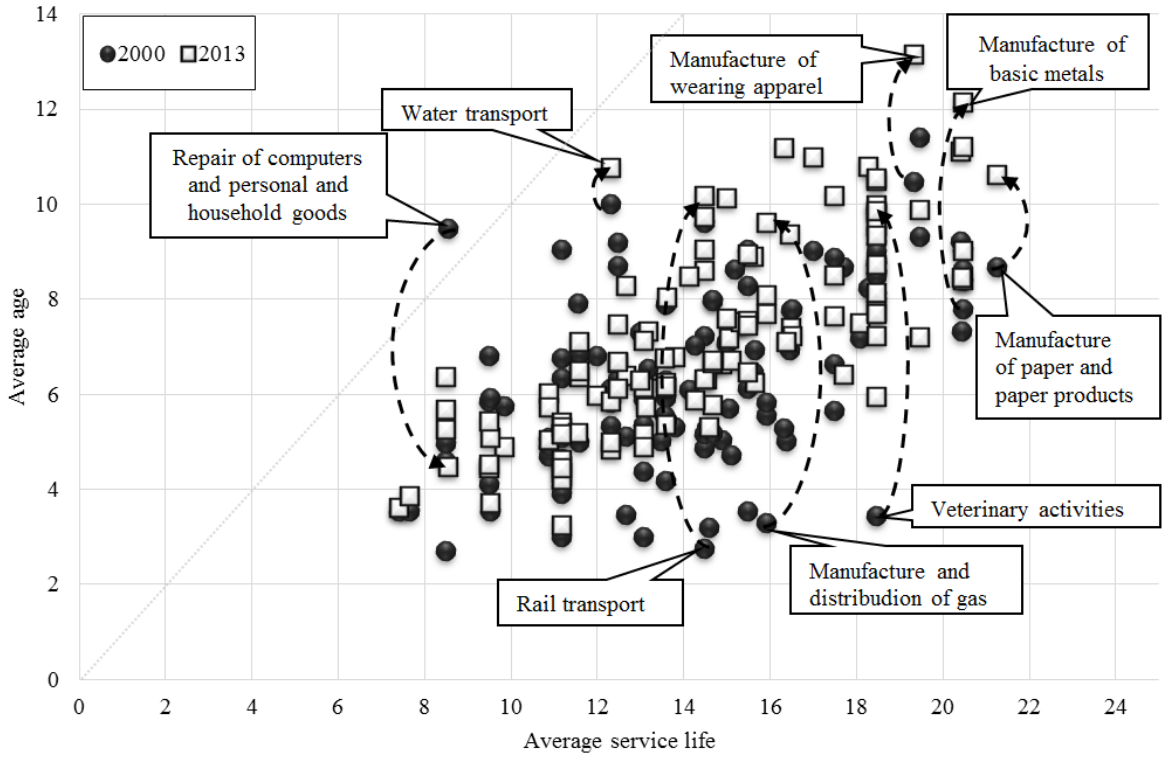


Figure 4 Comparison of Average service life and Average age of Other machinery and equipment in industries, 2000 and 2013

4 Conclusion

This paper dealt with the estimation of average age of the machinery and equipment in the whole Czech economy. Average age of assets was estimated by the adjusted perpetual inventory method based on Markov chains. To depict the development of average age, the age must be compared with average service life. Otherwise, the industries with higher service life could be interpreted wrongly as more obsolete.

In this paper we focus on the estimation of average age and identification of extreme industries with the stress on the biggest changes of average age in the examined period. Industry of Activities of membership organisation performed the biggest increase of average age (from both points of view - absolute change and relative to the service life). Industry of Legal and accounting activities performed biggest drop of average age, which could be interpreted as growing industry together with the modernisation process. Nevertheless, the age dynamics is different for transport equipment and other machinery and equipment.

As far as the current publication contains machinery and equipment under twenty thousand CZK, which was the threshold for the fixed capital before the revision [5], we consider as necessary to actualise the official service lives. Despite the fact we propose the average age of equipment as supportive indicator of modernisation, there are qualitative aspects of the statistics on investment, which cannot be full reflected by the model estimation. It is possible to buy a new (in terms of not used) but already obsolete asset. For the model purposes such asset will be "newborn" with the same age as the modern high tech asset. To support our approach, this is partly included in the PIM and thus our model also. As described in the materials and methods section, the amount of assets is not expressed by the number of pieces but by the gross value. Then the difference between the obsolete and modern new asset is differentiated by the price.

References

[1] Australian Bureau of Statistics: *5204.0 - Australian System of National Accounts, 2013-14*. [online], Australian Bureau of Statistics, Sydney, 2014. Available http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5204.02013-14?OpenDocument [6 February 2015].

[2] Czech Statistical Office: *Gross National Income Inventory*. [online], Czech Statistical Office, Prague, 2002. Available: http://apl.czso.cz/nufile/GNI_CZ_en.pdf. [26 November 2012]

[3] Czech Statistical Office: *Annual National Accounts*. [online], Czech Statistical Office, Prague, 2015. Available: http://apl.czso.cz/pll/rocenka/rocenka.indexnu_en [1 February 2015]
[4] Czech Statistical Office: *Classifications*. [online], Czech Statistical Office, Prague, 2015. Available: http://www.czso.cz/eng/redakce.nsf/i/classifications [4 February 2015].
[5] Czech Statistical Office: *Mimořádná revize národních účtů* [online], Czech Statistical Office, Prague, 2015. Available: http://apl.czso.cz/pll/rocenka/rocenka.avizo_revize [4 February 2015].
[6] Diewert, W. E.: Issues in the Measurement of Capital Services, Depreciation, Asset Price Changes and Interest Rates. In: *Measuring Capital in the New Economy* (Corrado, C., Haltiwanger, J., Sichel, D. eds.). University of Chicago Press, Chicago, 2005, 479-542.
[7] Doran, G. T.: There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review* **70** (1981), 35-36.
[8] Eurostat, United Nations, International Monetary Fund, OECD, World Bank: *System of National Accounts 2008*, New York. 2009.
[9] Harper, P.: An integrated system of capital flow and stock statistics. Measuring Capital – Beyond the Traditional Measures. In: *Proceedings of the Seminar Session of the 2007 Conference of European Statisticians*. United Nations, New York and Geneva, 2008, 41-48.
[10] Hulten, Ch., R., Wykoff, F., C.: Issues in the measurement of economic depreciation introductory remarks, *Economic Inquiry* **34** (1996), 10-23.
[11] Krejčí, I.: Computation of Stocks and Consumption of Fixed Capital with use of Markov Chain. In: *Proceedings of the 28th International Conference on Mathematical methods in Economics* (Houda, M., Friebelová, J. eds.). University of south Bohemia in České Budějovice, České Budějovice, 2010, 365–370.
[12] Krejčí, I., Sixta. J.: Využití alternativních metod při odhadech stavů a spotřeby fixního kapitálu. *Politická ekonomie*, **60** (2012), 780-800.
[13] Lawler, J., Bilson, A.: *Social Work Management and Leadership: Managing Complexity with Creativity*. Routledge, New York, NY, 2013.
[14] Ministry of Agriculture of the Czech Republic: *Rural Development Programme of the Czech Republic for 2007 – 2013*. [online]. Ministry of Agriculture of the Czech Republic, Prague, 2008. Available: http://eagri.cz/public/web/file/10574/RDP_November_2008.pdf [24 November 2014].
[15] Ministry of Education, Youth and Sport: *Národní program rozvoje vzdělavání v Ceske republice, Bila kniha*. [online], Institute for Information on Education, Prague, 2001. Available: http://www.msmt.cz/reforma-terciarniho-vzdelavani/bila-kniha/narodni-program-rozvoje-vzdelavani-v-ceske-republice-bila-kniha-2001?highlightWords=b%C3%ADl%C3%A1+kniha [20 February 2015]
[16] Ministry of Education, Youth and Sport: (2014) Strategie vzdělávací politiky České republiky do roku 2020. [online], Ministry of Education, Youth and Sport, Prague, 2014. Available: http://www.msmt.cz/uploads/Strategie_2020_web.pdf [3 April 2015].
[17] OECD: *Measuring Capital - OECD Manual 2009*: *Second edition*, OECD Publishing, Paris, 2009.
[18] Tijms, H. C.: *A First Course in Stochastic Models.* John Wiley & Sons, Chichester, 2003.
[19] U.S. Bureau of Economic Analysis: *How does BEA estimate the average age of net stocks of fixed assets and consumer durables?* [online], U.S. Bureau of Economic Analysis, Washington, DC, 2013. Available: http://www.bea.gov/faq/index.cfm?faq_id=1031 [20 January 2015]
[20] U.S. Bureau of Economic Analysis: *Current-Cost Average Age at Yearend of Private Equipment by Industry*. [online], U.S. Bureau of Economic Analysis, Washington, DC, 2014. Available: http://www.bea.gov/iTable/iTableHtml.cfm?reqid=10&step=3&isuri=1 &1003=108&1004=2000&1005=2013&1006=a&1011=0&1010=x [2 February 2015].

Gold, currencies and market efficiency

Ladislav Kristoufek[1], Miloslav Vosvrda[2]

Abstract. Gold and currency markets form a unique pair with specific interactions and dynamics. We focus on the efficiency ranking of gold markets with respect to the currency of purchase. By utilizing the Efficiency Index (EI) based on long-term memory, fractal dimension and approximate entropy on a wide portfolio of 142 gold price series for different currencies, we construct the efficiency ranking based on the extended EI methodology we provide. We uncover rather unexpected results as the gold prices in major currencies lay among the least efficient ones whereas very minor currencies are among the most efficient ones. We attribute such counterintuitive results partly to a unique period of examination (2011-2014) characteristic by quantitative easing and rather unorthodox monetary policies, as well as other factors discussed in some detail.

Keywords: efficient market hypothesis, gold, currencies

JEL classification: C58, G14, Q02

1 Introduction

For decades, the efficient market hypothesis (EMH) has been a building block of financial economics. In his fundamental paper, Fama [8] summarizes the then actual empirical findings following the theoretical papers of Fama [7] and Samuelson [23]. Fama [9] then recalls various issues of the hypothesis and reviews newer literature on the topic. The capital market efficiency is standardly parallelized with the informational efficiency so that the markets are efficient as long as all the available information is fully reflected into market prices [8]. Depending on the level of information availability, the EMH is usually separated into three forms – weak (historical prices), semi-strong (public information), and strong (all information, even private) [9]. Even though the theory has been repeatedly challenged on both theoretical [16] and empirical [4] grounds, it still remains a popular and fruitful topic of financial research.

Studies of the foreign exchange rates efficiency, in the same way as of the other assets, primarily focus on testing whether a given currency or a set of currencies may or may not be considered efficient. To reflect this point, Kristoufek & Vosvrda [13] introduce the Efficiency Index (EI) which can be used to rank assets according to their efficiency. In addition, the index is very flexible and it can incorporate various measures of the market efficiency. In the original study, Kristoufek & Vosvrda [13] study 41 stock indices and find the Japanese NIKKEI to be the most efficient one. From a geographic perspective, the most efficient indices are localized in Europe and the least efficient ones in Asia and Latin America. Kristoufek & Vosvrda [15] further focus on the index specification and show that approximate entropy adds a significant informative value to the index. Kristoufek & Vosvrda [14] then study efficiency across various commodity futures and uncover that energy commodities are the most efficient ones whereas the other agricultural commodities such as cattle and hogs are the least efficient ones. Here we focus on efficiency ranking of the gold market with respect to a currency used for the purchase, and we also contribute to the discussion on statistical properties of the Efficiency Index.

2 Methods

Coming back to the roots of the efficient market hypothesis in 1965, the treatment has been split into two main branches – based on the random walk hypothesis [7] and following the martingale specification [23].

[1]Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU, kristouf@utia.cas.cz

[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU, vosvrda@utia.cas.cz

We follow the latter approach as it is less restrictive and it assumes the returns of the efficient market to be only serially uncorrelated and with finite variance. This straightforward treatment enables us to use various measures of market efficiency and use them to construct the Efficiency Index, which allows to rank financial assets according to their efficiency. In this section, we briefly describe the Efficiency Index, its components and its statistical treatment.

2.1 Capital market efficiency measure

Kristoufek & Vosvrda [13, 14, 15] define the Efficiency Index (EI) as

$$EI = \sqrt{\sum_{i=1}^{n} \left(\frac{\widehat{M_i} - M_i^*}{R_i} \right)^2}, \tag{1}$$

where M_i is the ith measure of efficiency, $\widehat{M_i}$ is an estimate of the ith measure, M_i^* is an expected value of the ith measure for the efficient market and R_i is a range of the ith measure. EI is thus a distance from the efficient market situation. The index can comprise various efficiency measures but these need to be bounded, which turns out to be rather restrictive. We utilize three efficiency measures, which meet such criterion and which are frequently used in market efficiency studies [2, 3, 6, 5, 27, 26, 18] – Hurst exponent H as a measure of long-range dependence with an expected value of 0.5 for the efficient market ($M_H^* = 0.5$), fractal dimension D with an expected value of 1.5 ($M_D^* = 1.5$), and the approximate entropy with an expected value of 1 ($M_{AE}^* = 1$). Hurst exponent and fractal dimension share their range for stationary processes whereas approximate entropy does not. For this point, we need to rescale the approximate entropy part of the Efficiency Index so that we have $R_{AE} = 2$ and $R_D = R_H = 1$.

Specifically for Eq. 1, we utilize two measures of long-range dependence, two measures of fractal dimension and a measure of approximate entropy. For the long-range dependence estimation, we use the local Whittle estimator and the GPH estimator, which are suitable for short time series with possible weak short-term memory, and they are consistent and asymptotically normal [10, 1, 22, 24, 25, 19]. For the fractal dimension, we utilize the Hall-Wood and Genton estimators, which share desirable statistical properties for short time series [11, 12]. For the approximate entropy, we follow Pincus [20] and Pincus & Kalman [21]. For more details about the Efficiency Index construction and its components, please refer to Refs. [13, 14, 15].

2.2 Statistical inference

The original Efficiency Index [13] is a point estimate of the true index value. This poses problems when discussing the results and their statistical validity. We tackle this issue by introducing a new approach to estimating EI which stems in the following steps:

1. Obtain the estimated components $\widehat{M_i}$ of the Efficiency Index according to Eq. 1.
2. Shuffle the underlying return series.
3. Estimate the components of the Efficiency Index for the shuffled series, and label these as $\widehat{M}_{i,shuffle}$.
4. Use $\widehat{M}_{i,shuffle}$ in place of M_i^* in Eq. 1.
5. Obtain $\widehat{EI}$ based on the previous steps.
6. Repeat N times.
7. Obtain necessary statistics based on these N estimates.

This way, we obtain an estimate of the Efficiency Index which controls for the potential finite sample bias and the influence of distributional properties of the analyzed series. For purposes of our study, we set $N = 50$.

3 Results and discussion

We study the efficiency ranking of the gold prices quoted in different currencies. The portfolio of study consists of 142 worldwide currencies, which are listed in Table 1. The dataset has been obtained from oanda.com, which provides a large set of FX pairs as well as gold (and other precious metals) prices in various currencies. The covered period ranges between 1.1.2011 and 30.11.2014, which totals 1430 observations for each of the 142 analyzed currencies. These currencies cover almost all available and traded fiat currencies in addition to Bitcoin, the most popular and used cryptocurrency.

For the efficiency ranking, we use the Efficiency Index (Eq. 1) with adjustments described in Sec. 2.2. Specifically, we utilize two measures of long-range dependence – the local Whittle estimator and the GPH estimator –, two measures of fractal dimension – the Hall-Wood estimator and the Genton estimator – and the approximate entropy as proposed by Pincus & Kalman [21]. Using 50 repetitions (shuffling), we obtain the estimated Efficiency Index as a median value with a corresponding standard error for more information about the estimate precision.

The resulting ranking of gold prices with respect to the used currency is presented in Table 1. The ranking is rather unexpected or even surprising. Practically all of the most liquid currencies – the US dollar, the British pound, the Australian dollar, the New Zealand dollar, the Japanese yen, the Euro, the South Korean won, and the Norwegian krone – are among the least efficient gold markets. Among these, also the Bitcoin currency lays at the very bottom of the ranking. On the other side of the ranking, the Top 5 is formed by the Seychellois rupee, the Liberian dollar, the Comorian franc, the Maldivian rufiyaa, and the Belize dollar. The differences between levels of EI are stunning as the most efficient markets share the index between 0.11 and 0.15 whereas the least efficient ones jump above 0.3. Such divergence is further accentuated by very low standard errors of the estimates usually around 0.01 (medians and standard errors are reported in Table 1 as well). These unexpected results, which are contrary to the quite expected ones found for the stock markets [13, 15] and other commodities [14], only highlight the specific connection between the gold market and the currency markets, and it calls for further treatment of the causes.

The analyzed period between years 2011 and 2014 covers very unorthodox times with regards to monetary policies of the developed world as reactions to the Global financial crisis, the Eurozone crisis, the Greek crisis and connected phenomena. Various waves of the quantitative easing (QE) in the USA and the UK, together with parallel actions of the European Central Bank eventually leading to the quantitative easing as well, have formed an enormous pressure on the relevant currencies and their depreciation. The first two waves of QE in the USA pushed the gold prices upwards as these rallied till the end of 2011. The last wave of the USA QE, which was much weaker than the previous two, had no significant effect on the USD gold prices. The connection between the currency depreciation and the consequent gold price (in the given currency) boosts together with a long-term effect of QE known in advance form a perfect environment for the gold market inefficiency. This is well in hand with most of the currencies the central banks of which participated in QE or other forms of practical money-printing being among the least efficient markets. It also further puts forward the gold's speculative asset status during the QE periods.

Such reasoning is further supported by gold being used as a hedge against inflation [17]. During the initial stages of QE, there was a serious concern about uncontrolled inflation as a reaction to the virtual money-printing. As investors were hedging against expected inflation by purchasing gold, its price was pushed further up. In time, the concern slowly vanished as there were no signs of dangerous inflation pressures. Nonetheless, the predictability and inefficiency of the gold markets under QE currencies come out as the final effect.

Following complementary explanation of the ranking structure can be quite counterintuitive in the efficient market logic. The fact that a central bank or a central authority of a country is transparent and holds up to its word, can actually lead to market inefficiency. Consider a central bank announcing a new wave of QE. If the central bank is trustworthy, the investors will start behaving accordingly and maximize their profit by acting upon it. However, the QE process is a gradual one and it thus does not affect the market instantly but in steps. Putting these factors together leads us to quite well predictable market behavior with relatively low risk assuming the authority holds up to its promises. From the other side, the authorities which are not too trustworthy are prone to change the announced policy repeatedly so that the shocks to the currency market are unpredictable. Such unpredictability leads to higher efficiency.

To summarize, the combination of gold prices and currencies thus forms a very interesting and unique

structure, the dynamics of which is much different from other assets such as stock or commodity markets. We have shown that the least efficient gold prices are mostly the ones quoted in major currencies such the US dollar, the Euro, and the British pound. On the other side of the spectrum, the most efficient gold prices are the ones quoted in smaller and less traded currencies. From the practitioners' perspective, we have two possibilities of utilizing the results. We can either speculate on gold prices in the major currencies, or we can hedge gold prices using the minor currencies to obtain stable and efficient market position.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. FP7-SSH-612955 (FinMaP) and the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics".

References

[1] Beran, J. (1994). *Statistics for Long-Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall.

[2] Cajueiro, D. and B. Tabak (2004). Ranking efficiency for emerging markets. *Chaos, Solitons and Fractals 22*, 349–352.

[3] Cajueiro, D. and B. Tabak (2005). Ranking efficiency for emerging equity markets II. *Chaos, Solitons and Fractals 23*, 671–675.

[4] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1(2)*, 223 – 236.

[5] Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative Finance 7(1)*, 21–36.

[6] Di Matteo, T., T. Aste, and M. Dacorogna (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance 29*, 827–851.

[7] Fama, E. (1965). The behavior of stock market prices. *Journal of Business 38*, 34–105.

[8] Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance 25*, 383–417.

[9] Fama, E. (1991). Efficient Capital Markets: II. *Journal of Finance 46(5)*, 1575–1617.

[10] Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4(4)*, 221–238.

[11] Gneiting, T. and M. Schlather (2004). Stochastic Models That Separate Fractal Dimension and the Hurst Effect. *SIAM Review 46(2)*, 269–282.

[12] Gneiting, T., H. Sevcikova, and D. Percival (2010). Estimators of fractal dimension: Assessing the roughness of time series and spatial data. Technical report, Department of Statistics, University of Washington.

[13] Kristoufek, L. and M. Vosvrda (2013). Measuring capital market efficiency: Global and local correlations structure. *Physica A 392*, 184–193.

[14] Kristoufek, L. and M. Vosvrda (2014a). Commodity futures and market efficiency. *Energy Economics 42*, 50–57.

[15] Kristoufek, L. and M. Vosvrda (2014b). Measuring capital market efficiency: long-term memory, fractal dimension and approximate entropy. *European Physical Journal B 87*, 162.

[16] Malkiel, B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives 17(1)*, 59?82.

[17] Narayan, P. K., S. Narayan, and X. Zheng (2010). Gold and oil futures: Are markets efficient? *Applied Energy 87*, 3299–3303.

[18] Ortiz-Cruz, A., E. Rodriguez, C. Ibarra-Valdez, and J. Alvarez-Ramirez (2012). Efficiency of crude oil markets: Evidences from informational entropy analysis. *Energy Policy 41*, 365–373.

[19] Phillips, P. C. B. and K. Shimotsu (2004). Local whittle estimation in nonstationary and unit root cases. *The Annals of Statistics 32(2)*, 659–692.

[20] Pincus, S. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America 88*, 2297–2301.

[21] Pincus, S. and R. Kalman (2004). Irregularity, volatility, risk, and financial market time series. *Proceedings of the National Academy of Sciences of the United States of America 101(38)*, 13709–13714.

[22] Robinson, P. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23(5)*, 1630–1661.

[23] Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review 6*, 41–49.

[24] Taqqu, M., W. Teverosky, and W. Willinger (1995). Estimators for long-range dependence: an empirical study. *Fractals 3*(4), 785–798.

[25] Taqqu, M. and V. Teverovsky (1996). On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. In *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*.

[26] Zunino, L., B. M. Tabak, F. Serinaldi, M. Zanin, D. G. Pérez, and O. A. Rosso (2011). Commodity predictability analysis with a permutation information theory approach. *Physica A 390*, 876–890.

[27] Zunino, L., M. Zanin, B. Tabak, D. Pérez, and O. Rosso (2010). Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A 389*, 1891–1901.

Table 1 Estimated Efficiency Index for gold prices in worldwide currencies

Rank	Country	EI	Rank	Country	EI	Rank	Country	EI
1	Seychelles	0.1120±0.0108	49	Paraguay	0.2140±0.0085	97	Eritrea	0.2571±0.0081
2	Liberia	0.1190±0.0111	50	Albania	0.2167±0.0096	98	Lithuania	0.2582±0.0079
3	Comoros	0.1195±0.0118	51	Pakistan	0.2198±0.0083	99	Armenia	0.2582±0.0121
4	Maldives	0.1380±0.0077	52	Phillipines	0.2207±0.0078	100	Croatia	0.2602±0.0085
5	Belize	0.1447±0.0093	53	Bangladesh	0.2225±0.0071	101	Argentina	0.2609±0.0135
6	Rwanda	0.1494±0.0154	54	Lebanon	0.2259±0.0118	102	India	0.2612±0.0074
7	Mauritius	0.1512±0.0088	55	Fiji	0.2268±0.0146	103	Turkey	0.2640±0.0055
8	Chile	0.1536±0.0136	56	Laos	0.2284±0.0101	104	Singapore	0.2644±0.0091
9	Somalia	0.1539±0.0109	57	Namibia	0.2285±0.0097	105	Norway	0.2645±0.0078
10	Domin. Rep.	0.1560±0.0149	58	Mexico	0.2300±0.0105	106	Morocco	0.2653±0.0089
11	Costa Rica	0.1569±0.0098	59	Swaziland	0.2327±0.0084	107	Dominica	0.2658±0.0096
12	Mauritania	0.1592±0.0097	60	Kenya	0.2335±0.0085	108	Qatar	0.2668±0.0090
13	Burundi	0.1621±0.0166	61	Aruba	0.2369±0.0097	109	Uganda	0.2680±0.0097
14	S. T. & Princ.	0.1630±0.0142	62	Yemen	0.2374±0.0081	110	Oman	0.2680±0.0103
15	French Polyn.	0.1663±0.0093	63	Tunisia	0.2379±0.0098	111	Saint Helena	0.2681±0.0167
16	Macedonia	0.1664±0.0106	64	Jamaica	0.2395±0.0143	112	Bulgaria	0.2684±0.0088
17	Djibouti	0.1666±0.0096	65	Russia	0.2398±0.0078	113	North Korean	0.2686±0.0095
18	Indonesia	0.1671±0.0095	66	Botswana	0.2407±0.0106	114	Denmark	0.2687±0.0082
19	Cape Verde	0.1779±0.0137	67	Lesotho	0.2422±0.0107	115	Poland	0.2689±0.0083
20	Iraq	0.1803±0.0127	68	Bahrain	0.2425±0.0102	116	USA	0.2695±0.0077
21	Sierra Leone	0.1806±0.0133	69	Saudi Arabia	0.2428±0.0100	117	Cuba	0.2704±0.0094
22	Iceland	0.1810±0.0080	70	Thailand	0.2429±0.0089	118	Czech Rep.	0.2715±0.0077
23	Uzbekistan	0.1812±0.0130	71	Malaysia	0.2430±0.0084	119	Canada	0.2719±0.0067
24	Cambodia	0.1840±0.0096	72	Tajikistan	0.2432±0.0091	120	Congo (DRC)	0.2720±0.0102
25	Brazil	0.1871±0.0103	73	Kuwait	0.2441±0.0104	121	Hungary	0.2720±0.0065
26	Israel	0.1887±0.0090	74	Peru	0.2450±0.0127	122	South Korea	0.2737±0.0122
27	Ghana	0.1887±0.0092	75	Solomon Isl.	0.2452±0.0150	123	Hong Kong	0.2773±0.0098
28	Madagascar	0.1902±0.0106	76	Sri Lanka	0.2458±0.0077	124	Guyana	0.2775±0.0129
29	Libya	0.1913±0.0129	77	Sweden	0.2460±0.0098	125	EU	0.2778±0.0071
30	Suriname	0.1914±0.0113	78	Guatemala	0.2461±0.0091	126	Switzerland	0.2838±0.0082
31	Ukraine	0.1922±0.0175	79	Trin. & Tob.	0.2472±0.0128	127	Serbia	0.2895±0.0064
32	Mozambique	0.1930±0.0138	80	Barbados	0.2476±0.0119	128	Papua N. G.	0.2907±0.0185
33	Colombia	0.1938±0.0135	81	Neth. Antilles	0.2478±0.0087	129	South Africa	0.2938±0.0112
34	Mongolia	0.1943±0.0132	82	Ethiopia	0.2480±0.0071	130	Japan	0.2948±0.0093
35	Tonga	0.1954±0.0186	83	Romania	0.2484±0.0066	131	Honduras	0.2955±0.0106
36	Nepal	0.1955±0.0084	84	Panama	0.2485±0.0099	132	Australia	0.2959±0.0118
37	Nicaragua	0.1968±0.0106	85	China	0.2491±0.0112	133	New Zealand	0.2976±0.0135
38	Cayman Isl.	0.1969±0.0087	86	Taiwan	0.2501±0.0112	134	Nigeria	0.2984±0.0133
39	Syria	0.1983±0.0079	87	Macau	0.2506±0.0095	135	Falkland Isl.	0.3016±0.0119
40	Bahamas	0.2012±0.0125	88	Kyrgyzstan	0.2529±0.0100	136	Vanuatu	0.3257±0.0105
41	Algeria	0.2016±0.0110	89	Azerbaijan	0.2531±0.0123	137	Belarus	0.3378±0.0084
42	Moldova	0.2045±0.0084	90	Georgia	0.2540±0.0078	138	Egypt	0.3442±0.0188
43	Afghanistan	0.2057±0.0069	91	UAE	0.2540±0.0088	139	UK	0.3480±0.0123
44	Samoa	0.2096±0.0117	92	Turkmenistan	0.2540±0.0096	140	Gambia	0.3515±0.0134
45	Kazakhstan	0.2107±0.0063	93	Angola	0.2561±0.0110	141	Bitcoin	0.3524±0.0078
46	Haiti	0.2119±0.0121	94	Vietnam	0.2561±0.0177	142	Gibraltar	0.3561±0.0140
47	Uruguay	0.2123±0.0054	95	Jordan	0.2566±0.0098			
48	Brunei	0.2123±0.0122	96	Tanzania	0.2568±0.0102			

On the simulation study of the properties of MSE estimators in small area statistics

Małgorzata K. Krzciuk[1]

Abstract. In the paper, the problem of estimation of mean squared errors in small area statistics is considered. The analysis is conducted based on the model proposed by Fay and Herriot [4], which belongs to the class of area models. Two estimators of the mean squared error are analyzed. The first (called naive) does not take into account the influence of estimating model parameters on the prediction accuracy while the second does. Their biases are of different orders which depend on the number of domains. We analyze properties of both MSE estimators for different number of domains using Restricted Maximum Likelihood (REML) to estimate superpopulation model parameters. Considerations are supported by the simulation analysis based on the real data from the Local Data Bank of Polish Central Statistical Office. The simulation study is prepared in R language (R development Core Team [11]).

Key words: estimators of MSE, Best Linear Unbiased Predictor, Empirical Best Linear Unbiased Predictor, Fay-Herriot model, simulation.

JEL Classification: C150

AMS Classification: 97K80

1. Introduction

In the paper we raise the issue of estimation of the mean square error in predicting the domain mean. The type A model (called Basic Area Level Model), proposed by Fay and Herriot [4], is considered. We present two estimators of mean square error – naive estimator and the classic estimator considered by e.g. Datta and Lahiri [2]. In last section we show results of simulation study comparing properties of both estimators and the influence of the increase of the number of domains.

2. Superpopulation model

The general linear mixed model, including its special case considered by Fay and Herriot [4], can be written as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Y}$ is the random vector of values of the dependent variable, $\mathbf{X}$ is the matrix of the auxiliary variables and $\boldsymbol{\beta}$ is the vector of unknown parameters. Additionally, $\mathbf{X}$ and $\mathbf{Z}$ are known. Furthermore, $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are the vectors of random effects and stochastic disturbances with variance-covariance matrices denoted by $\mathbf{G}$ and $\mathbf{R}$, respectively (Jiang [6], p.1-2).

Fay-Herriot model belongs to the class of models of type A which are useful when auxiliary information is available only on the area level. Models from this class are assumed for direct estimators in domains (Pfeffermann, Rao [9], p. 255). The model for means in domain have the following form (cf. Prasad, Rao [10], p.164; Datta, Lahiri [2] pp. 615-616; Lahiri, Rao [8], p. 759):

$$\hat{\mu}_d = \mu_d + e_d \tag{2}$$

where the mean in the d-th domain can by written by:

$$\mu_d = \mathbf{x}_d^T\boldsymbol{\beta} + v_d \tag{3}$$

$\hat{\mu}_d$ is the direct estimator of the mean in the d-th domain and $\mathbf{x}_d^T$ is the vector of p values of auxiliary variables in the d-th domain. Additionally, $\boldsymbol{\beta}$ is the vector of the p unknown parameters and e_d is the error

[1] University of Economics in Katowice, 50,1 Maja Street, Katowice, Poland, malgorzata.krzciuk@uekat.pl.

associated with the sampling design. Furthermore, e_d and v_d for $d=1,...,D$ are mutually independent, $e_d \overset{iid}{\sim} N(0,W_d)$ and $v_d \overset{iid}{\sim} N(0,A)$. It is assumed that W_d is known. Hence, for the model with assumptions (2) and (3): $\mathbf{R}=diag_{1\le d\le D}(W_d)$ and $\mathbf{G}=A\mathbf{I}_{D\times D}$, where $\mathbf{I}_{D\times D}$ is identity matrix of size $D\times D$.

This model is popular in small area estimation, because – by building the linking models for the direct estimators with the use of auxiliary data and through borrowing strength from other domains – it produces reliable small area statistics (Datta, Rao, Smith [3], p.184). The advantage of the model is its elasticity due to linking data from various sources (Rueda, Mendez Gomez [12], p.571).

3. Best Linear Unbiased Predictor and Empirical Best Linear Unbiased Predictor

In the Henderson's theorem [5] the problem of prediction of the linear combination of vectors $\mathbf{v}$ and $\boldsymbol{\beta}$ given by $\theta=\mathbf{l}^T\boldsymbol{\beta}+\mathbf{m}^T\mathbf{v}$ under the model (1) is considered. Additionally, variance and covariance matrices $\mathbf{G}=\mathbf{G}(\boldsymbol{\delta})$ and $\mathbf{R}=\mathbf{R}(\boldsymbol{\delta})$ are assumed to be known. This matrices are therefore the function of vector variance components, which is assumed to be known, too. For the model with assumptions (2) and (3) we have: $\boldsymbol{\delta}=A$. The Best Linear Unbiased Predictor, the predictor with the smallest variance of prediction errors in the class ξ-unbiased predictors of θ, is given by:

$$\hat{\theta}^{BLUP}=\mathbf{l}^T\hat{\boldsymbol{\beta}}+\mathbf{m}^T\hat{\mathbf{v}} \tag{4}$$

where

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \tag{5}$$

and variance-covariance matrix of the vector $\mathbf{Y}$ has form:

$$\mathbf{V}=\mathbf{R}+\mathbf{Z}\mathbf{G}\mathbf{Z}^T. \tag{6}$$

Additional $\hat{\mathbf{v}}$ is equal $\mathbf{G}\mathbf{Z}\mathbf{V}^{-1}\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)$. We should note that $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$ are function of $\boldsymbol{\delta}$. If we replace $\boldsymbol{\delta}$ by its estimator ($\hat{\boldsymbol{\delta}}$), we obtain two stage predictor – the Empirical Best Linear Unbiased Predictor $\hat{\theta}^{EBLUP}$. It should be noted that when the assumption (1) is met and the expected value of the EBLUP is finite, $\hat{\boldsymbol{\delta}}$ is any even and translation invariant estimator and distributions of e_d and v_d are symmetric about zero, $\hat{\theta}^{EBLUP}$ is ξ-unbiased (Kackar, Harville [7], pp. 1258-1259).

The special case of the Henderson's theorem (Henderson [5]) for the considered Fay-Herriot model gives the following formula of the BLUP (cf. Lahiri, Rao [8], p. 759; Prasad, Rao [10], p.164):

$$\hat{\mu}_d^{BLUP}=\hat{\mu}_d-B_d(A)\left(\hat{\mu}_d-\mathbf{x}_d^T\hat{\boldsymbol{\beta}}\right) \tag{7}$$

where:

$$B_d(A)=W_d\left(A+W_d\right)^{-1} \tag{8}$$

and

$$\hat{\boldsymbol{\beta}}=\left(\sum_{d=1}^{D}\frac{B_d(A)}{W_d}\mathbf{x}_d\mathbf{x}_d^T\right)^{-1}\left(\sum_{d=1}^{D}\frac{B_d(A)}{W_d}\mathbf{x}_d\hat{\mu}_d\right) \tag{9}$$

The Mean Squared Errors (*MSE*) of (7) is given by formula:

$$MSE_{\xi}\left(\hat{\mu}_d^{BLUP}\right)=g_{1d}(A)+g_{2d}(A) \tag{10}$$

where:

$$g_{1d}(A) = AW_d(A+W_d)^{-1} \tag{11}$$

and

$$g_{2d}(A) = W_d^2(A+W_d)^{-2}\mathbf{x}_d^T\left(\sum_{d=1}^{D}(A+W_d)^{-1}\mathbf{x}_u\mathbf{x}_u^T\right)^{-1}\mathbf{x}_d \tag{12}$$

The predictor given by (7) where the parameter *A* is replaced by its estimator is called the Empirical Best Linear Unbiased Estimator. For the Empirical Best Linear Unbiased Predictor under the Fay-Herriot model the Mean Squared Errors (*MSE*) can be written by (cf. Yoshimori, Lahiri [14], pp. 282-283):

$$MSE_{\xi}\left(\hat{\mu}_d^{EBLUP}\right) = g_{1d}(A) + g_{2d}(A) + g_{3d}(A) + o\left(D^{-1}\right) \tag{13}$$

where $g_{1d}(A)$ and $g_{2d}(A)$ are given the above formulas (11) and (12). The last element is given by:

$$g_{3d}(A) = 2W_d^2(A+W_d)^{-3}\left(\sum_{d=1}^{D}(A+W_d)^{-2}\right)^{-1} \tag{14}$$

if the model parameter A is estimated using Restricted (Residual) Maximum Likelihood method.

4. Estimators of Mean Squared Errors

We study two estimators of the Mean Squared Error (*MSE*) – the naive estimator and the classic estimator presented e.g. by Datta and Lahiri [2].

The naive estimator has the form of the mean square error of BLUP, where A is replaced by its estimators – $\hat{A}$ (Das K., Jiang J., Rao J.N.K [1] p. 820):

$$\hat{MSE}_N\left(\hat{\mu}_d^{EBLUP}\right) = g_{1d}\left(\hat{A}\right) + g_{2d}\left(\hat{A}\right) \tag{15}$$

It should be noted that this estimator of *MSE* does not take into account the influence of estimating model parameters on the prediction accuracy (and its bias is order $O(D^{-1})$).

The second considered estimator given by the formula (cf. Datta, Lahiri [2], p. 619; Prasad, Rao [10], p.167):

$$\hat{MSE}_{PR}\left(\hat{\mu}_d^{EBLUP}\right) = g_{1d}\left(\hat{A}\right) + g_{2d}\left(\hat{A}\right) + 2g_{3d}\left(\hat{A}\right), \tag{16}$$

is taking into account the decrease in prediction accuracy resulting from the estimation of model parameters and it has the bias of order $o(D^{-1})$. In this case $g_{1d}(\hat{A})$, $g_{2d}(\hat{A})$ and $g_{3d}(\hat{A})$ are calculated respectively from formulas (11), (12) and (14). The MSE estimator (16) is asymptotically unbiased in the following sense: $E_{\xi}\left(\hat{MSE}_{\xi}\left(\hat{\mu}_d^{EBLUP}\left(\hat{A}\right)\right)\right) - MSE_{\xi}\left(\hat{\mu}_d^{EBLUP}\left(\hat{A}\right)\right) = o\left(D^{-1}\right)$. In the simulation study presented in the section 5, we will analyze the properties of both estimators for increasing number of domains.

5. Simulation Study

The purpose of the simulation study is the Monte Carlo analysis of the pace of the decrease of biases of two MSE estimators (given by (15) and (16) respectively) with the increase of the number of domains (included the problem of model misspecification).In the simulation studies we use real data from the Local Data Bank (Polish Central Statistical Office). Population elements are Polish regions called poviats in year 2013. The division of the population of poviats (of size $N = 379$) into $D = 16$ subpopulations is made according to the larger regions – voivodships. In the analyzed model (2) in the simulation $\hat{\mu}_d$ is the average expenditure on health care in the domain calculated as the sample mean in the d-th domain. The auxiliary variable in the model is the average population in the domain in the previous year. The intercept is included in the model. Taking into account the assumption of independence of random components e_d we draw a sample from the population as the stratified sample without replacement, where domains are strata. We assume approximate proportional allocation (c.a. 15% elements from each strata).

In the simulation, values of $\hat{\mu}_d$ are generated according to (2) where $\boldsymbol{\beta}$ is calculated from the formula (9) for the whole population data set. Random effects (v_d) and stochastic disturbance (e_d) are generated using three distributions: normal, shifted exponential and shifted gamma distribution with expectation 0 and variances respectively A and W_d. In the case of shifted gamma distribution we assume skewness equal 5. We include different distributions to check the properties of MSE estimators in the case of this type of model misspecification. Other types of model misspecification such as correlation of random effects and/or random componentes can be included as well (cf. Żądło T. [16]). The values of W_d we calculate using the following formula (Żądło T. [15] p.110):

$$W_d = \frac{N_d - n_d}{N_d n_d} \frac{1}{N_d - 1} \sum_{i=1}^{N_d} \left(y_i - N_d^{-1} \sum_{i=1}^{N_d} y_i \right)^2 \cdot \tag{17}$$

In practice these values are often replaced by its estimator (c.f. Żądło T. [15] p.107) or by smoothed value of this estimators (e.g. Wolter [13]). The parameter A is calculated using Restricted (Residual) Maximum Likelihood method (*REML)* based on the real data. The number of iterations in each case is 20.000.

In order to investigate the influence of the number of domains on biases of MSE estimators we consider: three distributions, two estimators of *MSE* – naive (given by (15)) and the classic one (given by (16))and four cases when the number of domains is equal 16 (original data), 32 (the original data are enlarged two times), 64 (the original data are enlarged four times) and 128 (the original data are enlarged eight times).

The simulation study was prepared using R language (R Development Core Team (2015)).

Figures 1-3 present distributions of relative biases of both *MSE* estimators for all of the domains. They are computed in the simulation study for different distributions and number of domains. Boxplots for naive MSE estimators are white, for classic MSE estimators they are grey.

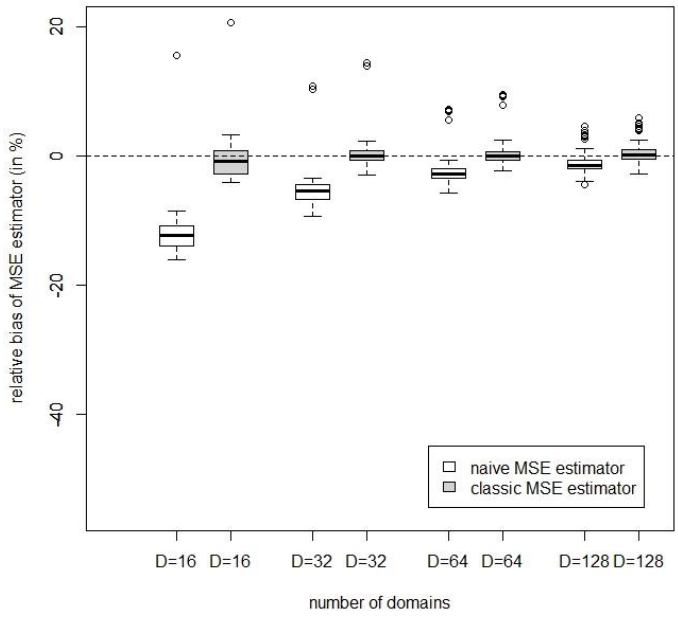


Figure 1 Values of relative biases of classic and naive MSE estimator for normal distribution

Figure 1 presents results for the first part of simulation, when random effects and stochastic disturbances are generated from normal distribution, $N(0, A)$ and $N(0, W_d)$ respectively. In all cases relative biases for classic *MSE* estimator are very close to 0. For $D = 16$ relative bias for naive estimator of *MSE* has median -12% but for $D = 128$ the value of the median bias over domains is smallest than 2%.

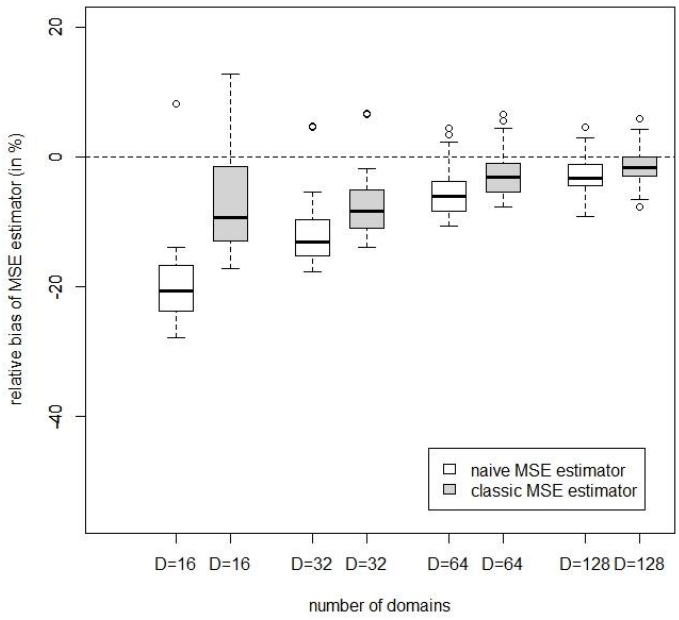


Figure 2 Values of relative biases of classic and naive *MSE* estimator for shifted exponential distribution

Second figure present results for the case when v_d and e_d are generated from shifted exponential distribution. For this part of the simulation interquartile ranges of relative biases in all of domains for both estimator have greater values than in the case of the normal distribution. For increasing number of domains we obtain results closer to zero and the difference between classical and naive estimator of *MSE* becomes smaller.

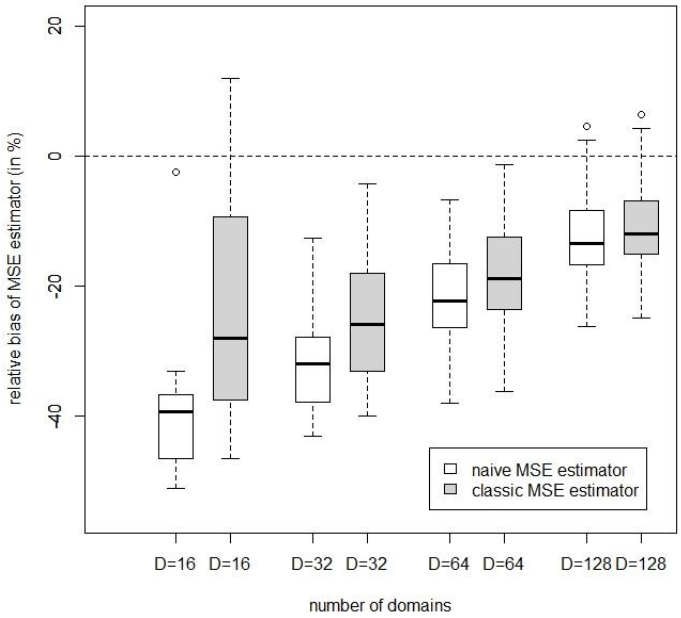


Figure 3 Values of relative biases of classic and naive *MSE* estimator for shifted gamma distribution

Similarly to the previous cases, for the shifted gamma distribution of both random effects and random components we obtain similar values of biases of both MSE estimators if the number of domains is sufficiently large. It is worth noting that median of relative biases for $D=16$ is close to -40% for naive MSE estimator and -30% for classic MSE estimator, and for $D=128$ both values of medians are close to -12%.

6. Summary

In the paper we consider the problem of properties of *MSE* estimators in small area estimation. We study, in the Monte Carlo simulation, the influence of the increase of the number of domains on properties of classic and naive *MSE* estimator for three distributions of random effects and stochastic disturbances: normal (what means that the model is specified correctly), shifted exponential and shifted gamma (what means that the model is misspecified). In the simulation study we show that (for the considered real data and Fay-Herriot model) biases of both MSE estimators are similar if the number of domains equals 64 even if the model is misspecified.

References

[1] Das K., Jiang J., Rao J.N.K.: Mean Squared Error of Empirical Predictor, *The Annals of Statistics*, **32**, 2,(2004), 818-840

[2] Datta G.S., Lahiri P.: A unified measures of uncertainty of estimated best linear unbiased predictors in small area estimation problems, *Statistica Sinica*, **10** (2000), 613-627.

[3] Datta G.S., Rao J.N.K., Smith D.D.: On measuring the variability of small area estimators under a basic area level model, *Biometrica*, **92** (2005), pp.183-196.

[4] Fay R.E., Herriot R.A.: Estimation of Incomes for Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74** (1979), 269-277.

[5] Henderson C.R.: Estimation of genetic parameters (Abstracts), Annals of Mathematical Statistics, **21** (1950), pp. 309-310.

[6] Jiang J.: *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Science+Business Media, New York, 2007.

[7] Kackar R.N., Harville D.A., Unbiasedness of two-stage estimation prediction procedures for mixed linear models, Communications in Statistics, Series A, **10** (1981),1249-1261.

[8] Lahiri P., Rao J.N.K.: Robust Estimation of Mean Squared Error of Small Area Estimators, *Journal of the American Statistical Association*, **90**,430 (1995), pp. 758-766.

[9] Pfeffermann D., Rao C.R. (2009), *Handbook of Statistics, Volume 29B Sample Surveys: Inference and Analysis*, Elsevier.

[10] Prasad N.G.N., Rao J.N.K.: The Estimation of the Mean Squared Error of Small-Area Estimators, *Journal of the American Statistical Association*, **85**, 409 (1990), 163-171.

[11] R Development Core Team: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2015.

[12] Rueda C., Mendez J.A., Gomez F.: Small area estimators on restricted mixed models, *Sociedad de Estadística e Investigación Operativa*, **16** (2010), 558-579.

[13] Wolter K.M., *Introduction to variance estimation.* Springer-Verlag, New York, 1985.

[14] Yoshimori M., Lahiri P.: A new adjusted maximum likelihood method for the Fay–Herriot small area model, *Journal of Multivariate Analysis*, **124** (2014), 281–294.

[15] Żądło T.: On prediction of the domain total under some special case of type A general Linear Mixed Models, *Folia Oeconomica*, **228** (2009), 105-112.

[16] Żądło T.: On prediction for correlated domains in longitudinal surveys, *Communications in Statistics – Theory and Methods*, **44** (4) (2015), 683-697.

An experimental comparison of Value at Risk estimates based on elliptical and hierarchical Archimedean copulas

Michael Kubát[1], Jan Górecki[2]

Abstract. In this paper, we estimate Value at Risk for a selected portfolio using elliptical and hierarchical Archimedean copulas, where the latter is based on a recent approach to estimation of hierarchical Archimedean copulas based on the Kendall correlation matrix. The estimates are compared using the Kupiec's test for three periods of time: a period of significant movements of index prices, a period of relative calm in the stock market and a long term period which includes the both situation in the stock market. Our experimental results show that the estimates based on elliptical copulas are more accurate in the periods of relative calm in the stock market, whereas the estimates based on hierarchical Archimedean copulas are more accurate in the period of significant movements in the stock market.

Keywords: copula, elliptical copulas, hierarchical Archimedean copulas, Value at Risk, returns modeling, backtesting

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Modeling portfolio returns and the subsequent estimation of the portfolio risk is one of the basic activities of financial institutions. Value at Risk (VaR) is frequently used to estimate the risk of future portfolio losses. Estimation of VaR is legally mandatory - banks are obliged by Basel II and insurance companies by Solvency II. The VaR is usually estimated based on historical data, and assuming constant development and constant characteristics of the portfolio or assets. Copulas, which appeared in connection with the Sklar's theorem [21], are often used for modeling the portfolio. The theorem states that given a multivariate probability distribution, it can be broken down into two components: i) its univariate marginal distributions and ii) the function describing the dependence among the variables, i.e., its copula.

For modeling i), there exist a number of options, e.g., the normal distribution, the skewed Student's distribution [10], a mixture of normal distributions and the Lévy processes [5]. As financial time series often do not meet the conditions of the normal and the Student's probability distributions of nonzero skewness (in case of the normal distribution also fixed kurtosis), the Lévy's models are frequently used for these purposes, see, e.g., [1]. Hence, in this article, we work with one of the Lévy's models - the normal inverse Gaussian (NIG) distribution - which is, as suggested in [4], appropriate for financial time series modeling.

For modeling ii), *elliptical copulas* (ECs) are the most popular choice, see, e.g., their applications to finance described in [20] or an application dealing with modeling and testing portfolio using ECs described in [14]. However, as ECs are radically symmetric, they are not suitable for specific applications, e.g., they can fail to adequately capture dependence between extreme events, see [2, 18] for financial examples. To overcome some of these restrictions, there appeared several multivariate alternatives, e.g., *Archimedean copulas* (ACs) or their asymmetric generalization, *hierarchical Archimedean copulas* (HACs). For example, a successful application of HACs to collateralized debt obligations is reported in [11], which shows their advantages to ECs.

Our research presented in this paper extends the research presented in [11] and aims to experimentally compare VaR estimates based on ECs and HACs. The VaR estimates are computed for three periods of time: a period of significant movements of index prices, a period of relative calm in the stock market and a long period which includes the both situation in the stock market. The results are compared using the widely known Kupiec's test suggested in [17]. Note that in the estimation processes that involve HACs, we use the approach to estimation of HACs based on the Kendall correlation matrix that was recently proposed in [7, 9], which we implemented in Matlab. We choose this HAC estimation approach as it has shown desirable properties when compared to other HAC estimation approaches, e.g., see the experimental comparisons reported in [6, 9].

[1] Department of Informatics, SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic, O130076@opf.slu.cz
[2] Department of Informatics, SBA in Karvina, Silesian University in Opava, Karvina, Czech Republic, gorecki@opf.slu.cz

The paper is structured as follows. Section 2 recalls the NIG distribution, ECs, HACs and backtesting, Section 3 presents the experimental comparison of VaR estimates based on ECs and HACs, and Section 4 concludes.

2 Methodology

2.1 Normal inverse Gaussian model

Normal inverse Gaussian (NIG) model is one of the Lévy processes with parameters $0 < |\beta| < \alpha$, $-\infty < \mu$ and $\delta > 0$, and its density is given by

$$f_{NIG}(x;\alpha,\beta,\mu,\delta) = \frac{\alpha \exp\big(\zeta + \beta(x-\mu)\big) K_1(\alpha\delta q\left(\frac{x-\mu}{\delta}\right))}{\pi q\left(\frac{x-\mu}{\delta}\right)}, x \in \mathbb{R},$$

where $\zeta = \delta\sqrt{\alpha^2-\beta^2}$, $q(y) = \sqrt{1+y^2}$ and K_1 is the modified Bessel function of the third order and index one. For more details on how to estimate its parameters, see, e.g., [1].

2.2 Copulas

Definition 1 [19] A *d*-dimensional copula is a *d*-dimensional multivariate distribution function with standard uniform univariate margins.

Copulas establish a connection between multivariate distribution functions and their univariate margins, which is well-known due to Sklar's Theorem.

Theorem 1 (Sklar's Theorem [21]) Let F be a d-dimensional multivariate distribution function with univariate margins $F_1, \ldots, F_d$. Then there exists a copula C: $[0,1]^d \rightarrow [0,1]$ such that

$$F(x_1,\ldots,x_d) = C\big(F_1(x_1),\ldots,F_d(x_d)\big), \tag{1}$$

holds for all $(x_1,\ldots,x_d) \in \overline{\mathbb{R}}^d$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty,+\infty\}$. Such a function C is uniquely determined, if $F_1,\ldots,F_d$ are all continuous. Conversely, if C is a copula and $F_1,\ldots, F_d$ are univariate distribution functions, then the function F given by (1) is a multivariate distribution function with margins $F_1,\ldots, F_d$.

2.3 Elliptical copulas

ECs are based on existing multivariate elliptical distributions. The Gaussian copula is based on the multivariate normal distribution and the Student *t*-copula is based on the multivariate Student *t*-distribution. Formally, a *Gaussian* copula is given by [19]

$$C_{\Sigma}^{Ga}(u_1,\ldots,u_d) = \phi_{\Sigma}\left(\phi^{-1}(u_1),\ldots,\phi^{-1}(u_d)\right),$$

where ϕ is the cumulative density function of the univariate normal distribution and ϕ_{Σ} is the cumulative density function of the multivariate normal distribution with a correlation matrix $\sum$. A *Student t*-copula is given by [19]

$$C_{v,\Sigma}^{t}\big(u_{1,\ldots,}u_d\big) = t_{v,\Sigma}\big(t_v^{-1}(u_1),\ldots,t_v^{-1}(u_d)\big),$$

where t_v is the cumulative density function of the univariate Student *t*-distribution with *v* degrees of freedom and $t_{v,\Sigma}$ is the cumulative density function of the multivariate Student *t*-distribution with a correlation matrix $\sum$ and *v* degrees of freedom.

2.4 Archimedean and hierarchical Archimedean copulas

Definition 2 [11] An *Archimedean generator* (simply, *generator*) is a continuous, non-increasing function $\psi : [0, \infty] \rightarrow [0,1]$ that satisfies $\psi(0) = 1, \psi(\infty) = \lim_{t\rightarrow\infty}\psi(t) = 0$ and that is strictly decreasing on $[0, \inf\{t \mid \psi(t) = 0\}]$.

Definition 3 [11] Any *d*-copula *C* is called *Archimedean copula* (AC), if it admits the form

$$C(\mathbf{u}) \coloneqq C\ (\mathbf{u};\psi) \coloneqq\ \psi\big(\psi^{-1}(u_1) + \cdots +\ \psi^{-1}(u_d)\big), \qquad \mathbf{u}\ \in\ [0,1]^d,$$

where ψ is a generator and ψ^{-1}: [0,1] $\rightarrow$ $[0,\infty]$ is defined by $\psi^{-1}(s) = \inf\{t \mid \psi(t) = s\}, s \in [0,1]$.

To derive an explicit form of an AC, we need explicit generators. In this work we use the three popular families of ACs presented in Table 1.

Table 1 The three considered one-parametric Archimedean copula families with the corresponding parameter ranges and forms [19]

Family	θ	$\psi\ (t)$
Clayton (C)	[-1, ∞) / {0}	$(1+t)^{-1/\theta}$
Frank (F)	(-∞, ∞) / {0}	$-\log\left(1-\left(1-\ \mathrm{e}^{-\theta}\right)\exp(-t)\right)/\theta$
Gumbel (G)	[1, ∞)	$\exp\ (-t^{1/\theta})$

As it follows from the construction of ACs that all multivariate margins of the same dimensions are equal, which is mostly considered too restrictive in high-dimensional applications, there appeared a generalization of ACs, hierarchical Archimedean copulas, which allow for partial asymmetry in multivariate margins.

Definition 4 [12] A *d*-dimensional copula *C* is called a *hierarchical Archimedean copula* (HAC) if it is either an Archimedean copula, or if it is obtained from an Archimedean copula through replacing some of its arguments with other hierarchical Archimedean copulas.

It follows from Definition 4 that every AC is a HAC (but not vice versa). For more details on HACs, see, e.g., [11].

2.5 Backtesting

Backtesting is a procedure in which the ability of a given model that estimates the future loss is evaluated. In the financial industry, the most commonly used model for estimation of the risk of future losses is Value at Risk (VaR). It expresses the maximum potential loss on the dependability of a given VaR confidence level. It is defined by

$$Pr\big(\Delta\Pi_{t+\Delta t} \le -VaR_{c,\Delta t}\big) = 1 - c,$$

where *Pr* is the probability measure, $\Delta\Pi_{t+\Delta t}$ expresses the change in the value of a time series Π_t at time *t* over a time period Δt and $VaR_{c,\Delta t}$ is the value of the maximum loss for the time period Δt at a given VaR confidence level $c \in [0,1]$. In this paper, Π_t is portfolio wealth at time *t*.

Having different $VaR_{c,\Delta t}$ estimators, their comparison is often based on so-called *01-sequence*, see, e.g., [15, 16], given for $t\ \in \{1, \ldots, n\}$ by

$$I_t = \begin{cases} 1, if\ \Delta\Pi_{t+\Delta t} \le -\overline{VaR}_{c,\Delta t} \\ 0, if\ \Delta\Pi_{t+\Delta t} > -\overline{VaR}_{c,\Delta t} \end{cases},$$

where $\overline{VaR}_{c,\Delta t}$ an is an estimated VaR value for time *t*. A widely known test based on this sequence, the Kupiec's test [17], is a two-sided test that tests the fit of the estimated VaR model with respect to the underestimation and overestimation of the risk. Given a 01-sequence (I_1, …, I_n), the test uses the *Kupiec's likelihood ratio statistic* (LR) given by

$$LR = -2\log\left[\frac{\pi_{ex}^{n_1}(1-\pi_{ex})^{n_0}}{\pi_{obs}^{n_1}(1-\pi_{obs})^{n_0}}\right],$$

where $\pi_{ex} = 1 - c$ is the expected probability of exception occurring, $\pi_{obs} = \frac{n_1}{n_0+n_1}$ is the proportion of exceptions, n_0 is the number of non-exception days (the number of "zeros" in the 01-sequence), n_1 is the number of exception days (the number of "ones" in the 01-sequence) and $n_0 + n_1 = n$. Under the null hypothesis that $\pi_{ex} = \pi_{obs}$, the statistic *LR* is asymptotically χ^2 (chi-squared) distributed with one degree of freedom. Using this fact and given *n*, the non-rejection interval for a number of exceptions n_1 can be obtained for a given significance level α. Throughout this paper, we use α = 5%.

3 Experiments

Firstly, we present the data that we analyze in our study. It is a part of the Dow Jones Industrial Average (DJIA) dataset. We arbitrarily selected 6 well-known US companies of the dataset, namely, MSFT (Microsoft Corporation
CSCO (Cisco Systems), NKE (Nike, Inc.), MCD (McDonald's), CAT (Caterpillar Inc.) and IBM (International Business Machines Corporation). The comparison is performed for the following three periods:

- Period A - 1998/29/11-2010/31/12 – the long term period (n = 3042 days)
- Period B - 2000/21/11-2004/16/11 – the period of significant movements (n = 1000 days)
- Period C - 1994/24/07-1998/07/07 – the period of relative calm (n = 1000 days)

For each of these periods, given a VaR confidence level c and a family of copulas, we used the following procedure based on the Monte Carlo method for $t = 501, \ldots, n+500$ to estimate corresponding VaR values:

1. For time t, choose the daily returns $\Delta\Pi_{t-500}, \ldots, \Delta\Pi_{t-1}$;
2. Estimate the parameters of the six NIG distributions using the method of moments, see [1];
3. Assuming an elliptical copula, estimate its parameters via the Matlab's Statistical toolbox function *copulafit*. Assuming a HAC, estimate its parameters via our implementation of Algorithm 3 proposed in [9];
4. Sample 10 000 observations distributed according to the multivariate distribution with the estimated copula and the estimated univariate marginal distributions using the approach proposed in [11];
5. For each of the observations sampled in the previous step, calculate the portfolio wealth $\Pi_t^i, i = 1, \ldots, 10000$ with uneven weights (= 1/6) of the assets;
6. Order $\Pi_t^i, i = 1, \ldots, 10000$;
7. Choose Π_t^j as the VaR estimated for t such that $j = [10\,000(1-c)]$ ($[y]$ returns the integer part of $y \in \mathbb{R}$);
8. Continue to $t+1$.

After this process, we computed three 01-sequences corresponding to each of the considered time periods, in the way described in Section 2.5.

3.1 Results

All in all, we computed 45 different 01-sequences (5 copulas families (Gaussian, Student, HAC Clayton, HAC rank and HAC Gumbel) * 3 VaR confidence levels c = 85%, 95%, 99% * 3 time periods). Their characteristics corresponding to the Kupiec's test are shown in Figures 1, 2 and 3. The figures show the differences between the expected ($[n\,(1-c)]$) and the estimated (n_1) number of exceptions, and the bounds of the corresponding non-rejection intervals (dashed lines) for the null hypothesis in the Kupiec's test corresponding to the significance level α = 5%, see Section 2.5.

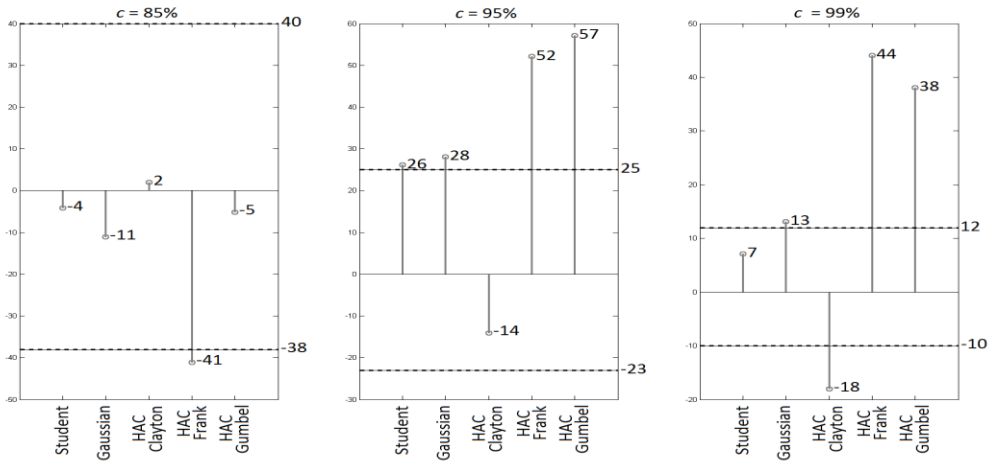


Figure 1 The differences between the expected and the estimated number of exceptions for Period A

In Figure 1, we observe that only the HAC Frank based VaR model is rejected for c = 85%, i.e., the corresponding n_1 - $[n\,(1-c)]$ is outside of the interval (-38, 40). For the remaining VaR models, the differences n_1 - $[n\,(1-c)]$ are close to 0. The HAC Clayton based VaR model, which is the closest to zero for c = 85%, is the only VaR model that is not rejected for c = 95%. The remaining VaR models overestimates the risk for this VaR confidence level. For c = 99%, only the Student based VaR model is not rejected. Also observe, that the HAC Clayton based VaR model is the only model that underestimated the risk for c = 95% and c = 99%.

Generally, for this period, the EC based VaR models performed better than HAC based VaR models in the sense that the differences n_1 - $[n\,(1-c)]$ are closer to zero for the EC based VaR models than for the HAC based VaR models.

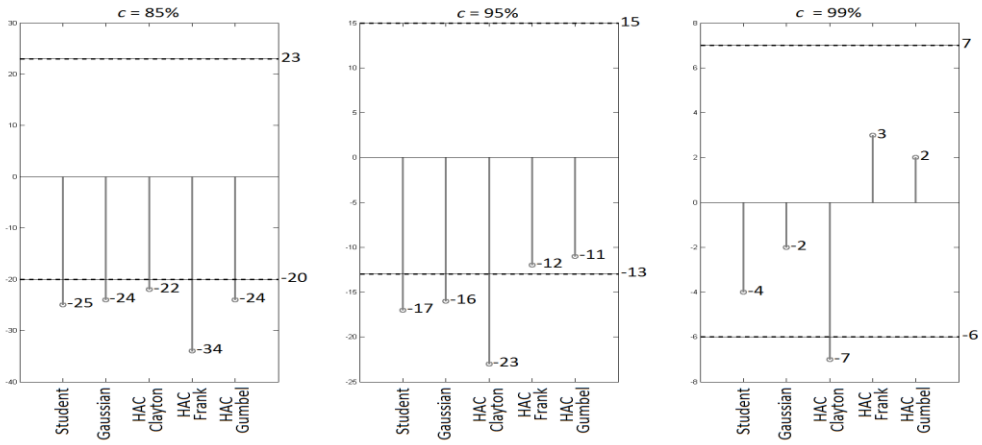


Figure 2 The differences between the expected and the estimated number of exceptions for Period B

In Figure 2, for c = 85%, we see that all the VaR models underestimate the risk and rejected. For c = 95%, only the HAC Frank and the HAC Gumbel based VaR models are not rejected, but again, all the considered VaR models underestimate the risk. For c = 99%, only the HAC Clayton based VaR model is rejected. We also observe that the HAC Clayton based VaR model underestimates the risk for c = 95% and c = 99%, similarly to Period A. Generally, on the contrary to Period A, the HAC based VaR models performed better than elliptical copula based VaR models, more precisely, the best (the most close to zero) or the second best (for c = 99%) VaR model is always a HAC based VaR model.

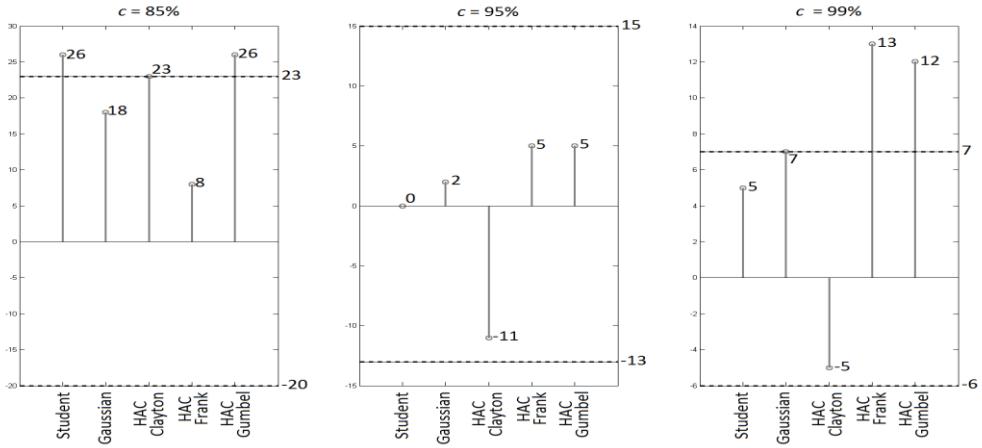


Figure 3 The differences between the expected and the estimated number of exceptions for Period C

In Figure 3, for c = 85%, we see that almost all VaR models overestimate the risk. We also observe that only the Gaussian and HAC Frank based VaR models are not rejected. For c = 95%, all the models are not rejected and the EC based VaR models score better that the HAC based VaR models. For c = 99%, the Gaussian, HAC Frank and HAC Gumbel based VaR models are rejected. The HAC Clayton based VaR model again underestimates the risk for c = 95% and c = 99%. Generally, for this period, the EC based VaR models performed better than the HAC based VaR models, similarly to Period A.

4 Conclusion

This paper presented an experimental comparison of VaR models based on elliptical and hierarchical Archimedean copulas. The main finding, which confirm the findings reported in [11], is that that the HAC VaR models have been performing better for the period of significant movements (Period B). For example, the HAC Gumbel based VaR model is in average the best performing for this period. This finding, on the one hand, suggests that hierarchical Archimedean copulas are more convenient for modeling VaR in times of significant movements in the stock market. On the other hand, EC based VaR models were more accurate for Period A and

Period C, i.e., for the periods in which the days of relative calm dominates over the days of significant moves in the stock market. Another observation is that the HAC Clayton based VaR model more or less underestimates the risk for the two higher VaR confidence levels for all three periods. The question closely related to these conclusion, i.e., how to decide when to use elliptical based and when to use HAC based VaR models, however, remains open, and we will consider it in further research. Also, we want to consider in further research other copula estimation approaches, e.g., the consistent HAC estimator suggested in [8].

Acknowledgement

The research of this work was funded by project SGS/21/2014 – Advanced methods for knowledge discovery from data and their application in expert systems, Czech Republic.

References

[1] Barndorf-Nielsen, O. E.: *Normal inverse Gaussian distributions and the modeling of stock returns*. Aarhus, Dissertation. Aarhus University, 1995.
[2] Bae, K.-H., Karolyi, G. A., and Stulz, R. M.: A new approach to measuring financial contagion. *Review of Financial studies* **16**(3):717–763, 2003.
[3] Cherubini, G., Luciano, E. and Vecchiato, W.: *Copula Methods in Finance*. Chichester : Wiley, 2004.
[4] Cont, R.: Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **1**(2), 223-236, 2001.
[5] Cont, R. and Takov, P.: *Financial modeling with Jump Processes*. Chapman&Hall, London, 2003.
[6] Górecki, J., Holeňa, M.: Structure determination and estimation of hierarchical Archimedean copulas based on Kendall correlation matrix. In: *A. Appice, M. Ceci, C. Loglisci, G. Manco, E. Masciari, Z.W. Ras (eds.) New Frontiers in Mining Complex Patterns*, Lecture Notes in Computer Science, 2014, 132–147. Springer International Publishing.
[7] Górecki, J., Holeňa. M.: An alternative approach to the structure determination of hierarchical Archimedean copulas. In: *Proceedings of the 31st International Conference on Mathematical Methods in Economics* (MME 2013), 2013, 201 – 206, Jihlava, Czech Republic
[8] Górecki, J., Hofert, M., Holeňa, M.: On the consistency of an estimator for hierarchical Archimedean copulas. In: *Proceedings of the 32nd International Conference on Mathematical Methods in Economics*, pp. 239 – 234, Olomouc, Czech Republic, 2014.
[9] Górecki, J., Hofert, M. and Holeňa, M..: *An Approach to Structure Determination and Estimation of Hierarchical Archimedean Copulas and Its Application in Bayesian Classification*. Journal of Intelligent Information Systems, pp. 1 - 39, Springer Science+Business Media New York, 2015.
[10] Hansen, B.E.: *Autoregressive conditional density estimation*. International Economic Review, 1994.
[11] Hofert, M.: *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. 1102 Suedwestdeutscher Verlag fuer Hochschulschriften, 2010.
[12] Hofert, M.: A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation* **82**(9), 1239–1255, 2012. DOI http://dx.doi.org/10.1080/00949655.2011.574632
[13] Jarque, C. M. and Bera, A. K. 1987: *A test for normality of observations and regression residuals*. International Statistical Review, 1987, **55**(2), 163-172.
[14] Kresta, A.: Backtesting of market risk estimation assuming various copula functions. In: *Proceedings of 30th International Conference Mathematical Methods in Economics*. Opava: Slezská univerzita v Opavě, 2012, 484-489. ISBN 978-80-7248-779-0.
[15] Kresta, A. and Tichý, T.: Backtesting of portfolio risk in term sof ordinary Lévy copula model.In: *Finance and the performance of firms in science,education and practice.* E. Jiříčková et al (Eds.). Zlín: Univerzita Tomáše Bati ve Zlíně, 2011.
[16] Kresta, A. and Tichý, T.: International Equity Portfolio Risk Modeling: The case of NIG model and ordinary copula functions. Finance a úvěr – *Czech Journal of Economics and Finance* **62**(2), 141–161, 2012.
[17] Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivative* **3**(2), 73-84, 1995.
[18] Longin, F. and Solnik, B.: Extreme correlation of international equity markets. *The Journal of Finance* **56**(2), 649–676, 2001.
[19] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. New York: Springer, 2006.
[20] Rank, J.: Copulas: *From Theory to Application in Finance*. London: Risk Books, 2006.
[21] Sklar, A: *Fonctions de répartition à n dimensions et leurs marges*. Publications de l'Institute de statistique de l'Université Paris, 229-231, 1959.

The Activity of the Banking and Shadow Banking Sector vs. Lending Standards in Selected Euro Area Countries: A Wavelet Approach

Zuzana Kučerová[1], Jitka Poměnková[2]

Abstract. Shadow banking activities can be associated with the potential financial instability because these activities are realised outside the regulated banking sector. The aim of the paper is to assess the relationship between the banking and shadow banking activities and lending standards imposed by banking institutions in nine Euro Area countries using a method of regression and correlation analysis and a wavelet approach. We use quarterly data over the period 2003-2013. When measuring the shadow banking activity at national level, the shadow banking activities are most extensive in Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. We conclude that banking institutions originated fewer loans in reaction to tighter lending standards (imposed by these banking institutions on their loans). However, the expected positive impact of tighter lending standards on loans provided by shadow banking institutions is not confirmed. We can also admit flexible reactions of OFIs to a current economic situation during the analysed period while the behaviour of MFIs followed a long-term strategy, i.e. banks did not react to short-term objectives as frequently as non-banks. In this context, we can conclude that banks met their stabilising role in the economy while non-banks tried to fill up gaps in the financial markets.

Keywords: wavelet co-spectrum, comovement, shadow banking sector, banking sector, market-based system, bank-based system.

JEL Classification: E42, E44, E51
AMS Classification: 62P20, 62M10, 91B84

1 Introduction

Generally, the financial sector comprises many types of institutions, among others shadow banking institutions. A topic of the shadow banking sector was discussed especially after the financial crisis. It is connected with the market-based financial system where firms prefer financing through financial markets than via the banks. Shadow banking institutions take part in the securitisation process and use sophisticated securitisation techniques in order to intermediate credit and produce structured financial products. Shadow banks could be regarded as an important source of potential financial instability because these activities are realised outside the regulated banking sector and as such they are an important source of potential financial instability. Last but not least, globally integrated financial markets increase the probability of a severe financial contagion. Therefore, it is of great importance to monitor these activities and measure the interconnectedness between banking (regulated) and shadow banking (unregulated) sector. The Financial Stability Board (FSB) recommends national authorities to enhance their monitoring framework in order to uncover potential risks hidden in the shadow banking sector by means of the application of a stylised monitoring process (see FSB [10]).

According to Bakk-Simon et al. [7], shadow banking *"…refers to activities related to credit intermediation, liquidity and maturity transformation that take place outside the regulated banking system*". Claessens *et al*. [4] restrict the definition of shadow banking to "*…those activities that are economically most bank-like: involve risk transformation and a focus on reducing counterparty risks*". Commonly used definition of the shadow banking sector is that of the FSB [10] which describes the shadow banking system"*…as the system of credit intermediation that involves entities and activities fully or partially outside the regular banking system, or non-bank credit intermediation in short*". In our paper, we use the definition of FSB [10], i.e. we focus on such activities realised fully or partially outside the traditional or regulated banking system.

[1] VSB-Technical University of Ostrava, Faculty of Economics, Dept. of Economic Policy, Sokolská tř. 33, Ostrava, Czech Republic, e-mail: zuzana.kucerova@vsb.cz.
[2] Brno University of Technology, Faculty of Electrical Engineering and Communication, Dept.of Radio Electronics, Technická 12, 616 00 Brno, pomenkaj@feec.vutbr.cz.

The aim of the paper is to measure the relationship between shadow banking sector activities and credit lending standards imposed by banking institutions in nine Euro Area countries in period 2003-2013. The data concerning the banking and shadow banking activities are drawn from the European Central Bank (ECB) online database and from the ECB Bank Lending Survey. The analysis is conducted only for 9 Euro Area countries: Austria, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Spain, and Portugal. We use the method of regression, correlation and wavelet analysis. The structure of the paper is as follows. The second section contains a literature review. Section three describes data, indicators and section four methods used in our paper. In the fifth section, we present the results of our analysis. Section six brings conclusion.

2 Literature Review

The co-movement analysis was traditionally performed in the time domain based on the correlation and similar methods. The most recent empirical analysis was applied using the stochastic approach with frequency uncertainty. This approach was extended by frequency domain which provided an additional view of the problem of time series character as well as co-movement. In the last decade, the research focused on the time-frequency domain perspective because the combination of time and frequency domains provides a more efficient means of statistical analysis. The time-frequency representation of time series can be estimated via several methods. The main approaches are wavelet analysis, the multiple window method using Slepian sequences, time-frequency varying autoregressive process spectrum estimation and time-frequency Fourier transform estimation. We can find the comparison of these main methods in the study of Blumenstein et al. [3] or Klejmová [13].

In our paper, we use a wavelet analysis. As Jiang and Mahadevan [12] denote, the wavelet transform can be applied to non-stationary time series due to the simultaneous time-frequency decomposition of inputs; it provides more natural results because it uses shorter windows for higher frequencies compared to the Fourier transform. Deutsche Bundesbank [6] uses the wavelet analysis to assess the relationship between deposits by non-bank financial intermediaries contained in M2 and (one by one) the gross domestic product, the harmonised index of consumer prices and a share price index in the Euro Area countries. There is a significant relationship between the annual growth rate of M2 (a deposit variable) and the share price index (a financial market variable) for fluctuations with periods between six and ten years. Analysis of economic indicators comovement via wavelets is proposed by Berdiev and Chang [2] who find that the strength of business cycle synchronization fluctuates across frequencies and over time. The study of Ftiti et al. [11] puts in evidence the existence of both long run and short-run comovements between OECD countries with the US and Europe. Fidrmuc et al. [9] study globalization and business cycles in China and G7 countries and show the existence of significant relationship between the time-varying wavelet measure of synchronization and trade only for business-cycle frequencies. In-deep insight into the structure of time series comovement is proposed by Maršálek et al. [16]. Through the continuous wavelet transform, its inverse and the comovement measurement in the time-frequency domain, it enables to uncover a detailed development of the business cycle synchronization in time.

3 Data and Indicators

3.1 Data

The European flow-of-funds data do not provide sufficient data to analyse the shadow banking in the Euro Area. However, there is a possibility to combine data from two ECB online databases – monetary statistics and euro area accounts (EEA) – and construct a proxy which enables us to roughly measure shadow banking activities (Bakk-Simon et al. [1]). The EEA data are divided into the following sectors (see ECB [7]): financial corporations, non-financial corporations, general government, and households. Financial corporations comprise monetary financial institutions (MFIs), other financial intermediaries (OFIs) and insurance corporations and pension funds (ICPFs). MFIs cover the regulated banking sector and include central banks, credit institutions and money markets funds. The OFIs are regarded as a part of the sector of the shadow banking activities and we focus on these institutions as representatives of the shadow banking sector. We also use data from the ECB Bank Lending Survey (BLS); these data are available since 2003:Q1 (see ECB [8]). The dataset of the survey works with the answers for selected Euro Area countries; banks in these countries provided quarterly information on credit standards for approving loans and credit terms and conditions (i.e. information from the supply side).

3.2 Indicators

Several categories of financial indicators can be used as a proxy of the size of the shadow banking sector.[3] In our paper, we measure the activities of the banking and shadow banking sector using *the difference of the cumulative*

[3] See e.g. Bakk-Simon *et al*. [1] or Kučerová [15].

level of long-term loans provided by the relevant sector. Differences are used in order to focus on the dynamics (i.e. net loans provided in the period) and not on the overall level of the outstanding amount of loans. As far as MFIs are concerned, the highest share of intermediation (measured by the volume of long-term loans) through the MFIs in 2013 was measured in Germany (28%), France (21%) but also in Spain (14%), Italy (13%) and in the Netherlands (9%). The highest share of non-bank intermediation (i.e. through OFIs) was in Luxembourg (35%), in the Netherlands (21%) and in Ireland (14%). In Italy and France, OFIs also had a relatively noticeable position (7%, respectively 6%). OFIs in the rest of the Euro Area produced some 5% of total long-term loans.[4] So, the highest activity of both MFIs (banking institutions) and OFIs (shadow banking institutions) is apparent in these nine Euro Area countries: Austria, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. Therefore, we focus only on these countries (except for Belgium, see below).

As a measure of lending standards, we use a measure defined by ECB [8] defined as the diffusion index; it is based on the responses to questions concerning lending standards and is measured as the weighted difference between the share of banks reporting that lending standards have been tightened and the share of banks reporting that they have been eased (banks who have answered "considerably" are given score 1 and banks that answered "somewhat" are given score 0.5). Positive values of the diffusion index indicate that a larger proportion of banks have tightened lending standards and *vice versa*. However, the data about lending standards were not available in case of Belgium. Therefore, this country was dropped out of our data sample.

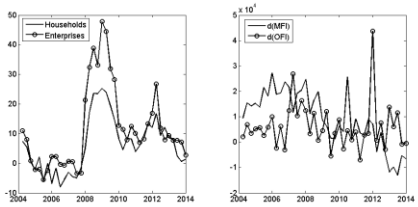


Figure 1 Time series representation, Source: ECB [7], ECB [8]

4 Methods

The time series *s(t)* can be transformed from time to time-frequency domain via continuous wavelet transform (CWT) which can be defined as

$$S_{CWT}(a,b) = \int_{-\infty}^{\infty} s(t)\frac{1}{\sqrt{b}}\psi\left(\frac{t-a}{b}\right)dt, \quad b>0, a \in R, \tag{1}$$

where *a* is the time position (time shift), *b* is the parameter of dilatation (scale) of the mother wavelet $\psi(\cdot)$. The dilatation is related to the Fourier frequency and numerator of the fraction $\sqrt{b}$ ensures the conservation of energy. The CWT transforms input time series from the time representation to the time-scale domain.

The analysis of the relation between two time series can be measured via several quantities. In the frequency domain we can use dynamic correlation, the coherence or cohesion (Jiang and Mahadevan [12]). Similar analysis could also be done in the time-scale domain. In our approach, we use the wavelet cross spectrum between two inputs $s_1(t)$ and $s_2(t)$ for their time-scale representation $S_{CWT\text{-}1}(a,b)$ and $S_{CWT\text{-}2}(a,b)$ which can be defined as

$$S_{12} = SO\left(S_{CWT-1}(a,b)S_{CWT-2}(a,b)\right), \tag{2}$$

where *SO* is the smoothing operator (Jiang and Mahadevan [12]).

5 Results

The analysis of the relationship between the credit lending standards imposed by MFIs on long-term loans provided to households (denoted as Households) and to enterprises (denoted as Enterprises) and the volume of new long-term loans (expressed using the first difference) provided by both MFIs and OFIs follows three basic steps. The results are presented predominantly in figures organised in columns in correspondence to the methods used in our paper. All figures keep the following denotation: Panel A - dependence between Households and the first difference of MFI, Panel B - dependence between Enterprises and the first difference of MFI, Panel C - for dependence between Households and the first difference of OFIs and Panel D - dependence between Enterprises and the first difference of OFIs.

[4] See also Kučerová [14].

In the first step, we perform a regression analysis to identify the above mentioned direction of dependency. In case of loans originated by MFIs, a negative relationship between lending standards and new long-term loans is expected, i.e. higher lending standards imposed by MFIs on their loans would lead to a decrease in the level of new loans originated by these institutions in the following period and *vice versa*. However, OFIs can be regarded as substitutes for MFIs, because they also offer some forms of financing as an alternative to bank loans. Therefore, it is also expected that OFIs would increase the supply of new loans in reaction to tighter lending standards of MFIs. This effect would emerge as a positive relationship between lending standards (imposed by MFIs on their loans) and new long-term loans (originated by OFIs). Results in table 1 show a significant negative relationship between MFIs and Households (correlation coefficient -0.33, 5% level significance). The regression between MFIs and Enterprises suggests negative relationship (i.e. the expected negative relationship is confirmed) but the slope of regression line is insignificant. Other regression analyses reveal rather stable slightly negative relationship between OFIs and Households/Enterprises (regression model) and insignificant correlation, i.e. the expected positive relationship is not confirmed. Visualisation of regression provides figure 2.

	Coefficient	**P-value**	**DW**	**F**	**P-value**	**R^2**
Dependent: d(MFI)						
Const.	11960,172	0,0000	0,5615	4,6643	0,0372	0,1093
Households	-392,1462	0,0372				
Dependent: d(OFI)						
Const.	6490,5206	0,0007	2,0765	0,1312	0,7191	0,0034
Households	-57,1625	0,7191				
Dependent: d(MFI)						
Const.	11262,2623	0,0000	0,5199	1,5560	0,2199	0,0393
Enterprises	-156,1441	0,2199				
Dependent: d(OFI)						
Const.	6255,7171	0,0020	2,0857	0,0117	0,9145	0,0003
Enterprises	-11,3414	0,9145				

Table 1 Linear regression models

In the second step, we proceed with a moving correlation to have a detailed view of the evolution of dependency measured by the classical correlation. The moving correlation consists of the calculation of static correlation on the moving time window which we moved per one observation till the last observation in the data sample; the length of correlated subsamples is 16 observations. The results provide a quick view of the evolution of correlation with respect to time and can reveal a structural break which can occur. Results (fig. 3) show the existence of a period with a significant correlation for both Households and MFIs in 2004q1-2008q2 and in 2004q4-2010q2. In case of Households and OFIs, there is only one significant period, i.e. 2006q4-2010q4. Moving correlation between Enterprises and MFIs is significant in the period 2004q4-2010q4 which confirms expectation from the regression analysis. Any significant relationship between Enterprises and OFIs is not proved. In all cases, the significant correlation is negative. Except for the relationship between OFIs and Households, the moving correlation has an increasing tendency, i.e. higher lending standards led to the drop of loans provided by MFIs (in case of both Households and Enterprises) and by OFIs in case long-term loans to Enterprises.

The last step focuses on wavelet co-spectral analysis combining both time and frequency domain of dependence. From the family of mother wavelet functions, we use the Morlet wavelet function in our paper. Results are depicted in figure 4. The cospectrum is denoted on the *z*-axis for specific periods (*x*-axis) and periodicities (*y*-axis). The figure shows a two-dimensional projection of three-dimensional charts. The intensity of different contours denotes the relative importance of different periodicities and time. In all panels (fig. 4, A-D), the wavelet co-spectrum shows several areas of comovement bounded by the time and frequency. The lower the frequency of inputs is (i.e. the higher and closer to the value of 36 the figure on the *y*-axis), the longer the cyclical component. For example, business cycles are defined between 6 quarters (rapid moving periodic component) and 8 years (slow moving periodic component).

Comparing Panel A and Panel B (fig. 4), we can see similar results in the identification of significant comovement differing primarily in the level of cospectra. Let us say that more significant results are achieved for Panel A. Here, for frequencies denoting long cycles (in this case 32-36 quarters) we can find significant area in periods 2005-2007 and 2010-2012. In addition, for Panel B, we can find also comovement in long cycles in 2008-2009, i.e. significant comovements between Enterprises (i.e. lending standards for long-term loans to enterprises) and long-term loans provided by MFIs (i.e. banks). Generally, we can admit the existence of permanent component represented by long cycles likely caused by financial crisis (before and after 2008). In both

panels (but especially in case of Enterprises) we can also find a comovement in middle cycles (16-22 quarters) in 2008-2010, in 2011 and in 2013. In low frequencies (below 12 quarters), we do not identify any significant comovement.

In case of Panel C and Panel D (fig. 4), we can see a significant comovement for long cycles similarly to the Panel A and B, but the time period is a little bit different; the comovements is identified later, i.e. in 2005-2007 and in 2009-10. Contrary to the Panel A and B, the comovement identified after 2010 is not such significant in the long cycles; it moved to lower frequencies (around 26 quarters). Additionally, we can identify a comovement in middle cycles (in this case particularly in the period of 14-18 quarters) after 2008, 2010 and 2012 in both Panel C and Panel D. Compared to panel A and B, this comovement moves in time (it is observed approx. one year earlier); the same holds for a frequency (shorter cycles). In both cases (Panel C and D) we find important comovements in period 2007-2009 in lower frequencies (around 8 quarters) which represent very rapid changes. This is not observed in Panel A and B.

To sum up, we can confirm the existence of a permanent component in all panels, but in the case of Panel C and Panel D, we can find the additional comovement representing a reaction to short term events in economy. Therefore, we can admit flexible reactions of OFIs to current economic conditions compared to the behaviour of MFIs which can be considered following rather a long-term strategy than meeting short-term objectives.

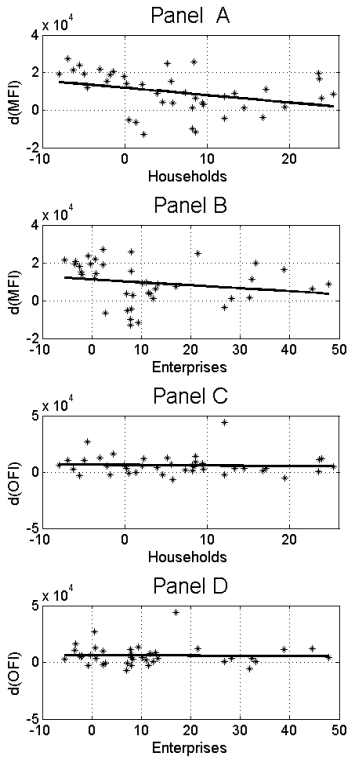


Figure 2 Linear regression

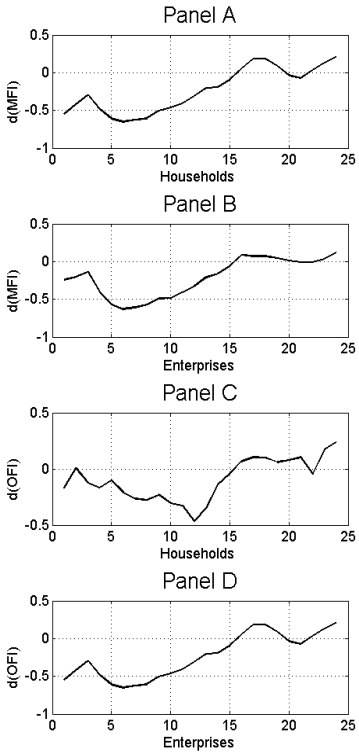


Figure 3 Moving correlation

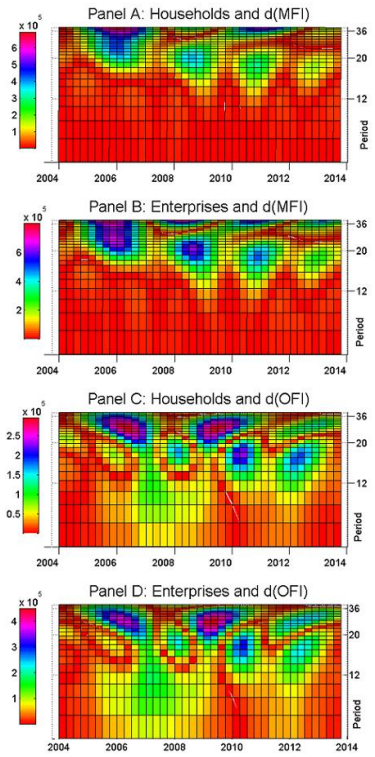


Figure 4 Wavelet co-spectrum

6 Conclusion

The paper was focused on the problem of interconnectedness of the banking sector and the shadow banking sector with a view to credit standards and the supply of loans. The aim was to measure the relationship between shadow banking sector activities and credit lending standards imposed by banking institutions on their loans.

First, we measured the shadow banking activity using the indicator of long-term loans in the Euro Area; the shadow banking activities are the most extensive in Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. However, data concerning credit lending standards were not available in the case of Belgium. Therefore, we dropped this country out of our sample.

We conclude that banking institutions (i.e. MFIs) in the analysed countries originated less loans in reaction to tighter lending standards imposed by these banking institutions on their loans which is consistent with expected results; this finding is also consistent with findings of Demiroglu *et al*. [5]. However, the expected positive impact of tighter lending standards (concerning bank loans) on loans provided by shadow banking institutions (i.e.

OFIs) has not been confirmed. In other words, shadow banking institutions (OFIs) did not react to the drop of long-term loans provided by MFIs by the increased supply of loans in the analysed period.

We also confirmed the existence of a permanent component in all analysed relationships (between MFIs and OFIs on the one hand and lending standards for long-term loans to households and enterprises on the other hand), but in the case of loans provided by OFIs (i.e. non-banks), we found the additional comovement representing a reaction to short term events in economy. Therefore, we can admit flexible reactions of OFIs to a current economic situation. Contrary to this finding, the behaviour of MFIs followed a long-term strategy, i.e. banks did not react to short-term objectives as frequently as non-banks. In this context, we can conclude that banks met their stabilising role in the economy while non-banks tried to fill up gaps in the financial markets.

Acknowledgements

This paper was supported by the VSB-Technical University of Ostrava, Faculty of Economics under SGS Grant [nr. SP2015/115]. Research described in this paper was financed by Czech Ministry of Education in frame of National Sustainability Program under grant LO1401. For research, infrastructure of the SIX Center was used.

References

[1] Bakk-Simon, K., Borgioli, S., Giron, C., Hempell, H., Maddaloni, A., Recine, F., and Rosati, S.: Shadow banking in the Euro Area: An Overview [online]. *European Central Bank Occasional Paper* **133** (2012), [cit. 2014-07-12]. Available at: http://www.ecb.europa.eu/pub/pdf/scpops/ecbocp133.pdf.

[2] Berdiev, A. N., and Chang, Ch.-P.: Business cycle synchronisation in Asia-Pacific: New evidence from wavelet analysis. *Journal of Asia Economics* **37** (2015), 20-33.

[3] Blumenstein, J., Poměnková, J., and Maršálek, R.: Comparative Study Of Time- Frequency Analysis Approaches With Application To Economic Indicators. In: *Proceedings of the 26th European Conference on Modelling and Simulation ECMS 2012* (Troitzsch, K., Möhring, M., and Lotzmann, U., eds.). University Koblenz Landau, Koblenz, 2012, 291-297.

[4] Claessens, Z., Pozsar, Z., Ratnovski, L., and Singh, M.: 2012. Shadow Banking: Economics and Policy [online]. *International Monetary Fund Discussion Note* **SDN/12/12** (2012), [cit. 2014-07-12]. Available at: https://www.imf.org/external/pubs/ft/sdn/2012/sdn1212.pdf.

[5] Demiroglu, C., James, Ch., and Kizilaslan, A.: Bank lending standards and access to lines of credit. *Journal of Money, Credit and Banking* **44** (2012), 1063-1089.

[6] Deutsche Bundesbank: The Shadow Banking System in the Euro Area: Overview and Monetary Policy Implications. *Deutsche Bundesbank Monthly Report* **March** (2014), 15-34.

[7] European Central Bank: *Statistical data warehouse: Euro Area accounts* [online database]. (2015a), [cit. 2015-02-12]. Available at: http://sdw.ecb.europa.eu/reports.do?node=1000002340.

[8] European Central Bank: *Statistical data warehouse: Bank lending survey* [online database]. (2015b), [cit. 2015-02-10]. Available at: https://sdw.ecb.europa.eu/browse.do?node=9484572.

[9] Fidrmuc, J., Korhonen, I., and Poměnková, J.: Wavelet spectrum analysis of business cycles of China and G7 countries. *Applied Economic Letters* **21** (2014), 1309-1313.

[10] Financial Stability Board: *Global Shadow Banking Monitoring Report 2013* [online]. Financial Stability Board, Basel, 2013, [cit. 2014-07-12]. Available at: http://www.financialstabilityboard.org/publications/r_131114.htm.

[11] Ftiti, Z., Tiwari, A., and Belanés, A.: Tests of Financial Market Contagion: Evolutionary Cospectral Analysis V. S. Wavelet Analysis. *Computational Economics* **1** (2014), DOI: 10.1007/s10614-014-9461-8.

[12] Jiang,W., and Mahadevan, S.: Wavelet spectrum analysis approach to model validation of dynamic systems. *Mechanical Systems and Signal Processing* **25** (2011), 575–590.

[13] Klejmová, E.: Comparison of parametric and non parametric methods for signal time-frequency description. In: *Proceedings of the 25th International Conference Radioelektronika 2015* (Pidanic, J., ed.). Univerzita Pardubice, 2015, 272-275.

[14] Kučerová, Z.: Banking and Shadow Banking Sector and Credit Lending Standards in the Euro Area. Paper presented in the 18th International Scientific Conference "*Enterprise and the Competitive Environment*": 5.-6. 3. 2015. Mendel University in Brno, Brno, 2015.

[15] Kučerová, Z.: Monitoring the Shadow Banking Sector in the Euro Area. In: *Proceedings of the 12th International Scientific Conference Economic Policy in the European Union Member Countries: 16.-18. 9. 2014* (Tvrdoň, M., and Majerová, I., eds.). Silesian University in Opava, Opava, 2014, 462-472.

[16] Maršálek, R., Poměnková, J., and Kapounek, S.: A Wavelet-Based Approach to Filter Out Symmetric Macroeconomic Shocks. *Computational Economics* **44** (2014), 477-488.

Empirical Analysis of the Dependence of the Foreign Direct Investments on the chosen Macroeconomic Factors

Kuchina Elena[1]

Abstract. The current paper focuses on the analysis of the impact of some vital macroeconomic indicators on the FDI. The present study investigates such macroeconomic factors as: GDP, Inflation, General Government Gross Debt, Strength of Legal Rights Index and Total Tax Rate. The current paper examines the following main research questions: What is the link between the foreign direct investments and the mentioned macroeconomic indicators? What is the character and intensity of this relationship? The econometric analysis of these research questions is done with the help of the panel data tools. The analysis was applied to different income categories; only for GDP the null hypothesis of its statistical insignificance can be rejected for each income group on the 5% (for the group 1 on the 10%) level of significance. Analyzing the dependence of FDI on GDP the time effect was examined, it was shown that only for years 2006-2008 the null hypothesis of the statistical insignificance can be rejected on the 5% level of significance. The empirical part was ended by the cluster analysis dividing the countries into groups based on the scenario of the FDI' development during the analyzed period.

Keywords: foreign investments, macroeconomic indicators, panel data, cluster analysis

JEL Classification: C51
AMS Classification: 91B84

1 Introduction

Foreign direct investments play a crucial role in the global business. For a host country it can provide the impulse for the growth by attracting skills, financing; by simplifying the access to new technologies. It can help host companies to be more involved in the international business. Due to such a significant impact of the foreign direct investment, it is important to examine the relationship between it and other vital macroeconomic factors, which can determine the macroeconomic politics as a whole.

The research question of this paper is to analyze the impact of such macroeconomic variables as: GDP, General Government Gross Debt, Inflation, Total Tax Rate, Strength of Legal Right Index on the Foreign Direct Investment based on the data spanning 123 countries, which are divided into five categories: high Income, OECD; high Income, non OECD; low income; lower middle income; upper middle income. It is fair to assume that for the different groups the character of the relationship may differ, which indicates the necessity to examine these relationships between the vital macroeconomic factors and the foreign direct investments also for each group separately. The data analyzed in this paper spanning the period from 2005 to 2013. Due to the fact that for the different countries the scenario of the development of the Foreign Direct Investments during the examined period (2005-2013) may be different, the cluster analysis is applied in the second part of the research, which aim is to divide countries into groups based on the similar scenario of the development of the Foreign Direct Investments.

The article is structured into three parts, which are data set and the methodology, determining the continuity of the steps, then the empirical modeling and main findings will be presented.

[1] University of Economics, Prague, Econometrics dep., W Churchill Sq. 4, Czech Republic, e-mail: xkuce10@vse.cz

2 Data Set

This section discusses the input data used in the analysis. The final panel data set consists of 123 countries for each of one there are annual data spanning the period from 2005 to 2013. For each country there are 9 observations over the analized period of the the following macroeconomic variables:

- Foreign Direct Investment (FDI) measured in current US$ [6].
- Inflation (INF) measured by the consumer price index (annual %) [6].
- Gross Domestic Product (GDP) measured at purchaser's prices in current US$ [6].
- Strength of Legal Rights Index (SLRI) measures the degree at which the collateral and bankruptcy laws protect the rights of borrowers and lenders, ranges from 0 to 12 (0 = weak to 12 = strong) [6].
- Total Tax Rate (TTR) measures the amount of taxes and mandatory contributions payable by business after accounting for allowable deductions and exemptions, measured as % of the commercial profit [6].
- General Government Gross Debt (GGD) measured as % of GDP [6].

The data about FDI, Inflation Rate, GDP, Strength of Legal Right Index, Total Tax Rate and group income are retrieved from the database of the World Bank [6] and the source of the General Government Debt was the database of the International Monetary fund [3].

For the purpose of this analysis the data about 123 countries were used. These countries can be spitted into five categories based on the income group:

- High income, OECD countries - 30 countries
- High income, non OECD countries - 14 countries
- Low income - 19 countries
- Lower middle income - 27 countries
- Upper middle income - 33 countries

The categorical variable with five categories was created, where the value 1 indicates the OECD countries with high income, 2 - non OECD countries with high income, 3 - countries with low income, 4 - countries with lower middle income and 5 - countries with upper middle income.

3 Methodology

In this section the procedure applied in the empirical part will be defined. The whole statistical and econometrical analysis was hold with the help of the software Stata 11 and SPSS 16.

Panel data set can be short or long. A short panel has many entities and few time periods, whereas a long panel has many time periods and few entities [1]. In the current analysis the panel data set, consisting of 123 countries (n) and 9 time periods (T), can be considered as a short balanced (all countries have the measurements in all time periods) and fixed (the same countries are observed for each period) panel.

Panel data analysis is a tool which enables to examine the group effects, time effects or both, these effects can be fixed or random. In case of the fixed effect model the intercept's varying across group or time periods is examined, whereas the random effect model examines the differences in error variance component across individual or time period [4].

A panel data regression can be written as [2]

$$y_{it} = \mathbf{x}'_{it}\,\boldsymbol{\beta} + \mathbf{z}'_{i}\,\boldsymbol{\alpha} + \varepsilon_{it},$$

where i = 1, 2, ...n; t = 1, 2, ..., T. K is the number of explanatory variables. The individual effect is $\mathbf{z}'_{i}\,\boldsymbol{\alpha}$ where $\mathbf{z}_{ic}$ contains a constant term and a set of individual specific variables.

If $\mathbf{z}_i$ contains only a constant term, then the pooled OLS can be applied. If the individual effect is not zero then the panel data models can provide the efficient tools to deal with this heterogeneity. A fixed effect model can be applied if the heterogeneity can be captured with the individual specific intercepts and the individual effect may possibly be correlated with any regressors, while a random model should be tried when the individual heterogeneity is captured in the disturbance term and the individual effect is not correlated with any regressors [4].

For the verification of the appropriate model to be used, firstly Breusch and Pagan LM and F-test should be conducted, where the rejection of the null hypothesis of LM-test indicates that the random effect model is more appropriate than the pooled OLS; if the null hypothesis of F-test is rejected, then the fixed effect model is favored over pooled OLS [2].

If the null hypothesis of *LM* and *F*-test is rejected, then conducting Haussmann test enables to decide if fixed or random effect model is more appropriate. In case of the rejection of the null hypothesis the fixed effect model is favored over the random effect model [2].

The countries arrangement between the different clusters is done with the help of the cluster analysis, where the *hierarchical cluster analysis* was applied, *Ward's* method with *Squared Euclidean distance* was used. More details can be found in [5].

4 Empirical Modeling

Following the methodology chapter the first step is to decide which model is more preferable, which means that for the relationship and each macroeconomic variable *F* and *LM*-tests were performed. Based on the results in the table 1, the null hypothesis can be rejected for both *F* and *LM*-tests on the 1% level of significance, which indicates that fixed effect model is favored over pooled OLS and random effect model is more preferable than pooled OLS. The next step is to apply the most appropriate model for each case , which can be done based on Haussmann test (table 1). Following the values of Haussmann test and the related *p*-values, it can be concluded that for the modeling the relationship between FDI and GDP, General Government Gross Debt and Inflation the null hypothesis can be rejected on the 1% level of significance, which indicates that the fixed effect model is more appropriate. For the modeling the relationship between FDI and the remained two macro economical variables (Strength of Legal Rights Index and Total Tax Rate) the null hypothesis cannot be rejected on the 1% level of significance and the random effect model is more preferable.

Variable	***F*-test**	***LM* - test**	**Haussmann test**
FDI → GDP	10.49 (0.00)	1103.25 (0.00)	20.03 (0.00)
FDI → GGD	29.46 (0.00)	2493.45 (0.00)	7.29 (0.01)
FDI → INF	28.75 (0.00)	2419.80 (0.00)	9.69 (0.00)
FDI → SLRI	28.41 (0.00)	2488.62 (0.00)	1.68 (0.20)
FDI → TTR	29.25 (0.00)	2524.36 (0.00)	1.43 (0.23)

Table 1 F-test, LM-test, Haussmann test

The table 2 provides the results of estimation β coefficient of the fixed and random effect model correspondingly (based on the results in the table 1). To control for heteroskedasticity and serial correlation the cluster robust standard errors were calculated.

Variable	β	*Robust Std. Err*	*p-value*	*rho*
FDI → GDP	0.029	0.004	0.00	0.67
FDI → GGD	-1.37e+08	8.64e+07	0.12	0.78
FDI → INF	1.49e+08	6.18e+07	0.02	0.77
FDI → SLRI	1.47e+09	6.16e+08	0.02	0.75
FDI → TTR	-8842458	2.31e+07	0.70	0.76

Table 2 Fixed effect model "within", random effect, robust, cluster (id)

Based on *p*-value in the table 2 for the General Government Gross Debt and Total Tax Rate the null hypothesis of the statistical insignificance cannot be rejected on the 5% level of significant. In case of the remained variables the null hypothesis can be rejected on the 5% level of significance, and based on the sign of the estimated parameters in all case the dependence is positive. The same analysis should be conducted for each income group separately in order to see if the character and the intensity of the relationship between the dependent and independent variables differ for the different income categories. The results of the parameters' estimation of the dependence of the foreign direct investments on GDP for different income groups can be found in the table 3.

FDI →GDP	*β*	*Robust Std. Err*	*p-value*	*rho*
Group 1	0.015	0.008	0.066	0.375
Group 2	0.029	0.007	0.001	0.821
Group3	0.014	0.004	0.005	0.719
Group 4	0.014	0.004	0.002	0.445
Group 5	0.035	0.000	0.000	0.515

Table 3 Fixed effect model "within", robust, cluster (id), sort by the income group, *FDI → GDP*

Based on the results in the table 3 the null hypothesis of the statistical insignificance of GDP can be rejected for group 1 on the 10% level of significance, while for the remained income groups on the 5% level of significance. Based on the estimation of *β*, the GDP has the greatest impact for the upper-middle income group.

Analyzing the relationship between FDI and Inflation for each income group, it can be seen that based on *p*-values, for any income group the null hypothesis of the statistical insignificance of *INF* cannot be rejected on the 5% level of significance.

FDI → INF	*β*	*Robust Std. Err*	*p-value*	*rho*
Group 1	8.89e+08	8.83e+08	0.32	0.69
Group 2	3.44e+08	3.05e+08	0.28	0.85
Group3	-3378434	2930131	0.26	0.70
Group 4	6.45e+07	5.66e+07	0.27	0.82
Group 5	1.75e+08	1.12e+08	0.13	0.84

Table 4 Fixed effect model "within", robust, cluster(id), sort by the income group, *FDI → INF*

Analogously table 5 contains the results for the examining the relationship between the FDI and Strength of Legal Rights Index where it can be observed that the null hypothesis of the statistical insignificance can be rejected on the 5% level of significance only for high income OECD and for low income.

FDI → SLRI	*β*	*Robust Std. Err*	*p-value*	*rho*
Group 1	6.46e+09	3.10e+09	0.04	0.67
Group 2	-4578718	2.91e+08	0.99	0.82
Group3	3.34e+07	1.23e+07	0.01	0.70
Group 4	3.92e+08	2.82e+08	0.17	0.82
Group 5	2.44e+09	2.38e+09	0.31	0.85

Table 5 Random effect model, robust, cluster(id), sort by the income group, *FDI → SRLI*

Following the results stating above, the relationship between FDI and GDP was shown for all income groups in contrast with other variables. The question of time effect can be examined with the two-way fixed effect with dummy variables relating to each time period where the basement category, with which all others are compared, was chosen 2013.

FDI → GDP	*β (p-value)*		*β (p-value)*
D1	4.03e+09 (0.13)	***D5***	2.83e+08 (0.83)
D2	6.25e+09 (0.01)	***D6***	2.83e+09 (0.17)
D3	9.85e+09 (0.00)	***D7***	1.93e+09 (0.15)
D4	6.60e+09 (0.02)	***D8***	-1.05e+09 (0.26)
GDP	0.03 (0.00)		

Table 6 Two-way fixed ("within" estimation), robust, cluster(id), *FDI → GDP*

Based on the results in the table 6, only for time-dummy variables relating to the beginning of the examined period (2006-2008) the null hypothesis of the statistical insignificance can be rejected on the 5% level of significance.

Cluster analysis

In all previous steps all countries were examined as a whole or split into five categories based on the income group. The purpose of this part is to divide countries into different categories based on the scenario of the development of FDI during 2005-2013. *Hierarchical cluster analysis* was applied, *Ward's* method with *Squared Euclidean distance* was used. All countries were divided into different clusters based on the values of FDI in each year. The following tables (table 7) shows the numbers of countries in each cluster for the cases: 2-5 clusters.

	C5	**C4**	**C3**	**C2**
1	101	101	101	121
2	19	19	20	2
3	1	2	2	0
4	1	1	0	0
5	1	0	0	0

Table 7 Cluster analysis (countries)

From the table 7 it can be seen that first cluster in all cases covers the majority of countries, which does not provide the countries' classification for appropriate analysis of each cluster separately. Due to this fact it can be concluded that the course of the development of FDI during the examined period was not shown as sufficient to cluster the countries, the qualitative factors as well as other macroeconomic factors should be taken into account.

5 Conclusion

The current paper has constituted the econometric analysis of the dependence of the foreign direct investments on some vital macroeconomic factors as: GDP, General Government Gross Debt, Inflation, Strength of Legal Right Index and Total Tax Rate. The panel data approach was used to estimate the relationship between FDI and the mentioned variables. For the GDP, General Government Gross Debt and Inflation the preferable model is fixed effect model, while for Strength of Legal Right Index and Total Tax Rate the random effect model. The GDP, Strength of Legal Right Index and Inflation were shown as statistically significant on the 5% level of significant. Then the same analysis was applied to different income categories; only for GDP variable the null hypothesis of its statistical insignificance can be rejected for each group on the 5% level of significance (for the group 1 on the 10% level of significance). Analyzing the relationship between FDI and GDP was extended by examining the time effect by including time dummy variables. It was shown that only for time-dummy variables relating to the beginning of the examined period (2006-2008) the null hypothesis of the statistical insignificance can be rejected on the 5% level of significance. The empirical part was ended by the cluster analysis, which purpose was to divide the countries into groups based on the scenario of the FDI' development during the analyzed period. As a consequence, it was shown that countries' clustering should be performed based on other macroeconomic and categorical characteristics.

Finally, the current analysis should be considered as a preliminary analysis of the analyzing and modeling of FDI, as it enables to examine the character and intensity of the relationship between the foreign direct investments and other vital macroeconomic variables separately. Due to it the main limitation of this analysis is the fact that only insulated effect is examined and the possible mutual interactions between the variables are not taken into account.

Acknowledgements

I gratefully acknowledge the support of this project by the research grant No. VŠE IGA F4/24/2014, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Cameron, C. A. and Trivedi P. K.: *Microeconometrics. Methods and applications.* Cambridge University Press, New York, 2005.
[2] Green, W.H.: *Econometric analysis*. Prentice Hall, Upper Saddle River, New Jersey, 2002.
[3] International Monetary Fund. *World Economic Outlook database*. [Online]. [cit. 2015-04-15]. Retrieved from www.imf.org.
[4] Park, Hun Myoung. 2011. Practical Guides To Panel Data Modeling: A Step-by-step Analysis Using Stata. *Tutorial Working Paper*. Graduate School of International Relations, International University of Japan, 2011. [Online]. [cit. 2015-04-03]. Retrieved from http://www.researchgate.net/publictopics.PublicPostFileLoader.html?id=55004b32d685cc141c8b46a3&key=69e88cf0-f685-4259-89c9-ab86d1b2a99f.
[5] Řezanková, H., Húsek, D., Snášel, V.: *Shluková analýza dat*. 2. vydání. Professional Publishing, Praha, 2009.
[6] World Bank. *Indicators.* [Online]. [cit. 2015-03-30]. Retrieved from http://data.worldbank.org/indicator.

Stock Market Common Movement Investigation via Cross Spectral Analysis

Jaromír Kukal[1], Quang Van Tran[1]

Abstract. Cross spectrum analysis is a frequency based technique which is capable of measuring the strength of the relationship and the lead lag characteristics among stock markets while they incorporate the price related information. In this research experiment, we try to use this technique to study the co-movement of the Czech stock market with international stock markets and examine the link between them at a quiet time as well as at a turbulent period. For this purpose, besides the Czech stock market index we choose six stock market indices series of the most important stock markets in the world for a period from 1.2005 to 12.2009. The results we obtain show that there is a strong interconnection among markets and a possible contagion propagation is considerable.

Keywords: cross spectrum, coherence, stock market indices, co-movement, contagion

JEL classification: G01 G15
AMS classification: 90A09 91G70

1 Introduction

Behavior of stock markets has always been a focal interest of many practitioners as well as researchers. In a globalized world, markets are interlinked with each other for many reasons and a price shock at international markets may have an effect on stock price in a local market and vice versa. So far while the traditional methods applicable in the time domain are widely used for this purpose, the use of frequency based methods is still seldom seen. The first application of a spectral method was dated back at the early sixties by Granger [4] as well as the more powerful methods based on cross spectrum introduced some time later [3]. Several studies have been conducted to investigate co-movement among countries within a continent or across continents. Others have explored a possible lead lag relationship among closely linked economies. However, to our knowledge, no work has dealt with a possible connection of the Czech stock market with the rest of the world.

The Prague Stock Exchange is a relatively small local stock market. Therefore, it is expected to co-move with key stock markets around the world. Also it seems to be affected by big players and probably it cannot have any substantial influence on them. Thus the co-movement is one-sidedly incited. This qualitatively known causality effect needs to be quantified as the extent of external effect on price movements at Prague stock market can provide its participants valuable information. In our opinion, cross spectrum analysis can be a most suitable tool to solve this problem and have not been used yet for this purpose. To fill this gap in the literature, we use data on Czech stock market index and data on other six main stock markets in the world from 1.2005 to 12.2009. In order to measure the impact of price movement of main markets on the ones of the Czech market at different times with distinguished nature, we divide dataset from the whole period into two subsets for two subperiods. The first half of this period is considered to be the non-turbulent time while the second half which overlaps with the financial crisis in 2008 represents the time of turbulence and we evaluate the comovements of Prague stock market with world markets in these two different periods.

[1]Czech Technical University in Prague, Department of Software Engineering, Trojanova 13, Praha 2, Czech Republic, {jaromir.kukal,tranvqua}@fjfi.cvut.cz

2 The cross spectrum analysis

Like in spectral analysis, cross spectrum analysis[1] is an alternative way to characterize time series. In both cases a series is considered to be a mixture of a great number of sine and cosine wave(let)s at different frequencies with random amplitudes. According to the spectral representation theorem (Hamilton, [6]), the cross spectrum function of two covariance-stationary processes with zero mean x_t and y_t can be defined as the finite Fourier transform of cross covariance function as follows:

$$f_{xy}(\omega) = \sum_{s=-\infty}^{\infty} \gamma_{xy}(s) \mathrm{e}^{-2\pi \mathrm{i}\omega s}, \tag{1}$$

where cross covariance function γ_{xy} is defined as:

$$\gamma_{xy} = \mathrm{E}(x_t y_{t+s}).$$

As cross covariance function is not symmetric, the cross spectrum is a complex value function, therefore it cannot be graphically represented directly. Instead, let define the so called complex coherency of the stationary series x_t and y_t:

$$P_{xy}(\omega) = \frac{f_{xy}(\omega)}{[f_{xx}(\omega) f_{yy}(\omega)]^{\frac{1}{2}}},$$

where $f_{xy}(\omega)$ is the cross spectrum of the two series, and $f_{xx}(\omega)$ and $f_{yy}(\omega)$ are their respective spectrum. Then the squared coherence spectrum[2] is defined as

$$C_{xy}(\omega) = \frac{\mid f_{xy}(\omega) \mid^2}{f_{xx}(\omega) f_{yy}(\omega)}. \tag{2}$$

It is apparent that coherence is similar to R squared coefficient and takes the value from interval [0,1]. It measures the strength of relationships between corresponding frequency components in the two series x_t and y_t. Since two series are jointly stationary, components with different frequencies are uncorrelated, thus the coherence captures the full strength of the relationships between the two series.

The gain G_{xy} indicates how much the spectrum of series x_t has been amplified to approximate that component of series y_t.

$$G_{xy}(\omega) = \frac{\mid f_{xy}(\omega) \mid}{f_{xx}(\omega)}. \tag{3}$$

The phase F_{xy} is a measure of the timing between the two series. It is measured in the fraction of a cycle that series y_t leads series x_t as follows:

$$F_{xy}(\omega) = \frac{1}{2\pi} \arg(f_{xy}(\omega)), \tag{4}$$

where $\arg(f_{xy}(\omega))$ is the argument of complex value from $(-\pi, \pi]$. The frequencies will be defined in Nyquist form as

$$\omega^* = \tfrac{\omega}{2\pi}.$$

3 Data and results

To evaluate the impact of the movement of prices of world markets on the prices at the Prague Stock Exchange we choose the series of Prague stock market index and six main stock market indices from 1.2005 to 12.2009. The main indices are: CAC40 (Paris, France), DAX (Franfurt, Germany), FTSE (London, Great Britain), HSI (Hong Kong) and S&P 500 and NASDAQ (USA). These series are divided into two parts: the first one is from 1.2005 to 6.2007 and this part is considered to be the part of non-turbulent time. The second part is from 7.2007 to 12.2009 and it is assumed to be the turbulent part of stock market movement. The descriptive statistics of original series used for our investigation are shown in Table 1. After that these series are transformed into logarithmic returns series and their corresponding

Series	CAC	DAX	FTSE	HSI	PX	SP500	NASD
Mean	4554.92	5841.69	5479.29	18644.57	1365.40	1231.82	2187.89
Median	4628.94	5735.75	5546.75	17914.30	1430.85	1263.45	2212.87
Maximum	6168.15	8105.69	6732.40	31638.22	1936.10	1565.15	28.59.12
Minimum	2519.29	3666.41	3512.10	11015.84	628.50	676.53	1268.64
Std. Dev.	896.24	1137.86	760.04	4287.17	317.27	201.23	317.85
Skewness	-0.2456	0.2772	-0.4587	0.6162	-0.3349	-0.6335	0.6067
Kurtosis	2.0822	2.0451	2.3085	2.6225	2.2735	2.6652	3.0911
Observs	1209	1209	1209	1209	1209	1209	1029

Table 1: Descriptive statistics of original series

Series	CAC	DAX	FTSE	HSI	PX	SP500	NASD
Mean	0.000022	0.000279	0.000101	0.000341	0.000057	-0.000054	0.000051
Median	0.000404	0.001197	0.000500	0.000637	0.000890	0.000793	0.000683
Maximum	0.105946	0.107975	0.096404	0.142431	0.123641	0.109572	0.111594
Minimum	-0.094715	-0.077391	-0.092645	-0.146954	-0.161855	-0.094695	-0.095877
Std. Dev.	0.015781	0.015412	0.014444	0.019900	0.019191	0.015349	0.016207
Skewness	0.175234	0.106958	-0.042636	-0.109388	-0.713252	-0.276802	-0.099160
Kurtosis	11.51038	11.09040	12.26260	12.67253	15.36977	12.42167	10.06291
Observs	1208	1208	1208	1208	1208	1208	1208

Table 2: Descriptive statistics of transformed series

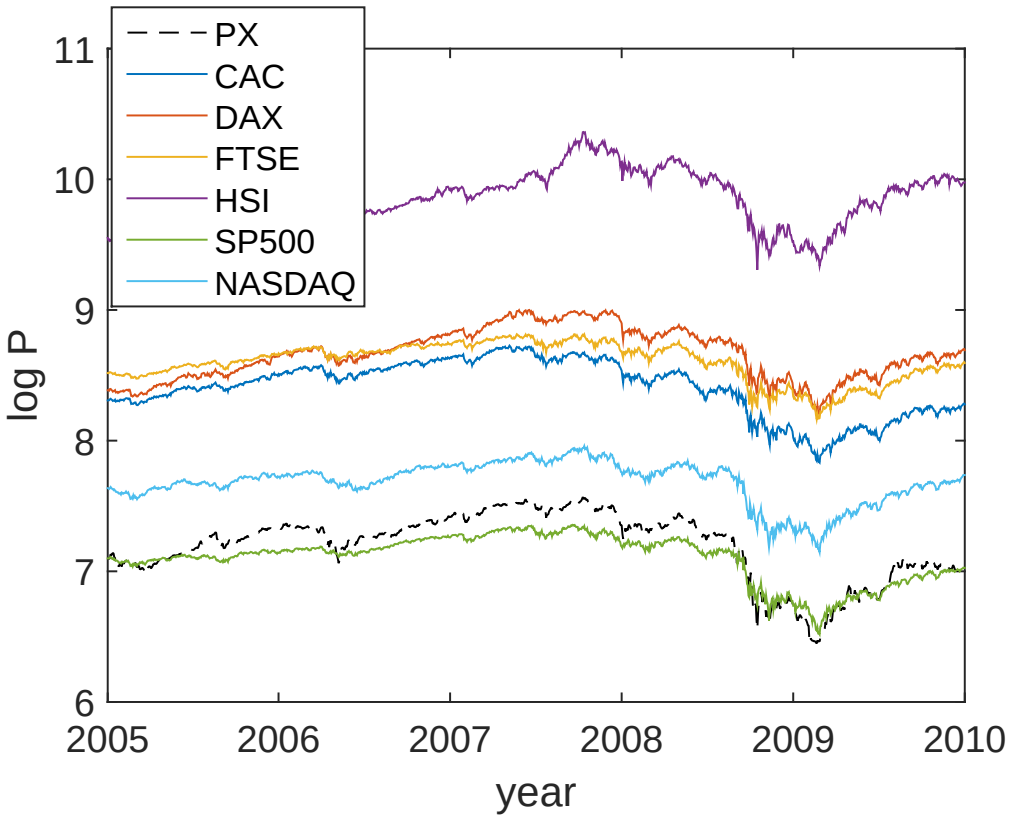


Figure 1: The (log-) development of stock market indices in 2005 – 2009

descriptive statistics are shown in Table 2. The (log-) development of indices is displayed in Figure 2 and a similarity of their advancement over the whole examined period can be observed.

First, we estimate the pairwise squared coherence of PX return series with one of return series of other stock market indices both in a non-turbulent time (the first half of investigated series) and in the turbulent time. As shown in (2), in order to estimate the coherence, one has to estimate both power spectrum density of return series of PX and the corresponding return series and the cross spectrum of them. For this purpose we use Welch method[3] described in [10]. When estimating power spectrum, the time series is divided into k successive blocks with an accordingly zeropadding adjustment and periodograms of each block are calculated using fast discrete Fourier transform. The spectrum is the average of periodograms across the blocks. We estimate the spectrum and cross spectrum of series PX and the corresponding return series for $k = 10$. Then the squared coherence are calculated and results are displayed in Figure 2.

In Figure 2 the red line displays the coherence of index PX with other indices at a time of quiet and the blue line shows it at the time of turbulence. The horizontal line is significance level for estimated values of coherence. Also the five highest values of squared coherence in both times and their corresponding frequencies are reported in Tables 3 and 4. The results show that there is no substantial difference in comovement of Prague stock market index PX with the most important markets in the world both at the quiet time as well as at the turbulent time. The size of the coherence is roughly the same in both periods. The highest coherence at the quiet time appears at frequency of (approximately) 0.2 which is equivalent to 5-day period, except with Hong Kong stock market index HSI with frequency of a half of size. At a time of turbulence while correlation of the Czech stock market with North American and Asian stock markets stay at the same frequency, its comovement with neighboring markets (France, Germany) happens at two thirds lower frequency. Further at figure 1 we can observe a statistically significant peak for coherence of index PX with Europeam stock market indices at a frequency close to 0.5 for the non-turbulent time which approximately corresponds to a two-day period. We can interpret it as for a calm time, market participants at Prague stock market try to adjust the stock prices accordingly to the world trend faster while when the time is tumultuous, they take a wait-and-see stance and adjust the prices later.

SC	CAC	DAX	FTSE	HSI	SP500	NASD
1	0.7515	0.6628	0.8108	0.6439	0.6627	0.6668
	0.2051	0.2017	0.1982	0.1017	0.2034	0.2052
2	0.7999	0.6604	0.8100	0.6414	0.6604	0.6644
	0.2034	0.2000	0.1965	0.1000	0.2017	0.2089
3	0.7461	0.6600	0.8068	0.6400	0.6596	0.6615
	0.2069	0.2034	0.2000	0.1034	0.2052	0.2034
4	0.7722	0.6534	0.8048	0.6330	0.6535	0.6528
	0.2017	0.1982	0.1948	0.0983	0.2000	0.2086
5	0.7328	0.6518	0.7977	0.6299	0.6500	0.6498
	0.2086	0.2051	0.2017	0.1051	0.2068	0.2017

Table 3: Quiet time squared coherence values between PX and other indices and their corresponding frequencies for five most significant couplings

We also calculate phase and gain at the frequency with highest value of coherence, both at the time of calmness as well as at the turbulent time and the results are shown in table 5. The results show that at the non-turbulent time a stock price move at Prague stock market caused by movement of world stock prices is affected in a similar way while in the turbulent time the influence of the Anglo-Saxon markets are much stronger than the one of Continental Europe. The shift measuring how much the world stock market prices lead prices of stocks traded at Prague stock exchange expressed in the fraction of a cycle shows that the adaptation to the Continental Europe prices is faster than to what happens at the Anglo-Saxon stock markets.

[1]Sometimes it is also called co-spectrum.
[2]Sometimes it is also called squared coherence or just coherence.
[3]This method is also called periodogram method.

SC	CAC	DAX	FTSE	HSI	SP500	NASD
1	0.7664	0.7790	0.7712	0.6755	0.7817	0.7514
	0.0672	0.0690	0.1948	0.1258	0.2103	0.2103
2	0.7655	0.7778	0.7705	0.6732	0.7808	0.7504
	0.0690	0.0672	0.1966	0.1276	0.2086	0.2012
3	0.7600	0.7742	0.7694	0.6723	0.7732	0.7411
	0.0655	0.0706	0.1931	0.1241	0.2121	0.2086
4	0.7568	0.7696	0.7675	0.6651	0.7709	0.7379
	0.0707	0.0655	0.1983	0.1293	0.2069	0.2138
5	0.7480	0.7610	0.7643	0.6635	0.7553	0.7206
	0.0638	0.0724	0.1934	0.1224	0.2138	0.2090

Table 4: Turbulent time squared coherence values between index PX and other indices and their corresponding frequencies for five most significant couplings

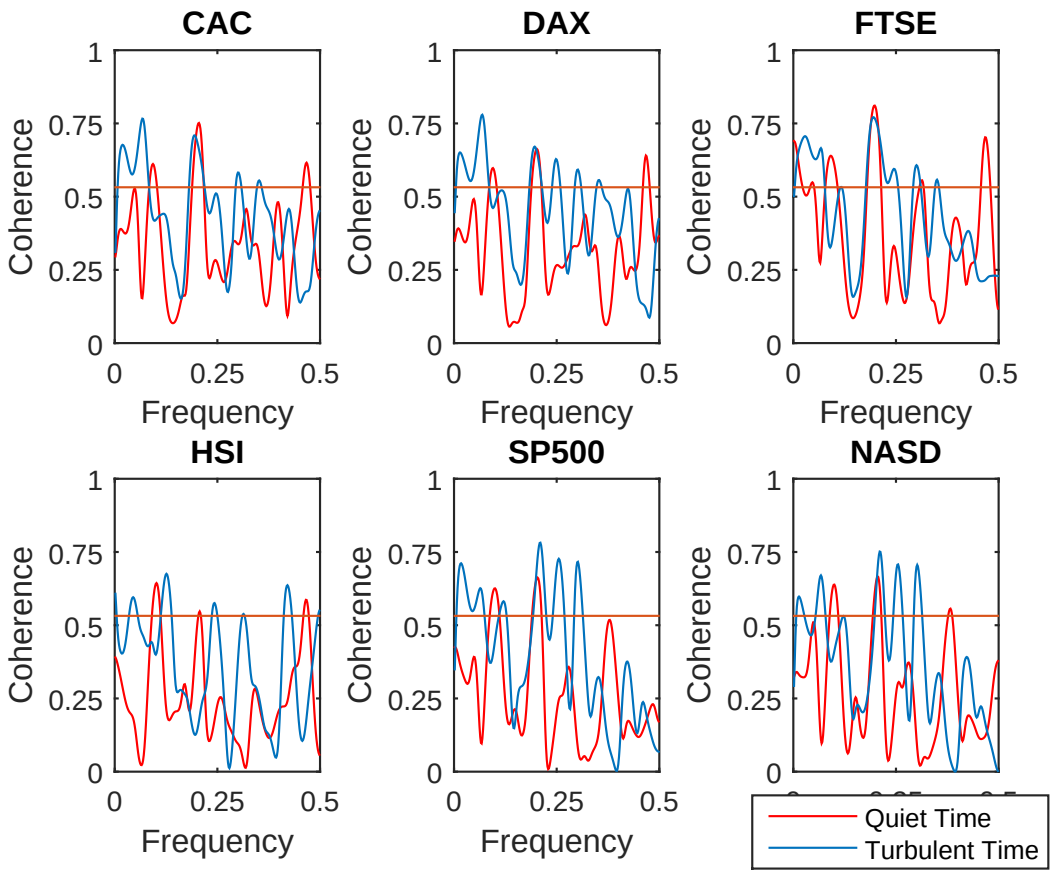


Figure 2: The relationship of squared coherence of returns of index PX with other world stock market indices and frequency

Indices	Quiet time		Turbulent time	
	Gain	Shift	Gain	Shift
CAC	0.6846	0.0479	0.5745	-0.0239
DAX	0.6678	0.0581	0.5880	0.0012
FTSE	0.5612	0.0486	0.9099	0.0488
HSI	0.5230	-0.0707	1.0336	-0.0469
SP500	0.5100	0.1945	1.0409	0.2129
NASD	0.6561	0.1776	0.9718	0.1903

Table 5: Gains and shifts of market indices related to index PX during the turbulent time for the highest ω^*

4 Conclusion

We study the comovement of stock prices at Prague Stock Exchange with stock prices of major stock market in the world using cross spectrum analysis of stock market indices of these markets from 1.2005 to 12.2009. During this period, which is divided into two periods representing one quite period (the first half) and one turbulent period (the second half) the results of our analysis show that price movement at Prague Stock Exchange is closely related with movement at other important stock markets in the world both in the quite period. However, the stock market in Prague is closer to Continental Europe markets than to the rest of our investigated markets. We also find that the market in Prague tends to react quicker in a calm time rather than in a turbulent time. As the Prague stock market probably cannot affect the world markets, this indicate that in a turbulent time the contagion may spread from the main stock markets to a local market slower as its participants take a cautious reaction to the unclear event.

Acknowledgements

The authors acknowledge the financial support of CTU in Prague under Grant SGS14/208/OHK4/3T/14.

References

[1] Cuthbertson, K.: *Quantitative financial economics*. Wiley, New York, 1996.

[2] Forbes, K., and R. Rigobon. No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance* **57** (5) (2002), 2223–61.

[3] Granger, C. W. J. and Hatanaka, M.: *Spectral Analysis of Economic Time Series*. Princeton University Press, Princeton, 1964.

[4] Granger, C. W. J. and Morgenstern, O.: Spectral Analysis of Stock Market Prices, *Kyklos* **14** (1963), 1–27.

[5] Gray, D.: Central European foreign exchange markets: a cross-spectral analysis of the 2007 financial crisis. *The European Journal of Finance* **20** (6) (2014), 550–567.

[6] Hamilton, J. D.: *Time Series Analysis* . Princeton University Press, Priceton, 1994.

[7] Hau, H.: The exchange rate effect of multi-currency arbitrage. *CEPR Discussion Paper* **7348**, (2009).

[8] Orlov, A.: A cospectral analysis of exchange rate co-movements during the Asian financial crisis. *Journal of International Financial Markets, Institutions and Money* **19** (5) (2009), 742–58.

[9] Smith, K.: Pre- and post-1987 crash frequency domain analysis among Pacific Rim equity markets. *Journal of Multinational Financial Management* **11**(1), (2001), 69–87.

[10] Welch, P. D.: The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio Electroacoustics* AU-15 (1967), 70—73.

Optimization Models Used for the Electricity Supplier Selection

Martina Kuncová[1], Jana Sekničková[2]

Abstract. The choice of the electricity suppliers in the Czech Republic is not so easy as it may seem. The electricity market transformation caused the increasing number of suppliers offering this commodity to households and companies. The final price for the electricity consumption is not influenced only by the suppliers' conditions and the consumption itself but also by the distributor. As the distributor is given for each of the 3 regions it is necessary to compare the tariffs inside the region. Prices change every year according to the regulations of the Energy Regulatory Office and also with respect to the decisions of each supplier. Considering the possible changes in the annual consumption of the consumer we look for the best tariffs and ranges for the consumption for each region via an optimization model. The results of two tariff rates (D25d and D02d) are compared.

Keywords: Electricity prices, Suppliers, Consumption, Optimization model.

JEL Classification: C44, C63, O13
AMS Classification: 90C15

1 Introduction

Electricity belongs to the commodities that are essential for our lives and also for the economic development. The expansion of modern technologies and the increase of the electronic equipment usage to ease the work, to relax, to study, etc. causes the non-decreasing demand for electricity. The transformation of the electricity market in the Czech Republic started in 2002 and for households in 2006. After this process each household and company can choose the supplier of the electricity. This liberalization has led to the increasing number of suppliers entering the electricity market. Other subjects on this market in the Czech Republic are distributors, Energy Regulatory Office (ERU) and operator of the market (OTE) [3]. The Czech Republic is divided into three network parts operated by three distributors (PRE, CEZ, E.ON.) and in each distributors area a lot of suppliers is working. Each household has its tariff rate according to the supplier's conditions.

The complete list of suppliers and their tariffs and prices is changing every year. The selection of the suppliers depends on the contract conditions but mainly on the prices. With respect to all these it is a hard task for consumer to find the best supplier. Various techniques and methods can be used to model the situation on the market – like in [4], [5], [7], [8].

In this paper we use the optimization approach, in contrast to simulation approach presented in [4] and [5] or multicriteria decision making approach published in [1] and [7]. The optimization consists in the search for the electricity consumption ranges for each product just to minimize the annual electricity consumption cost. We analyze the problem focused on the electricity consumption of one household (tariff rate D25d) to compare the final prices for all suppliers and their products in all regions in 2015 and we compare it with the results from the previous research [5] and also with the results for the tariff rate D02d. For each acceptable supplier's product we set the range of total energy consumption that corresponds to this product to be selected as the best one.

2 Czech Electricity Market

The electricity market in the Czech Republic was specific till 2002 when the transformation process oriented at the fully liberalization has started. In the first phase (till 2006) only the companies could enter the market but since 2006 also the households can choose the electricity supplier on the retail market. Except of customers and retailers there are also other subjects on the electricity market, especially distributors, the Energy Regulatory Office (ERU) and the Operator of the market (OTE). OTE predicates the whole market consumption and analyses the differences [6], ERU regulates the prices of the transfer and distribution of the electricity [3]. The high number of suppliers and their products on the retail market embarrasses the position of the households. Accord-

[1] University of Economics, Prague, Department of Econometrics, W.Churchill Sq. 4, 13067, Prague 3, e-mail: martina.kuncova@vse.cz .
[2] University of Economics, Prague, Department of Econometrics, W.Churchill Sq. 4, 13067, Prague 3, e-mail: jana.seknickova@vse.cz

ing to this situation it is hard to follow the rules and the price changes on the market and so it is hard to choose the best (cheapest) product.

The product selection is influenced mainly by the electricity take-off amount and by the prices for the electricity consumption. The final price is given by more factors such as consumption, fixed fees or taxes. Generally the price can be divided into two components. The first one is the controlled charge for services related to electricity transport from the generator to the final customer. This charge is annually given by Energy Regulatory Office [3]. It covers:

- monthly lease for the circuit breaker,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT),
- price per system services,
- price for the support of the renewable energy purchase,
- charges for the electricity market operator,
- electricity ecological tax (28,30 CZK per 1 MWh).

The second part of the total price is given by the electricity supplier. It covers:

- fixed monthly fee for the selected product,
- price per megawatt hour (MWh) in high tariff (HT),
- price per megawatt hour in low tariff (LT).

The final price is increased by VAT that is 21% from 2013.

3 Data and Methods

In our previous analysis [5] we compare the offered 57 products for the tariff rate D25d with the electricity consumption about 10 MWh anually, 45% energy in high tariff and 55% in low tariff and with the circuit breaker from 3x20A to 3x25A. This tariff rate is given to household when the electricity is used also for the accumulative heating and hot water heating for lower and middle yearly offtake with operative management of the validity period of the low tariff for 8 hours. It is so-called dual tariff rate as it has 2 periods (high tariff, low tariff) during the day. Another tariff rate that can be used in this situation (with higher electricity consumption) is the tariff rate D02d where no low tariff is possible. We have chosen the situation with higher electricity consumption because there is higher chance to switch to another supplier and product because of the price difference. The same situation was mentioned in [2] where the authors wrote: "*The impact of overall electricity costs and knowledge about these is particularly important for the latter decision, while respondents that perceive relatively high search and information costs are less likely to switch to an alternative electricity supplier*".

According to the ERU calculator [3] in this paper we use data for 60 products (offered by 32 suppliers) in all three distribution areas. All data display prices for year 2015. For illustration we selected 3 districts, each from one area: Chrudim (for distributor CEZ), Prague (for distributor PRE) and Pelhrimov (for distributor E.ON).

To find the best product we can use various methods and techniques. One of them (that we also have used for the previous comparison [5]) is Monte Carlo simulation of the total energy consumption. The ranges for the electricity consumption in each month were set (at about 900 kWh per month on average), the high tariff is used in 45% from the whole consumption. Our simulated consumption has been generated for each month from the normal distribution with 20% of the average taken as the standard deviation. In all Monte Carlo simulations 1000 experiments have been tried to randomly select consumption for each month and afterwards the annual costs are calculated. The formula for the annual cost calculation for each supplier's product is following:

$$COST_{ij} = (1 + VAT) \cdot \left[12 \cdot \left(mf_{ij} + mf_j\right) + c\left(0.45 \cdot \left(ph_{ij} + ph_j\right) + 0.55 \cdot \left(pl_{ij} + pl_j\right) + (os + t)\right)\right] \quad (1)$$

where

i … tariff, $i = 1, \ldots, 60$,
j … distributor, $j = 1, \ldots, 3$,
VAT … value added tax ($VAT = 0.21$),
mf … fix monthly fee,
c … yearly electricity consumption in MWh,
ph … price in high tariff per 1 MWh,
pl … price in low tariff per 1 MWh,
os … price for other services per 1 MWh,
t … electricity tax per 1 MWh ($t = 28.3$ CZK).

In the first step the value of this function has been calculated for each tariff $i = 1, \ldots, 57$ and each distributor j with respect to generated yearly electricity consumption. In the second step for each distributor j the tariff with the lowest value of cost function is chosen as the best tariff (BT_j),

$$BT_j = \arg\min_i COST_{ij}, j = 1, 2, 3. \tag{2}$$

The second approach for selection of the best product is using of multi-criteria evaluation of alternatives and we used it in [7]. We has applied three multi-criteria methods – WSA, TOPSIS and MAPPACC with respect to three criteria (price in high tariff, price in low tariff and fixed monthly fee). Each suggested method has been applied on data for each distributor and the best product has been selected. It is obvious that only non-dominated alternatives were selected as the best ones.

In the case we use only one criterion the third approach can be used and we introduce it in this paper. The decision problem is based on the fact that the consumer does not know his electricity consumption and so this variable is stochastic. In the case of deterministic consumption the problem leads to mono-criteria problem where the objective function for analysis is the cost function. The total costs can be calculated for each distributor and each tariff according to formula (1) where c denotes known yearly electricity consumption and the best product has the lowest value of total costs – it can be selected using formula (2). It is obvious that dominated product (in the multi-criteria sense) cannot be selected for any value of consumption. And this idea can be used also in stochastic case. For non-dominated suppliers (or product) we are able (based on formula (1)) to calculate the values of energy consumption c that ensure winning of the given supplier. The mathematical model of this problem for each product $i = 1, \ldots, 60$ and distributor $j = 1, 2, 3$ can be written as following:

$$\mathbf{min}(\mathbf{max})\ z = c$$

subject to

$$COST_{ij} \leq COST_{kj}, \tag{3}$$

$$COST_{ij} = (1 + VAT) \cdot \left[12 \cdot (mf_{ij} + mf_j) + c\left(0.45 \cdot (ph_{ij} + ph_j) + 0.55 \cdot (pl_{ij} + pl_j) + (os + t)\right)\right],$$

for all $i = 1, \ldots, 60, j = 1, 2, 3$ and $k = 1, \ldots, 60$.

The formula (3) is valid for the tariff rate D25d. In case of the tariff rate D02d (with no low tariff possibility), the pl_{ij} and pl_j values are set to zero and the probability of the high tariff usage is set to 1.

This model corresponds to 180 linear optimization problems with 60 linear constraints in simple case (with two extremes). However the solution of this set of problems can be simplified by dominance that was mentioned above. For dominated product i (with respect to distributor j) problem (3) has no feasible solution. In 2015 there are only 19 non-dominated products for E.ON, 8 non-dominated products for PRE and 9 non-dominated products for CEZ distributor and so only 36 optimization problems have to be formulated. We can solve each problem twice (the first for minimal consumption and the second for maximal consumption) or we can use the theory of alternative solution in core points. Using this theory we can also reduce the number of solved problems.

4 Results

Using model (3) from previous section we analyzed 60 products for all three distributors in 2015 with respect to tariff D25d. We use also this model on 57 products from year 2014 and 61 products in 2015 with respect to tariff D02d. All results are compared in this section.

4.1 Tariff D25d in year 2015

Table 1 summarized the results from analysis described in previous section based on data for year 2015 and tariff D25d. Only four of products can be selected as the best one. For the household with energy consumption less than 4.4 MWh per year Fonergy Premium is the best tariff regardless of distributor. Table 1 also shows that households with high energy consumption (more than 14 MWh per year) should select CARBOUNION KOMODITY in each region. We can so conclude that Fonergy Premium is the best tariff for households with low energy consumption, ST Energy standard for households with middle consumption and CARBOUNION KOMODITY for households with high energy consumption. Note that product no. 43 Fosfa is dominated by product no. 48 Fonergy Premium in PRE and CEZ region and it is the winner only in the case of zero consumption.

In comparison to previous analysis [3] the household with total consumption about 10 MWh per year will select CARBOUNION KOMODITY in E.ON region and ST Energy standard in PRE and CEZ region.

	product no.	*name of product*	*consumption in MWh*		*costs in CZK*	
			from	*till*	*from*	*till*
E.ON.	43	Fosfa	0.0000	0.0000	1379.40	1379.40
	48	Fonergy Premium	0.0000	4.3969	1379.40	14067.25
	40	ST Energy standard	4.3969	7.4227	14067.25	22558.52
	60	CARBOUNION KOMODITY	7.4227	*unbounded*	22558.52	*unbounded*
PRE	43	Fosfa	0.0000	0.0000	1481.04	1481.04
	48	Fonergy Premium	0.0000	5.0087	1481.04	15743.00
	40	ST Energy standard	5.0087	14.0390	15743.00	40827.96
	60	CARBOUNION KOMODITY	14.0390	*unbounded*	40827.96	*unbounded*
CEZ	43	Fosfa	0.0000	0.0000	1597.20	1597.20
	48	Fonergy Premium	0.0000	5.0087	1597.20	16498.54
	40	ST Energy standard	5.0087	16.8562	16498.54	50921.67
	60	CARBOUNION KOMODITY	16.8562	*unbounded*	50921.67	*unbounded*

Table 1 Ranges for the best products of 2015

4.2 Tariff D25d in year 2014

The results from section 4.1 can be compared with data from 2014. Note that after testing only 8 non-dominated alternatives stay for E.ON, 11 for PRE and 12 for CEZ distributor. Table 2 represents ranges for the best products of 2014 with respect to tariff D25d calculated by solving of problem (3) and using formula (1).

	product no.	*name of product*	*consumption in MWh*		*costs in CZK*	
			from	*till*	*from*	*till*
E.ON.	40	Fosfa	0.0000	0.0000	1306.80	1306.80
	45	Fonergy Premium	0.0000	4.6412	1306.80	15337.03
	57	CARBOUNION KOMODITY	4.6412	*unbounded*	15337.03	*unbounded*
PRE	40	Fosfa	0.0000	0.0000	1422.96	1422.96
	45	Fonergy Premium	0.0000	4.6412	1422.96	15362.50
	57	CARBOUNION KOMODITY	4.6412	*unbounded*	15362.50	*unbounded*
CEZ	40	Fosfa	0.0000	0.0000	1524.60	1524.60
	45	Fonergy Premium	0.0000	4.6412	1524.60	15926.23
	57	CARBOUNION KOMODITY	4.6412	*unbounded*	15926.23	*unbounded*

Table 2 Ranges for the best products of 2014

Product Fosfa can be chosen again only in the case of zero energy consumption so only two products are efficient in all three distributors areas. Fonergy Premium is the best tariff for households with energy consumption less than 4.64 MWh per year and for households with higher consumption the best alternative is CARBOUNION KOMODITY. The boundary value is the same for all regions in 2014, one year later it changes its value among regions.

From tables 1 and 2 we can see that situation on the energy market changes in the time. The list of products involves almost the same products and Fonergy Premium and CARBOUNION KOMODITY stay the best ones. However the boundary values for products change in the time and for the household with energy consumption e.g. about 6 MWh per year it is profitable to change CARBOUNION KOMODITY (2014) to ST Energy standard in 2015 with respect to minimize total energy costs. The total costs in 2015 by ST Energy standard

are 18566 CZK in E.ON, 18497 CZK in PRE and 19379 CZK in CEZ region in comparison with 18624 CZK in E.ON, 18671 CZK in PRE and 19575 CZK in CEZ region for CARBOUNION KOMODITY. We can so see that the suggested change of product is efficient in all distributor areas and the household save 58 – 196 CZK, i.e. 0.3 – 1% of total energy costs.

4.3 Tariff D02d in 2015

As we did the same analysis for data from 2015 in tariff D02d it is possible to compare these two tariffs. In the table 3 (in comparison with table 1) we can see that higher number of products is selected as the best. Note firstly that product no. 13 Fosfa is evaluated as well as product no. 35 Fonergy Premium for all distributors (numbers of the products belong to the given year and tariff rate as we have taken them from [1] and so they are not the same for the same product in different years). These two products are so indifferent from the decision making point of view. Note also that product no. 5 Pražská energetika Komfort+Garant 2015 is the best tariff only in the case of zero energy consumption and this alternative is dominated by no. 13 Fosfa.

From the table 3 we can see that for households with energy consumption lower than 1.97 MWh per year the best choice is no. 13 Fosfa or no. 35 Fonergy Premium product and the best product for households with consumption up to 7.26 MWh per year is no. 31 Europe Easy Energy Duo 24 regardless of distributor. For higher consumption the best product differs with respect to distribution area but households with consumption about 7.5 – 12 MWh per year will select product no. 28 Europe Easy Energy Company in all areas and consumption 27 – 60 MWh per year corresponds to product no. 57 EP Energy Trading Klasik. Also tariff no. 51 Bohemia Energy Home Standard could be selected – especially in the cases with high energy consumption in PRE and CEZ distributor areas.

	product no.	*name of product*	*consumption in MWh*		*costs in CZK*	
			from	*till*	*from*	*till*
E.ON.	5	Pražská energetika Komfort+Garant 2015	0.0000	0.0000	914.76	914.76
	13	Fosfa FEE Standard 24	0.0000	1.9714	914.76	8870.40
	35	Fonergy Premium	0.0000	1.9714	914.76	8870.40
	31	Europe Easy Energy Duo 24	1.9714	7.2632	8870.40	29776.81
	28	Europe Easy Energy Company	7.2632	27.0000	29776.81	107298.81
	57	EP Energy Trading Klasik	27.0000	*unbounded*	107298.81	*unbounded*
PRE	5	Pražská energetika Komfort+Garant 2015	0.0000	0.0000	1030.92	1030.92
	13	Fosfa FEE Standard 24	0.0000	1.9714	1030.92	8985.80
	35	Fonergy Premium	0.0000	1.9714	1030.92	8985.80
	31	Europe Easy Energy Duo 24	1.9714	7.2632	8985.80	29890.16
	28	Europe Easy Energy Company	7.2632	27.0000	29890.16	107404.51
	57	EP Energy Trading Klasik	27.0000	116.0000	107404.51	455650.28
	51	Bohemia Energy Home Standard	116.0000	*unbounded*	455650.28	*unbounded*
CEZ	5	Pražská energetika Komfort+Garant 2015	0.0000	0.0000	1030.92	1030.92
	13	Fosfa FEE Standard 24	0.0000	1.9714	1030.92	9218.35
	35	Fonergy Premium	0.0000	1.9714	1030.92	9218.35
	31	Europe Easy Energy Duo 24	1.9714	7.2632	9218.35	30746.94
	28	Europe Easy Energy Company	7.2632	12.0000	30746.94	49909.16
	57	EP Energy Trading Klasik	12.0000	60.0000	49909.16	242518.12
	51	Bohemia Energy Home Standard	60.0000	*unbounded*	242518.12	*unbounded*

Table 3 Ranges for the best products of 2015 in tariff D02d

4.4 Comparison for model household

In our previous analysis [5] we started with the problem of selection of the best product for household with the electricity consumption about 10 MWh annually, 45% energy in high tariff and 55% in low tariff and with the circuit breaker from 3x20A to 3x25A. The household uses the tariff rate D25d. Now we compare the results from this paper for such model household.

In year 2014 from table 2 we can see that the best tariff for such household was product no. 57 CARBOUNION KOMODITY with costs about 30782 CZK in E.ON, 30703 CZK in PRE and 31800 CZK in CEZ distribution area.

One year later the best product in E.ON area stays the same but costs decrease to 29686 CZK. For PRE and ČEZ distributors the best choice (see table 1) is product no. 40 ST Energy standard with costs about 29608 CZK, or 31001 CZK respectively. We can so conclude that the correct change of product leads to savings and the total costs will be lower in 2015 than in 2014. On the other hand we do not take into account the conditions of the switching to new product or supplier and they might be complicated or the previous supplier can ask for high compensation price in case of switching.

In the case the household will change the tariff rate to D02d (this change is not usual) the best product is no. 28 Europe Easy Energy Company in all three areas with costs higher than 40000 CZK (see table 3). From this we can see that such change is disadvantageous.

5 Conclusion

Selection of the best electricity supplier can be a complicated task. As authors in [8] written*, "the ability of consumers to choose the best alternative supplier is limited even in a relatively simple and transparent market"*. It is possible to calculate the ranges for consumption where the non-dominated supplier wins. We suggested the model that calculates minimal and maximal energy consumption needed for selection of given product as the cheapest and the results obtained express that dominated offers should not be selected. Our analysis also has shown that the supplier with the lowest annual cost for one year do not have to be the cheapest one in the next year (for the same consumption). On the other hand the choice is also influence by the region where the household is as for the same conditions and different regions we have found out different suppliers to be the best ones. We also showed the selection of the best supplier is dependent on the time.

Acknowledgements

This work was supported by grant no. 13-07350S of the Czech Science Foundation.

References

[1] Diakoulaki, D., Karangelis, F.: Multi-criteria decision analysis and cost–benefit analysis of alternative scenarios for the power generation sector in Greece. *Renewable and Sustainable Energy Reviews*, **11/4** (2007), 716–727.

[2] Ek, K., Sőderholm, P.: Households' switching behavior between electricity suppliers in Sweden. *Utilities Policy*, **16/4** (2008), 254–261.

[3] ERU - Energy Regulatory Office: *Price calculator* [online], available at: http://kalkulator.eru.cz/VstupniUdaje.aspx [cit. 2015-05-05]

[4] Hegazy, Y.G., Salama, M.M.A. and Chickhani, A.Y.: Adequacy assessment of distributed generation systems using Monte Carlo Simulation. *Power Systems - IEEE Transactions* **18/1** (2003), 48-52.

[5] Kuncová, M., Sekničková, J. 2014. Analysis of the efficiency of the electricity supplier selection depending upon the price changes. In: *Mathematical Methods in Economics* (MME2014) [CD]. Palacký University in Olomouc, 2014, 542–547.

[6] OTE - *Basics* [online], available at: https://www.otecr.cz/o-spolecnosti/zakladni-udaje [cit. 2015-05-05]

[7] Sekničková, J., Kuncová, M.: Evaluation of the locality influence on the electricity supplier selection. In: *Mathematical Methods in Economics* (MME2014) [CD]. Palacký University in Olomouc, 2014, 879–883.

[8] Ventosa, M., Baíllo, Á., Ramos, A. and Rivier, M.: Electricity Markets modeling trends, *Energy Policy* **33**, 897–913.

[9] Wilson, Ch., M., Waddams Price C.: Do consumers switch to the best supplier? *Oxford Economic Papers* **62/4** (2010), 647-668.

Choice of optimization routine for multi-agent models: A case of viral video marketing campaign

Michal Kvasnička[1]

Abstract. Very few agent-base computational models are optimized because the usually used optimization routine, the genetic algorithm, is extremely time-consuming. This paper explores how much precision is lost if a simpler optimization routine, mutational hill climber, is used instead. It shows on the case of a viral-video marketing model that even though the standard genetic algorithm is slightly more precise, the mutation hill climbing could be used as an approximate optimization routine for robustness check and scenario analysis.

Keywords: optimization, genetic algorithm, mutation hill climbing, simulation, agent-based model, social network, viral video marketing.

JEL classification: C61, C63, D85, M31
AMS classification: 68U20, 37N40, 05C82, 90B15

1 Introduction

Agent-based computational models (ABM) allow us to study many real-world phenomena which are difficult to explore with the standard tools of economic theory. The models can also be used to optimize the explored processes. However, since the structure of the ABM models does not allow us to use the standard optimization routines and available techniques (e.g. genetic algorithms) are computationally extremely demanding, most researchers do not try to optimize their models, and even those few exceptions that do usually stick with one or few particular settings of their model's exogenous parameters, and do neither provide robustness checks, nor explore systematically how the optimal parameters depend on the exogenous parameter of the model.

For example, consider the ABM literature on viral marketing. (The ABM is a natural tool to study this phenomenon because the explicit description of the social network is necessary, the agents interact locally within the network, and their behavior is non-continuous.) Most of this wide literature (for a review see e.g. [1]) is only exploratory: it studies how knowledge about a new product diffuses over a social network. The understanding of the diffusion process is interesting as such, but it also creates an opportunity for marketers to actively utilize the spontaneous knowledge diffusion as an advertising tool. To do it, the marketer must decide how many and which consumers "infect" with the knowledge of the product and how to do it. This, however, requires to optimize the model. Presently, there are only few papers trying to optimize the ABM viral marketing models because the optimization is extremely time consuming even in a case of a small stylized social network. If we neglect the papers with unrealistic assumptions that marketers know the complete description of the social network and are able to optimize over it, there are only two papers trying to optimize a viral marketing campaign, [6] and [2].

Both these papers assume that the marketer has only local knowledge about the social network, such as the number of agents in the network, the agents' degree (the number of their connections), their clustering, their willingness to share the knowledge etc. The marketer's problem is then to seed the network, i.e. to choose the agents she initially "infects" with the knowledge of a product, to maximize her objective function. Both these papers also describe a seeding strategy as a vector of the number of initially seeded agents and weights placed on their desirable properties. Both the papers use genetic algorithms implemented in BehaviorSearch [4, 5] to optimize the seeding strategy too. The papers differ in their underlying assumptions about the diffusion process ([6] assumes the traditional "word-of-mouth" marketing, while [2] assumes viral-video marketing) and the marketer's objective function ([6] assumes that the marketer maximizes the discounted number of "infected" agents, while [2] assumes she maximizes the profit from the marketing campaign). The authors of [6] optimized two scenarios on five stylized networks, which took about 462 CPU-days. [2] optimized only one scenario, which took about 480 CPU-days.

[1]Masaryk University, Faculty of Economics, Department of Economics, Lipová 41a, Brno, 602 00, michal.kvasnicka@econ.muni.cz

The above mentioned case shows that using genetic algorithms (which are supposed to be the most sophisticated optimization tool available, see [4, 5]) to optimize an ABM model practically precludes a more systematic exploration of the optimal setting of the model, at least at the present state of the computing power. To do so, it is necessary to use faster (yet perhaps less sophisticated) optimization routines. The mutation hill climbing seems to be a promising substitute. However, it is not known how much precision is lost when the genetic algorithms are substituted with the mutation hill climbing. The goal of this paper is explore this question. Since the ABM does not allow us to study the problem in general, it is assessed on the particular case of the ABM model of viral-video marketing described in [2]. Our procedure is the following: 1) we implement the model, 2) we find its optimal parameters both using genetic algorithms and mutation hill climbing (the first routine is given reasonably more resources than the later one), and 3) we compare the results.

2 Model

The model used to test the relative performance of the standard genetic algorithm (GA) and the mutation hill climbing (MHC) is a slightly modified version of [2]. It differs from the other ABM models of viral marketing in its agents' activation: it assumes that people share videos over Internet and that their decision to share the video is independent from their decision to adopt the product advertised by the video (if any). It differs from the model described in [2] in three respects: 1) there are fewer agents (250 instead of 1 000), 2) the distribution of followers' links is independent on the distribution of the friends' links, and 3) the agents' properties seen by the marketer were slightly modified too. The model consists of three parts: 1) the activation mechanism (i.e. how the agent share the video and get "infected"), 2) an explicit description of the used social network, and 3) the seeding strategy (i.e. which and how many agents get initially infected by the marketer).

2.1 Activation mechanism

I call every agent that has viewed the video *infected* with no regard whether she has adapted the advertised product or not. An infected agent's decision whether and with whom to share a video and the receiver's decision to view the video are modeled as a simple probabilistic act: an infected person shares the video with each of her neighbors with some probability. If she shares it with a person, the person gets infected. Each infected person shares the video only once.

The precise mechanism of the activation is following: At the initialization, every agent i draws a probability p_i that she shares the video; p_i is drawn from the continuous uniform distribution $U(0, v)$ where v is the maximal *virality* of the video. Then she creates a list l_i of agents to share the video with: she adds each of her neighbors to the list l_i with probability p_i. When agent i is first infected at time t, she shares the video with all agents in her list l_i at time $t + 1$, and each of these agents get instantly infected. Agent i shares the video with no one after time $t + 1$.

2.2 Network structure

An agent's neighborhood is defined by a social network in which the agent is located. There are two kinds of relations between the agents in the model: friendship and following. Friendship is a symmetric relationship between two agents that can share videos with each other. Following is an asymmetric relationship between a followed person and her follower, e.g. between a celebrity and her fan. The followed person can share videos with the follower but not vice versa. I assume that no person can be at the same time one's friend and follower.

Each model network consists of 250 agents and its parameters are selected to resemble the properties of the empirical networks. Each network is created in two steps. First, the symmetric small world network which represents the friendship is created by algorithm described by [8]; the code has been adapted from [10]. The agents are arranged into a circle. Each agent initially has 10 friends, 5 agents to the left of her and 5 agents to the right of her. Each link representing friendship is then rewired with probability 10 %, which creates the initial small world network. Second, the asymmetric power network of directed links which represents the following is created over the friends' network. The algorithm has been adapted from [11]: one follower connection is added at a time, each agent is selected randomly as the follower with an equal probability and one other agent is selected randomly as the followed one with the probability proportional to each agent's number of the previously created followers. Two followers' links per agent are added. The set of agent i's neighbors used in the activation mechanism described above is the union of the set of her friends and the set of her followers.

2.3 Marketer's seeding strategy and profit

The marketer's seeding strategy is based on the approach developed by [6]. The agents are included into the seed because they have some desirable properties. Since there are multiple desirable properties, they are weighted. A seeding strategy (S, w) thus consists of two parts: the seed size S, i.e. how many agents the marketer initially infects, and weights w_j placed on measures of some desirable properties f_j that determine which agents are selected into the seed. An index $\sum w_j f_j$ is calculated for each agent and S agents with the highest value of the index are included into the seed. I use six measures f_j: 1) f_1 = the number of agent's friends divided by the maximal number of friends in the population, 2) f_2 = the number of agent's followers divided by the maximal number of followers in the population, 3) f_3 = the number of agent's friends and followers divided by the maximal number of friends and followers in the population, 4) f_4 = the agent's sharing probability p_i divided by the maximal sharing probability in the population, 5) f_5 = 1− (agent's clustering ratio / the maximal clustering ratio in the population), and 6) f_6 is a random number drawn from $U(0, 1)$. The first three fs are various measures of degree—the more connections an agent has, the better she can share the video. (Only f_1 and f_2 are used here, i.e. $w_3 = 0$. Property f_3 is used only for comparison, see below.) The f_4 measures how likely the agent is to share the video and how likely are her recipients to watch it—the higher, the better she can share the video. The f_5 indicates the agent's absence of clustering—the less likely an agent's neighbors are to be neighbors themselves, the better the video can spread through the population. The f_6 allows including agents into the seed randomly, which may be beneficial e.g. when the agents with the highest degree are connected together.

Since an agent's decisions to share a viral video and to adopt the product advertised by the video are independent, the expected revenue from the viral marketing campaign is then equal to $\rho\sigma N$, where N is the number of agents infected during the campaign, ρ is the probability that an infected agent adopts the product because of the campaign, and σ is the profit from one adopter. The cost of the campaign is $\gamma S + F$ where γ is a cost of seeding one agent, S is the seed size, and F is a fixed cost. The marketer's problem is then to select the seeding strategy (S^*, w^*) that maximizes her expected profit from the campaign, $\Pi = \rho\sigma N - \gamma S - F$. The seeding strategy (S^*, w^*) maximizing Π also maximizes $\pi = N - cS$ where $c = \gamma/(\rho\sigma)$ is the relative cost of seeding one agent. (Since I do not explicitly address the problem of setting the optimal budget for the video creation, I treat F, ρ, and v as constants.)

2.4 Simulation and implementation

Each simulation consists of two parts: the initialization and the run. In the initialization, the agents are created and connected within the model network. Each agent i is assigned the probability p_i that she shares the video with her friends and followers. She then creates the list l_i of the agents which she shares the video with if infected. At the end of the initialization, the initial agents are infected. Each simulation run proceeds in discrete steps. In each step, each agent i infected in the previous step infects the agents in l_i. The run ends when there are no infected agents that have not yet shared the video. Each model is simulated for 250 agents, the activation and network described in sections 2.1 and 2.2, maximal virality $v = 20$ %, and relative seeding cost $c = 10$. The model was implemented in NetLogo 5.1 [9].

2.5 Optimization

To assess the relative performance of the standard genetic algorithm (GA) and the mutation hill climbing (MHC) (as implemented in BehaviorSearch 1.0 [4]), the optimal seeding strategy (S^*, w^*) was searched by both these methods. The seed size S was searched on domain of $1, 2, \ldots, 20$ and each weight w_j on domain $[0, 1]$ (the model automatically normalizes $\sum w$ to unity). All variables were encoded as Gray binary chromosomes. The fitness of each individual strategy was evaluated by the mean relative profit π of 50 independent replications of the simulation. The whole search for the optimal parameters was repeated 50 times for both the GA and MHC optimizer (i.e. 50 optimal seeding strategy candidates were produced by each optimizer).

The genetic algorithm started with the initial population of 50 randomly chosen strategies. The standard generational evolution steps (one-point crossover rate 0.7, mutation rate 0.03, and tournament selection with tournament size 3) were performed on the population of the 50 strategies for 80 generations. Total number of model runs was 200 000 (excluding checking replicates) per one search. The time needed for the GA optimization was about 61 CPU-days. The mutation hill climber started with one random strategy. The mutation rate was set to 0.03, the stalled searches were restarted after 300 attempts. The total number of model runs was 30 000 (excluding checking replicates). The mean time needed for the MHC optimization was about 11 CPU days (the mean is computed from 72 similar optimization runs carried for a related project).

3 Results

The performance of the GA and MHC optimizer is compared in two ways: 1) the "optimal" seeding strategy parameters are compared and 2) the profits realized from the two strategies are compared with each other and with the profits from the simple seeding strategies. The performance analysis was carried in R 3.2.0 [3].

Table 1 presents the parameter of the optimal seeding strategies found by the standard genetic algorithm and the mutation hill climber respectively. The statistics are computed from 50 candidates for each optimizer. The full distributions of the parameters are presented in Figures 1 and 2. The parameters are very similar, with the exception of the sharing weight w_4 and clustering weight w_5 (the mean values of these two parameters differ significantly between the respective optimizers at the significance level 5 %). The standard genetic algorithm is slightly more efficient—the standard deviations of the parameters found by GA are in general lower, i.e. individual runs of the optimizer produce candidates that are closer to each other than the candidates produced by MHC. However, the difference between the variances of the candidates' parameters is statistically significant at 5 % only in the case of the seed size and the weight of followers w_2.

	mean		sd	
	MHC	GA	MHC	GA
seed size	3.180	3.080	0.720	0.274
w_1 friends	0.216	0.241	0.106	0.106
w_2 followers	0.245	0.261	0.098	0.065
w_4 sharing	0.323	0.362	0.086	0.083
w_5 clustering	0.161	0.088	0.094	0.066
w_6 random	0.055	0.048	0.037	0.030

Table 1: The optimal seeding strategy candidates's parameters found by the standard genetic algorithm (GA) and the mutation hill climber (MHC). The statistics are computed from 50 candidates for each optimizer.

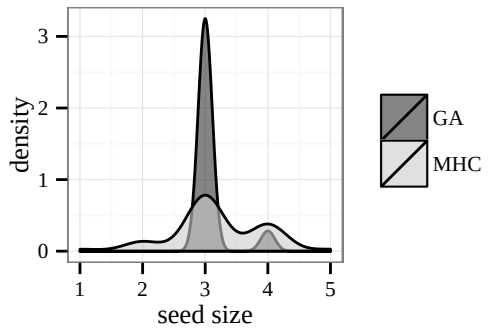


Figure 1: Distribution of the optimal seed size found by the standard genetic algorithm (GA) and the mutation hill climber (MHC). The kernel densities are computed from 50 candidates for each optimizer.

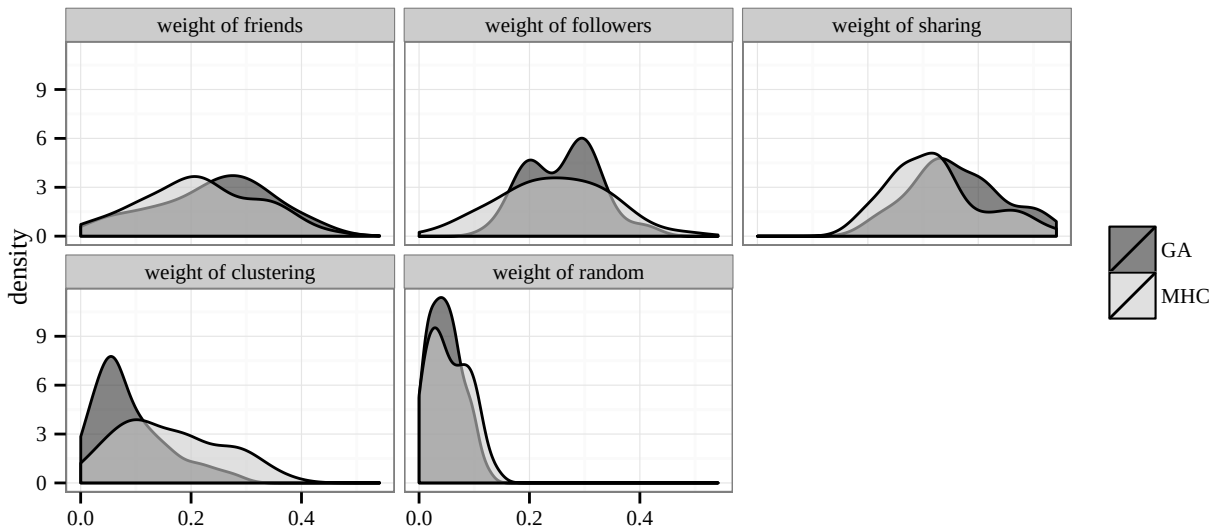


Figure 2: Distribution of the optimal seeding strategy candidates' weights found by the standard genetic algorithm (GA) and the mutation hill climber (MHC). The kernel densities are computed from 50 candidates for each optimizer.

The previous results in optimization of the agent-based viral marketing models ([6, 2]) show there might be a plateau in the objective function (here profit) around the optimal strategy, i.e. that there is some range of parameters that produce very similar outcomes as the optimal seeding strategy. Furthermore, the individual parameters could

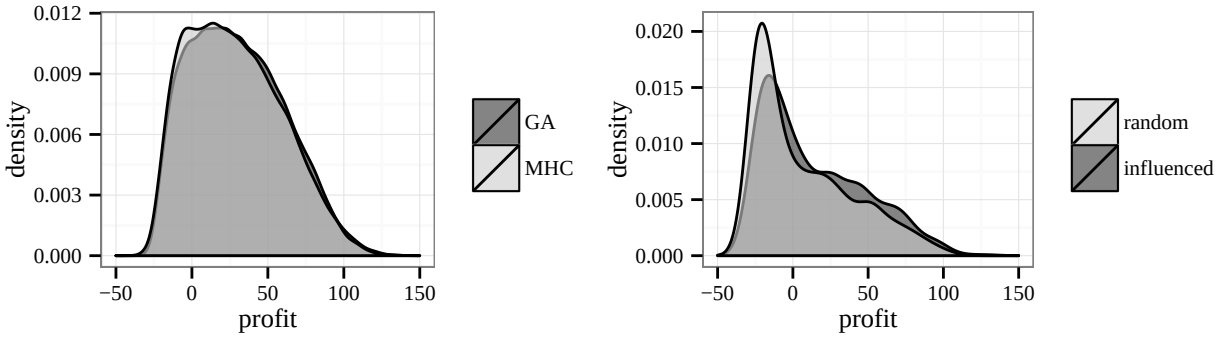


Figure 3: Distribution of profits of the optimal strategy candidates. The kernel densities are computed from 1 000 replication for each candidates; there are 50 candidates for GA and MHC and 1 candidate for each simple strategy.

partially substitute for each other. Thus the difference between the optimal strategy candidates found by the GA and MHC is not a sufficient proof that one routine is inferior to the other, especially since we do not know the truly optimal parameters. Thus it is necessary to compare the profits yielded by the strategies.

For this reason, I simulated the model 1 000 times for each of the 50 optimal strategy candidates and 1 000 times for other four strategies. First, for the mean GA and MHC optimal strategies; their parameters were computed as a mean of the respective parameter of the 50 candidates found by the respective optimizer and are the same as values presented in the first two columns of Table 1 with the exception of the seed size which was rounded to the nearest integer. Second, for two simple strategies created for a comparison: strategy $(S^R, 0, 0, 0, 0, 0, 1)$ that selects the agents into the seed randomly (i.e. $w_6 = 1$) and strategy $(S^F, 0, 0, 1, 0, 0, 0)$ that selects into the seed the agents that can influence most other agents, i.e. the agents with the maximal number of friends and followers together (i.e. $w_3 = 1$). These two simple strategies are natural candidates for a marketer that does not want to optimize—she would either seed the agents randomly or seed the agents with most connections. The seed sizes $S^F = S^R = 3$ were selected because they maximized the expected profit.

The results of these simulations are presented in Table 2 and Figure 3. The average profit of the GA candidates is slightly higher than the average profit of the MHC candidates though their distributions are very similar. The profits of the mean GA and mean HMC optimal strategies are virtually the same. The comparison with the profits of the simple strategies (random and influenced) show that this is not due the plateau in the profit function—both these simple strategies perform much worse than both the GA and MHC candidates.

seed strategy	mean profit	sd profit
GA	30.18	30.87
MHC	28.24	30.69
mean GA	30.27	31.64
mean MHC	30.90	31.08
random	6.86	32.65
influenced	14.21	33.78

Table 2: Profits of the strategy candidates. The statistics for the standard genetic algorithm (GA) and the mutation hill climber (MHC) were computed from 50 000 independent replications (1 000 replications for each of 50 optimal strategy candidates), the statistics of the mean and simple strategies were computed from 1 000 replications.

4 Conclusions

The results show that the standard genetic algorithm yields slightly better optimal strategy candidates—individual candidate strategies yield slightly higher profit and vary less among themselves than the optimal strategy candidates found by the mutation hill climbing. However, the loss of precision is relatively small even with the individual candidates and virtually disappears when strategies constructed as averages of the individual candidates are used. The result seems to be genuine and not caused by a too huge plateau in the objective function. Definitely, both GA

and MHC optimal strategies perform much better than simple strategies that do not require much optimization, e.g. the random strategy or the strategy of including the agents with most connections. Hence the conclusion could be made that the MHC performs almost as well as the standard genetic algorithm but is much faster since it does not need so many replications. Thus it seems that the mutation hill climbing could be safely used as an approximate optimization routine for robustness check and scenario analysis—at least in the case of the agent-based viral marketing studies.

Acknowledgements

The paper has been created as a part of specific research no. MUNI/A/1203/2014 at the Masaryk University.

Computational resources were provided by the MetaCentrum under the program LM2010005 and the CERIT-SC under the program Centre CERIT Scientific Cloud, part of the Operational Program Research and Development for Innovations, Reg. no. CZ.1.05/3.2.00/08.0144.

References

[1] Kvasnička, M.: Viral video diffusion in a fixed social network: an agent-based model, Procedia Economics and Finance, 12(2014), 334342.

[2] Kvasnička, M.: Search for the optimal strategy to spread a viral video: An agent-based model optimized with genetic algorithm, 32nd International Conference Mathematical Methods in Economics Conference Proceedings, (2014), 548553.

[3] R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. http://www.R-project.org/.

[4] Stonedahl, F. BehaviorSearch, computer software, 2010. http://www.behaviorsearch.org/.

[5] Stonedahl, F.: *Genetic Algorithms for the Exploration of Parameter Spaces in Agent-Based Models*, Ph.D. thesis, 2011, available at http://forrest.stonedahl.com/thesis/forrest_stonedahl_thesis.pdf.

[6] Stonedahl, F., Rand, W., and Wilensky, U.: Evolving viral marketing strategies. In: *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, ACM, 2010, 1195–1202.

[7] Watts, D. J., and Dodds, P. S.: Influentials, networks, and public opinion formation. *Journal of consumer research* **34**(4), 2007, 441–458.

[8] Watts, D. J., and Strogatz, S. H.: Collective dynamics of 'small-world' networks. *Nature* 393, 1998, 400–442.

[9] Wilensky, U.: NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999. http://ccl.northwestern.edu/netlogo/.

[10] Wilensky, U.: NetLogo Small Worlds model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 2005. http://ccl.northwestern.edu/netlogo/models/SmallWorlds.

[11] Wilhite, A.: Economic activity on fixed networks. In: *Handbook of Computational Economics* (Tesfatsion, J., and Judd, K. L., eds.), vol. 2. Elsevier, 2006, 1013–1045.

Two phase approach for large public service system design

Marek Kvet [1], Jaroslav Janáček[2]

Abstract. This paper deals with the two phase method of the public service system design suggested for solving large problem instances. The necessity to solve large instances of the weighted *p*-median problem has led to the approximate approach based on a radial formulation, which enables to solve bigger instances in admissible time making use of a universal IP-solver. Nevertheless, huge instances of the *p*-median problem may resist to attempts to solve them by the approach due to big size of the possible service center location set. To comply with this phenomenon, we suggest the two phase method. The first phase consists of the process, when bisection process applied on the D^{max} problem is used to reduce the set of possible service centers. We search for such D^{max} to obtain suitable set of possible locations for the following public service design problem. Then the second phase, which is based on the radial approach, solves the public service system design problem with reduced set of possible locations. Hereby, we study the impact of the parameters of the two-phase decomposition on the solution accuracy and saved computational time.

Keywords: *p*-median problem, D^{max} problem, problem size reduction, radial formulation

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

A public service system designer must often face the dilemma of choice between a near to optimal solution of the service center deployment and a heuristic solution. At first sight, the near to optimal solution seems to be the best choice, because such solution yields the best satisfaction of an average user demand. Nevertheless, the attempt at obtaining optimal or near to optimal solution of a huge service system design problem is dearly paid for by unacceptable long term of IP-solver computing or an effective metaheuristic development. The second choice does not mean that a bad design is accepted. It means that the design must be produced in a given acceptable term. In this case, the designers cannot depend on long term developing some tailor made software system, but they have to use some common readymade IP-solver. IP-solvers are usually equipped with branch and bound technique adapted for solving a common case of integer programming problems and that is why they fail in processing huge instances of special type of location problems. Within this paper, we focus on suggestion of a two phase method, which enables to reduce huge size of a public service system design problem to such extent, which is solvable by a standard method of the weighted *p*-median problem with usage of a commercial IP-solver. The first phase of the suggested method is based on a bisection process applied on some version of the set covering problem, which is easy to solve using a common IP-solver. The set covering problem known also as the D^{max} problem in the location theory selects the minimal set of locations from the huge set of all possible service center locations so that no user is distant more than distance D^{max} from the nearest location from the set of selected locations. The bisection process searches for such D^{max} that the set of selected locations represents a suitable reduced set of possible center locations for the second phase of the suggested method. The second phase makes use of the radial approach [3], [4] to solve middle sized weighted *p*-median problem. The remainder of the paper is organized as follows. The description of scheme of a huge instance of the weighted *p*-median problem size reduction is introduced in Section 2. Section 3 is devoted to brief description of the radial formulation of the *p*-median problem and Section 4 contains the results of numerical experiments, which were used to choose the most effective way of the reduction. The results are summarized in Section 5.

[1] Ing. Marek Kvet, PhD., University of Žilina, Faculty of Security Engineering, 1.mája 32, 010 26 Žilina, Slovak Republic, marek.kvet@fbi.uniza.sk
[2] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, e-mail: jaroslav.janacek@fri.uniza.sk

2 Bisection reduction of the set of possible center locations

The original public service system design problem is defined on a transportation network, where two subsets of nodes are distinguished. There are specified a huge set I of possible service center locations and a set J of users' locations, where b_j denotes the number of users sharing the location j. The network distance between locations i and j is denoted by d_{ij}. The objective of the public service system design problem is to place p service centers at the possible locations from the set I so that the sum of all products of b_j and the distance from the users' location j to the nearest located service center is minimal.

To reduce the huge weighted p-median problem broadly discussed in [5], [6], we take into consideration the sequence $d_0 < d_1 < \ldots < d_u$ of all unique distance values from the matrix $\{d_{ij}\}$ for $i \in I$, $j \in J$. The strategic decision in the problem concerns location of centers at possible center locations from the set I. To model this decision at particular location, we introduce a zero-one variable $y_i \in \{0, 1\}$, which takes the value of 1, if a center should be located at the location i, and it takes the value of 0 otherwise. Using the introduced variables and a given value d_s from the above sequence, we can formulate the D^{max} problem for $D^{max} = d_s$ in the following way. We introduce a zero-one constant a_{ij}^s for triple $[i, j, s]$, where $i \in I$, $j \in J$ and $s \in [0 \ \ldots \ u]$. The constant a_{ij}^s is equal to 1, if the distance d_{ij} between the user location j and the possible center location i is less than or equal to d_s. Then the problem can be formulated as follows.

$$\textit{Minimize} \quad \sum_{i \in I} y_i, \tag{1}$$

$$\textit{Subject to:} \quad \sum_{i \in I} a_{ij}^s y_i \geq 1 \quad j \in J, \tag{2}$$

$$y_i \in \{0,1\} \quad i \in I. \tag{3}$$

The solution of the problem (1)-(3) known also as the set-covering problem minimizes the number of located centers so that the longest distance from a user to the nearest located service center is less than or equal to the value $D^{max} = d_s$. The suggested reduction of the huge set of possible locations I consists of application a solving procedure to the problem (1)-(3). This procedure is further denoted as *CMethod*(s), which produces couple (m_1, I_1), where m_1 is the number of located centers and the set I_1 contains the associated center locations. If general positive weight w_i is associated with each possible center location, the weighted covering problem can be formulated as follows.

$$\textit{Minimize} \quad \sum_{i \in I} w_i y_i, \tag{4}$$

Subject to: (2) *and* (3).

The associated procedure is further denoted as *WCMethod*(s), and it produces triple (w, m_1, I_1), where w is the resulting objective function value of (4), m_1 is the number of located centers and the set I_1 contains the associated center locations.

The bisection method using procedure *CMethod*(s) for reduction of the set I to m elements with possible deviation e can be described by the following steps:

Step 0. Initialize the lower bound ls and upper bound us on the searched subscript as $ls = 0$ and $us = u$.
Step 1. Repeat the following steps until $|m_1 - m| \leq e$ or $us - ls \leq 1$
Step 2. Set $s = us + ls \ div \ 2$ and compute $(m_1, I_1) = CMethod(s)$.
Step 3. If $m_1 > m$ then update $ls = s$, otherwise update $us = s$.

The same scheme can be applied to the procedure *WCMethod*(s) to obtain reduced set I_1. As concerns the weights w_i, they may reflect the number of users covered by the cluster of users' locations, which is defined as $\{j \in J: a_{ij}^s = 1\}$. We suggest the following formula (5) for the weight determination.

$$w_i = \sum_{j \in J} a_{ij}^s b_j d_{ij} \tag{5}$$

3 Radial formulation of the *p*-median problem

The reduced problem of the *p*-center location at places from the set *I* of possible center locations so that the sum of distances from each user to the nearest service center is minimal, can be formulated in several ways [1], [2], [3], [8]. We present here the radial approach, which has proved to be very flexible as concerns its implementation with a commercial IP-solver. The approach makes use of the above defined set *I* of possible center locations, set *J* of users' locations with b_j users sharing the location $j \in J$ and the matrix $\{d_{ij}\}$ of distances between each element from *I* and *J*. The approach also uses the sequence $d_0 < d_1 < \ldots < d_u$ of all distance values from the matrix. The sequence is partitioned into $v+1$ zones by so called dividing points $D_1, D_2 \ldots D_v$ chosen from the sequence, where $0 = d_0 = D_0 < D_1$ and also $D_v < D_{v+1} = d_u$. Possible ways of the dividing points selection have been broadly discussed in [5], [6], [7]. The zone *s* corresponds to the interval $(D_s, D_{s+1}]$. The length of the *s*-th interval is denoted by e_s for $s \in V$, where $V = \{0, 1 \ldots v\}$.

The basic decisions made in the problem concern location of centers at possible center locations from the set *I*. To model these decisions at particular locations, we introduce a zero-one variables $y_i \in \{0, 1\}$ for $i \in I$. The variable y_i takes the value of 1, if a center should be located at the location *i*, and it takes the value of 0 otherwise. In addition, we introduce auxiliary zero-one variables x_{js} for $s \in V$. The variable x_{js} takes the value of 1, if the distance of the user at location *j* from the nearest located center is greater than D_s and it takes the value of 0 otherwise. Then the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + \ldots + e_vx_{jv}$ stands for upper bound of the distance d_{j*} from user location *j* to the nearest located center. If the distance d_{j*} belongs to the interval $(D_s, D_{s+1}]$, then the value of D_{s+1} is the upper estimation of d_{j*} with the maximal possible deviation e_s. Let us introduce a zero-one constant a_{ij}^s for each triple $[i, j, s]$, where $i \in I$, $j \in J$ and $s \in V$. The constant a_{ij}^s is equal to 1, if the distance d_{ij} between the user location *j* and the possible center location *i* is less than or equal to D_s, otherwise $a_{ij}^s = 0$. Then the radial-type weighted covering model can be formulated according to [6], [7] as follows.

$$\textit{Minimize} \quad \sum_{j \in J} \sum_{s \in V} b_j e_s x_{js}, \tag{6}$$

$$\textit{Subject to:} \quad x_{js} + \sum_{i \in I} a_{ij}^s y_i \geq 1 \quad j \in J, \ s \in V, \tag{7}$$

$$\sum_{i \in I} y_i \leq p, \tag{8}$$

$$x_{js} \geq 0 \quad j \in J, \ s \in V, \tag{9}$$

$$y_i \in \{0,1\} \quad i \in I. \tag{10}$$

The objective function (6) gives the upper bound of the sum of weighted distance values. The constraints (7) ensure that the variables x_{js} are allowed to take the value of 0, if there is at least one center located in radius D_s from the user location *j*. The constraint (8) puts a limit *p* on the number of located centers.

4 Computational study

Within this section, we present the results of numerical experiments, which are aimed at ascertainment of the characteristics of the suggested approaches from the viewpoint of computational time and the solution accuracy.

Especially, we focus our effort on examination of the reduced problem from the point of starting state of the following problem processing. We take into consideration that the D^{max} problem (1)-(3) solved in the first phase does not reflect the objective function of the *p*-median problem and so the unsuitably selected possible center locations may burden the following optimization process. We follow the hypothesis that also the rate of reduction influences quality of the resulting design computed by the second phase. To prove the hypothesis, we organized the first series of experiments so that we solved mediate instances of the original problem by exact and approximate (radial) methods and compared their results to the solutions obtained by the suggested approach, where the rate of reduction varied from 20 to 80 percent of the original size.

The experiments were performed on the pool of benchmarks obtained from the real emergency health care system, which was originally designed for each self-governing region of Slovak Republic. The structure of the system is given by the deployment of ambulance vehicles at nodes of the associated road network. The instances are organized so that they correspond to the administrative organization of Slovakia. For each self-governing

region (Bratislava - BA, Banská Bystrica - BB, Košice - KE, Nitra - NR, Prešov - PO, Trenčín - TN, Trnava - TT and Žilina - ZA) all cities and villages with corresponding number of inhabitants b_j were taken into account. The coefficients b_j are rounded to hundreds. The number of possible service center locations $|I|$ is the same as the number of user locations $|J|$ in all solved instances. It means that each community (even the smallest) may represent a possible service center location. The value of parameter p corresponds to the original number of ambulance vehicles deployed in the given self-governing region. The reduction process in the starting series of experiments consisted of solving the problem (1) – (3) and the results are reported in the Table 1.

The table is organized into two sections, where the upper one contains brief description of the instance. The description consists of the number of possible service center locations $|I|$ and the number of centers which are to be located. The second section contains deviations of the obtained solution objective function value from objective function value of the exact solution. The deviation is given in percentage, where the base is the objective function value of the exact solution. The rows denoted as Rate-xx correspond with used reduction rate, which is given in percentage xx. It means that the rate 20% denotes the situation, when the original size $|I|$ of the problem is reduced to $0.8*|I|$ in the first phase. The rate 0 denotes the result obtained by radial approach applied on the original set of possible service center locations.

Region	**BA**	**BB**	**KE**	**NR**	**PO**	**TN**	**TT**	**ZA**
$\|I\|$	87	515	460	350	664	276	249	315
p	9	52	46	35	67	28	25	32
Rate - 0	1.97	0.02	1.00	0.00	0.00	1.13	0.00	0.11
Rate - 20	1.63	3.92	5.58	2.29	8.46	7.96	3.21	0.71
Rate - 30	13.64	3.92	5.58	6.15	8.46	7.96	3.21	0.71
Rate - 40	13.64	6.13	16.91	6.15	14.16	13.01	19.63	13.90
Rate - 50	18.21	37.12	16.91	20.87	14.16	13.01	15.40	23.30
Rate - 60	34.58	38.77	37.70	16.08	14.96	42.77	23.08	23.30
Rate - 70	27.55	29.38	40.72	48.98	33.59	63.29	31.68	40.34
Rate - 80	45.95	49.07	46.10	22.88	50.12	70.50	30.72	41.13

Table 1 Deviations of the obtained solution objective value from objective value of the exact solution in percentage of the exact solution for individual benchmarks and given rates of reduction. In the reduction phase the objective function (1) was used.

As the deviations for higher rates in the previous experiments exceed acceptable size, we repeated the experiments for the reduction process consisted of solving the problem (4), (2), (3) and the results are reported in the Table 2.

Region	**BA**	**BB**	**KE**	**NR**	**PO**	**TN**	**TT**	**ZA**
$\|I\|$	87	515	460	350	664	276	249	315
p	9	52	46	35	67	28	25	32
Rate - 0	1.97	0.02	1.00	0.00	0.00	1.13	0.00	0.11
Rate - 20	1.97	0.19	0.88	0.69	0.49	1.21	0.26	0.40
Rate - 30	0.00	0.56	0.88	1.11	0.49	1.21	0.26	0.64
Rate - 40	0.33	0.56	5.90	2.19	1.98	2.28	0.26	0.64
Rate - 50	0.60	3.30	13.95	2.19	4.70	3.65	0.49	0.85
Rate - 60	0.60	3.35	25.91	1.24	8.32	8.05	2.16	1.31
Rate - 70	27.53	7.42	25.69	6.30	12.17	10.00	4.40	5.33
Rate - 80	41.65	19.20	41.77	10.64	17.50	16.68	14.87	8.86

Table 2 Deviations of the obtained solution objective function value from the objective function value of the exact solution in percentage of the exact solution for individual benchmarks and given rates of reduction. In the reduction phase the objective function (4) was used.

All experiments were performed using the optimization software FICO Xpress 7.7 (64-bit, release 2014). The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with the parameters: 2.0 GHz and 8 GB RAM. The computational time is not reported here, but we note that the computational time for any instance solved by radial approach does not exceed 3 seconds whereas exact solution, where the location-allocation was used, was completed at most in 1 minute.

The average deviations are compared on the following figure, where the right bar corresponds to the results obtained by reduction process consisted of solving the problem (1) – (3) and the left one corresponds to the reduction process consisted of solving the problem (4), (2), (3).

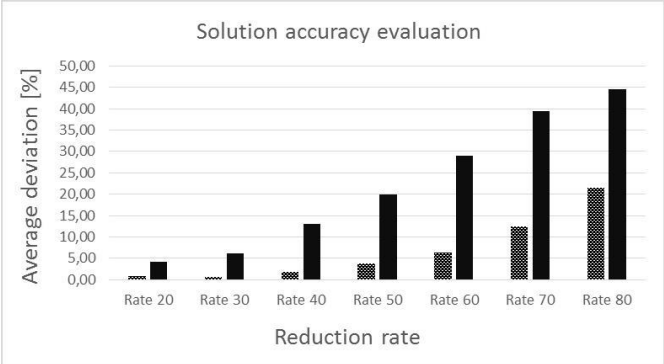


Figure 1 Comparison of two approaches to the reduction process

The first series of experiments proved that only reduction using the model (4), (2), (3) yields acceptable accuracy of the resulting system design. To verify the suggested approach on larger instances of the problem we organized the second series of experiments using the benchmark derived from the whole real emergency health care system of Slovak Republic. The cardinality of original set of possible service center locations $|I|$ is 2916. Due to this size, the attempts on solving the problem exactly failed when the location-allocation model was used. That is why we compare the obtained result to the approximate solution by radial approach. The experiments were performed for the rate of reduction from 20 to 80 percent of the original size and the results are reported in the Table 3. Organization of this table differs from the previous ones in the columns Phase 1 (reduction) and Phase 2 (radial), where computational times of the associated phases in seconds are reported. The column Slovakia contains the above mentioned deviation.

Region	**Slovakia**	**Computational time**	
$\|I\|$	2916	Phase 1	Phase 2
p	292	[s]	[s]
Rate - 0	0	-	34
Rate - 20	0.10	102	17
Rate - 30	1.75	103	9
Rate - 40	1.75	103	9
Rate - 50	4.87	101	14
Rate - 60	8.27	100	6
Rate - 70	10.45	98	5
Rate - 80	29.19	96	2

Table 3 Deviations of the obtained solution objective function value from the objective function value of the approximate radial solution in percentage of the approximate radial solution for given rates of reduction and computational times in seconds.

5 Conclusions

The paper is focused on mastering larger instances of the public service system design problem using commercial IP-solver, which can fail in solving large *p*-median problem. The two phase method has been designed and examined to cope with the drawback. The suggested method is based on exploitation of easily solvable covering problem to reduce the original big *p*-median problem to the same problem of a smaller size. The attention has been paid to the impact of the reduction phase on the deterioration of the objective function value of the resulting service system design. It has been found that the reduction based on simple covering problem caused extremely big loss of accuracy and made the resulting design inefficient. Contrary to this result, weighted covering used for the reduction proved to be much more efficient and made the suggested two phase method usable for designing public service systems on larger networks. We found that the original size of the public service system design problem can be reduced to thirty percent of the original size with expected ten percent objective value deviation from the solution of the non-reduced problem. Thus we can conclude that we present a useful tool for the large public service system design, which can be easily implemented using common commercial optimization software. Further research in this field can be focused on finding some better weights of the weighted covering problem of the reduction phase to improve the two phase method accuracy.

Acknowledgement

This work was supported by the research grants VEGA 1/0518/15 "Resilient rescue systems with uncertain accessibility of service", VEGA 1/0339/13 „Advanced microscopic modeling and complex data sources for designing spatially large public service systems" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks". This paper was also supported by the project: University Science Park of the University of Žilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the European Regional Development Fund. We would like to thank to „Centre of excellence for systems and services of intelligent transport" (ITMS 26220120050) for built up the infrastructure, which was used.

References

[1] Avella, P., Sassano, A., Vassil'ev, I..: *Computational study of large scale p-median problems*. In: Mathematical Programming 109, pp. 89-114, (2007).

[2] Current, J., Daskin, M., Schilling, D.: *Discrete network location models*. In: Drezner Z (ed) et al. Facility location. Applications and theory. Berlin: Springer, pp. 81-118, (2002).

[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, (2011).

[4] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: Proceedings of the 1st International Conference ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, (2008).

[5] Janáček, J., Kvet, M.: *Approximate solving of large p-median problems*. In: ORP3-Operational research peripatetic post-graduate programme: proceedings, Cádiz, Spain, September 13-17, 2011, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, 2011, ISBN 978-84-9828-348-8, pp. 221-225, (2011).

[6] Janáček, J., Kvet, M.: *Relevant Network Distances for Approximate Approach to Large p-Median Problems*. In: Operations research proceedings 2012: selected papers of the international conference on operations research: September 4-7, 2012, Hannover, Germany, ISSN 0721-5924, ISBN 978-3-319-00794-6, pp 123 – 128, (2012).

[7] Janáček, J., Kvet, M.: *Public service system design with disutility relevance estimation*. In: Proceedings of the 31st international conference:"Mathematical Methods in Economics", September 11-13, 2013, Jihlava, Czech Republic, ISBN 978-80-87035-76-4, pp 332 – 337, (2013).

[8] Marianov, V., Serra, D.: *Location problems in the public sector*. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp. 119-150, (2002).

Uncovering Factors Underlying Retail Investor Stock Investment Intentions

Mimi Lounio[1], Mikael Collan[1], Sanna-Katriina Asikainen[1]

Abstract. Well known findings from behavioral finance have challenged the traditional assumptions that individuals make investment decisions solely based on risk and return. In that vein this paper draws from the consumer behavior literature to develop a structural equation model (SEM) to explain retail investor stock investment intentions and the interrelationships between five identified factors underlying these intentions. The model is tested with new Finnish data specifically collected for the purposes of this research.

The results suggest that compatibility is the main driver of retail investor investment intentions and that investment intentions are stronger, when not only economic, but also functional and emotional value is expected from investing. Further findings suggest that expected effort needed, psychological risk, and the risk of being cheated by unethical actors, in addition to financial risk contribute negatively to investment intentions. The study derives valuable implications that can assist in public decision-makers and financial institutions in their attempts to promote household investing.

Keywords: Investment behavior, investment decisions, consumer choice

JEL Classification: G11, M3
AMS Classification: 62H15, 62H99

1 Introduction

Traditionally consumer behavior and investment behavior have been studied separately within the distinctive academic fields of marketing and finance [3], yet scholars have recently pointed out that it may not be sensible setting investment decisions apart from other consumer choices [8, 11, 22]. It has been suggested that theories and techniques from the discipline of marketing could help stimulate and complement investment research [3, 8]. In this vein, the purpose of this paper is to contribute to investment research by presenting a set of explanatory variables with regards to stock investment intentions, formulating a set of hypotheses with regards to said investment intentions, and testing the hypotheses by using a structural equation model (SEM) [7, 2] with a Finnish dataset collected for the purpose. The variables used and the hypotheses tested have been selected, based on consumer behavior literature, see [12, 18].

This paper sheds some much needed light on the question of "what motivates / demotivates" individuals to invest and reveals connections between variables that explain investor behavior. The main contribution of the paper is to the cross-disciplinary area between finance and marketing, where issues of behavioral finance find home and are usable for helping financial institutions in better tailoring their products and marketing efforts to the needs and preferences of potential investors.

The paper goes forward as follows: first, the variables used, the hypotheses, and the structural model are shortly presented; second, the data used is shortly presented, the model is tested, and the results presented; third, the paper closes with conclusions.

2 The variables, the hypotheses, and the model

In classical financial literature the *expected investment value* (or expected value from an investment) is understood strictly in terms of monetary value, however, recent research has proposed that the consumers perceive the value from investing to go beyond the economic value alone. In fact, the perceived value can also include functional, emotional, and symbolic aspects [17]. Similarly, the *expected sacrifice* of stock investing is not limited to the financial risk alone, but may also include monetary costs, time costs, and costs in terms of effort, together with financial, social, source, and psychological risks [9, 14]. Based on this notion that there is more than just the expected return that guides investment intentions we construct a model to explain "investment intentions" with

[1] Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 32, 53851 Lappeenranta, Finland, mimilounio@hotmail.com

five additional variables and test a set of hypotheses on the connections between the variables. The variables, with some supporting references, are shortly presented in Table 1.

Expected sacrifice	Expected sacrifice refers to the consumer's anticipation of the total costs of investment including all components of "cost"
Expected investment value	Expected or perceived value has been considered as a key variable affecting consumer choice; generally buyers base their purchase decision on their evaluation of an offers value, in this case the expected value of the investment alternative, see [21]
Subjective investment knowledge	Subjective knowledge is a mixture of knowledge and self-confidence [16]. Several studies within the field of consumer behavior have recognized that consumers with higher product knowledge use different evaluative strategies and decision processes than consumers with less knowledge, and therefore evaluate products differently [5, 6, 18]
Perceived behavioral control	We define perceived behavioral control as the consumer's perception of the sufficiency of his or her financial assets to invest. Ajzen [1] defines perceived behavioral control as the "self-efficacy in relation to the behavior" .
Compatibility	We define compatibility as the degree to which an investment alternative (stocks) fits the consumer's needs, values, and past experience, see [19].
Investment intention	Investment intention refers to the intention to invest in a given investment alternative. The construct is adapted from the concept of behavioral intention, which is most often defined as the individual's expectancies about the likelihood to act [12], or as an anticipated future behavior [20].

Table 1. The variables used with supporting references

We construct eight hypotheses with regards to the interconnections among and the type of mutual effect between the six variables, the hypotheses are shortly explained and listed in table 2.

H1	Expected sacrifice has a direct and negative effect on expected investment value - the bigger the sacrifice connected to investing in stocks the lower the perceived value from the stocks is
H2	Expected investment value has a direct and positive effect on investment intention - the more value is expected from stocks the more likely the consumer is to invest in stocks
H3	Subjective investment knowledge has a direct and positive effect on expected investment value – the more the consumer knows about investing in stocks the better the expected value from stocks is
H4	Perceived behavioral control has a direct and positive effect on investment intention - more assets to invest cause the likelihood to invest in stocks to be higher
H5	Perceived behavioral control has a direct and positive effect on compatibility – more assets to invest cause investing in stocks to be more compatible with the consumers situation
H6	Expected investment value has a direct and positive effect on compatibility – the higher the perceived value from stocks is the more compatible stocks are to the consumers situation
H7	Expected sacrifice has a direct and negative effect on compatibility – the bigger the sacrifice connected to investing in stocks the lower the fit of investing in stocks is to the consumers situation
H8	Compatibility has a direct and positive effect on investment intention – the better stocks fit to the consumers situation the more likely the consumer is to invest in stocks

Table 2 The formulated hypotheses with short explanations

Based on the above listed variables and to test the hypotheses an initial model visible in Figure 1was constructed.

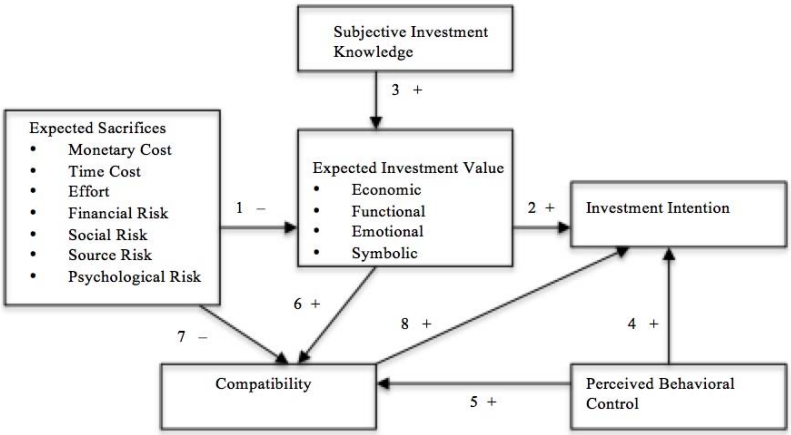


Figure 1 Initial model with the hypotheses

Structural equation modeling (SEM) allows the testing of several causal (direct, indirect, and total) relationships between variables simultaneously [7]. Measurement assessment was first conducted using confirmatory factor analysis (CFA) [7]. This was followed by the 2-step structural model evaluation [2]. The analyses were run with LISREL 8.80 software, with maximum likelihood estimation (MLE). The measurement items were adapted from prior research to the extent possible. The list of used measures for each variable is available from the corresponding author at request.

3 Data, model testing, and results

The data used in the analysis was collected with an email survey from Finnish consumers aged between 45 and 65 years, the Finnish age group with the highest net worth and the highest amount of deposited capital per person. The number of questionnaires sent out was 2400 and the obtained data consists of 244 responses (response rate of 11,3%), of which 154 were fully completed, thus making the final sample n=154 as listwise deletion was the selected approach to missing values.

3.1 Assessment of the measurement model

Due to the small sample size compared to the number of variables, a confirmatory factor analyses (CFA) was performed in three parts. Three measurement models were used: one for the variable "expected investment value", one for the variable "expected investment sacrifices", and one for the remaining four research variables.

The "final" models performed well; all standardized loadings were above 0.6 and t-values above 1.64. The goodness-of-fit statistics for the measurement models are presented in table 3. Most fit-indices were within, or near the desired range, indicating either a good or a mediocre fit [15]. However, the p-values might indicate problems with the models' fit, as chi-square ($\chi 2$) should not be significant (indicates the existence of differences between the model and the data). Chi-square should, however, not be used "too strictly", as it is sensitive to sample size, departs from multivariate normality, and makes the assumption that the model fits the population perfectly [15].

Measurement model	$\chi 2$	d.f.	p	RMSEA	NNFI	CFI	GFI
Expected investment value	65.149	44	0.0208	0.0560	0.984	0.989	0.934
Expected sacrifices	155.339	84	<0.001	0.0745	0.973	0.981	0.887
Remaining variables	52.903	25	0.0009	0.0854	0.959	0.959	0.920

Table 3 Goodness-of-fit statistics

We also tested the reliability and the discriminant validity of the measures by computing a composite reliability (CR) and average variance extracted (AVE) measures for each latent variable. All values for our measures were above identified cut off-levels, CR>0.60 [10] and AVE>0.5 [13]. Moreover, AVE's were greater than the shared variances and thus discriminant validity is supported [13].

3.2 Second order confirmatory factor analyses

As theory suggests that expected investment value and expected sacrifice would have a higher-order, multidimensional structure, we conducted second order confirmatory analyses by using item parcels from the first order CFA as indicators [4].

Independent (exogenous) variable	Dependent (endogenous) variable	Standardized GAMMA (Y)	T-value
Expected Investment Value	Econ	0,828	10.627
	Func	1.00	13.269
	Emot	0.602	7.375
Expected Sacrifices	Effo	0.834	11.003
	Fina	0.796	10.101
	Source	0.709	8.478
	Psych	0.752	9.460

Table 4 Paths in the second order CFA

Based on the examination of the modification indices and standardized residuals of the second order CFA, five factors were dropped one by one from the higher order model. The results suggest that the variable "expected investment value" consists of the factors "economic value" (Econ), "functional value" (Func), and "emotional value" (Emot), whereas the variable "expected sacrifices" includes the factors "effort"(Effo), "financial risk"(Fina), "source risk" (Source), and "psychological risk" (Psych). Table 4 shows the path coefficients (Y) and significance (t-values) between the first and the second order latent variables.

$\chi 2$	d.f.	p	RMSEA	NNFI	CFI	GFI
12.853	13	0.459	0.0000	0.978	0.999	0.977

Table 5 Goodness-of-fit statistics of the second order measurement model

The fit statistics of the second order measurement model, see Table 5, indicate that "expected investment value" and "expected sacrifices" are adequate measures to be used in the final step of the analyses, structural modeling, as all fit-indices are within the desired range.

3.3 Evaluating the structural model

Due to the sample size, item parcels were used for the variables "expected investment value" and "expected sacrifices" (the factor-structures were already tested in the second order factor analysis).

Independent (exogenous) variable	Dependent (endogenous) variable	Standardized GAMMA (Y)	T-value	Hypothesis
Subjective Investment Knowledge	Expected Investment Value	0.524	6.486	H3 accept
Expected Sacrifice	Expected Investment Value	0.027	0.286	H1 reject
Expected Sacrifice	Compatibility	-0.157	-2.428	H7 accept
Perceived Behavioral Control	Compatibility	0.419	6.292	H5 accept
Perceived Behavioral Control	Investment Intention	0.169	1.966	H4 accept
Dependent (endogenous) variable	**Dependent (endogenous) variable**	**Standardized BETA (ß)**	**T-value**	**Hypothesis**
Expected Investment Value	Investment Intention	-0.013	-0.121	H2 reject
Expected Investment Value	Compatibility	0.637	9.548	H6 accept
Compatibility	Investment Intention	0.683	5.120	H8 accept
Independent (exogenous) variable	**Dependent (endogenous) variable**	**Indirect effect**	**T-value**	**Relationship**
Expected Sacrifice	Compatibility	0.032	0.285	insignificant
Expected Sacrifice	Investment Intention	-0.226	-1.729	negative, significant
Subjective Investment Knowledge	Compatibility	0.281	5.646	positive, significant
Subjective Investment Knowledge	Investment Intention	0.228	4.466	positive, significant

Perceived Behavioral Control	Investment Intention	0.296	4.098	positive, significant
Dependent (endogenous) variable	**Dependent (endogenous) variable**	**Indirect effect**	**T-value**	**Relationship**
Expected Investment Value	Investment Intention	0.793	4.519	positive, significant

Table 6 Results from testing the structural model

The results of testing the structural model are shown in Table 6. Table 6 also shows the indirect effects, which are multiplications of the un-standardized parameter estimates of the intervening variables [10].

Structural model	**χ2**	**d.f.**	**p**	**RMSEA**	**NNFI**	**CFI**	**GFI**
Stocks	86.949	38	<0.001	0.0918	0.950	0.965	0.897

Table 7. Goodness-of-fit statistics of the structural model

The goodness-of-fit analysis results are presented above, in Table 7. The fit statistics indicate that the model does not fit the data perfectly as p<005, GFI is slightly below 0.9, and RMSEA is between 0.8-0.10, indicating mediocre fit (0.0918). The other indices are within the desired range.

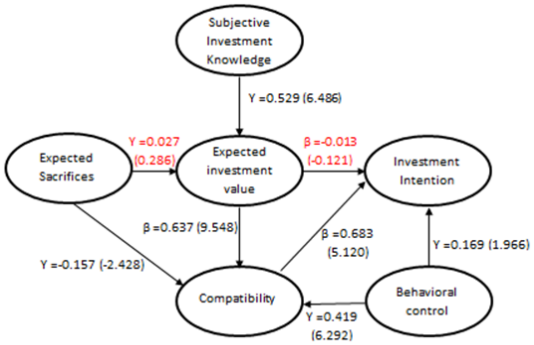


Figure 2 The final structural model with the paths between latent variables (t-values)

Figure 2 shows the final structural model with the remaining paths between the variables.

4 Conclusions

The results of the research indicate that "compatibility" that is, the degree to which stocks fit the consumer's needs, values, and past experience, is the main driver of consumer investment intentions in investing to stocks. Unexpectedly the relationship between the expected value of investing in stocks and investment intention in stocks is not statistically significant. However, expected investment value of stocks affects stock investment intentions indirectly, via compatibility. This is an interesting finding that calls for more research on the subject, as it is a strong sign for actors involved, e.g., in marketing investing in stocks to emphasize the fit of stock investment over the expected value from stocks. These findings also have implications on what kind of consumer products are engineered for investment in stocks.

Moreover, the results suggest that expected investment value from stocks among retail investors does not only consist of economic value, but also of functional and of emotional value. This is also interesting from the point of view of promoting investing in stocks and seems to support findings of behavioral finance in general. A finding that can also be considered surprising was that having more wealth (perceived behavioral control) has only a small impact on increasing the intentions of consumers to invest in stocks. Also this finding merits more research, as it is not clear if this is a Finnish peculiarity, or somehow specific to the data used.

We also found that consumers seem not to avoid investing in stocks only due to financial risk, but also due to effort, psychological risk, and the risk of being cheated by unethical actors (source risk). This is also interesting from the point of view of marketing and financial engineering.

Our results also suggest that subjective investment knowledge has a positive effect on expected investment value. Generally the result supports the findings from within the field of consumer behavior, which have indicated that consumers with differing levels of product knowledge use different evaluative strategies and therefore assess products differently.

Overall the results strengthen the idea that investment decisions are not only a matter of risk and return, but rather a combination of several factors with complex relationships. Improving the understanding of these relationships will help both financial institutions as well as public authorities in their attempts to promote household investment in stocks.

References

[1] Ajzen, I.: Consumer attitudes and behavior. In: *Handbook of Consumer Psychology (*Haugtvedt, C. P.*, et al.*, eds.). Lawrence Erlbaum Associates, New York, USA, 2008, 525- 48

[2] Anderson, J. C. and Gerbing, D. W.: Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin* **103** (1988), 411-23.

[3] Aspara, J. and Tikkanen, H.: Consumers' stock preferences beyond expected financial returns: The influence of product and brand evaluations. *International Journal of Bank Marketing* **28** (2010), 193-221.

[4] Bandalos, D. L. and Finney, S. J.: Item parceling issues in structural equation modeling. In: *New developments and techniques in structural equation modeling (*Marcoulides, G. A. and Schumacker, R. E., eds.). Lawrence Earlbaum Associates Inc., New Jersey, USA, 2001,

[5] Bettman, J. R. and Park, C. W.: Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: a protocol analysis. *Journal of Consumer Research* **7** (1980), 234-48.

[6] Biswas, A. and Sherrell, D. L.: The influence of product knowledge and brand name on internal price standards and confidence. *Psychology and Marketing* **10** (1993), 31-46.

[7] Byrne, B. M.: *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Lawrence Erlbaum Associates. Mahwah, NJ, USA, 1998.

[8] Clark-Murphy, M. and Soutar, G. N.: What individual investors value: Some Australian evidence. *Journal of Economic Psychology* **25** (2004), 539-55.

[9] Diacon, S. and Ennew, C.: Consumer perceptions of financial risk. *The Geneva Papers in Risk and Insurance: Issues & Practice* **26** (2001), 389-409.

[10] Diamantopoulos, A. and Siguaw, J. A.: *Introducing Lisrel*. Sage Publications. London, 2000.

[11] East, R.: Investment decisions and the theory of planned behavior. *Journal of Economic Psychology* **14** (1993), 337-75.

[12] Fishbein, M. and Ajzen, I.: *Belief, Attitude and Behavior: an Introduction to Theory and Reseach*. Addison- Wesley. Reading, 1975.

[13] Fornell, C. and Larcker, D. F.: Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research* **18** (1981), 39-50.

[14] Huber, F., Herrmann, A. and Morgan, R. E.: Measuring customer value and satisfaction in services transactions, scale development, validation and cross-cultural comparison. *International Journal of Consumer Studies* **31** (2001), 554–64.

[15] Jöreskog, K. G. and Sörbom, D.: *LISREL 8 Users reference guide*. Scientific Software International, Inc. Lincolnwood, IL., USA, 1996.

[16] Park, C. W. and Lessig, V. P.: Familiarity and its impact on consumer decision biases and heuristics. *Journal of Consumer Research* **8** (1981), 223-30.

[17] Puustinen, P., Maas, P. and Karjaluoto, H.: Development and validation of the perceived investment value (PIV) *Journal of Economic Psychology* **36** (2013), 41-54.

[18] Rao, A. and Monroe, K. B.: The moderating effect of prior knowledge in cue utilization in product evaluation. *Journal of Consumer Research* **15** (1988), 253-63.

[19] Rogers, E. M.: *Diffusion of Innovations*. The Free Press. New York, 2003.

[20] Swan, J. E. and Trawick, J. F.: Disconfirmation of expectations and satisfaction with a retail service. *Journal of retailing* **57** (1981), 49-66.

[21] Zeithhaml, V. A.: Consumer perceptsions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of marketing* **52** (1988), 2-22.

[22] Zhou, R. and Pham, M. T.: Promotion and prevention across mental accounts: when financial products dictate consumers' investment goals. *Journal of Consumer Research* **31** (2004), 125-35.

Variational Formulation of Option Pricing Problem a Platform for Finite Element Method in Finance

Ladislav Lukáš[1]

Abstract. The paper deals with variational formulation of option pricing problems. We start from the well-known case, the Black-Scholes model for a put option with strike price and maturity given, which assumes the underlying asset to follow a geometric Brownian motion. This problem provides a reasonable basic framework to follow basic steps of derivation of variational formulation of option pricing problem. In general, variational formulation consists of finding a continuous function defined on the time interval with the values in a properly defined functional space. Finite element method applied to option pricing problem in finance yields usually a system of ordinary differential equations if discretization process applies to space domain of underlying asset only. Finite element linear basis functions span a finite dimensional space to approximate solution of American option pricing problem.

Keywords: American option pricing, variational inequality, finite element method.

JEL Classification: C63, G13
AMS Classification: 91G80

1 Introduction

There is well-known, options are derivative contracts between a buyer, or holder, and a seller, writer or issuer, of the contract where future payoffs to the buyer are determined by the price of another security, called also an underlying asset, such as a common stock. An option is a contract that gives the holder the right but not the obligation to buy or sell a certain financial asset by a certain date T, for a given price K. We know, there are two main types of options – i) a call option, which is a option to buy security, and ii) a put option, which gives the right to sell it. So, at least five following features must be specified within the option contract precisely:

- the type of the option, i.e. call or put;
- the underlying asset, usually common stocks, shares, bonds, commodities, stock indexes, currencies, etc.;
- the amount of the underlying asset;
- the expiration date, or maturity date T;
- the exercise price, or strike price K.

In **Fig. 1**, typical pay-off functions $h(S; K)$ are plotted for call and put options, respectively, assuming zero transaction costs..

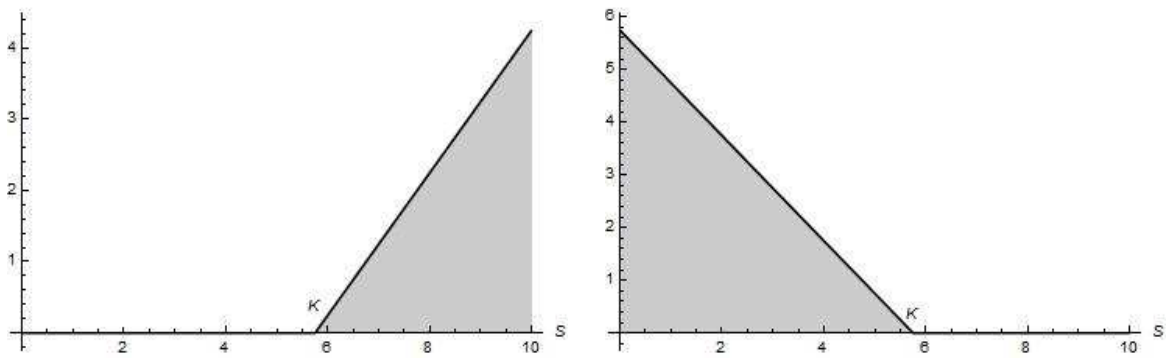


Figure 1 Pay-off functions: $h(S; K) = \max(S - K, 0)$ for call (left), $h(S; K) = \max(K - S, 0)$ for put (right)

The option pricing theory is available in many famous textbook, e.g. [5]. Since the value of the option depends on the security value at the exercise day, which is a priori unknown, the underlying asset price is to be modeled by a stochastic process. If the underlying asset price follows a geometric Brownian motion, then the

[1] University of West Bohemia/Faculty of Economics, Dept. of Economics and Quantitative Methods, Husova 11, 306 14 Pilsen, lukasl@kem.zcu.cz.

value of the European option is described by a deterministic backward partial differential equation (PDE) of the second order and of parabolic type, in particular.

The classical option pricing theory is based on a continuous model assuming one risky asset A and one risk-less asset A^0 both undergoing time evolutions satisfying the following equations. The price S^0 of risk-less asset A^0 is governed by ordinary differential equation

$$\mathrm{d}S_t^0 = rS_t^0\mathrm{d}t, \quad t \geq 0, \tag{1}$$

where $r > 0$ is the risk-less interest rate, and subscript t denotes the time dependence, in general.

Whereas, the price S of risky asset A is governed by stochastic differential equation

$$\mathrm{d}S_t = S_t(\mu\mathrm{d}t + \sigma\mathrm{d}B_t), \quad t \geq 0, \tag{2}$$

where μ and $\sigma > 0$ is drift rate and volatility of price S, respectively, adopting S_t to denote precisely a spot price of asset A at time instant t, as usual, and $\{B_t\}_{t\geq 0}$ is a standard Brownian motion.

Assuming complete markets we know, the pay-off of every option can be duplicated by properly composed portfolio consisting of an investment in the assets A and A^0. Considering the simplest case, i.e. European options with pay-off functions $h(S_T; K)$ at the maturity T, we recall the basic steps.

Let $V = V(t, S_t)$ denote the value of European option at t depending upon S_t, which is called the pricing function, too. At $t = T$, we have $V(T, S_T) = h(S_T; K)$, and for $0 < t < T$, assuming the function V is smooth, dropping subscript t for transparent formula reading as usual, and using (2), we get by Ito's formula

$$\begin{aligned} \mathrm{d}V &= \frac{\partial V}{\partial t}\mathrm{d}t + \frac{\partial V}{\partial S}\mathrm{d}S + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\mathrm{d}t \\ &= \left(\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)\mathrm{d}t + \sigma S\frac{\partial V}{\partial S}\mathrm{d}B. \end{aligned} \tag{3}$$

Now, considering a portfolio consisting of the option and of a short position having Δ shares in the asset A, we write the portfolio value $\Pi = V - \Delta S$, and its self-financed dynamics takes consequently the following form

$$\begin{aligned} \mathrm{d}\Pi &= \mathrm{d}V - \Delta\mathrm{d}S \\ &= \left(\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} - \Delta\mu S\right)\mathrm{d}t + \sigma S\left(\frac{\partial V}{\partial S} - \Delta\right)\mathrm{d}B. \end{aligned} \tag{4}$$

We know quite well that such typical portfolio composition is called Δ-hedging principle, which provides the random component in expression (4) may be eliminated by choosing $\Delta = \partial V/\partial S$, thus providing a portfolio deterministic increment only

$$\mathrm{d}\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2}\right)\mathrm{d}t. \tag{5}$$

Finally, by arbitrage-free arguments in complete markets, the rate of return of that risk-less portfolio $\Pi = V - (\partial V/\partial S)\, S$ must be equal to the rate of return given by (1), which gives

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} = r\Pi, \tag{6}$$

thus leading to the celebrated Black-Scholes PDE

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} - rV = 0, \quad \text{for } (t, S) \in [0,T) \times (0, +\infty)\,. \tag{7}$$

The terminal condition for pricing function V is defined is following

$$V(T, S_T) = h(S_T; K). \tag{8}$$

The problem, to find a solution $V(t, S_t)$ satisfying (7) and (8) is called *terminal value problem* (note: final value problem does appear as an alternative term for TVP often) for backward in time parabolic PDE. In general, suitable qualitative conditions are also required upon functions μ, σ, and r within technical more details.

However concluding, in case μ, σ, and r are given constants, this final value problem is solvable analytically depending upon the option type, i.e. either call or put, thus providing the famous Black-Scholes formulas for European call and put options, respectively.

2 Variational formulation of option pricing problem

In [4], there is given a short overview of applications of finite element method (FEM) in option pricing. Further, [1] and [3] bring more details as for the variational formulations of option pricing problems as well as for numerical details of finite difference method and FEM, in particular.

The subject of variational inequalities has its origin in the calculus of variations associated with minimization of energy functional, e.g. total potential energy of elastic body under given loads being supported as to eliminate any rigid body motion. However, in the case of unilateral support, the corresponding variational formulation for determination of stress state of the elastic body converts into variational inequality problem.

Using numerical methods in financial engineering and in option pricing in particular, is rapidly increasing. Except traditional finite difference methods and Monte-Carlo simulation techniques, the finite element method becomes popular, too. The main advantages of FEM in option pricing, as given in [1], are large flexibility in construction of approximation functions and fulfilling given boundary conditions on one side, and support by a strong theory on the other one, as it is based upon variational formulation of the problem. This one enables also deeper mathematical analysis as for proving existence and uniqueness of solution, as for building error estimations and other qualitative properties of the solution.

The theory of variational formulations of parabolic PDEs is well known, and in [1] there are given ones. The framework is particularly useful when classical, or strong, solution do not exist either because of some singularity in the data or the domain boundary or the coefficients or nonlinearity. Even when the boundary value problem for PDE has a classical solution, the variational theory is interesting for several reasons:

- it provides global estimates of solution;
- it has strong connections with FEM;
- it is the most natural way to study obstacle problems, e.g. American option pricing.

There is well-known, variational formulations of parabolic PDE rely on suitable functional spaces, known as Sobolev spaces. Following [1], we are going to mention the Sobolev space useful particularly for solving the final value problem (7) and (8), which is posed with single asset price variable *S*.

Let $L^2(\mathbb{R}_+)$ denotes the Hilbert space of square integrable functions on $\mathbb{R}_+$ endowed with the norm $\|.\|$ and the inner product (.,.) as usual

$$\| v \| = \left(\int_{\mathbb{R}_+} v(S)^2 \mathrm{d}S\right)^{\frac{1}{2}}, \quad (v,w) = \int_{\mathbb{R}_+} v(S)w(S)\mathrm{d}S\,. \tag{9}$$

Functional space which is well-suited to option pricing problem (7) and (8) is

$$W = \left\{ v \in L^2(\mathbb{R}_+) \mid S\frac{\mathrm{d}v}{\mathrm{d}S} \in L^2(\mathbb{R}_+) \right\}, \tag{10}$$

where the derivative d*v*/d*S* has to be understood in the sense of distributions on $\mathbb{R}_+$. The space *W* being endowed with the norm $\| v \|_W = \sqrt{(v,v)_W}$ is a Hilbert space, as stated in [1], provided a natural scalar product in space *W* has the following form

$$(v,w)_W = (v,w) + \left(S\frac{\mathrm{d}v}{\mathrm{d}S}, S\frac{\mathrm{d}w}{\mathrm{d}S}\right). \tag{11}$$

2.1 European option pricing

The weak formulation of problem for the Black-Scholes PDE follows. Since we try to concentrate ourselves on the main steps in theory, not being burden with technicalities here, we focus upon well-known and simple example, still keeping line with [1].

For illustration, consider a vanilla put option, i.e. such one where the underlying assert *A* doesn't pay any dividend, having maturity *T* and the pay-off function $h_P(T, S_T)$. Make time substitution $\tau = T - t$, in order to convert the terminal value problem (7) and (8) into an initial value problem (IVP) , then one may write

$$\frac{\partial u}{\partial \tau} - rS\frac{\partial u}{\partial S} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} + ru = 0, \quad \text{for } (\tau, S) \in [0,T) \times \mathbb{R}_+, \quad u(0,S) = u_0(S), \text{ for } S \in \mathbb{R}_+, \tag{12}$$

where we just adopted to denote $u_0(S) = h_P(T, S_T)|_{\tau = T - T = 0}$.

In general, following steps are very usual when deriving weak formulation of problems within the framework of PDE theory. Let us simply multiply (12) by a smooth real valued function *w* defined on $\mathbb{R}_+$, and next, inte-

grate the product integrand being constructed in the variable S on $\mathbb{R}_+$. Further, assuming the integration by parts is allowed, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\int_{\mathbb{R}_+} u(\tau,S)w(S)\mathrm{d}S\right) + a_\tau(u,w) = 0\,, \tag{13}$$

where the bilinear form a_τ is defined by following expression

$$a_\tau(v,w) = \int_{\mathbb{R}_+}\left(\frac{1}{2}\sigma^2 S^2\frac{\partial v}{\partial S}\frac{\partial w}{\partial S} + rvw\right)\mathrm{d}S + \int_{\mathbb{R}_+}\left(-r+\sigma^2+\sigma S\frac{\partial\sigma}{\partial S}\right)S\frac{\partial v}{\partial S}w\,\mathrm{d}S\,. \tag{14}$$

Now, precise theory needs to bring some technical assumptions regarding:

- function $\sigma(\tau, S)$ with its values to be squeezed within a layer defined by its bottom level and upper one given by a couple of positive constants for all $\tau \in [0, T]$, and $S \in \mathbb{R}_+$;
- all non-negative values of absolute value function $|S\,\partial\sigma(\tau, S)/\partial S|$ to be bounded from up with another positive constant, in similar way as well.

These assumptions maintain desired regularity of integrands in (14), which further lead to continuity and another desired properties of bilinear form $a_\tau(v, w)$ on W, which enable us to formulate weak form of the IVP (12) precisely, by recalling Lemma 1.1 from [1], as follows

$$\begin{aligned}&\text{Find } u \in C^0([0,T]; L^2(\mathbb{R}_+)) \cap L^2(0,T; W) \text{ with } \frac{\partial u}{\partial t} \in L^2(0,T; W'), \text{ and } u_{|t=0} = u_0 \text{ in } \mathbb{R}_+,\\ &\quad\text{and for a.e. } t \in (0,T), \text{ holds } \left(\frac{\partial u}{\partial t}(t)\,, w\right) + a_t(u(t),w) = 0,\ \forall w \in W.\end{aligned} \tag{15}$$

Here, W' denotes a dual space to W, and t replaces τ, as it is usual to do in the literature very often. In variation formulation framework, the space W is called space of test functions w, which fulfill some regularity conditions and boundary conditions in particular forms, as well.

As given in [1], the solution of weak formulation (15) of IVP (12) exists and is unique, and further, it yields also a good estimation of the solution. Another importance of variational formulation (15) is that it provides a consistent platform for numerical solution of option pricing problem using finite element method, in general.

However, it is also important to note, that we restrict ourselves with vanilla put option. So, as strictly speaking, for the call option pricing care must be hold, since one has either to use put-call parity for calculation, or reformulate the problem in similar way but using different Sobolev space properly adopted to handle a pay-off function of call option, which is not a function of $L^2(\mathbb{R}_+)$, in general.

It is well-known, that put-call parity for European vanilla options has the following form

$$C(t,S) = S - K\exp\left(-\int_0^t r(\tau)\mathrm{d}\tau\right) + u(t,S)\,, \tag{16}$$

where $u(t, S)$ is pricing function of a vanilla put option with strike price K, being issued as solution of variational problem (15), and $C(t, S)$ is a solution of IBV (12) with initial condition $C(0, S) = \max(S - K, 0)$, which is identified as usual pay-off function of vanilla call option just being transformed from common terminal form into initial one.

2.2 American option pricing

One of the well-known applications of variational inequalities in finance is valuation of American options. In order to keep reasoning parallel to the previous section, we consider the plain vanilla option on complete market, i.e. without arbitrage opportunity, again. However, we concern with the American option style, which means that early exercise of option is possible, so the option holder need not to wait for selling the underlying asset till the expiration date. Again, let assume the maturity, or expiration date T, strike price K, and the pay-off function of American put option to be $\max(K - S_t, 0)$, which gives different value than $\max(K - S_T, 0)$ for European put option with the same T and K, in general.

Now, we will follow [4], [5], and [8], mainly. In fact, for American put option, when stock price, i.e. underlying asset price, falls below a certain point, one should exercise the option immediately. For example, if at time t, $S_t < K(1 - \exp(-r(T-t))$, and keeping in mind that the pay-off at the option expiration date will never exceed K in any case, then the option holder can get the immediate gain in amount

$$K - S_t > K - K(1 - \exp(-r(T-t)) = K\exp\big(-r(T-t)\big), \tag{17}$$

and by depositing the gain in a saving account, the total pay-off will exceed K at $t = T$. Therefore one can conclude that there exists a point at t which is known as an optimal exercise point.

We know that option price is never lower than the pay-off function, i.e. inequality (18) holds, since otherwise there would exist an arbitrage opportunity.

$$V(t, S_t) \geq \max(K - S_t, 0), \ \forall\, t \in [0, T]. \tag{18}$$

Concluding, for an American put option with expiration date $t = T$, there exist two sub-regions, say Σ_1 and Σ_2, respectively, covering $[0,T] \times \mathbb{R}_+$, i.e. $\Sigma_1 \cup \Sigma_2 = [0,T] \times \mathbb{R}_+$, which share a common piece of their boundaries, say $\Sigma_1 \cap \Sigma_2 = \Gamma$, which is called optimal exercise boundary. It is defined by mapping Γ: $[0, T] \rightarrow \Gamma(t)$ with range $[0, S_t]$, which is not given a priori and has to be determined together with option pricing function $V(t, S_t)$ by solving the corresponding option pricing problem. Providing (18) holds, these sub-regions are defined as follows

$$V(t, S_t) > \max(K - S_t, 0), \ \forall\, (t, S_t) \in \Sigma_1, \quad \Sigma_1 = \{(t, S_t) | \Gamma(t) \leq S_t < +\infty, \ \forall\, t \in [0, T]\}, \tag{18a}$$

$$V(t, S_t) = \max(K - S_t, 0), \ \forall\, (t, S_t) \in \Sigma_2, \quad \Sigma_2 = \{(t, S_t) | 0 \leq S_t \leq \Gamma(t), \ \forall\, t \in [0, T]\}, \ \Gamma(t) < K, \ \forall\, t \in [0, T), \tag{18b}$$

where Σ_1 is called the continuation sub-region, since the pay-off is zero when $S_t \geq K$, and the holder should continue to keep the option, whereas Σ_2 is called stopping sub-region.

Now, following the same reasoning procedure based on construction of self-financing portfolio, Δ-hedging principle, and Ito formula, one can infer that $V(t, S_t)$ satisfies the Black-Scholes PDE in the sub-region Σ_1

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0, \quad \text{for } (t, S) \in \Sigma_1, \ \text{ assuming } V(t, S_t) \in C^{1,2}(\Sigma_1). \tag{19}$$

On the optimal exercise boundary Γ,

$$V(t, S_t) = \max(K - S_t, 0), \quad \frac{\partial V(t,S)}{\partial S} = -1\,, \quad \forall\, t \in [0, T). \tag{20}$$

Further, the terminal condition at $t = T$, and the condition when $S_t \rightarrow +\infty$, are both the same as for European put option

$$V(T, S_T) = \max(K - S_T, 0), \qquad \lim V(t, S_t) = 0, \ S_t \rightarrow +\infty, \quad \forall\, t \in [0, T). \tag{21}$$

The problem: find $V(t, S_t)$ and $\Gamma(t)$ satisfying (19)-(21) in Σ_1, where $\Gamma \subset \partial(\Sigma_1)$, is called free boundary value problem for American put option pricing.

In the stopping sub-region Σ_2 assuming (18b), the function $V(t, S_t)$ is coincident with pay-off function thus enabling direct computation of $\mathcal{L}\, V(t, S_t)$

$$V(t, S_t) = \max(K - S_t, 0), \quad \mathcal{L}\, V(t, S_t) = \mathcal{L}\,(K - S_t) = -rK < 0, \quad \forall\, (t, S_t) \in \Sigma_2, \tag{22}$$

where $\mathcal{L}$ denotes the Black-Scholes operator defined by its application on any function $G(t, S) \in C^{1,2}(\Sigma)$, as usual

$$\mathcal{L}\, G(t, S) = \frac{\partial G}{\partial t} + rS\frac{\partial G}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 G}{\partial S^2} - rG = 0, \quad \text{for } (t, S) \in \Sigma = [0,T) \times \mathbb{R}_+. \tag{23}$$

Combining (18a) and (19), with (18b) and (22), and adding (21), we get American put option pricing problem in the region $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Gamma = [0,T) \times \mathbb{R}_+$ formulated in form of variational inequality in strong sense, see [8], p.126,

$$\min\{-\mathcal{L}\, V(t, S_t), \ V(t, S_t) - \max(K - S_t, 0)\} = 0, \ \forall\, (t, S_t) \in \Sigma\,, \ \&\& \ (21) \tag{24}$$

Finally, the problem formulation in variational inequality form in weak sense is given in [1], and [7], with all technical details. Here, we just emphasize the main difference of variational inequality formulation against the variational formulation of European put option pricing problem (15). In principle, instead of variational equation $(\frac{\partial u}{\partial t}(t), w) + a_t(u(t), w) = 0, \ \forall w \in W$, the variational inequality (25) appears therein

$$(\frac{\partial u}{\partial t}(t), w - u(t)) + a_t\big(u(t), w - u(t)\big) \geq 0, \ \forall w \in W. \tag{25}$$

3 Finite element method - basics

Two types of discretization methods are used for numerical solution of continuum problems defined by PDE's in order to reduce infinite dimensional problems to finite-dimensional ones. FEM (Finite Element Method) concerns with construction of suitable finite-dimensional space of approximation functions, e.g. [1], [2], [4], [5], and [10], while FDM (Finite Difference Method) uses difference schemes to approximate differential operators.

First, we have to shrink the range $\mathbb{R}_+$ of possible values S_t to $[0,\bar{S}]$, $\bar{S} > 0$, finite, and to set boundary condition $V(t, \bar{S}) = 0$, $\forall\, t \in [0, T)$, which corresponds with (21).

Semi-discretization using FEM means to discretize range $[0,\bar{S}]$, only. Following [5], we divide $[0,\bar{S}]$ into m+1 subintervals, each having length $h = \bar{S}/(m+1)$, providing discrete set of nodes $s_i = ih$, $i = 0, 1, .., m+1$. The space of approximation functions W_h, which serves also as a space of test functions in variational problem formulation, is constructed from finite element basis functions. Let $\phi(s) = (s+1)\mathrm{ind}(-1 \le s \le 0)+(1 - s)\ \mathrm{ind}(0 < s \le 1)$, where ind(.) is a set indicator function. Define piecewise linear finite element basis functions: $\phi_{h,i}(s) = \phi((s - s_i)/h)$. The space W_h is spanned by basis functions $\{\phi_{h,1}(s), \ldots, \phi_{h,m}(s)\}$. An alternative approach promotes finite element shape functions as being more intuitive ones. These are defined by polynomials defined on elements by proper interpolation conditions. We seek a finite element approximation $u_h(t, s) \in W_h$ to the solution (25) in form

$$u_h(t,s) = \sum_{i=1}^{m} u_i(t)\phi_{h,i}(s), \quad u_i(t) \ge 0, \ 1 \le i \le m, \quad t \in [0,T). \tag{26}$$

Substituting (26) and computing integrals of type (14) in (25), we obtain (27) to solve

$$(\boldsymbol{w} - \boldsymbol{u}(t))^T[\mathbf{M}\,\dot{\boldsymbol{u}}(t) + \mathbf{A}\,\boldsymbol{u}(t) + \mathbf{F}] \ge 0, \quad \forall\, \boldsymbol{w} \ge 0. \tag{27}$$

4 Conclusions

- Framework of variational formulation of option pricing problem was briefly discussed.
- Formulation of American option pricing problem leads to variational inequality. The optimal early exercise boundary must be determined together with pricing function by solving variational inequality numerically.
- Basics of FEM are discussed for semi-discretization procedure briefly.
- Near future research will be focused on utilization of higher order polynomial FEM shape functions and triangular elements with curved boundary for approximation of early exercise boundary.

Acknowledgements

The research project was supported by the grant no. 15-20405S of the Grant Agency, Prague, Czech Republic, and the grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Achdou, Y., Pironneau, O.: *Computational methods for option pricing*. Frontiers in Applied Mathematics, vol. 30, SIAM, Philadelphia, 2005.
[2] Berridge, S.J., Schumacher, J.M.: An irregular grid approach for pricing high-dimensional American options. *J. of Computational and Applied Mathematics* **222** (2008), 94–111.
[3] Duffy, D.J.: *Finite Difference Methods in Financial Engineering – A Partial Differential Equation Approach*. John Wiley & Sons Inc., Chichester, England, 2006.
[4] Feng, L., Kovalov, P., Linetsky, V., Marcozzi, M.: Chapter 7, Variational Methods in Derivatives Pricing. *Handbooksin OR & MS*. Vol. 15 (Birge, R. and Linetsky, V. (Eds.)), Elsevier, New York, 2008.
[5] Feng, L., Linetsky, V., Morales, J.L., Nocedal, J.: On the Solution of Complementarity Problems Arising in American Options Pricing. *Optimization Methods and Software* **26 (4-5)** (2011), 813–825.
[6] Fries, Ch.: *Mathematical finance – theory, modeling, implementation*. John Wiley & Sons, Hoboken, New Jersey, 2007.
[7] Jaillet, P., Lamberton, D., Lapeyre, B.: Variational inequalities and the pricing American options. *Acta Applicandae Mathematicae* **21** (1990), 263–289.
[8] Jiang, L.: *Mathematical Modeling and Methods of Option Pricing*. World Scientific Publ., Singapore, 2005.
[9] Karatzas, I., Shreve, S.E.: *Methods of mathematical finance*. Springer-Verlag, New York, 1998.
[10] Memon, S.: Finite element method for American option pricing: a penalty approach. *Int. J. of Numer. Analysis and Modeling, Series B* **3** (2012), 345–370.
[11] Shiryaev, A.N.: *Essentials of Stochastic Finance – Facts, Models, Theory*. World Scientific Publ. Co., Singapore, 2008.
[12] Stojanovic, S.: *Computational Financial Mathematics using Mathematica*. Birkhaeuser, New York, 2003.

Estimation and Bootstrapping of the Bradley-Terry model: Preliminary results on forecasting Tennis Match Outcomes

Štefan Lyócsa[1], Tomáš Výrost[2]

Abstract. We apply the minorization-maximization (MM) algorithm of Hunter (2004) to estimate the Bradley-Terry model (Bradley & Terry, 1952) for predicting tennis match outcomes. The predictive ability of the Bradley-Terry model was tested using a data set of 24 238 tennis matches played from December 31st, 2002 to April 19th, 2014 as part of the Association of Tennis Professional's World tour. The MM algorithm of Hunter (2004) leads to short computation times, which allows us to design simple bootstrap-based betting system.

Keywords: forecasting, tennis, minorization-maximization algorithm, Bradley-Terry model.

JEL Classification: Z23, G13, G14
AMS Classification: 90-08

1 Introduction

In this paper we utilize the McHale & Morton's (2011) version of the Bradley-Terry (BT) type model (Bradley & Terry, 1952) to test for the profitability of several betting systems. Estimating the BT models using the MM algorithm of Hunter (2004) leads to short computation times, which allows us to design simple bootstrap-based betting systems.

The BT model was already used for the purpose of forecasting tennis match outcomes by Glickman (1999) and also by McHale & Morton (2011). Our forecasting exercise is more extensive; in terms of the size of the sample, and different, new type of betting systems. Contrary to some betting systems presented in earlier studies of tennis-betting markets, the bootstrap-based betting systems suggest betting for a relatively large portion of all matches, which makes it a useful tool for bookmakers. The use of a large data set allows us to document that "price fishing" (searching for the best available price, i.e. odds) is extremely important for making profitable bets. This is of particular interest to the betting market, as with the introduction of online betting, the information costs have decreased for bettors, enabling easier exploitation of more advantageous prices.

2 Data

Our initial sample consisted of tennis matches played on the ATP tour from January 3, 2000 to April 19, 2014. The data necessary for the purpose of this study (including market odds from several bookmakers) was retrieved from www.tennis-data.co.uk, a source used in a number of studies covering the tennis betting market, e.g. Forrest & McHale (2007), Scheibehenne & Bröder (2007). Due to their similarity and the relatively low number of matches played on carpet, the hard-court and carpet surfaces were considered as one type of a surface.

The out-of-sample data set consists of 30 013 matches played from December 31, 2002 to April 19, 2014. However, due to some restrictions put on the data set required by the BT model (explained below) and/or availability of the market odds, we are only predicting the outcome for 24 238 matches on all surfaces.

3 Methodology

3.1 Bradley-Terry Model

The Bradley & Terry (1952) model for player abilities is used in situations when $P \in \mathbb{N}$ individuals (players) are repeatedly compared in pairs. In a Bradley-Terry model, each player $i \in \{1, 2, \ldots, P\}$ is assigned a non-negative number α_i, called "ability". Based on ability estimates we are able to calculate probabilities of winning a game in a match and subsequently probabilities of winning a match. The probability of player i winning a game over player j ($i, j \in \{1, 2, \ldots, P\}$) is defined as:

[1] University of Economics, Department of Quantitative Methods, Tajovského 13, 04001 Košice, stefan.lyocsa@euke.sk.
[2] University of Economics, Department of Corporate Finance, Tajovského 13, 04001 Košice, tomas.vyrost@euke.sk.

$$P(i \text{ beats } j) = \frac{\alpha_i}{\alpha_i + \alpha_j} \tag{1}$$

To calculate the probability of winning a match we used the approach of Barnett & Clarke (2002). Using a data set of past match results we can estimate the abilities of players by maximizing the likelihood:

$$L_0(\boldsymbol{\alpha}) = \prod_{i=1}^{P}\prod_{j=1}^{P}\left(\frac{\alpha_i}{\alpha_i + \alpha_j}\right)^{b_{i,j}} \tag{2}$$

where $b_{i,j} \in \mathbb{N} \cup \{0\}$ is the overall number of games won by player i over player j in the historical data set.

The initial model may be enhanced by adjusting for the changing form of the players by introducing exponential weighing of past observations as proposed by Dixon & Coles (1997). Our model controls for the surface that the match was played on similarly to McHale & Morton (2011). Define $g_{i,j,k}$ as the number of games won by player i over player j in match k, setting $g_{i,j,k} = 0$ if the match k was played by players other than i and j. Let $\varepsilon \in \mathbb{R}^+$ be the half-life parameter ($HL = ln(2)/\varepsilon$) of the exponential weighing scheme, $S_k \in [0,1]$ be the surface parameter, $t_k \in \mathbb{N}$ the time index at which the match k was played and $M \in \mathbb{N}$ the number of matches played. For fixed values of the half-life and surface parameters ε and S_k, we define a pseudo-likelihood function of the enhanced model over all matches up to time $t = \max\{t_k: 1 \le k \le M\}$ as:

$$L(\boldsymbol{\alpha}) = \prod_{k=1}^{M}\prod_{i=1}^{P}\prod_{j=1}^{P}\left(\frac{\alpha_i}{\alpha_i + \alpha_j}\right)^{g_{i,j,k} S_k \exp[\varepsilon(t-t_k)]} \tag{3}$$

Maximizing $L(\boldsymbol{\alpha})$ is equivalent to maximizing its pseudo-log-likelihood:

$$l(\boldsymbol{\alpha}) = \sum_{k=1}^{M}\sum_{i=1}^{P}\sum_{j=1}^{P}\left(g_{i,j,k} S_k e^{\varepsilon(t-t_k)}\left[\ln\alpha_i - \ln\left(\alpha_i + \alpha_j\right)\right]\right) \tag{4}$$

With a large data set of matches consisting of many players using general non-linear optimization routine for the optimization of this likelihood leads to unfeasible computational times and the time required to solve the optimization proved to become a limiting issue as the betting systems we employ include bootstrap procedures, raising the number of times the optimization have to be performed.

The MM algorithm replaces the problem of maximizing an arbitrary real-valued function $f(\theta)$ by finding a minorization, or surrogate function $g(\theta|\theta^{(l)})$ defined using a fixed value of $\theta^{(l)}$, which satisfies the following conditions (Hunter & Lange, 2004):

$$g(\theta|\theta^{(l)}) \le f(\theta) \quad \text{for all } \theta \tag{5}$$

$$g(\theta^{(l)}|\theta^{(l)}) = f(\theta^{(l)}) \tag{6}$$

The minorizing function $g(\theta|\theta^{(l)})$ is chosen as a function than can be optimized more easily than $f(\theta)$. The MM algorithm performs a maximization of $g(\theta|\theta^{(l)})$, thus iteratively obtaining estimates for maximizing $f(\theta)$ by setting

$$\theta^{(l+1)} = \arg\max_{\theta} g(\theta \mid \theta^{(l)}) \tag{7}$$

A minorization of the pseudo-log-likelihood can be obtained by the supporting hyperplane property (Hunter & Lange, 2004), as for any convex function f it is possible to obtain a minorization $f(v) \ge f(u) + df(u)/du\,(v - u)$. By setting $f(u) = -\ln(u)$ for $u \ge 0$ (Lange et al., 2000) we obtain the minorization of the pseudo-log-likelihood:

$$g(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}_l) = \sum_{k=1}^{M}\sum_{i=1}^{P}\sum_{j=1}^{P}\left(g_{i,j,k} S_k e^{\varepsilon(t-t_k)}\left[\ln\alpha_i - \ln\left(\alpha_i^{(l)} + \alpha_j^{(l)}\right) - \frac{\alpha_i + \alpha_j - \left(\alpha_i^{(l)} + \alpha_j^{(l)}\right)}{\alpha_i^{(l)} + \alpha_j^{(l)}}\right]\right) \tag{8}$$

By taking the derivatives of the surrogate function, it is easy to show that the update formula for the MM algorithm is:

$$\alpha_i^{(l+1)} = \frac{\sum_{k=1}^{M}\sum_{j\neq i} g_{i,j,k} S_k e^{\varepsilon(t-t_k)}}{\sum_{k=1}^{M}\sum_{i\neq j} \frac{S_k e^{\varepsilon(t-t_k)}(g_{i,j,k}+g_{j,i,k})}{\alpha_i^{(l)}+\alpha_i^{(l)}}} \tag{9}$$

3.2 Estimation algorithms

Assume we want to estimate abilities of players at day $t \in \mathbb{N}$, to forecast the outcome of all tennis matches played in the next 30 days. This is the forecast window and the set of forecasted matches will be denoted as $F_t = \{k: t \leq t_k < t + 30\}$, $k \in \mathbb{N}$. The BT model described in the previous section allows for the estimation of abilities, given a constant value of the half-life and the surface parameter. These parameters cannot be estimated from the pseudo-log-likelihood. The choice of a suitable half-life and surface parameters is based on an out-of-sample analysis on a set of matches played during the arbitrarily chosen window of 1080 days prior to day *t*. Our approach is as follows:

1. Using a set of matches $EW_{t,v} = \{k: t - 1080 + 30(v - 1) \leq t_k < t - 540 + 30(v - 1)\}$, where $v \in \{1, \ldots, 18\}$, estimate player's abilities via pseudo-log-likelihood (8) using values of half-life and S_k selected from the grid of possible values. The value of 540 days is the size of the estimation window and is chosen arbitrarily.
2. Based on estimated player's abilities we calculate the probability of winning a match for a set of matches $CW_{t,v} = \{k: t - 540 + 30(v - 1) \leq t_k < t - 540 + 30v\}$, where $CW_{t,v}$ is the out-of-sample set of matches with respect to $EW_{t,v}$ at a given v. The value of 30 days is the so-called drift parameter and it is chosen arbitrarily. Based on the match outcome forecasts and the actual outcome, we calculate the ratio of successful bets.
3. Steps 1 and 2 are repeated for different values of v. This corresponds to testing the forecasting ability of the model at a given half-life and S_k for a sample of matches played during 540 days.
4. Steps 1 to 3 should be repeated for all combinations of half-life and S_k, i.e. in our case 657 times (explained in a later section). The combination which achieved the highest overall ratio of successful bets is chosen for the estimation of player's abilities.

The previous steps lead to the calculation of abilities for given surface parameters, and all half-life parameters. A different set of surface parameters is needed depending on what surface the forecasted match is being played on. Therefore:

5. Repeat steps 1 to 4 for the prediction of all possible surfaces, i.e. 3 times altogether.
6. The previous steps 1-5 were used to estimate three kinds of player's abilities, one for each surface. These are used to predict the outcome of matches (F_t) played in the forecast window.

The entire procedure is repeated by drifting the data window and forecast period 30 days ahead. This was repeated until the final forecasting window included April 19, 2014, the end of our sample.

The values of the half-life *HL* and surface parameters S_k were chosen to maximize the ratio of successful bets placed on each match during the calibration periods. The set of parameters considered was created from a grid defined by half-lifes and surface weights. The arbitrarily selected half-life values considered were $HL \in \{30, 60, 90, 240, 360, 540, 720, 1080, 1440\}$. Define $wh, wc, wg \geq 0$ as weights for hard-court/carpet, clay, and grass surfaces, respectively. Say a match played on grass is analyzed. In that case, the optimal weights were selected from the sets $wg \in \{0.40 + r0.05, r = 0, 1, ..., 12\}$, $wh, wc \in \{0.00 + s0.05, s = 0, 1, \ldots, 8\}$, with the restriction that[3] $wh + wc + wg = 1$. This approach leads to a grid of 657 combinations of possible values for *HL* and S_k (9 *HL*s and 73 surface weights).

The surface weights and half-life were also calibrated out-of-sample. As a consequence, although we use data from as early as January 3rd, 2000, our first out-of-sample predicted abilities and thus matches used to test for market efficiency/profitability of betting systems start on December 30, 2002. Previous observations of nearly three years are used to estimate and calibrate the forecasting model with respect to surface weights and half-life.

[3] Due to the estimation issues, if one of the weights is set to 0, it is adjusted by a small amount to be positive, i.e. 0.0001.

3.3 Betting

In this version of the paper we present results from 3 betting systems, where the bettor can utilize information from the BT model and existing market odds. Two betting rules are based on the bootstrapped samples. Hunter (2004) discusses the formulation of standard error estimates of the BT model via the MM algorithm. One of the possibilities Hunter (2004) considered is using a bootstrap approach in which all matches are resampled with replacement and the BT model is fitted using the bootstrapped data set (Hunter, 2004, p. 402). A possible issue with this approach is that it is very likely that the bootstrapped sample of comparisons will violate the requirement that each player should have at least one win and loss (Hunter, 2004). Our approach is as follows:

1. Resample the data set using the standard resampling with replacement bootstrapping approach of Efron & Tibshirani (1994).
2. Remove all players who do not have minimum of 8 matches and or do not have at least one win and loss. This step is repeated until the data set conforms to these requirements.
3. Estimate the model (8) based on the bootstrapped data set. Based on the preferred half-life and S_k, estimate and store player's abilities, probabilities of winning a match, and the difference between the probabilities of winning a match of the two players. In this case, the bootstrapped probability of the player with lower initial (unbootstrapped) ability is subtracted from bootstrapped probability of the player with higher initial (unbootstrapped) ability. This variable is denoted as $PWHL_{kb}$, $b = 1,\ldots,B$ (Probability of Winning of a Higher ability player minus the Lower ability player).
4. Steps 1-3 should be repeated 1000 times.

Let p_i denote the probability of player i with higher initial ability winning the match k. Let $q_{percentile}(PWHL_k)$ denote the percentile of the distribution of $PWHL_k$. Now, the 3 betting systems are as follows:

Bet on player i if:

- S01: $p_i > p_j$
- S02: $q_{0.000}(PWHL_k) > 0 \wedge o_i < o_j$
- S03: $q_{0.250}(PWHL_k) > 0 \wedge o_i < o_j$

The S01 betting system is a simple unit bet on all favourites (as suggested by the BT model), regardless of the uncertainty related to player's ability. Betting systems S02, S03 are also unit bets, which rely on the information derived from the bootstrapped samples and on market odds (o_i and o_j). We assumed that the more certain our estimate is of the outcome of the match, the less likely it is that we observe negative values of $PWHL_k$.

4 Results and conclusion

Returns based on forecasting matches on all surfaces combined are reported in Table 1. For example, the value of 0.0004 in the first row corresponds to an average of +0.04% return using the respective betting rule (in this case, S01). Compared to this mean return, the standard deviation of 0.7750 is rather high, which however is the consequence of the betting market contract structure, as all lost bets result in a return of –1 (–100%).

The column #B denotes the number of bets using the given betting system during the out-of-sample period, e.g. 24 238 bets for the first betting system. The 80.76% corresponds to the proportion of bets compared to the whole sample of 30 013 matches played in a given period.

Results from bootstrap-based betting provide evidence against the market efficiency hypothesis. All strategies led to positive average returns, with highest average returns of 0.70%. Even the least restrictive strategy has led to positive returns, while still betting on 17 217 (57.37%) matches. This suggests that together with the market odds known at the time of placing bets, the model can be used on a much larger scale.

The returns using average odds are naturally much smaller. The "*price-fishing*" effect is considerable. Our results agree with that of McHale & Morton (2011) in that the forecasting model by itself is as important as betting with the best available odds.

	Best market odds	Average market odds	#B
	mean [stand. dev.]	mean [stand. dev.]	
Always bet	0.0004 [0.7750]	-0.0373 [0.7338]	24238 80.76%
BT-Min/Max	0.0059 [0.4661]	-0.0194 [0.4394]	4792 15.97%
BT-0.250	**0.0070** [0.6263]	-0.0249 [0.6001]	17217 57.37%

Table 1 Average betting returns for all surfaces combined

Notes: BT – Returns using the BT model. ATP Rank – Returns using the official ATP ranking approach to forecast the winner on the same set of matches as forecasted by the BT model and corresponding betting system. #B – is number of considered matches (bets), the percentage in brackets are calculated from all 30 013 matches played during our out-of-sample database in a given period. Bolded coefficients denote statistical significance at least at the 10% significance level.

Acknowledgements

The authors also acknowledge funding support from the Slovak Grant Agency for Science (VEGA project No. 1/0392/15).

References

[1] Hunter, D. R.: MM Algorithms for Generalized Bradley-Terry Models. *Annals of Statistics* 32 (2004), 384-406.

[2] Bradley, R.A., and Terry, M.E.: Rank analysis of incomplete block design I: the method of paired comparisons. *Biometrica* **39** (1952), 324-345.

[3] McHale, I. G., and Morton, A.: A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting* **27** (2011), 619-630.

[4] Glickman, M. E.: Parameter Estimation in Large Dynamic Paired Comparison Experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **48** (1999), 377-394.

[5] Scheibehenne, B., and Bröder, A.: Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting* **23** (2007), 415-426.

[6] Barnett, T. J., and Clarke, S. R.: Using Microsoft Excel to model a tennis match. In: *6th* Conference *on Mathematics and Computers in Sport* (pp. 63-68). Bond University, Queensland, Australia.

[7] Dixon, M. J., and Coles, S. G.: Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **46** (1997), 265-280.

[8] Hunter, D. R., and Lange, K.: A tutorial on MM algorithms. *The American Statistician* **58** (2004), 30-37.

[9] Lange, K., Hunter, D. R., and Yang, I.: Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics* **9** (2000), 1-20.

[10] Efron, B. & Tibshirani, R. J. (1994). *An Introduction to Bootstrap*. Chapman and Hall.

Efficiency of Government Expenditure: A Comparison of OECD and EU Countries

Zuzana Machová[1], Igor Kotlán[2], Agata Drobiszová[3]

Abstract. Classification of government expenditure can play an important role. The article uses two kinds of classification of government spending a) traditional breakdown into productive and unproductive government spending, and b) into individual and collective services. Therefore, the aim of this article is to investigate the role of the classification of government expenditure in the case of their impact on economic growth. Empirical analysis is focused on both the OECD and EU countries and their comparison in the period 2000-2012. From a methodological point of view, GMM estimator is used. The results of empirical analyses show that in case of the traditional breakdown of government expenditure, the impact of productive expenditure on economic growth is negative in OECD countries and not significant in EU countries. However, in case of the classification of government expenditure on individual and collective services, results confirm that individual services affect economic growth negative and collective services promote economic growth, as the expenditure on collective services also includes spending on R&D unlike unproductive spending e.g. in case of expenditure on education.

Keywords: Economic Growth, Classification of Government Expenditure, Individual Services, Collective Services.

JEL Classification: H20, H50, O40
AMS Classification: 62J05

1 Introduction

How to promote economic growth by fiscal policy instruments? It is a question that is currently highly discussed in economic literature. To answer this question it is necessary to know how the individual components of the state budget affect economic growth. This issue is particularly important in the context of economic, financial or debt crises. For more on this topic from several interesting points of view, see also e.g. [6], [20] or [24]. Economists analyze either the influence of different types of taxes, or the influence of different types of government expenditure on economic growth, but nowadays the aggregate influence of all the fiscal variables on growth is studied above all. In the analyses, in accordance with contemporary literature, direct and indirect taxes are usually distinguished, as well as productive and unproductive government expenditures. The explanation of the categorization of taxes is that direct taxes have higher distortionary effects on the economy and thus harm economic growth more. In the case of government expenditures, there is an assumption that productive government expenditures are investment expenditures that support economic growth, while unproductive, consumption expenditures do not. However, the conclusions of empirical studies are contradictory. It has been neither significantly confirmed what the effects of separate types of taxes are, nor what the effects of particular types of expenditures are.

While the issues of the influence of taxes on growth were examined in several recent papers [13], [15] or [16] and are not of interest in this paper, the effects of government expenditures seems a rather unexplored area. Starting from economic theory, there are also doubts about the accuracy of the classification of expenditure into productive and unproductive parts, or rather about the accuracy of the assignment of specific types of expenditures to productive or unproductive parts. There appears to be a question as to whether another classification would be better, with respect to the analysis of government expenditures effects on growth. One possibility is to study the separate effects of expenditures on individual, and on collective services including expenditures on R&D.

[1] VSB-Technical University of Ostrava, National Economy, Sokolská 33, Ostrava, Czech Republic e-mail: zazana.machova@vsb.cz.
[2] VSB-Technical University of Ostrava, National Economy, Sokolská 33, Ostrava, Czech Republic e-mail: igor.kotlán@vsb.cz.
[3] VSB-Technical University of Ostrava, National Economy, Sokolská 33, Ostrava, Czech Republic e-mail: agata.drobiszova@vsb.cz.

Thus the aim of this article is to investigate the role of the classification of government expenditure in the case of their impact on economic growth. Empirical analysis is focused on both the OECD and EU countries and their comparison in the period 2000-2012.

2 Model of Growth: Theoretical Foundation

According to Barro [3] or Rebelo [21], the growth rate of an economy (γ) in a steady state is equal to the growth rate of physical capital (k), human capital (h), and the growth rate of consumption (c) of households, and depends on intertemporal elasticity of substitution in consumption ($1/\theta$) adjusted to time preference rate (ρ), and net marginal rate of return on capital (both human and physical):

$$\gamma = \frac{\dot{k}}{k} = \frac{\dot{h}}{h} = \frac{\dot{c}}{c} = \frac{1}{\theta}\left(r - \rho\right). \tag{1}$$

Government expenditures may be considered an input to private production in the event that they are of investment character (expenditures on infrastructure above all). In that case, government expenditures have a positive effect on economic growth, as they raise the net marginal rate of return on capital. Many empirical studies have been based on Barro's assumption, and this type of government expenditures has become labelled productive (see e.g [12], [11] or [7]). However, the studies differ in what type of real government expenditures they include into the productive part. For simplicity, Barro [3] assumes the broad concept of capital that includes both physical and human capital (consistently with Rebelo [21] and his AK model). With constant returns to scale, the production (Y) function thus shall be as follows:

$$Y = AK\left(\frac{G}{K}\right)^{\alpha} = K\phi\left(\frac{G}{K}\right), \tag{2}$$

where A is the technology level, K is capital, G is productive government expenditure, α is technological coefficient, and expresses a positive decreasing marginal product (i.e. ϕ ' > 0 and ϕ '' < 0). For the purposes of expressing the relationship between government expenditure and economic growth, it is possible to assume the existence of a flat tax rate on income τ , which is used to fund government expenditure: $G = \tau Y = \tau K\phi\left(G/K\right)$, in the case of a balanced budget. For simplicity, different rates for different types of taxes are not considered. The influence of different types of taxes is taken into account e.g. in [15]. They show the channels through which the taxes affect the growth variables. They also point out that some economic and legal aspects of labor markets must be considered in the analyses. For these aspects, see e.g. [23], [22], [5] or [18].

The net marginal return on capital is then equal to:

$$r = \left(1-\tau\right)\phi\left(\frac{G}{K}\right)\left(1-\phi'\frac{G}{Y}\right) = \left(1-\tau\right)\phi\left(\frac{G}{K}\right)\left(1-\eta\right), \tag{3}$$

where η is the sensitivity of Y to changes in G for a given K, and the growth rate of an economy is modified from the equation (1) to:

$$\gamma = \frac{1}{\theta}\left[\left(1-\tau\right)\phi\left(\frac{G}{K}\right)\left(1-\eta\right) - \rho\right]. \tag{4}$$

As already stated, the growth rate of the economy is thus positively affected by productive government expenditure, but the sensitivity of production to changes in government expenditure has, on the other hand, a negative influence on growth.

The most important study by Kneller, Bleaney and Gemmell [12] uses the standard Classification of the Functions of Government (COFOG), but distinguishes only between a group of productive expenditures (including general public services, defense, educational, health, housing, and transport and communication expenditures), and a group of unproductive expenditures (including social security and welfare expenditures, expenditures on recreation, and on economic services). The authors show that the group of productive expenditures really positively affect growth, while unproductive ones do not. Among similar studies, Denaux [7] must also be mentioned. He uses the breakdown of expenditures on government expenditures to physical capital accumulation (approximated

by stock of roads per cap.), and expenditures to human capital accumulation (approximated by government spending on schooling). He finds no significant effect of spending on physical capital, but he confirms the positive effect of spending on human capital, i.e. on higher education on economic growth. However, as already stated within the introduction, the results of empirical analyses are not consistent, especially when a more detailed level of classification of expenditures is used. For example, Drobiszová and Machová [8] also use the COFOG, but they include all the main types of expenditures into their analysis. They show that only spending on defense, education, health, and general public services have a positive effect on economic growth (for more about how institutional condition influence economic growth from several interesting points of view, see e.g. [9]). Unlike the abovementioned studies, the positive effects of expenditures on infrastructure were not confirmed. Recent experience from OECD countries shows that public wages, interest payments, subsidies and government consumption are less growth enhancing, while spending on education and health boosts growth [1]. It is also interesting to realize the consequences of the fact that Bhattacharya and Mukherjee [4] confirm that households in the OECD move from non-Ricardian to Ricardian behaviour as government debt reaches high levels and as uncertainty about future taxes increases. At first glance it may seem that expenditures on, e.g. education and health, i.e. human capital, are provably supporting economic growth, while the effects of expenditures on, e.g. infrastructure are questionable. But it is very important to realize that, according to the methodology of COFOG 98, all of the main categories of government expenditures also include corresponding expenditures on R&D. This may significantly distort the analyses. One of the ways to avoid this problem is to distinguish between expenditures on collective services, and individual services, which should be closer to Barro's approach above, assuming that expenditures on collective services have positive effects on growth, while expenditures on individual services do not. The classification and further explanation are described in following part.

3 Methodology, Data and Results of Empirical Analysis

According to the OECD and Eurostat [19], individual services are the services that general government provides to specific identifiable households. That is, services such as health and education, which are consumed by households individually. Collective services are those that the general government provides simultaneously to all members of the community. That is, services such as defense and public order and safety, which are consumed by households collectively. Because they are considered as part of gross fixed capital formation (according to the SNA 2008, and ESA 2010), the R&D for individual services are also included, which is the most important point of the classification. Collective services also include the overall policy-making, planning, regulatory, budgetary, coordinating and monitoring responsibilities of ministries overseeing individual services. These activities, unlike the services to which they relate, cannot be identified with specific individual households and are considered to benefit households collectively.

Based on the COFOG 98, individual services include expenditures on:

- Health (category 7), Recreation, culture and religion (8), Education (9), Social protection (10),

except for categories including R&D (7.5, 8.4, 8.5, 9.7, 10.8), and n.e.c. categories (7.6, 8.6, 9.8, 10.9).

The excepted categories as well as categories:

- General public services (1), Defense (2), Public order and safety (3), Economic affairs (4), Environment protection (5), Housing and community amenities (6),

make up the group of collective services. The extend of all the expenditure types is involved by national and international laws. However, in the case of environmental protection, the impact of laws is very strong (for more on the legal aspects, see e.g. [10]).

From a methodological point of view, the analysis was based on a dynamic panel model which used data for 34 OECD countries, or alternatively for 28 European Union countries in the period 2000-2012. Gross domestic product per capita in PPP (Y) was dependent variable. Independent variables included control growth variables, i.e. gross fixed capital formation as a percentage of GDP (K), and approximation of human capital formation (L). The second group of independent variables was formed by fiscal variables. The model included both the revenue and expenditure side of the budget; fiscal balance was omitted to avoid perfect collinearity of the variables. Regression coefficients of the fiscal variables may thus be interpreted as a change of the output in the case that the corresponding variable as well as the fiscal balance in opposite direction is changed by unit, ceteris paribus. Fiscal variables included direct taxes, indirect taxes, both expressed as corresponding tax quota, i.e. tax revenue-to-GDP ratio (alternative indices could be also used, e.g. the World Tax Index, see [14] or [17]), and finally government expenditures. At first, productive and unproductive expenditure in accordance with Kneller, Bleaney and Gemmell [12] were included, and secondly expenditures on individual services and collective services, all of them expressed as a percentage of GDP. The data was gained from Eurostat, the World Bank and OECD iLibrary, specifically the OECD Tax Database, and the database of national accounts.

As the estimation technique, a generalized method of moments (GMM) was used, which included the method of instrumental variables. This method uses the Arellano-Bond estimator [2], and the transformation method of first difference of each variable in the regression, which prevents potential problems with the stacionarity of the time series. Using a robust estimator in calculating the covariance matrices (White Period method) ensured that the estimation results of standard deviations of parameters and hypothesis tests were correct with regard to a possible occurrence of autocorrelation and heteroscedasticity. This estimation type ensures that the appropriate transformation process and using appropriate instruments eliminates the risk of endogeneity of the lagged values of the dependent variable and the independent variables with a random component. In the analyses below, the lagged values of the dependent variable were used as the instruments, starting from a lag of two. A Sargan test (see the J-statistics in the tables below) confirmed the validity of the instruments in both estimated models. For better interpretation of the regression coefficients, all the variables were used in logarithmic form.

Table (1) shows the results of the estimation of the model including productive and unproductive government spending according to Kneller, Bleaney and Gemmell [12]. Table (2) shows the results from the model where government expenditures were divided on expenditures on collective services, and on individual services.

Countries	OECD		EU	
Dependent variable	ln(GDP per capita)		ln(GDP per capita)	
Variable	Coefficient	t-Statistic	Coefficient	t-Statistic
ln(GDP per capita (-1))	0.895	465.53***	0.668	23.42***
ln(Physical Capital)	0.009	10.10***	0.035	2.22**
ln(Human capital)	0.032	29.46***	0.195	4.50***
ln(Direct Taxes (-1))	-0.031	-16.49***	-0.115	-6.56***
ln(Indirect taxes (-1))	-0.017	-16.74***	0.357	38.53***
ln(Productive expenditure (-1))	-0.001	-1.28	0.001	0.02
ln(Unproductive expenditure (-1))	-0.014	-25.54***	-0.056	-1.49
Instrument rank	35		27	
J-statistic	34.21		20.98	

*Notes: t-statistics are adjusted for heteroskedasticity and autocorrelation; standard deviations are calculated using robust estimates; *, **, *** indicate significance levels of 10 %, 5 %, and 1 %, respectively. Source: own calculations.*

Table 1 Effects of productive and unproductive expenditures on growth in the period 2000-2012

In both tables, a statistically significant positive effect of control growth variables was confirmed, with human capital having a quantitatively stronger impact on long-run growth, which corresponds to economic theory for developed countries. Also, in both tables, a statistically significant negative effect of direct taxes was confirmed. On the other hand, indirect taxes seem to have positive effects on growth. It may be a consequence of tax harmonization in the EU, which is successful just in case of indirect taxes, and following reduction of distortions. However, the influence of taxes on growth is not of interest of this paper. For more detail about impact of particular types of taxes on economic growth see, e.g. [13], [15] or [16].

Countries	OECD		EU	
Dependent variable	ln(GDP per capita)		ln(GDP per capita)	
Variable	Coefficient	t-Statistic	Coefficient	t-Statistic
ln(GDP per capita (-1))	0.897	270.15***	0.649	31.72***
ln(Physical Capital)	0.013	10.44***	0.043	5.55**
ln(Human capital)	0.036	22.67***	0.212	5.14***
ln(Direct Taxes (-1))	-0.026	-13.12***	-0.111	-11.49***
ln(Indirect taxes (-1))	-0.020	-31.93***	0.361	19.58***
ln(Collective services (-1))	0.006	5.89***	0.010	0.33
ln(Individual services (-1))	-0.015	-16.63***	-0.081	-2.25**
Instrument rank	35		27	
J-statistic	35.69		23.54	

*Notes: t-statistics are adjusted for heteroskedasticity and autocorrelation; standard deviations are calculated using robust estimates; *, **, *** indicate significance levels of 10 %, 5 %, and 1 %, respectively. Source: own calculations.*

Table 2 Effects of expenditures on collective and individual services on growth in the period 2000-2012

With regard to the aim of the paper, the results concerning government expenditures are, nevertheless, the most important. Table (1) shows, that the influence of unproductive expenditures, classified according to Kneller, Bleaney and Gemmell [12], on economic growth is negative in case of both group of countries. But empirical analysis shows that impact of unproductive expenditure on economic growth in the European Union countries is not significant. This may be due to the fact that EU countries also include the less developed countries, where

social security expenditure still creates a stimulating institutional environment. Productive expenditure affects economic growth negatively in OECD countries and positively in EU countries but in both case the effect on economic growth is not significant. This difference can be connected with level of development of countries and efficiency of the use of this expenditure.

On the other hand, if the expenditures are examined in categorization on expenditures on collective and individual services (table 2), it is shown that the impact of expenditures on collective services is positive, while the effect of expenditures on individual services on growth is negative and quantitatively is higher than the effect of unproductive expenditures because on this case expenditure on individual services not including expenditures on R&D.

4 Conclusion

The topic of fiscal policy and its impact on economic growth is very current, especially with regard to research questions that have arisen with the recent economic, financial and debt crises. Fiscal consolidation is inevitable in many countries, and their fiscal authorities are forced to make budget cuts, which is in direct conflict with the standard Keynesian approach. Furthermore, many empirical papers have confirmed the negative effects of most types of taxes on economic growth. Improving fiscal balance by raising taxes could thus be counterproductive. On the other hand, there also exist theoretical as well as empirical studies showing that certain types of government expenditures may have significant positive effects on economic growth. Restructuring the expenditures side of a budget could thus improve the performance of the economy without any negative effects on fiscal balance.

Most of the previous studies distinguish between the effect of so-called productive and unproductive government spending. However, unproductive expenditures also include expenditures on R&D in those studies, as well as expenditures on education and health which are of an individual character. These methodological discrepancies may be easily avoided using a classification distinguishing between expenditures on collective services, and expenditures on individual services, which is also in accordance with new national accounting methodology SNA 2008 and ESA 2010.This approach was also used in this paper, where a dynamic panel model for 34 OECD countries and 28 European Union countries in the period 2000-2012 was estimated, using a GMM and instrumental variables method. Four models were estimated – the first two included government expenditures classified as productive and unproductive in accordance with the literature, while the two other included expenditures on collective, and on individual services. In the first two models, unproductive government expenditures have a negative effect on growth, but they were significant only in the case of OECD countries. Productive expenditure influenced economic growth negatively in the OECD countries and positively in the EU countries, but in both cases it was not significant. It only supports the inconsistency of the results in the case of the breakdown of expenditures on productive and unproductive. On the other hand, in the two other models, it is shown that, in accordance with the assumptions, expenditures on collective services, including all the expenditures on R&D, positively affect long-run economic growth, while the expenditures on individual services, including individual expenditures on education and health, affect growth negatively.

The results support the idea that the conclusions of previous studies on this topic may be strongly distorted by the inappropriate classification of expenditures, typically in the case of expenditures on education and health. These are usually considered productive, but if their part of R&D expenditures is detached, their effect on growth is not clear in fact. Another conclusion of this article is also that differences in the breakdown of government expenditures are important for countries that have a larger share of expenditure on R&D, as is evident when comparing the models for the EU and OECD countries.

Acknowledgements

This paper was financially supported within the VŠB - Technical University SGS grant project No. SP2015/110 “The Influence of Fiscal Deficit on Economic Growth and the Possibilities of its Decreasing in the Czech Republic“.

References

[1] Afonso, A., and Jalles, J. T.: Fiscal Composition and Long-term Growth. *Applied Economics* **46,** 3 (2014), 349-358.
[2] Arellano, M., and Bond, S.: Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58** (1991), 277–297.
[3] Barro, R. J.: Government Spending in a Simple Model of Endogenous Growth. *The Journal of Political Economy* **98**, 5 (1990), 103–125.
[4] Bhattacharya, R., and Mukherjee, S.: Non-Kenynesian Effects of Fiscal Policy in OECD Economies: An Empirical Study. *Applied Economics* **45**, 29 (2013), 4122-4136.

[5] Birčiaková, N., Stávková, J., and Antošová, V.: Income (in)justice in the Czech Republic. *DANUBE: Law and Economics Review* **4**, 3 (2013), 185-199.
[6] Caba, S.: Economic crisis and globalization index dynamics. *Actual Problems of Economics* **1** (2013), 527-537.
[7] Denaux, Z. S.: Endogenous growth, taxes and government spending: Theory and evidence. *Review of Development Economics* **11,** 1 (2007), 124-138.
[8] Drobiszová, A., and Machová, Z.: The Impact of Fiscal Policy on Economic Growth in the OECD Countries. Politická ekonomie **63**, 3 (2015), 300-316.
[9] Drobiszová, A., and Machová, Z.: Optimal Mix of Fiscal Policy in the EU with a Special Reference to the Czech Republic. In: *Proceedings of the 25th International Business Information Management Association Conference* (Soliman, K. S. eds.) International Business Information Management Association (IBIMA), Amsterdam, Netherlands, 2015, pp. 2011-2019.
[10] Gaberščik, A.: The laws of nature also apply to human society. *DANUBE: Law and Economics Review* **3,** 4 (2012), 47-53.
[11] Gemmel, N., Kneller, R., and Sanz, I.: The Timing and Persistence of Fiscal Policy Impacts on Growth: Evidence from OECD Countries. *Economic Journal* **121**, 550 (2011), F33-F58.
[12] Kneller, R., Bleaney, M., and Gemmell, N.: Fiscal Policy and Growth: Evidence from OECD Countries. *Journal of Public Economics* **74** (1999), 171-190.3
[13] Kotlán, I., and Machová, Z.: The Issue of Using Tax Quota in the Growth Empiricism. In: *Proceedings of the 10th International Scientific Conference on Economic Policy in the European Union Member Countries: Selected Papers* (Tvrdoň, M., Majerová, I. eds.) Vendryně, 2012a, pp. 145-160.
[14] Kotlán, I., and Machová, Z.: World Tax Index: Methodology and Data. *DANUBE: Law and Economics Review* **3**, 2 (2012b), 19-33.
[15] Kotlán, I., and Machová, Z.: The Impact of the Tax Burden on the Living Standard in OECD Countries. Ekonomický časopis **61**, 9 (2013), 951-962.
[16] Kotlán, I., and Machová, Z.: Tax Policy Horizon in the OECD Countries. *Politická ekonomie* **62**, 2 (2014), 161-173.
[17] Machová, Z., and Kotlán, I.: World Tax Index: New Methodology for OECD Countries, 2000-2012. *DANUBE: Law and Economics Review* **4**, 2 (2013), 165-179.
[18] Nagy, Z. I., and Való, V.: The Lawful Execution of Employer Rights Is an Important Subsystem of Corporate Governance (a Labour Case Study). *DANUBE: Law and Economics Review* **4**, 3 (2013), 201-217.
[19] OECD and Eurostat: Eurostat-OECD Methodological Manual on Purchasing Power Parities (2012 Edition). Publications Office of the European Union, Luxembourg, 2012.
[20] Olteanu, D.C.: Overlending and Financial Instability in Eastern Europe During the Crisis. *Actual Problems of Economics* **142,** 4 (2013), 429-441.
[21] Rebelo, S.: Long-Run Policy Analysis and Long-Run Growth. *Journal of Political Economy* **99**, 3 (1991), 500-521.
[22] Schovánková, S.: Determinants of the Spatial Distribution of Foreign Employees on Different Skill Job Positions: The Case of the Czech Republic. *DANUBE: Law and Economics Review* **4**, 2 (2013), 115-139.
[23] Simonovits, A.: Does Higher Tax Morale Imply Higher Optimal Labor Income Tax Rate? *DANUBE: Law and Economics Review* **4,** 2 (2013), 97-114.
[24] Šišková, N.: The Impact of the Eurozone's Crisis on the Institutional Framework and National Law of Member States - Some Considerations. *DANUBE: Law and Economics Review* **4**, 4 (2013), 293-300.

Dynamic analysis of multi-criteria network systems

Renata Majovská[1], Petr Fiala[2]

Abstract. Many of today's systems are characterized by a network structure and evaluation of alternatives by multiple criteria. These systems operate in dynamic environment. The paper proposes an approach for analyzing these systems. The approach is based on combination of Decision Making Trial and Evaluation Laboratory (DEMATEL) and Analytic Network Process (ANP) methods. The DEMATEL is a comprehensive method for building and analyzing a structural model involving causal relationships between complex factors. The evaluation of relationships is based on expert judgements. The ANP is a method that makes it possible to deal systematically with all kinds of dependence and feedback. The structure of the ANP model is described by clusters of elements connected by their dependence on one another. These connections indicate the flow of influence between the elements. The approach proceeds in two steps. The DEMATEL method is used in the first step to filter of some parts in the model based on expert threshold values. In the second step, the ANP is applied for more precise evaluation of relations by multiple criteria. Dynamic analysis of the systems use time dependent priorities that can be expressed by combination of forecasting using pairwise comparison functions and predictions based on using of compositional data exponential smoothing. The combination of these methods gives a powerful instrument for dynamic analysis of network systems by multiple criteria. The proposed approach is applied for dynamic project portfolio management.

Keywords: Network systems, multiple criteria, dynamics, DEMATEL, ANP.

JEL Classification: C44
AMS Classification: 90B

1 Introduction

The network economy is a term for today's global relationship among economic subjects characterized by massive connectivity. The central act of the new era is to connect everything to everything in deep web networks at many levels of mutually interdependent relations, where resources and activities are shared, markets are enlarged and costs of risk are reduced. Many of today's systems are characterized by a network structure and evaluation of alternatives is based on multiple criteria. Network systems contain both positive and negative feedbacks. Network economy drives and is driven by dramatic acceleration in technological innovation, in information and communication technologies especially. New technologies provide a permanent feedback that enables activity modifications and quick responses and therefore fundamentally change business models. The analysis of possible effects of network economy is an opportune topic for challenging scientific research [4].

In the paper is proposed a hybrid procedure for operation in network dynamic environment. Selection of methods was influenced by graphical representation of the analyzed network systems. The procedure is based on a combination of DEMATEL and dynamic version of ANP. The DEMATEL (Decision Making Trial and Evaluation Laboratory) method [5], originated from the Geneva Research Centre of the Battelle Memorial Institute, is especially pragmatic to visualize the structure of complicated causal relationships. DEMATEL is a comprehensive method for building and analyzing a structural model involving causal relationships between complex factors. Analytic Hierarchy Process (AHP), developed by T. Saaty [10], is a very popular method for setting priorities in hierarchical systems. A variety of feedback processes create complex system behavior. For the network seems to be very appropriate Analytic Network Process (ANP) approach [10]. The ANP makes possible to deal systematically with all kinds of dependence and feedback in the system. Dynamic analysis of the systems use time dependent priorities that can be expressed by combination of forecasting using pairwise comparison functions and predictions based on using of compositional data exponential smoothing. Selected methods are shortly summarized in the paper with regard to their use for analysis of network systems. These methods have specific

[1] VŠB-Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, renata.majovska@vsb.cz
[2] University of Economics/Faculty of Informatics and Statistics, Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, pfiala@vse.cz

advantages in analyzing and evaluating network multi-criteria systems. The combination of these approaches gives a powerful instrument for analyzing network dynamic systems by multiple criteria.

Applications of the network problems are found in many areas. The unifying concept of global networks with associated methodologies allows to explore the interactions among such networks. This paper aims to verify the ability to model and solve the problem of dynamic project portfolio using a combination of DEMATEL and dynamic version of the Analytic Network Process (ANP) model. The key to success in project portfolio management is to select the right projects at the right time [6], [8]. The project selection process is considered a major component of project portfolio management. This should be accompanied by periodically repeated inspections of project portfolio, which would identify projects that should be terminated. Effective portfolio management helps to achieve outperformance, making strategy real through organizational change. Strategic project portfolio management enables present a framework for organization to complete significant strategic projects. Portfolio management is a process. This process must improve over time. Building feedback into every stage of the process is critical for the improvement.

2 DEMATEL method

The DEMATEL method can be summarized in the following steps:

Step 1. *Find the initial direct relation matrix.*

Suppose we have *m* experts in this study and *n* elements to consider. Each expert is asked to indicate the degree to which he believes an element *i* affects an element *j*. These pairwise comparisons between any two elements are denoted by a_{ij} and are given an integer score ranging from 0, 1, 2, 3 and 4 representing:

0 no influence,
1 low influence,
2 medium influence,
3 high influence,
4 very high influence.

The scores by each expert will give us a (*n*, *n*) non-negative answer matrix $X^k = \left[x_{ij}^k \right]$, with $1 \le k \le m$. The diagonal elements of each answer matrix X^k are all set to zero. We can compute the (n,n) average matrix A for all expert opinions by averaging the m experts' scores as follows:

$$a_{ij} = \frac{1}{m} \sum_{k=1}^{m} x_{ij}^k \tag{1}$$

The average matrix $A = \left[a_{ij} \right]$ is called the initial direct relation matrix. Matrix A shows the initial direct effects that an element exerts on and receives from other elements. Furthermore, we can map out the causal effect between each pair of elements in a system by drawing an influence map. DEMATEL can convert the structural relations among the elements of a system into an intelligible map of the system.

Step 2. *Calculate the normalized initial direct-relation matrix.*

The normalized initial direct-relation matrix D is obtained by normalizing the initial direct-relation matrix A in the following way:

Let
$$s = \max\left(\max_{1 \le i \le n} \sum_{j=1}^{n} a_{ij}, \max_{1 \le j \le n} \sum_{i=1}^{n} a_{ij} \right) \tag{2}$$

then
$$D = \frac{1}{s} A . \tag{3}$$

Since the sum of each row i of matrix A represents the total direct effects that element i gives to the other elements, $\max_{1 \le i \le n} \sum_{j=1}^{n} a_{ij}$ represents the total direct effects of the element with the most direct effects on others. Likewise, since the sum of each column j of matrix A represents the total direct effects received by element *j*, $\max_{1 \le j \le n} \sum_{i=1}^{n} a_{ij}$ represents the total direct effects received of the element that receives the most direct effects from

others. The positive scalar s takes the greater of the two as the upper bound, and the matrix D is obtained by dividing each element of A by the scalar s. Note that each element d_{ij} of matrix D is between zero and one.

Step 3. *Compute the total relation matrix.*

A continuous decrease of the indirect effects of problems along the powers of matrix D guarantees convergent solutions to the matrix inversion similar to an absorbing Markov chain matrix. Note that $\lim_{k\to\infty} D^k = O$ and $\lim_{k\to\infty}\left(I + D + D^2 + ... + D^k\right) = \left(I - D\right)^{-1}$, where O is the (n,n) null matrix and I is the (n,n) identity matrix.

The total relation matrix T is an (n,n) matrix $T = \left[t_{ij}\right]$, where $i, j = 1, 2, ..., n$ and is defined as

$$T = D + D^2 + ... + D^k = D\left(I + D + D^2 + ... + D^{k-1}\right) = D\left(I - D\right)^{-1} \text{ as } k \to \infty . \tag{4}$$

Step 4. *Set a threshold value and obtain the impact-relations-map.*

In order to explain the structural relation among the elements while keeping the complexity of the system to a manageable level, it is necessary to set a threshold value p to filter out some negligible effects in matrix T. While each element of matrix T provides information on how one element affects another, the decision-maker must set a threshold value in order to reduce the complexity of the structural relation model implicit in matrix T. Only some elements, whose effect in matrix T is greater than the threshold value, should be chosen and shown in an impact-relations-map [11].

3 The ANP method

The Analytic Network Process (ANP) is the method that makes it possible to deal systematically with all kinds of dependence and feedback. The well-known AHP theory is a special case of the Analytic Network Process that can be very useful for incorporating linkages. The structure of the ANP model is described by clusters of elements connected by their dependence on one another. A cluster groups elements share a set of attributes. At least one element in each of these clusters is connected to some element in another cluster. These connections indicate the flow of influence between the elements.

Setting priorities by ANP can be summarized in the following steps:

Step 1. *Supermatrix*
A supermatrix is a matrix of all elements by all elements. Paired comparisons with respect to control criteria are needed for all the connections in the model. The ANP derives ratio scale priorities by making paired comparisons of elements by using a 1 to 9 scale of absolute numbers. The weights from the paired comparisons are placed in the appropriate column of the supermatrix. The sum of each column corresponds to the number of comparison sets.

Step 2. *Weighted supermatrix*
The weights in the column corresponding to the cluster are multiplied by the weight of the cluster. Each column of the weighted supermatrix sums to one and the matrix is column stochastic.

Step 3. *Limited supermatrix*
Powers of weighted supermatrix can stabilize after some iterations to limited supermatrix. The columns of each block of the matrix are identical and we can read off the overall priority.

4 Dynamics of ANP Method

An important characteristic of the economic systems is dynamics. Time dependent priorities in the ANP model can be expressed by forecasting using pairwise comparison functions [11] or by predictions based on using of compositional data exponential smoothing [9].

Time-dependent comparisons

Dynamic extensions of ANP method can work with time-dependent priorities in a networked system. There are two approaches for time-dependent pairwise comparisons:

- structural, by including scenarios,
- functional, by explicitly involving time in the judgment process.

For the functional dynamics there are analytic or numerical solutions. The basic idea with the numerical approach is to obtain the time dependent principal eigenvector by simulation [11].

Judgment matrix in dynamic form

$$A(t)=\begin{pmatrix} a_{11}(t) & a_{12}(t)\dots & a_{1k}(t) \\ a_{21}(t) & a_{22}(t)\dots & a_{2k}(t) \\ \vdots & \vdots & \vdots \\ a_{k1}(t) & a_{k2}(t) & a_{kk}(t) \end{pmatrix}. \tag{5}$$

Forecasting using pairwise comparison functions brings a problem with keeping the consistency of paired comparisons. A procedure based on compositional data exponential smoothing was designed, which is suitable for short-term predictions [9].

Compositional Data Analysis

The compositional data are everywhere, where we need to work with data containing only relative information, which is useful for working with weights. Time series for compositional data were analyzed according Brundson and Smith [3].

The following operations are defined on the simplex space in [1].

$$S^k=\left\{x=(x_1,x_2,\dots,x_k), x_i>0, i=1,2,\dots,k, \sum_{i=1}^k x_i=1\right\}. \tag{6}$$

Closure operator $C(x)$: For any vector $x=(x_1,x_2,\dots,x_k)\in R_+^k$

$$C(x)=\left(\frac{x_1}{\sum_{i=1}^k x_i},\frac{x_2}{\sum_{i=1}^k x_i},\dots,\frac{x_k}{\sum_{i=1}^k x_i}\right). \tag{7}$$

Perturbation: For any two vectors from simplex space $x, x\in S^k$

$$x\oplus y=C(x_1y_1,x_2y_2,\dots,x_ky_k)\ . \tag{8}$$

Power transformation: For any vector from simplex space $x\in S^k$ and $\alpha\in R_+$

$$\alpha\otimes x=C\left(x_1^\alpha,x_2^\alpha,\dots,x_k^\alpha\right). \tag{9}$$

Difference:
$$x\ominus y=x\oplus(-1\otimes \mathrm{y}) \tag{10}$$

Exponential smoothing with compositional data can be used for predicting weights in a short time

$$w_t=(\mathrm{w}_{t1},\mathrm{w}_{t2},\dots,\mathrm{w}_{tk}), \mathrm{w}_{ti}>0, i=1,2,\dots,k, \sum_{i=1}^k w_{ti}=1. \tag{11}$$

Simple exponential smoothing

Vector of observations at time t

$$x_t=(x_{t1},x_{t2},\dots,x_{tk}), x_{ti}>0, i=1,2,\dots,k, \sum_{i=1}^k x_{ti}=1. \tag{12}$$

Vector of predictions at time t

$$y_t=(y_{t1},y_{t2},\dots,y_{tk}), y_{ti}>0, i=1,2,\dots,k, \sum_{i=1}^k y_{ti}=1. \tag{13}$$

The formula for simple exponential smoothing of compositional data

$$y_t=\alpha\otimes x_{t-1}\oplus(1-\alpha)\otimes y_{t-1}\ . \tag{14}$$

Double exponential smoothing

Classical double exponential smoothing is generally useful for trend modelling [2]. We introduce for trend modelling a vector of trend values u_t, a vector of slopes v_t, a smoothing constant $0\le\alpha\le1$, a trend constant $0\le\beta\le1$.

Formulas for double exponential smoothing of compositional data

$$u_t = \alpha \otimes x_t \oplus (1-\alpha) \otimes (u_{t-1} \oplus v_{t-1}), \tag{15}$$

$$\mathrm{v}_t = \beta \otimes (u_t \ominus u_{t-1}) \oplus (1-\beta) \otimes v_{t-1}, \tag{16}$$

$$y_t = u_{t-1} \oplus v_{t-1}. \tag{17}$$

Hybrid procedure

For the prediction of time-dependent priorities ANP method we propose a hybrid procedure that combines the benefits of long-term forecasting of pairwise comparison functions and short-term weight predictions using exponential smoothing compositional data. This procedure also mutually enriches both procedures obtaining more accurate data. Both procedures were presented in the previous sections and here we limit ourselves to a brief summary of the hybrid procedure steps.

Step 1: Formulation of pairwise comparison functions.
Step 2: Testing and improving consistency of pairwise comparisons.
Step 3: Collection of historical data by ANP priorities over time.
Step 4: Using of compositional exponential smoothing.
Step 5: Selection of the best coefficient α, β with lowest value of error.
Step 6: Forecasting of priorities for next time periods.
Step 7: Re-formulation of pairwise comparison functions based on short-run model. Go to Step 2.

5 Project portfolio management

Project opportunities come in time and it is necessary to decide which will be accepted for creating a dynamic portfolio of projects and which will be rejected. Project portfolio is set all projects that are implemented in the organization at that time. The basic objectives of the project portfolio management include: optimize the results of the entire project portfolio and not individual projects, the selection of projects to start, interruption or discontinuation of projects, defining priorities for projects, coordinate internal and external sources, organization learning from each other project.

It is generally expected that the portfolio should be designed in such a way as to maximize the possibility of achieving the strategic goals of the company. This is consistent with the notion that portfolio selection problem is a multi-criteria decision making. The main goal of each project is to increase the value of the organization, so most managers prefer financial criteria for project evaluation. The most commonly used indicators include net present value, internal rate of return, payback period, rate of return. In addition to these financial indicators, however, in selecting a portfolio of projects should be taken into account other characteristics, which include for example: the probability of completing the project on time, within budget and within the proposed quality; consistency between strategic and tactical plans; the balance between investment projects and maintenance projects; efficient use of resources; relations between projects; the scope of each project; time-dependent consumption of resources on projects; allocation of expenditure and resources for research and development; allocation of marketing spending and resources.

Lot of professionals tried to find sophisticated way to improve techniques for project management in different ways. The paper presents an approach for dynamic project portfolio management based on a combination of DEMATEL method and the dynamic version of the ANP model. The approach proceeds in two steps. The DEMATEL method is used in the first step to filter of some parts in the model based on expert threshold values. In the second step, the ANP is applied for more precise evaluation of relations by multiple criteria. The structure of the ANP model for dynamic project portfolio (DPP) is described by clusters of elements connected by their dependence on one another. A cluster groups elements (projects, resources, criteria, time) that share a set of attributes. At least one element in each of these clusters is connected to some element in another cluster. These connections indicate the flow of influence between the elements (see Fig. 1).

The ANP model consists of four basic clusters with their elements and influences:

Projects

This cluster consists of potential alternatives of projects of which will be selected a dynamic portfolio. There are priorities among projects for inclusion in the portfolio. The cluster has connections to all other clusters.

Resources

Resources are necessary for the implementation of projects. Main resources are human resources between which are relations important for creating project teams. The cluster has connections to all other clusters.

Criteria

Projects are evaluated according to criteria which include benefits, opportunities, costs, and risks (BOCR)**.** The cluster has connections to all other clusters.

Time

Time is measured in discrete units. Elements of other clusters vary in time and theirs values depend on the values in previous time periods.

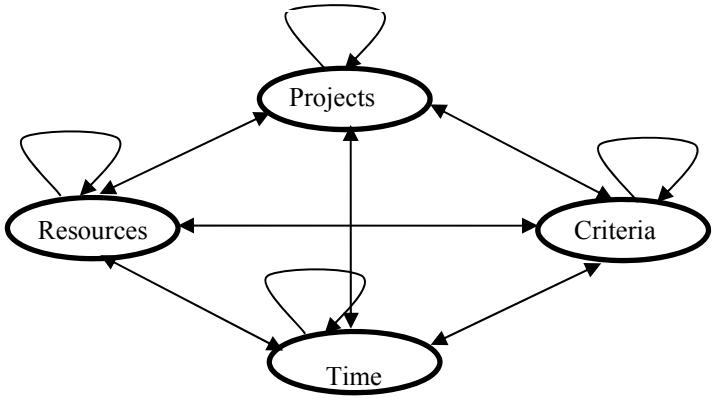


Figure 1 Flows of influence between the elements

Sub-networks

The basic ANP model is completed by specific sub-networks. The sub-networks are used to model important features of the DPP problem. The most important features in our ANP-based framework for DPP management are captured in sub-networks:

- dynamic flow of projects,
- time dependent resources.

Dynamic flow of projects

Project opportunities come in time and it is necessary to decide which will be accepted for creating a dynamic portfolio of projects and which will be rejected. The sub-network connects clusters: time and projects.

Time dependent resources

A specific sub-network is devoted to model time dependent amounts of resources. The time dependent amount of resources is given by. The sub-network connects clusters: time, resources and projects.

6 Conclusions

Network systems are very popular in theory and practice because many of today's systems are characterized by a network structure. Preference elicitation is the key feature for network systems and it is a complex problem. The proposed hybrid procedure is based on a combination of DEMATEL, ANP, and PROMETHEE approaches because these methods complement each other in multi-criteria analysis of network systems. The combination of such approaches can give more complex views on network systems. The presented approach is very flexible. Modifications of the approach are possible according to other structures of combination, other methods, a dynamization of the approach, fuzzy evaluations, and others. Such hybrid approaches can be used in many applications [7], [12].

The paper presents a proposed methodology for dynamic project portfolio management. The basic ANP model with clusters (projects, resources, criteria and time) was created. The proposed ANP model captures the relationships between the clusters and their elements. An important factor of the ANP model is time. The paper proposes a hybrid procedure for time-dependent priority setting. The procedure is based on a combination of pairwise comparison functions and exponential smoothing. The methodology is verified on the projects of an engineering company. The experimental results will affect the specification, completing and extending the model.

Acknowledgements

The research project was supported by the grant No. 13-07350S of the Grant Agency of the Czech Republic.

References

[1] Aitchison, J.: The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B* **44** (1982), 139-177.

[2] Brown, R. G., and Meyer, R. F.: The fundamental theorem of exponential smoothing. *Operations Research* **9** (1961), 673-685.

[3] Brundson, T. M., and Smith, T. M. F.: The time series analysis of compositional data. *Journal of Official Statistics* **14** (1998), 237-253.

[4] Fiala, P.: An ANP/DNP analysis of economic elements in today's world network economy. *Journal of Systems Science and Systems Engineering* **15** (2006), 131–140.

[5] Gabus, A., and Fontela, E.: World problems, an invitation to further thought within the framework of DEMATEL. Battelle Geneva Research Centre, Geneva, 1972.

[6] Levine, H. A.: *Project Portfolio Management*. Jossey-Bas, San Francisco, 2005.

[7] Ou Yang, Z. P., Shieh, H. M., Leu, J. D., and Tzeng, G. H.: A Novel Hybrid MCDM Model Combined with DEMATEL and ANP with Applications. *International Journal of Operations Research* **5** (2008), 160-168.

[8] Rad, P. F., and Levin, G.: *Project Portfolio Management*: *Tools and Techniques*. IIL Publishing, New York, 2006.

[9] Raharjo, H., Xie, M., and Brombacher A. C.: On modeling dynamic priorities in the analytic hierarchy process using compositional data analysis. *European Journal of Operational Research* **194** (2009), 834-846.

[10] Saaty. T. L.: *Decision making with Dependence and Feedback: The Analytic Network Process*. RWS Publications, Pittsburgh, 2001.

[11] Saaty T. L.: Time dependent decision making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables. *Mathematical and Computer Modelling* **46 (**2007), 860-891.

[12] Tzeng, G. H., Chiang, C. H., and Li, C. W.: Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert Systems with Applications* **32** (2007), 1028–1044.

Analytical Paths of Fisher Linear Discriminant Analysis: A Method for Processing Tourism Data

Lukáš Malec[1]

Abstract. Tourism is considered to be one of the significant factors of economic growth. In this study we compared several principal indicators based on overnight stays and arrivals, and consider the most-visited European countries. Annual time series data from the period following the global economic crisis were selected for Fisher linear discriminant analysis (LDA) and consequently processed. This standard approach was generalized to the analytical (resp. smooth) paths of eigenvalues and corresponding eigenvectors depending on the parameter. Such an arrangement gives us one boundary point equal to a standard Fisher discriminant analysis and the other one to eigenvalue decomposition of only the between-sets covariance matrix (plsLDA). While the standard approach in calculus also considers the variance and covariance relations within sets, the analytical paths cover the vanishing of such influence. Many practical outcomes demonstrating the applicability of the given approach are discussed herein. Due to scaling of input data, the distribution of values between maximum and minimum was studied. It was recognized that the countries behave differently considering the results of both boundary points. Theoretical notes on the functional relation by analytical paths form a substantial part of this study.

Keywords: multivariate analysis, path of eigenvalues, functional relation, tourism.

JEL Classification: C01, C33
AMS Classification: 62H30

1 Introduction

Tourism and other related activities are a very important part of economy and also have an impact on the social, cultural and educational sectors of the society. Despite some of the negative effects of tourism, the Tourism Satellite Account results give an indication of tourism's direct significant impact on gross domestic product and employment throughout the European countries. The economic conditions and environment, which act in many directions, can seriously influence the number of arrivals, overnight stay decisions and corresponding per-capita tourism expenditures. Europe is the main tourism destination, with more than 563.4 million (51.8%) international tourism arrivals in 2013 increasing of 5.4% over 2012 [12, 4]. Worldwide events such as global economic crisis (considering years 2008 and 2009) may have seriously disrupted long-term processes and also lead to increased competition between destinations. The effect of the crisis was intensified in 2009 due to the outbreak of the H_1N_1 influenza virus. The various events and their consequences especially affected the amount and variability of international tourists, mostly contributing to national economies. However, as a resilient sector of economy, the recovery of tourism in the changing global environment is relatively very fast.

The evaluation of tourism sector potential over territory and over time is the important task. Current research methods in tourism are very general, considering time series techniques, econometric models as well as some relatively new statistical models such as artificial intelligence techniques. A good summary on modelling and forecasting techniques is published by Song and Li [11]. Despite the fact that publicly available statistical databases in Europe operated by Eurostat or individual national statistical offices cannot serve as a basis for investigating the complete state of tourism, they are still useful to model complex relations and trends.

In this study the twelve most tourism-rich European countries in 2013 were examined. The observations were formed by annual post-crisis data, measured on three variables derived from nights spent and arrival parameters. The method of choice was a Fisher linear discriminant analysis to characterize the differences between sets of quantitative data formed by individual countries using their projection to the space of lower dimension. We extend the standard algorithm to its analytical (resp. smooth) paths dependent on the parameter. Considering such paths as a functional relation, one boundary point was LDA and the other point defined was its partial least

[1] University of Business in Prague/Department of Information Technologies and Analytical Methods, Spálená 76/14, 110 00, Prague 1 – New Town, Czech Republic, email: malec@vso-praha.eu.

squares variant (plsLDA). The partial least squares variant is a simplified version of a standard Fisher discriminant analysis considering only the eigenanalysis of between-sets covariance matrix. In this arrangement, the path of eigenvalues and corresponding eigenvectors describe the way in which the within-sets covariance matrix influence vanishing. In our practical tourism example, the results of both analyses were in some sense opposite. The partial least squares variant of LDA or the in-between point both boundaries can be set as the method of choice in cases of small-sample problem and collinearity or near-collinearity within sets of data.

2 Experimental

2.1 Tourism data

In this study we investigated data gathered by the European Statistical Office [4] covering nights spent and arrivals from the twelve most-visited European countries according to nights spent parameter in 2013.[2] The relatively short 2010–2013 (inclusive) annual time-series data was selected due to its influence on smoothness in the year scale, which provides a better understanding of the linear approach, despite the fact that higher-frequency data would increase the number of repeated observations. Hotels, holiday and other short-stay accommodation, camping grounds, recreational vehicle parks and trailer parks were selected as target tourist accommodation establishments. The individual variables were the number of nights spent (NGT), ratio of non-residents and residents nights spent data (NRR) and total arrivals (ARR). For proper interpretation of results, the input variables were scaled to equal minimum, resp. maximum values individually in all sets.

2.2 Methods

The Fisher linear discriminant analysis is a fundamental method of multivariate techniques applied to various types of data [5], with possible other data sources [2]. However, studying the analytical (resp. smooth) paths of its parameters opens up the possibility of accommodating novel approaches to practical data processing. On the basis of the paths of eigenvalues and corresponding eigenvectors, we demonstrated two boundary points in the dependency on a parameter where one corresponded to a standard Fisher discriminant analysis and the other one to its partial least squares variant. We studied the functional relation of both methods on the basis of classes of analytical (resp. smooth) functions. Similarly to between canonical correlation analysis and partial least squares (PLS-SVD) published by Malec [8], and excluding the inferential aspects, the input assumptions of LDA opposite to its partial least squares variant are very strong [6, p. 296], especially considering the within-sets relations demand. For such reasons, the partial least squares variant of Fisher discriminant analysis or the point in-between both boundaries considering their functional relation can be preferred in cases of small-sample problem, and collinearity or near-collinearity within sets of data. Note that the in-between point also considers one of the many regularization approaches to discriminant analysis.

Due to chosen scaling of input data, the analyses in this special application revealed the differences in the distribution of values between the maximum and minimum rather than studying the data in absolute numbers. The variable in individual set is significant if the extreme value (or, generally, more values) is distant from the others in the sense of being close to the outlying observation.

Discriminant analysis and its partial least squares variants are presently a studied topic [1, 9]. In the study of Barker and Rayens [1] a discriminant analysis was examined on modified form of only between-sets covariance matrix considering some other constraints whose approach is close to our solution. We examined the functional approach which covers the within-sets covariance matrix vanishing expressed as analytical (resp. smooth) paths of eigenvalues and corresponding eigenvectors. It is very important to note that both boundary points can be expressed as individual optimization problems considering the maximum of discriminant variable property. However, the plsLDA approach deals exclusively with only the multivariate mean differences between individual sets of data. In the special case of only two data sets and regular within-sets covariance matrix, just one simple and nonzero eigenvalue and corresponding eigenvector can be expressed on the basis of individual set means differences and inversion of within-sets covariance matrix expressed in short formulas. But the situation at more than two sets is more difficult where the application of spectral theorem is essential. The program is written by the author using the Matlab 7.1 (Mathworks, Natick, MA, USA) software platform.

[2] The list of countries with corresponding two-letter ISO codes (excepting Greece) and ranged in descending order of nights spent is as follows: France – FR, Spain – ES, Italy – IT, Germany – DE, Austria – AT, Netherlands – NL, Greece – EL, Croatia – HR, Poland – PL, Portugal – PT, Sweden – SE and Czech Republic – CZ.

Algorithm

We denote the data matrix centred by individual columns as X of type (n, p). Assume the observations are divided to s sets by individual frequencies n_i measured on p variables. The between-sets covariance matrix B and the within-sets covariance matrix W are introduced according to the standard definition [6, p. 295]. In the following, matrix W is considered regular.

We derive the functional relation between LDA and its partial least squares algorithm in a way close to the standard algebraic solution, see, e.g., [1; 6, pp. 294-306] in dependency on a parameter $\delta \in \langle 0,1 \rangle$

$$\max_{\boldsymbol{a}(\delta) \in R^p} \frac{\boldsymbol{a}'(\delta) B \boldsymbol{a}(\delta)}{\boldsymbol{a}'(\delta) W(\delta) \boldsymbol{a}(\delta)}, \tag{1}$$

where $W(\delta) = (1-\delta)W + \delta I$ and I denotes the identity matrix. At $\delta = 0$ this term provides the optimization problem corresponding to Fisher linear discriminant analysis $\max_{\boldsymbol{a} \in R^p} \frac{\boldsymbol{a}'B\boldsymbol{a}}{\boldsymbol{a}'W\boldsymbol{a}}$ and the boundary point $\delta = 1$ gives exactly the plsLDA solution $\max_{\boldsymbol{a} \in R^p} \frac{\boldsymbol{a}'B\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}$. Both the problems are scaled *a posteriori* to unit norms of coefficients $\boldsymbol{a}$ in this study to form discriminant variables $X\boldsymbol{a}$.

According to the Kuhn-Tucker theorem [7], the numbers λ (Lagrange multipliers) exist in such a way that the solution of (1) is a stationary point of the generalized eigenvalue problem by Krzanowski [6], p. 298

$$B\boldsymbol{a}(\delta) = \lambda(\delta) W(\delta) \boldsymbol{a}(\delta). \tag{2}$$

For the higher-order analyses of optimization problem (1) is simultaneously valid $\boldsymbol{a}_k'(\delta) W(\delta) \boldsymbol{a}_l(\delta) = 0$ considering all $1 \le l < k$ [10, p. 258, theorem 5.9]. We have $r = \min(p, s-1)$ positive eigenvalues which correspond to rank of matrix B. The eigenvalues are sorted as $\lambda_1(\delta) \ge \lambda_2(\delta) \ge \ldots \ge \lambda_r(\delta) > 0$.

The *k*-th order solution of (2) is given by following pairs (paths)

$$\left(\lambda_k(\delta), \boldsymbol{a}_k(\delta)\right), \tag{3}$$

which is herein achieved by applying spectral theorem to a symmetric system using Cholesky decomposition or square root of matrix.[3]

If all positive eigenvalues $\lambda_k(\delta)$, $k = 1, 2, \ldots, r$ are simple (distant) for all δ, then according to Bunse-Gerstner *et al.* [3], it is assured that the paths in (3) are analytical (resp. smooth) functions of a parameter. In case the eigenvectors in (2) satisfy the selected initial condition, the paths (3) are analytical functions on $\delta \in \langle 0,1 \rangle$ which are determined uniquely.

3 Results and Discussion

The input tourism data suffers from relatively great extent of the collinearity within sets, which is why it is suitable to apply a partial least squares variant of the Fisher linear discriminant analysis with proper interpretation of the results. Due to better properties of functional relation, we choose the previous-step within-sets covariance matrix as input to the following step of iteration process, which does not disrupt the investigated analytical (resp. smooth) paths. This approach is dependent on the choice of such a parameter step; we selected the value 0.01. Figure 1 demonstrates the path of eigenvalues which are not intersecting on the given interval. Because the eigenvalues are simple, the paths of corresponding eigenvectors are smooth.

[3] The symmetric problem can be solved by standard spectral or singular decomposition theories.

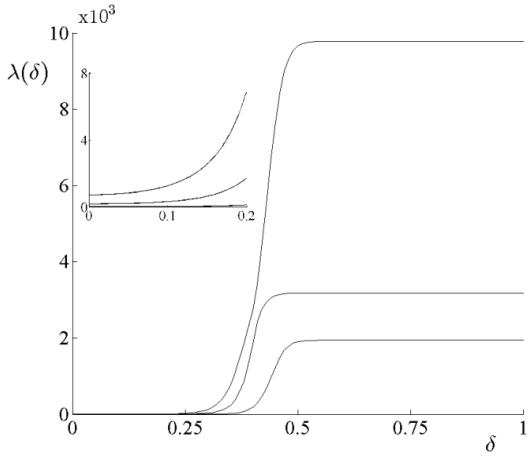


Figure 1 Path of eigenvalues

The coefficients of discriminant variables (DVs) of both the Fisher linear discriminant analysis and its partial least squares variant are introduced in Table 1. It is evident that the first two discriminant variables of LDA comprise a greater share of discriminant significance (98.53%) than its partial least squares variant (86.96%). The first-order coefficients of LDA demonstrate the greatest significance in discriminating the sets for nights spent opposite to arrivals. The second discriminant variable coefficients reveal the significance of nights spent opposite to non-residents and residents nights spent ratio. On the contrary, the situation at partial least squares variant is inversed considering the first two discriminant variables. The case of first discriminant variable is now turned to the second one of LDA, and the second to the first. If we deal with the most important separation, we can conclude that nights spent and arrivals were the most significant in the Fisher discriminant analysis, while the nights spent and the non-residents and residents nights spent ratio was the most important in the partial least squares variant.

	Signif. (%)	NGT	NRR	ARR	Signif. (%)	NGT	NRR	ARR
	LDA				plsLDA			
DV#1	78.77	-0.672	-0.051	0.738	65.62	0.538	-0.745	0.395
DV#2	19.76	0.638	-0.759	0.130	21.34	-0.455	0.137	0.880
DV#3	1.48	0.434	0.832	0.345	13.03	0.709	0.653	0.265

Table 1 Discriminant coefficients

Discriminant variables are a time series, as are the input variables they are obtained from. The score of the first two leading discriminant variables were analysed in this study (see Figures 2 and 3). We can see the separation of Italy on the first discriminant variable of LDA as well as the separation of France on the second one. Thus Italy is distant due to nights spent and arrivals, while France is distant due to the nights spent and the non-residents and residents nights spent ratio. On the other hand, the Czech Republic, Poland and Croatia were the most similar. The points corresponding to the years 2012 and 2013 in Italy, 2012 in Netherlands and 2013 in Sweden were considered the most outlying. The situation at the partial least squares variant was again inversed. The first discriminant variable separates mainly France from the rest, while on the second discriminant variable Italy is distant from the others.

4 Conclusion

It is demonstrated in this study that an analytical (resp. smooth) extension of the standard Fisher linear discriminant analysis works well in practical processing situation. Despite the great applicability of standard algorithm, in our study with a small number of observations and consequent near-collinearity within sets, the partial least squares variant should be a method of choice; preferably the combination of both methods can be interpreted. The Fisher discriminant analysis considers in computation the within-sets covariance matrix which is not the case in plsLDA, where only the differences of means between sets in multivariate arrangement are investigated. Our approach of scaling gives not the discriminant information on absolute values of input variables considering the individual sets of tourism data, but study the distribution of values between their maximum and minimum.

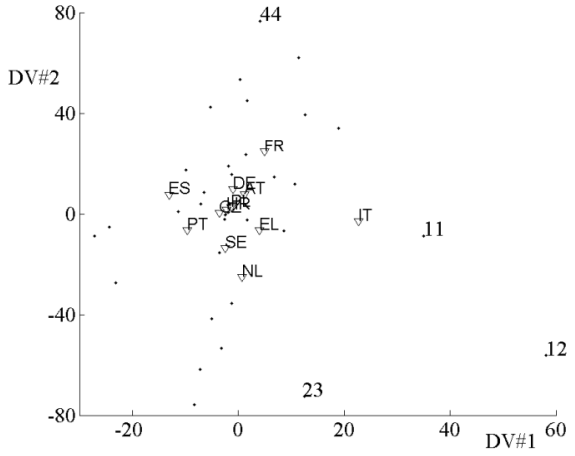


Figure 2 LDA score

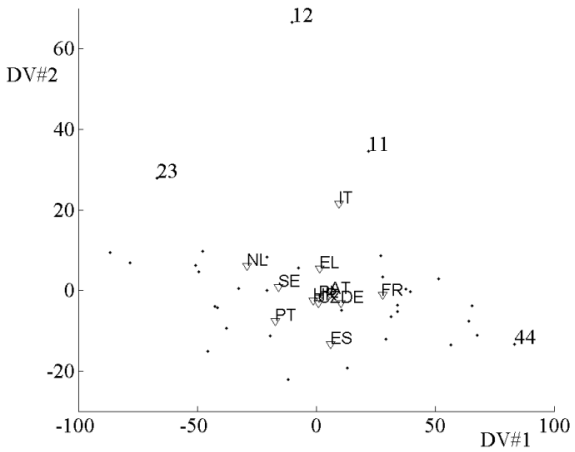


Figure 3 plsLDA score

The statistical benefit of partial least squares variant is comparable to the standard algorithm. If we deal with the most important discrimination considering our practical example, nights spent and arrivals are the most significant in LDA, while the nights spent and the non-residents and residents nights spent ratio are the most important in its partial least squares variant. The score of LDA reveals Italy as the most separated, while the partial least squares variant reveals France as the most distant from the others. Future studies can be directed to select the point in-between both boundaries on the basis of eigenvalue (resp. corresponding eigenvector) paths with a possible interpretation of results. Also the use of forms other than identity matrix in the denominator of optimization problem (1) as one boundary excluding the proper scaling is possible, e.g. considering only the diagonal of within-sets covariance matrix.

Acknowledgements

This study was supported by the internal grant FRV No. 3/2013 of the University of Business in Prague.

References

[1] Barker, M. and Rayens, W.: Partial least squares for discrimination. *Journal of Chemometrics* **17** (2003), 166–173.
[2] Bílková, D.: Development of wage distribution of the Czech Republic in recent years by highest education attainment and forecasts for 2011 and 2012. In: Löster, T. and Pavelka, T. (Eds.). 6th International Days of Statistics and Economics, Prague, Czech Republic, 2012, 162–182.
[3] Bunse-Gerstner, A., Byers, R., Mehrmann, V. and Nichols, N. K.: Numerical computation of an analytic singular value decomposition of a matrix valued function. *Numerische Mathematik* **60** (1991), 1–39.
[4] European Statistical Office. Eurostat database, Luxembourg, Luxembourg, 2015. Retrieved from http://ec.europa.eu/eurostat/data/database.
[5] Kiráľová, A. and Malec, L.: Communication and managerial competencies as a prerequisite for employability of graduates in tourism. In: Soliman, K. S. (Ed.). The 24th International Business Information Management Association Conference, Milan, Italy, 2014, 477–487.
[6] Krzanowski, W. J.: Principles of multivariate analysis: A user's perspective. (2nd ed.). Oxford University Press, New York, 2000.
[7] Kuhn, H. W. and Tucker, A. W.: Nonlinear programming. In: Neyman, J. (Ed.). Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, 1951, 481–492.
[8] Malec, L.: On the multivariate processing of rank-deficient tourism data. *European Journal of Tourism, Hospitality and Recreation* **4** (2013), 181–203.
[9] Pérez-Enciso, E. M. and Tenenhaus M.: Prediction of clinical outcome with microarray data by a partial least squares discriminant analysis (PLS-DA) approach. *Human Genetics* **112** (2003), 581–592.
[10] Seber, G. A. F.: Multivariate Observations. John Wiley & Sons, New York, 2004.
[11] Song, H. and Li, G.: Tourism demand modelling and forecasting – A review of recent research. *Tourism Management* **29** (2008), 303–320.
[12] United Nations World Tourism Organization. Tourism highlights, Madrid, Spain, 2014. Retrieved from http://dtxtq4w60xqpw.cloudfront.net/sites/all/files/pdf/unwto_highlights14_en_hr_0.pdf.

Stable Distributions and Czech Financial Market

Jiří Málek[1], Quang Van Tran[2]

Abstract. We investigate the thickness of tails of probability density function (pdf) of exchange rates CZK/EUR, CZK/USD and Prague stock market index PX on daily, weekly and monthly frequency using α-stable distributions. It is shown that α is in the all cases less than two, which means that their tails are thicker compared with the one of a normal distribution. For estimations we use two quantile methods and recursive regression algorithm. With two exceptions coefficient α increases with the increasing time interval. The lowest one (and therefore the thickest tail) was estimated by using recursive regression algorithm for the exchange rate CZK/USD on a daily basis namely 1.0456. Both quantile methods also estimate α CZK/USD on a daily basis as a low but higher than the recursive regression algorithm.

Keywords: stable distribution, α-parameter, fat tails, quantile methods, recursive regression algorithm.

JEL Classification: G10, G120
AMS Classification: 91B82, 91B28

Introduction

Normal distribution due to its simplicity and easy processing is often used for modeling the returns on financial instruments. However, it appears that the tails of the empirical distributions tend to be thicker than the normal ones, which in turn has implications in the management of financial risks: The probability of the loss is much higher than in a normal distribution.

The α-stable distribution can be a useful tool that replaces the normal distribution, since it allows to model variable power tails. In addition, according to the Generalized Central Limit Theorem these distributions may be the standardized limit of the sum independent identical distributed random variables (with the same tail index).
In this paper we analyze the log-returns of some instruments of Czech financial market (exchange rates CZK/EUR, CZK / dollar and stock index PX). There are made the estimates of the tail indexes at different time frequencies (day, week, month). The first part gives a definition of the α-stable distribution and its basic properties. In the next section some methods are applied and compared to estimate the alpha parameters.

1 Stable distributions

1.1 Definition of stable distribution

Let $X, X_1, X_2, X_3, \ldots, X_n$, i.i.d.
A random variable X is said to have the α-stable distributions if there is for any $n \geq 2$ a positive number c_n and a real number d_n such that

$$X_1 + X_2 + \cdots + X_n \overset{d}{=} c_n X + d_n \quad .$$

Thus, any sum of independent equally distributed random variables have the same distribution except for the „mean“ and „variance“. Unfortunately there is no general form of the probability density function, we know only the general form of the characteristic function:

$$\Phi\ t\ = \exp\left\{-\sigma^{\alpha}|t|\left(1-i\beta \operatorname{sign}\ t\ \tan\frac{\pi\alpha}{2}\right)+i\mu t\right\}\ \text{ for } \alpha \neq 1 \qquad (1)$$

$$\Phi\ t\ = \exp\left\{-\sigma^{\alpha}|t|\left(1+i\beta\frac{2}{\pi}\operatorname{sign}\ t\ \log|t|\right)+i\mu t\right\}\ \text{ for } \alpha = 1$$

,

[1] Department of Banking and Insurance, University of Economics Prague, malek@vse.cz.
[2] Department of Statistics and Probability, University of Economics Prague, tran@vse.cz.

where parameters $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\mu \in \mathbb{R}$ and $\sigma \in (0,\infty)$ [3], and will label this distribution as $S(\alpha,\beta,\sigma,\mu)$

So the α - stable distributions require four parameters. These parameters may be interpreted as

α…tail power tail index , as α decreases tail thickness increases

β…skewness prametr, determines asymetry, a positive β indicates that the right tail is fatter than the left tail and vice versa, β=0 corresponding to a symetric distribution

μ…location parametr (coresponding to a mean value for $\alpha > 1$)

σ…scale parametr generalized a standard deviation, for $\alpha = 2$ corresponding to a standard deviaton of normal distribution

Motivation for financial market

Let $S_1, S_2, \ldots S_n, S_{n+1}$ is time series of prices of traded instrument (for example stocks). The log-returns are $R_1, R_2, \ldots R_n$ where

$$R_i = \log \frac{S_{i+1}}{S_i}.$$

Then

$$R = \log \frac{S_{n+1}}{S_1} = \sum_{i=1}^{n} R_i$$

Examples (for the situation where the exact formula for pdf exists)

Normal distribution (for $\alpha = 2$)
Cauchy distribution (for $\alpha = 1, \beta = 0$)
Lévy distribution (for $\alpha = 0.5, \beta = 1$)
We will be interested mainly for the tail index which approximately means that[4] $P(X > x) \approx c_\alpha x^{-\alpha}$
One can show that for $\alpha \in (0,2]$ there is no a variance and for $\alpha \in (0,1]$ there is no a mean value.

1.2 Simulation

Since there is no explicit formula for the probability density, it is a problem to simulate α -stable distribution. Although it is possible by the numerical inverse Fourier transform methods to find the density, and then again by numerical integration to find the cumulative distribution function and consequently its inversion, but it introduces into the computation process too many errors. We introduce a method of Chambers et al. [2] , Weron [10]

for $\alpha > 1$:

Let γ is uniformly distributed on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $\gamma_0 = -\frac{\pi}{2}\beta\frac{\alpha-2}{\alpha}$, and W is independent exponentially distributed random variable with parameter $\lambda = 1$, then

$$X = \frac{\sin\alpha(\gamma-\gamma_0)}{\cos\gamma^{1/\alpha}}\left[\frac{\cos(\gamma-\alpha(\gamma-\gamma_0))}{W}\right]^{\frac{1-\alpha}{\alpha}}$$

has a distribution $S(\alpha,\beta,1,0)$. This distribution can be transformed at general stable distribution $S(\alpha,\beta,\sigma,\mu)$ (see Weron [10])

2 Parameters estimation

To estimate the parameters there are used first two quantiles method Fama-Roll [3], [4]) and McCulloch [7]. Fama-Roll Method is considered as a simple one and easy to use. First they estimate σ by

[3] There are other equivalent representations of the characteristic function, which may be advantageous in certain situations (see Nolan [9]).
[4] The exact formula for c_α can be found in Nolan [9].

$$\sigma = \frac{x_{0.72} - x_{0.28}}{1.654}$$

(where x_f referred to f-th quantile) and then he finds α so that cdf is equals f (we use $f = 0.9$) which we may write as :

$$S_\alpha\left(\frac{x_f - x_{1-f}}{2\sigma}\right) = f\ .$$

McCulloch extends the previous method and weakens the restrictions on the parameters. He defines

$$\upsilon_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}$$

which is independent of both σ and μ and

$$\upsilon_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.5}}{x_{0.95} - x_{0.05}}$$

υ_β is also independent of both σ and μ, so υ_α, υ_β are functions of α and β. This relations may be inverted by tabulated function $\alpha = \psi_1(\upsilon_\alpha, \upsilon_\beta)$, $\beta = \psi_2(\upsilon_\alpha, \upsilon_\beta)$ (see McCuloch [7], Weron [10]).

The third method is based on the sample characteristic function (see Koutrouvelis [5], [6], Weron [10]). From (1) we can derive

$$\log(-\log|\Phi(t)|^2) = \log(2\sigma^\alpha) + \alpha\log|t|\ .$$

Using an sample characteristic function we have linear regression relation

$$Y_k = a + \alpha X_k + \varepsilon_k$$

where $a = \log(2\sigma^\alpha)$, $X_k = \log|t|$ and t is a real number. This method is considered more accurate than quantile methods.

We used daily, weekly and monthly data on exchange rates CZK /EUR, CZK /USD and Prague stock market index PX.

2.1 Data description

We use three methods described in the previous section to estimate value of α from the Czech financial market. They are series of Prague Stock Market Index PX with three recording frequencies: daily (denoted as PXd), weekly (PXw) and monthly (PXm) from 1.1996 to 4.2015 and two series of spot exchange rate of Euro and UD dollar versus Czech crown: Euro vs Czech crown with the same frequencies: daily (EUd), weekly (EUw) and monthly (EUm) from 9.1999 to 4.2015, US Dollar vs Czech crown: daily (USd), weekly (USw) and monthly (USm) from 10.1995 to 4.2015. All original data series have been transformed into logarithmic differences series. The following two tables 1 and 2 provide basic descriptive statistics of them. Among others, it is clear that none of them is generated from a normal distribution.

	PXd	**PXw**	**PXm**	**EUd**	**EUw**
Mean	0.000182	0.000846	0.003838	-7.12E-05	-0.000349
Median	0.000518	0.002055	0.009831	-5.85E-05	-0.000417
Maximum	0.123641	0.155728	0.204956	0.027053	0.047867
Minimum	-0.161855	-0.304534	-0.316453	-0.022191	-0.067372
Std. Dev.	0.013980	0.031399	0.069135	0.002980	0.008591
Skewness	-0.449656	-1.079625	-0.890201	0.380675	-0.212733
Kurtosis	14.65269	13.81408	6.094168	11.32869	10.86853
Jarque-Bera	27517.98	5082.142	122.6582	11858.88	2100.872
Probability	0.000000	0.000000	0.000000	0.000000	0.000000
Observations	4835	1003	231	4069	812

Table 1 Descriptive statistics of investigated series I

	EUm	**USd**	**USw**	**USm**
Mean	-0.001392	-2.83E-06	-3.58E-05	-0.000146
Median	-0.001022	0.000000	-0.000848	-4.31E-05
Maximum	0.060144	0.076630	0.081625	0.141403

Minimum	-0.047791	-0.051279	-0.073802	-0.105031
Std. Dev.	0.016653	0.005979	0.017491	0.037320
Skewness	0.506696	0.426050	0.291897	0.328619
Kurtosis	4.566688	13.72495	4.399722	3.736040
Jarque-Bera	26.98143	34291.11	97.27241	9.453176
Probability	0.000001	0.000000	0.000000	0.008857
Observations	186	7110	1015	233

Table 2 Descriptive statistics of investigated series II

2.2 Estimation of alpha

	α	lower bound	upper bound
PXd	1.5749	1.5158	1.6316
PXw	1.6126	1.4773	1.7742
PXm	1.6136	1.2071	2
EUd	1.4429	1.3914	1.4856
EUw	1.5579	1.4319	1.6958
EUm	1.6034	1.3313	2
USd	1.3125	1.2771	1.3519
USw	1.6797	1.5434	1.9193
USm	1.7670	1.4679	2

Table 3 Estimation results for using method suggested by McCulloch

	α	lower bound	upper bound
PXd	1.5566	1.4892	1.6042
PXw	1.6689	1.5506	1.8564
PXm	1.7009	1.4694	2
EUd	1.4800	1.4885	1.5048
EUw	**1.5621**	1.4748	1.6894
EUm	**1.4898**	1.1828	1.7780
USd	1.2621	1.2257	1.3792
USw	1.6806	1.5546	1.8945
USm	1.7919	1.4890	2

Table 4 Estimation results for using method suggested by Fama and Roll (f =0.95)

	α	lower bound	upper bound
PXd	1.6407	1.5961	1.6856
PXw	**1.6880**	1.5986	1.7843
PXm	**1.6196**	1.1426	1.7940
EUd	1.3199	1.2623	1.4313
EUw	1.6380	1.2824	1.7533
EUm	1.6732	1.1648	1.8976
USd	1.0456	1.0192	1.0778
USw	1.7871	1.6962	1.8779
USm	1.8225	1.6414	2

Table 5 Estimation results for using recursive regression algorithm

We use quantile methods proposed by McCulloch and Fama and Roll to estimate α. Its corresponding confidence intervals are calculated by using bootstrap technique. The results are shown in Tables 3,4, and 5.

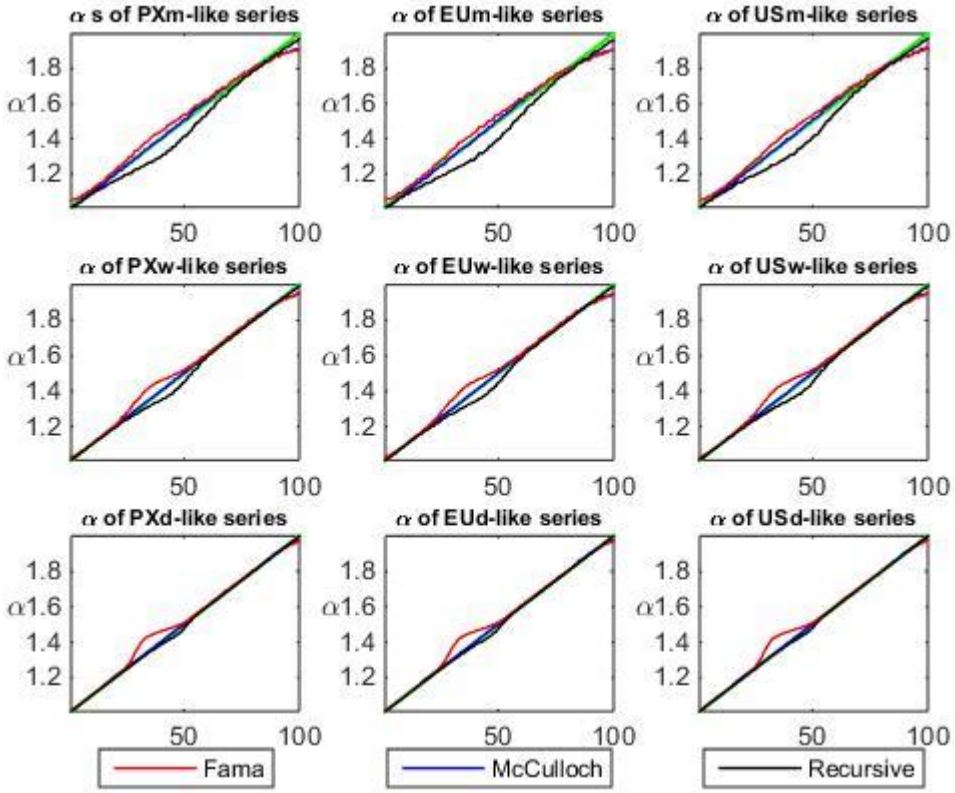


Figure 1 The accuracy comparison of three methods used to estimate α

Further, to examine the accuracy of three methods used to estimate parameter α, we generate a thousand series with the same length as well as the mean and standard error as a series of our interest. We use again these methods to calculate value of parameter α from each series. Then the mean and the standard error of estimated parameter α are calculated. The results are shown in Figure 1[5].

Regarding value of parameter α, with two exceptions, it always decreases with the increasing frequency of recording or data type" daily, weekly or monthly. This confirms the validity of hypothesis that the rate of return of a financial asset on a longer time scale is close to a normal distribution. However, the difference between weekly and monthly time intervals are usually smaller than the difference between the daily and weekly interval. The lowest (and therefore the thickest tail) was estimated by using recursive regression algorithm for the exchange rate CZK/USD on a daily basis namely 1.0456, which is very close to the value of α for Cauchy distribution, for which the probability of extreme events approximately decreases polynomially by order 2 while with a normal distribution it lowers exponentially. Value α of daily changes of exchange rate CZK/USD estimated by both quantile methods is low but higher than the recursive regression algorithm. Interestingly, the highest value of α is obtained from series of monthly changes of exchange rate CZK /USD and it is also determined by recursive regression algorithm (quantile methods confirm this one, though with a lower value). A possible explanation is that the market gets messy information (or so it interprets them in this way), which subsequently leads to the occurrence of more extreme values and these ones are gradually corrected over time. It may also be a short-term (exaggerated) effect of behavior of market participants, which is gradually dying out.

As far as the accuracy of each method used in this experiment is concerned, in Figure 1 it is clear that the quantile proposed by Fama and Roll (the red line in figure 1) tends to overestimate the values of α while the recursive method (the black line) often underestimate it in a wide range of value α. However, the interval of overestimation and underestimation narrows when more observations are available. The quantile method suggested by McCulloch provides a relatively accurate estimate of α over almost the whole interval (1,2). But as α increases toward 2 (approximately from 1.9), both quantile methods underestimate α and the recursive method gives us the most accurate values of α. Knowing this, we assume that the higher value of α estimated by Fama's method from

[5] Actually, there are four line in Figure 1. The green line is the line starting at the origin and capturing the true values of parameter α.

weekly data of exchange rate EUR/CZK series (1.56) and monthly series (1.49) in table 3 as well as the similar case of weekly and monthly Prague stock exchange index returns series (1.69 vs 1.62) in table 5 using recursive method may not be irregularities against hypothesis of closeness to normality of financial asset returns on a longer time interval. They may come from the inaccuracy of these two methods and the hypothesis as such may hold. The result shows that when estimating α, one may use one method for a certain range and another for different one.

3 Conclusions

In this research, we have applied the α-stable distribution, which should be a more appropriate representation for capturing extreme events than a normal distribution, on real data from the Czech financial market. They are daily, weekly and monthly data on exchange rates CZK /EUR, CZK /USD and Prague stock market index PX. We estimate values of α which characterizes the thickness of the tails of the distribution using three different methods and examine their accuracy. Our results show that all distributions have α less than 2 which means that they are not a normal distribution and the values of α usually decrease when data frequency increases. They also indicate that McCulloch's and the recursive methods are superior to the one suggested by Fama and Roll the exact choice of a method to be used to estimate α depends on the range α comes from.

Acknowledgements

Author Jiří Málek acknowledges the financial support of Czech Science Foundation with grant P402/12/G097 and Institutional support IP 100040/1020. Author Tran Van Quang is grateful for the financial support of grant GAČR P402/12/G097 of Czech Science Foundation.

References

[1] Belov I. A.: On the Computation of the Probability Density Function of Stable Distribution Chambers, *Mathematical Modelling and Analysis 2005*. Pages 333-341
[2] Chambers J. M., Mallows C. L., and Stuck B.W.: A method for simulating stable random variables, *J. Amer. Statist. Assoc.* **71** (1976), 340-344
[3] Fama, E. and Roll, R.: Some Properties of Symmetric Stable Distributions. *J. Amer. Statist. Assoc.* **63** (1968), 817–836.
[4] Fama, E.,and Roll R.: Parameter Estimates for Symmetric Stable Distributions. *J. Amer. Statist. Assoc*. 66 (1971), 331–338.
[5] Koutrouvelis, I. A.: Regression–Type Estimation of the Parameters of Stable Laws. *J. Amer. Statist. Assoc*. **75** (1980), 918–928.
[6] Koutrouvelis, I. A.: An Iterative Procedure for the Estimation of the Parameters of the Stable Law. *Commun. Statist.–Simula*. **10** (1981), 17–28.
[7] McCulloch, J. H.: Simple Consistent Estimators of Stable Distribution Parameters. *Commun. Statist.–Simula.* **15** (4) (1986), 1109–1136.
[8] McCulloch, J. H.: Measuring Tail Thickness to Estimate the Stable Index α: A Critique. *Journal of Business and Economic Statistics* **15**, No 1(1997)
[9] Nolan, J. P: Modeling Financial Data with Stable Distribution. In *Handbook of Heavy Tailed Distributions in Finance*, Edit by Rachev S.T., Elsevier science, 2003.
[10] Weron, R.: *Performance of the Estimators of Stable Law Parameters*, Wroclav University of Technology, 1995.

ARIMA models and exponential smoothing

Lubos Marek, Michal Vrabec[1]

Abstract. Exponential smoothing is one of many approaches to analyzing time series data. It is usually applied to modeling the trend component of time series (it can, however, be used for other components of time series as well) which are not stable over time and change rapidly. If this is the case, the usual decomposition cannot be used and other ways of modeling are sought, such as ARIMA or other stochastic models. There is, however, a connection between the two. In this paper we show that for most of the exponential smoothing models an ARIMA counterpart exists (and vice versa). The theoretical duality of these models is shown, including some inference. Both models are applied to simulated data in several software packages (SAS, eViews, and Statgraphics) and the results are compared. This way the issue is illuminated to the readers, who are then able to compare the considered software packages when working with similar models.

Keywords: time series, ARIMA, exponential smoothing.

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

We can choose from among several basic approaches to analyzing time series, selecting the suitable method according to the particular evolution of the time series in question. In many cases standard decomposition is sufficient, based on the systematic components of a time series (the trend and the seasonal and cyclic components). If the character of the time series is changing quickly and the behavior cannot be described by a single trend function over the entire period, an adaptive model of the trend can be employed – exponential smoothing or moving averages. Another, theoretically much more demanding, option uses a stochastic model for the time series. Here we have in mind AR, MA, ARMA, ARIMA or SARIMA models. The theoretical aspects of such methods make the two approaches seem to be diametrically different from each other.

The goal of the present paper is to show that such a sharp contrast only appears to exist, and the methods in fact do have a common core. Namely, there is a relationship between the exponential smoothing and the ARIMA method. For certain models of exponential smoothing, we are able to find their counterparts in the form of suitable ARIMA models. And vice versa, of course: certain ARIMA models can be transformed to exponential-smoothing ones by means of suitably chosen smoothing constants. We will describe the theoretical duality of both of these approaches and, on simulated data, we will show the results produced by different methods. More detailed theoretical information can be found, e.g., in [1], [2], and [6]. Statistical software packages SAS, eViews, and Statgraphics were used for practical analyses. In the entire text, the classical notation introduced by the authors in [2] and [6] will be used.

2 Methodology

2.1 Simple (Brown's) exponential smoothing

Let us consider model ARIMA (p, d, q) in the notation of [2] and set $p = 0$, $d = 1$, and $q = 1$ for its parameters. In other words, we have model ARIMA (0, 1, 1) in the form

$$(1-B)Y_t = (1-\theta B)\varepsilon_t \tag{1}$$

or equivalently

$$Y_t = Y_{t-1} + \varepsilon_t - \theta\,\varepsilon_{t-1}\,, \tag{2}$$

where $-1 < \theta < 1$. This process is sometimes also denoted by IMA (1, 1). If the theta parameter is set equal to zero, i.e., $\theta = 0$, we get the well-known model of the random walk. For $-1 < \theta < 1$ we can write

[1] University of Economics, Prague/Department of Statistics and Probability, W. Churchill sq. 4, Prague 3, Czech Republic, marek@vse.cz.

$$\frac{(1-B)}{(1-\theta B)} = (1-B)(1+\theta B+\theta^2 B^2+...) = 1+\theta B+\theta^2 B^2+...-B-\theta B^2-...= \\ =1-(1-\theta)B-(1-\theta)\theta B^2-(1-\theta)\theta^2 B^3-...=1-\alpha B-\alpha(1-\alpha)B^2-\alpha(1-\alpha)^2 B^3-... \tag{3}$$

where $\alpha=(1-\theta)$. Since $-1<\theta<1$, we get $0<\alpha<2$ for α. Hence it is

$$Y_t=\alpha\sum_{j-1}^{\infty}(1-\alpha)^{j-1}Y_{t-j}+\varepsilon_t. \tag{4}$$

The last formula is nothing else but writing the process in autoregression form. Estimating Y_t, we get

$$\hat{Y}_t=\alpha\sum_{j=1}^{\infty}(1-\alpha)^{j-1}Y_{t-j}. \tag{5}$$

A closer look at Formula (5) reveals that the estimated value of Y_t at time t is the weighted average of the past values of the time series Y_{t-1}, Y_{t-2}, ... with the weights exponentially decreasing with each step to the past. In fact, Formula (5) can be further re-written as

$$\hat{Y}_{t+1}=\alpha\sum_{j-1}^{\infty}(1-\alpha)^{j-1}Y_{t+1-j}=\alpha Y_t+(1-\alpha)\alpha\sum_{j=2}^{\infty}(1-\alpha)^{j-2}Y_{t+1-j} \\ =\alpha Y_t+(1-\alpha)\alpha\sum_{i=1}^{\infty}(1-\alpha)^{i-1}Y_{t-i}=\alpha Y_t+(1-\alpha)\hat{Y}_t. \tag{6}$$

This new estimate Y_{t+1} is therefore equal to the weighted average of the new observation at time t on the one hand, and the estimate of the new observation on the other hand. Coefficient α is called the level smoothing constant in the **simple (Brown's) exponential smoothing**. We can thus see that the Brown's exponential smoothing model can be replaced with the ARIMA (0, 1, 1) model.

2.2 Double (Brown's) exponential smoothing

Let us again consider an ARIMA model, now with parameters $p=0$, $d=2$, and $q=2$. It means

$$(1-B)^2 Y_t=(1-\theta B)^2\varepsilon_t \tag{7}$$

or equivalently

$$Y_t=2Y_{t-1}-Y_{t-2}+\varepsilon_t-(1+\theta)\varepsilon_{t-1}+\theta\varepsilon_{t-2}. \tag{8}$$

The last Formula can be re-written as

$$Y_t=2Y_{t-1}-Y_{t-2}+\varepsilon_t-\varepsilon_{t-1}+\theta(\varepsilon_{t-2}-\varepsilon_{t-1}). \tag{9}$$

This model is sometimes called ARIMA (0, 2, 2) in the literature, but this is inaccurate: the true ARIMA (0, 2, 2) model has two parameters of moving averages– cf. the following Section. Nevertheless, a suitable exponential-smoothing counterpart for the model can be found. Set $\alpha=(1-\theta)$. Because of invertibility, $-1<\theta<1$, and therefore also $0<\alpha<2$, must hold. Model (7) can hence be viewed as a model of **double (Brown's) exponential smoothing** – cf. [3], [4], and [5].

2.3 Holt's exponential smoothing

In the standard ARIMA (*p, d, q*) model, let us set $p=0$, $d=2$, and $q=2$. In other words, we consider the ARIMA (0, 2, 2) model, whose form is

$$(1-B)^2 Y_t=(1-\theta_1 B-\theta_2 B^2)\varepsilon_t \tag{10}$$

or equivalently

$$Y_t=2Y_{t-1}-Y_{t-2}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}. \tag{11}$$

Inequalities $-1<\theta_2<1$, $\theta_2+\theta_1<1$, $\theta_2-\theta_1<1$ *hold for the parameter values* (expressing the condition that the process should be invertible). The resulting process is sometimes denoted by IMA (2, 2).
By considerations similar to the case of the IMA (1, 1) model we establish that IMA (2, 2) can be written as a model of exponential smoothing. Setting

$$\theta_1=2-\alpha-\alpha\gamma \text{ and } \theta_2=\alpha-1, \tag{12}$$

we can express the IMA (2, 2) model in the form of **Holt's exponential smoothing** – cf., e.g., [4] and [5]. Parameter α is the smoothing constant for the level of the time series, and parameter γ is the smoothing constant for the trend of the time series.

2.4 General ARIMA (p, d, q) model in comparison with exponential smoothing

On a more general note, a relationship between an ARIMA (p, d, q) model and types of exponential smoothing other than those considered above can be proven (cf. [1]).

a) Let us consider a seasonal ARIMA model

$$(1-B)(1-B)^{L}Y_t = (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_{L+1}B^{L+1})\varepsilon_t \tag{13}$$

where L is the season length. Let assume for the theta parameters:

$$\theta_1 = 1-\alpha-\alpha\gamma,\ \ \theta_2=\theta_3=\ldots=\theta_{L-1}=-\alpha\gamma,\ \ \theta_L=1-\alpha\gamma-\delta(1-\alpha),\ \text{and}\ \ \theta_{L+1}=(1-\alpha)(\delta-1). \tag{14}$$

Model (13) can now be transformed to Winters' exponential smoothing if the seasonal model is additive, and parameters α, γ, δ are the standard smoothing constants in this model: α is the level smoothing constant, γ the trend smoothing constant, and δ the seasonality smoothing constant – see [5] for more details.

The requirement that the seasonal ARIME model should be invertible implies the following constraints on parameters α, γ, and δ:

$$\max(0,-L\alpha)<\delta(1-\alpha)<2-\alpha,\ \text{and}\ 0<\alpha\gamma<\left[2-\alpha-\delta(1-\alpha)\right](1-\cos\gamma). \tag{15}$$

b) For triple exponential smoothing, the equivalent ARIMA model is

$$(1-B)^3 Y_t = (1-\theta B)^3 \varepsilon_t .$$

c) For general *k*th degree exponential smoothing, the equivalent ARIMA model is

$$(1-B)^k Y_t = (1-\theta B)^k \varepsilon_t .$$

d) For exponential smoothing with the smoothing function given as

$$f(t) = \beta_0+\beta_1 t+\beta_2 sin(2\pi t/12)+\beta_3 cos(2\pi t/12)+\beta_4 t\ sin(2\pi t/12)+\beta_5 t\ cos(2\pi t/12)$$

the equivalent ARIMA model is

$$(1-B)^2(1-\sqrt{3}B+B^2)^2 Y_t = (1-\theta B)^2(1-\theta\sqrt{3}B+\theta^2B^2)^2\varepsilon_t$$ (1 − B)² (1 - √3 B + B²)² .

e) For exponential smoothing with the smoothing function given as

$$f(t) = \varphi^t \ \text{ pro } |\varphi|<1$$

the equivalent ARIMA model is

$$(1-\varphi B)Y_t = \left(1-\frac{\theta}{\varphi}B\right)(1-\theta B)\,\varepsilon_t .$$

f) For exponential smoothing with the smoothing function given as

$$f(t) = \beta_0+\beta_1\varphi^t \ \text{ pro } |\varphi|<1$$

the equivalent ARIMA model is

$$(1-\phi B)(1-B)Y_t = \left(1-\frac{\theta}{\phi}B\right)(1-\theta B)\,\varepsilon_t .$$

Equivalent ARIMA models could, in a similar way, be found for types of time series models other than exponential smoothing (such as certain seasonal models, etc.).

3 Examples

Theoretical relationships between exponential smoothing and ARIMA models are described in the Sections above. Such relationships have their theoretical validity, but a question remains, namely, what the practical applications of such models are. In particular, we would like to know to what extent, if any, the predictions based on a specific ARIMA model differ from those based on its exponential-smoothing counterpart. At the same time,

we compare performance of different statistical software packages in dealing with such problems. Namely, SAS, eViews, and Statgraphics are used to accomplish this. In all software packages we adhere to unified conditions as far as possible, choosing the same models, the same methods for parameter estimation (that is, conditional least squares), the same methods for building predictions, etc. Had our primary goal been comparison between software packages, we might compare a number of different characteristics of the model outputs. However, we restrict ourselves to a "user approach" and only compare parameter values and predictions one step ahead. Generated sequences of 200 observations will be used. The random component is the same in all of our examples, with the normal distribution of zero mean and constant dispersion.

3.1 ARIMA vs. simple Brown's exponential smoothing

In order to compare these two models we generate a time series of 200 observations from the ARIMA (0, 1, 1) model with parameter $\theta = 0.5$. This series is therefore determined by the formula

$$Y_t = Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}. \tag{16}$$

As stated above, this ARIMA (0, 1, 1) model has its counterpart in simple Brown's exponential smoothing with the smoothing constant equal to $\alpha = (1-\theta) = 1-0.5 = 0.5$. We will, however, set the smoothing constant in Brown's model equal to the estimated value of α calculated from the estimated value of θ according to the formulae mentioned above, depending on the respective software. Let us see how different packages coped with estimating θ and predicting one step ahead. These results are shown in Table 1.

	ARIMA		Brown (simple)	
	θ	prediction	α	prediction
eViews	0.53406	3.24650	0.46594	3.91637
Statgraphics	0.53371	3.24697	0.46629	3.79251
SAS	0.53215	3.24910	0.46785	3.91727

Table 1 Comparison of results

These results imply that the results of all packages are fully comparable with regard to estimating θ (differences only occur on the third decimal place). Therefore the values of smoothing constant α, calculated on the basis of θ, are also similar. For predictions the situation is different: while the ARIMA predictions are nearly identical, the Statgraphics prediction somewhat departs from the other two for Brown's exponential smoothing. However, even this departure is small and the results can be deemed comparable.

3.2 ARIMA vs. double Brown's exponential smoothing

Let us again consider a time series of 200 observations generated from the ARIMA model with parameter $\theta = 0.2$, namely

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - \varepsilon_{t-1} + 0.2(\varepsilon_{t-2} - \varepsilon_{t-1}). \tag{17}$$

The above-mentioned theoretical results say that double Brown's exponential smoothing corresponds to this ARIMA model, and the smoothing constant is $\alpha = (1-\theta) = 1-0.2 = 0.8$. As in Section 3.1, we will set the smoothing constant in Brown's model equal to the estimated value of α calculated from the estimated value of θ according to the formulae mentioned above, depending on the respective software.

Let us see again how different packages coped with estimating θ and predicting one step ahead. These results are shown in Table 2.

	ARIMA		Brown (double)	
	θ	prediction	α	prediction
eViews	0.17817	203.388	0.82183	203.527
Statgraphics	0.15242	202.699	0.84759	201.686
SAS	0.13098	201.370	0.86902	201.702

Table 2 Comparison of results

We can see that the results concerning estimating parameter θ are more or less comparable between the software packages. They do somewhat differ from each other, but these differences do not significantly affect the predictions. The same conclusion can be made about the values of smoothing constant α, as well as the prediction made within double Brown's exponential smoothing model. Mutual comparison of both types of models has a satisfactory outcome.

3.3 ARIMA vs. Holt's exponential smoothing

Let us generate a time series from model ARIMA (0, 2, 2), that is,

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - 0.3\varepsilon_{t-1} + 0.6\varepsilon_{t-2}\,. \tag{18}$$

Parameters of this model are $\theta_1 = 0.75$, $\theta_2 = -0.2$, i.e., the process is invertible. The theory says that the corresponding exponential-smoothing model is Holt's one with parameters

$$\alpha = (\theta_2 + 1) = -0.2 + 1 = 0.8 \text{ a } \gamma = (2 - \alpha - \theta_1)/\alpha = (2 - 0.8 + 0.75)/0.8 = 0.5625\,.$$

Let us have a look at the results of the procedures. Again, we do not use the parameter values used for the simulations but instead use their estimates, which may differ from each other in different software packages.

	ARIMA			Holt		
	θ_1	θ_2	prediction	α	γ	prediction
eViews	0.80069	-0.12780	-683.46524	0.87220	0.37504	-682.58259
Statgraphics	0.81219	-0.11816	-683.47800	0.88184	0.34697	-682.63700
SAS	0.80243	-0.12623	-683.46700	0.87377	0.37058	-682.60245

Table 3 Comparison of results

Our comments on Table 3 are similar to those made above. The results are comparable both for parameter estimates and for predictions one step ahead – whether from the viewpoint of the ARIM or exponential smoothing model.

4 Conclusions

The basic goal of this paper is to point out the relationship between ARIMA and exponential smoothing models. The methodological part of the paper studied the theoretical basis of this relationship. Another goal was a practical comparison between both of these approaches; that is, we want to verify whether these theoretically equivalent (even though apparently very different) models provide the same predictions. Three of the most frequently used exponential-smoothing models have been chosen for such a practical comparison: simple and double Brown's exponential smoothing models, and Holt's exponential smoothing. Our results are, in all three instances, mutually comparable and differences between the basic approaches are negligible. All practical calculations were carried out in three software packages: eViews, Statgraphics, and SAS. Comparison of the results with respect to the software used confirms that procedures in these packages provide comparable results, regarding both predictions and parameter estimates.

On the basis of our results is seems that it is not important whether an exponential smoothing model or an ARIMA one is employed. It may be true from the viewpoint of results as such, but in practical situations we recommend choosing a model with regard to the available data, interpretation ability of the model, and the ultimate goal of the analysis.

References

[1] Abraham, B. – Ledolter, J.: *Statistical Methods for Predictioning*, John Wiley & Sons, New York, 1983.
[2] Box, G. E. P. – Jenkins, G. M. – Reinsel, G.C.: *Time series analysis, predictioning and control.* Third edition. Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1994.
[3] Brown, R. G.: *Smoothing, Predictioning and Prediction.* The Pitt Building, Trumpington Street, Cambridge, United Kingdom: Englewood Cliffs: Prentice-Hall, 1963.

[4] Everette S. Gardner, J. (2006): *"Exponential smoothing: The state of the art."* International Journal of Predictioning 22(4): pp. 637–666.
[5] Fomby, T. B.: *Exponential Smoothing Models*, Department of Economics Southern Methodist University Dallas, TX June 2008.
[6] Wei, W. W. S.: *Time series analysis – Univariate and multivariate methods*. Redwood City, California, Addison-Wesley Publishing Company 1990.

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics."

High age Mortality Modeling Using Bayesian Approach

Martin Matějka [1], Tomáš Karel [2], Jan Fojtík[3], Pavel Zimmermann[4],

Abstract. Main problem of many demographical issues is the lack of observations especially in case of high ages. If the issue relates to small areas, such as country, region or micro-region, the impact of this problem is even more significant. The lack of the data has negative impact on the estimations of demographical indicator widely used in many other scientific fields. This article is focused on high age mortality models. The main idea in this work is based on Bayesian approach, which has not been explored much so far in this context. The solution used in this work combines two sources of information to improve the accuracy of the estimation. One source of information comes from the examined area and merges with the second information coming from surrounding areas as a prior information where the accuracy of new solution is examined. Results of this paper work are presented on real data collected from selected countries of the European Union.

Keywords: High age, Mortality models, Bayesian approach.

JEL Classification: C10,C11, J10
AMS Classification: 91D20, 62C12

1 Introduction

Models of mortality became a very popular task in disciplines such as demography, political sciences, actuarial sciences or economy. Due to lack of (or even nonexistence of) data, mortality in very high ages is usually modeled separately from mortality in adult ages. Popular models are usually formulated as one dimensional parametric regression functions of age ("mortality laws") that allows extrapolation to the rarely observed age ranges. Presently the most popular specifications are based on logistic or exponential specifications. An extensive list of specifications is provided in Burcin, Tesarkova, and Sidlo [2] or in Pitacco et al. [9].

There is only a limited number of publications on the topic of modeling the time dynamics of the (very) high age mortality, i.e. on extrapolation (or description of the historical evolution) in both the age and the time dimension. In Gavrilov [3] and later on in Gavrilova [4] long term trends in parameters of Gompertz-Makeham model (not only for high ages) were studied. Similar analysis was performed in Bongaarts [1] where the author assumed the logistic model. The 'growth' parameter of the model was assumed to be constant in time and hence parallel shifts of the force of mortality were assumed. In Huliková [7] this assumption was proved to be unsuitable for males especially for countries from beyond the former iron curtain and age specific shifts were suggested.

The problem of the lack of observations naturally appears especially in cases of modeling small populations. It is probably not reasonable to ignore the fact such small populations have its own specifics and assume that they evolve completely isolated from the nearest surrounding populations. Hence it seems reasonable to find suitable method to use external information collected from geographically and / or economically close areas which might be considered relevant for the population of our interest. One of the ways of incorporating the external information is application of Bayesian methods with informative prior. Bayesian general linear model is applied in this article to fit the high age mortality model for the Czech Republic. Data for prior are collected from the "Visegrad four" (V4) countries, which are considered reasonably close both economically as well as geographically to the Czech Republic.

In this article we focus purely on the logistic specification as it is presently one of the most popular models. It is also used for data extrapolating by one of the most popular world wide data source, the Human mortality database Wilmonth et al. [10].

[1] University of Economics, Prague, department of statistics and probability calculus, Winston Churchill sq. 4, Prague 3, matejka.vse@gmail.com.
[2] University of Economics, Prague, department of statistics and probability calculus, Winston Churchill sq. 4, Prague 3, tomas.karel@vse.cz.
[3] University of Economics, Prague, department of statistics and probability calculus, Winston Churchill sq. 4, Prague 3, xfojj00@vse.cz.
[4] University of Economics, Prague, department of statistics and probability calculus, Winston Churchill sq. 4, Prague 3 ,zimmerp@vse.cz.

2 Model specification and methodology

The logistic specification of the dependence of the force of mortality on age and time period is defined as

$$m(x,t) = \frac{1}{1+\exp(-(\beta_0 + \beta_1 x + \beta_2 t))} \text{ for } x \geq x_0 \quad (1)$$

where $m(x,t)$ denotes the force of mortality, x is the age, x_0 is the high age threshold and β_j are the parameters and t is the time variable. Formula (1) can also be written as

$$\ln\left(\frac{m(x,t)}{1-m(x,t)}\right) = \beta_0 + \beta_1 x + \beta_2 t \text{ for } x \geq x_0 \quad (2)$$

Therefore this regression model can be treated as a member of the broad class of generalized linear models. It is then assumed, that the number of deaths at a specific age x and time t, $D(x,t)$, has binomial distribution, i.e.

$$D(x,t) \sim Bi(E(x,t), m(x,t)) \quad (3)$$

where $E(x,t)$ is the exposure. To estimate the parameters β_j where j = 1,2,3 the maximum likelihood method is usually applied in the classical non-bayesian analysis. The main focus of this article is to incorporate information from external dataset of surrounding V4 countries into the model as a prior information. The most important criticism of the Bayesian approach is the subjectivity in selecting the prior distribution. Note that this problem is avoided using the empirical Bayesian approach. Hence the application of empirical Bayesian approach seems to be reasonable. Namely we use an informative prior distribution in the form of independent normal distributions

$$\beta_j \sim N(\mu_j, \sigma_j), \text{ for } j = 1,2,3 \quad (4)$$

where μ_j and σ_j are parameters estimated using classical GLM approach on the external dataset from the surrounding V4 countries.

Analytical formula for the posterior density (especially for the multivariate models) usually does not exist. Therefore Markov chain Monte Carlo (MCMC) simulations are used in order to get approximate posterior densities and its characteristics for both the parameters which are used for predictions. The methodology of the use MCMC simulation methods for simulation posterior densities is described in detail for example in Gelman et al [5] and Koop [8] .

3 Data

Data used in this work presents number of death and exposure to risk for four countries: Czech Republic, Poland, Slovakia and Hungary. Deaths represent sum of death over gender and country. Same procedure was used for exposure to risk. The data are selected for years 2000-2009 in ages from 80 to 100 year olds. Data comes from two sources providing detailed mortality and population data sets. Specifically the Demography section in Eurostat database and The Human Mortality Database (HMD). Methods of calculation exposure to risk and missing deaths counts are presented in Wilmoth, et al. [10].

4 Results

Estimates presented in this article were obtained with the help of the package Data Analysis Using Regression and Multilevel/Hierarchical Models (arm) in statistical freeware R. See the documentation Gelman et al. [6] for details. The data from the surrounding V4 countries are referred here as prior data.

As stated above, independent normal distributions were assumed as the prior distributions for the parameters β_j. The data from the surrounding V4 countries were used to fit the parameters of the prior distributions. The comparison of the prior and posterior distributions for each of the three parameters is displayed in Figure 1. Table 1 contains the prior and posterior mean values and standard deviations of these parameters.

Coefficients	Prior		Posterior	
	Exp. value	Stdev.	Exp. value	Stdev.
β_0	26,39737	0,56821	27,32272	0,38104
β_1	0,11606	0,00018	0,11809	0,00017
β_2	-0,01905	0,00028	-0,01955	0,00019

Table 1 Comparison of prior and posterior densities of parameters

Table 2 contains classical estimates based purely on the Czech data. The bayesian posterior estimates were shifted from the classical estimates by the information obtained from the other V4 countries (prior information). Despite the fact that the prior distribution itself was informative, a certain reduction of variability can still be observed in the posterior distributions in comparison with the prior distributions.

Coefficients	Exp. value	Stdev.
β_0	33,8055	1,2102
β_1	0,1277	0,0004
β_2	-0,0232	0,0006

Table 2 Classical GLM estimates for the Czech data.

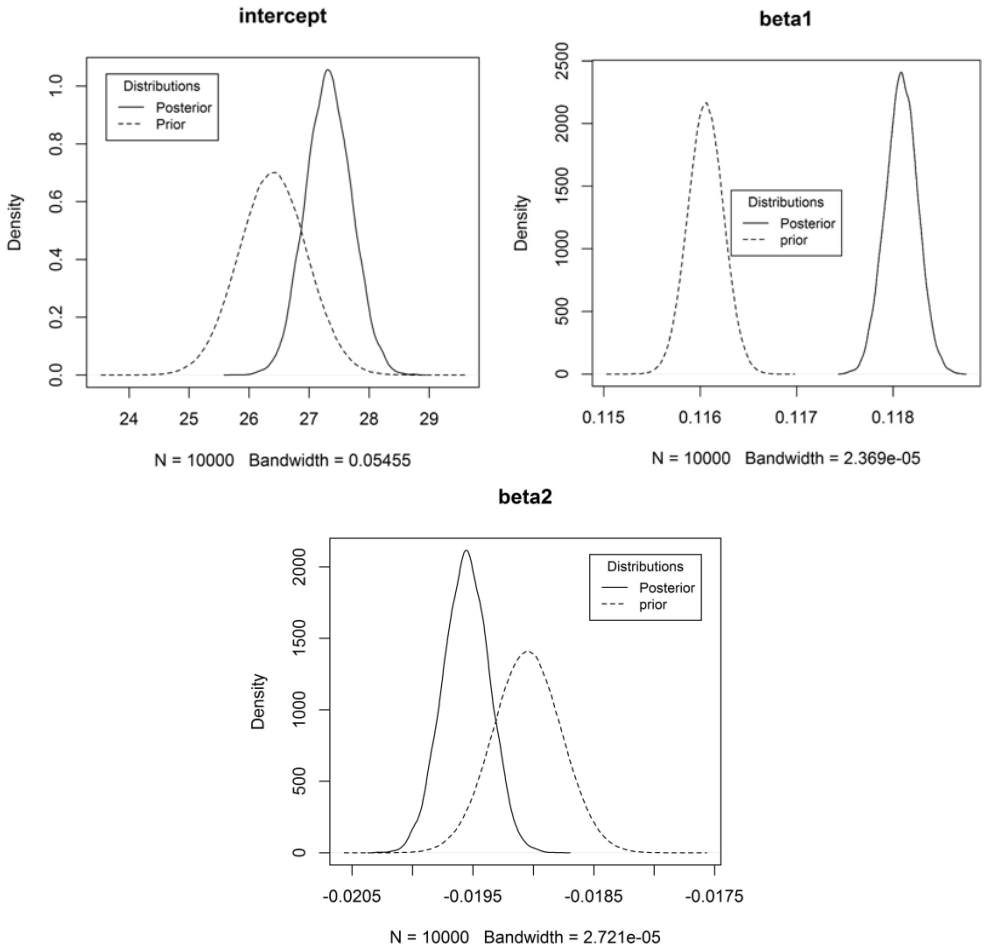


Figure 1 Prior and Posterior densities of parameters β_j .

Prior and posterior densities of parameters β_j for j = 1,2,3 are illustrated on Figure 1. The comparison of classical (frequentist) and bayesian prediction (with prior observations and Czech observations) is presented in Figure 2. We can observe the difference between these two predictions caused by the weight of prior information from external dataset. In cases of sufficient number of observations in the region of our interest is the influence of prior information avoided and the bayesian approach convergates to classical GLM prediction.

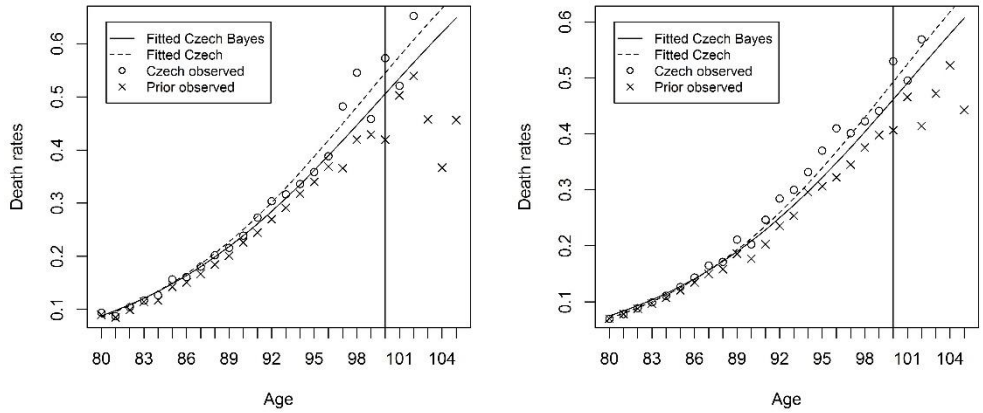


Figure 2 Comparison of Bayesian and classical prediction for 100+ years death rates in 2000 (left side) and 2009

The graphical illustration of prediction over age (i.e. from 101 to 105 years) and time (i.e. for 2010 and 2011) using Bayesian approach is shown in Figure 3. The surface represents the predicted values by the model. The observed death rates are represented by the points.

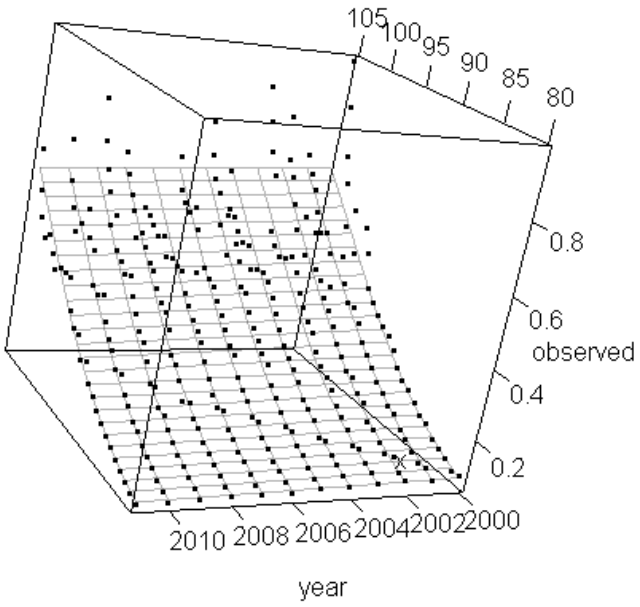


Figure 3 Bayesian prediction over age and time

5 Conclusions

The problem of lacking data was successfully reduced by using Bayesian methods that allowed incorporating external datasets from surrounding countries into the estimates of mortality dynamics in the Czech Republic. On one hand, demographic models are typically set up on the country level. On the other hand, high age data will always be scarce in small areas and it is obvious that areas reasonably similar (both geographically, as well as economically) will also have similar mortalities. Therefore it is natural to base the country level estimates not only on the country of interest data but also on the data collected in the surrounding areas. The bayesian methods applied in the previous section decreased the variability of the posterior distribution of all parameters considered.

Using the informative prior distribution estimated on data observed reduced the subjectivity in selecting priors only on the choice of the distribution family, in this case normal family, and on the choice of the prior data, but not on the parameters of the priors itself.

Acknowledgements

This article was supported by the grant IG 410025 "Využití bayesovských metod pro modelování úmrtnosti"of the internal grant agency of the University of economics, Prague.

References

[1] Bongaarts, J.: *Long-range trends in adult mortality: Models and projection methods*. Demography, 2005, 42(1):23-49.
[2] Burcin, B., Tesárková, K., and Šídlo, L.: *Nejpoužívanější metody vyrovnání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci*. Revue pro výzkum populačního vývoje, 2010, 52:77-89.
[3] Gavrilov, LA., and Gavrilova, NS: *Historical perspective on the kinetics of human mortality*. Proc. Acad. Sci. USSR [Doklady Akademii Nauk SSSR], 1979, 245(4):1017-1020.
[4] Gavrilova, NS., and Gavrilov, LA.: *Ageing and longevity: Mortality laws and mortality forecasts for ageing populations.* Demografie, 2011, 53(2):109-128.
[5] GELMAN, A.: *Bayesian Data Analysis*. CRC Press. 2003. ISBN 978-1-4398-4095-5
[6] GELMAN, A., et al.: Data Analysis Using Regression and Multilevel/Hierarchical Models, version 1.8-4, 2015. http://cran.r-project.org/web/packages/arm/arm.pdf
[7] Hulíková Tesárková, K..: *Selected methods of mortality analysis focused on adults and the old-age.* PhD thesis, Ph. D. dissertation, Dept. Demography, Charles Univ., Prague, CZE, 2012.
[8] KOOP, G.: *Bayesian Econometric*. John Wiley & Sons. 2003. ISBN 0-470-84567-8
[9] Pitacco, E.,Denuit, M., Haberman, S., and Olivieri, A.: *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, 2009.
[10] Wilmoth, J. R., Shkolnikov, V., and Barbieri. M.: Human mortality database [electronic resource], 2012. URL http://www.mortality.org.

Using exploratory factor analysis for determination of tourist satisfaction factors in the Czech Republic

Lyaysan Mavlikaeva[1], Kateřina Ryglová[1]

Abstract. Tourist satisfaction is one of the most important factors in tourism industry. Tourist satisfaction assessment can help regulate efforts on visitors' travel experiences, shaping the quality of services and chase the guidelines for effective destination marketing strategy. Attention should be paid to factors of tourists' satisfaction because these factors influence the tourist's decision about the destination.

The main purpose of this study is determining the most important satisfaction factors for tourists in the Czech Republic using exploratory factor analysis. The exploratory factor analysis (EFA) was performed for reducing the number of variables and determined significantly correlated factors. This study have found the most important factors of tourist satisfaction: "transport and food", "entertainment", "price", "service in the accommodation", "uniqueness of the Czech Republic", "service in restaurant and cafe". Managerial recommendations for increasing tourist satisfaction can be offered. The research is conducted using primary data collected through questionnaire survey of Russian tourists, who are perceived to be one of the most important source markets of Czech tourism.

Keywords: tourist satisfaction, Czech Republic, Russian tourists, EFA

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

International tourism is a profitable and developed industry in the Czech Republic. In total, 7 851 865 tourists arrived in the Czech Republic in 2013, which was about 2.67% more than in 2012 [6]. The dominant groups of international tourists are from Germany, Slovakia, Poland and Russia [5]. However, Russian tourists have the highest average number of nights and expenditures compared to other foreign tourists in the Czech Republic [1]. Also compared to other nationalities, they are more interested in visiting the Czech Republic [1]. According to the investigation that was done by marketing agency "Stem/mark", 16% of Russian respondents have not visit the Czech Republic yet, but they are planning to do that during the next five years [1]. In spite of the fact, that the number of Russian tourists has decreased by 47.8% [5] in the first quarter of 2015 (due to the devaluation of the rouble and the crisis in Ukraine), the group of Russian tourists is still one of the largest and economically advantageous groups for Czech foreign tourism development for next several years [12].

Despite the importance of Russian tourists for developing Czech international tourism, there has been relatively little discussion about Russian tourists in the Czech Republic. Satisfaction factors of Russian tourists have been the subject of numerous studies and analysis which can be found in [2][7][9]. However, these studies was done for specific coastal destinations (Antalya, Turkey), while cities in the Czech Republic are more popular destinations for cultural and heritage tourism. Reasons, motives and also satisfaction factors of tourists by vacation can be differ depending on the destinations. Because of that, the purpose of this paper was to determine the significantly correlated satisfaction factors for Russian tourists in the Czech Republic. Also, the presented paper addresses this gap by exploring the travel consumption behaviour and satisfaction level of Russian tourists towards the Czech destination products and services. It identifies the extent to which the Czech Republic delivers the attributes that tourists are seeking. The major purposes of the paper are:

1) to analyse the most important satisfaction factors for Russian tourists in the Czech Republic;
2) to determining which themes should be emphasised in marketing.

By better understanding of the Russian market and more in-depth knowledge of the views of Russian travellers, should not only help increase the number of Russian tourists, but also improve the quality of their travel experience in the Czech Republic. In addition, the gained significant satisfaction factors will create a

[1] Mendel University in Brno, Department of Marketing and Trade, Zemědělská 1, {xmavlika@node., katerina.ryglova@}mendelu.cz.

support for a further researches of satisfaction in tourist destinations which concentrate on the cultural and heritage tourism.

2 Literature review and Methodology

In a saturated marketplace, the analysis and increasing the level of tourist satisfaction is crucial for development of the tourist destination because it influences the visitors' travel experiences and the decision to return and/or recommend destination to other tourists [3][4][19]. Moreover, recommendations by previous visits (WOM) are one of the most popular and reliable information for potential tourists [4][19].

Williams and Soutar [18] defined tourist satisfaction as an individual's cognitive–affective state derived from a visitor experience in a destination. However, tourists experience a medley of services such as hotels, restaurants, shops, attractions and they may evaluate each service element separately [4]. Therefore, overall satisfaction with a hospitality experience is viewed as a function of individual attribute-level evaluations of a destination [3][4][8][9][11][13][14]. Also empirical studies indicated that tourists' satisfaction with individual components of the destination leads to their satisfaction with the overall destination [4][10][11]. Consequently, the attention should be paid to factors of tourists' satisfaction because these factors influence the tourist's decision about the destination.

Many previous studies investigated the satisfaction factors of destination for tourist. For example, Chi and Qu [4] distinguished seven key factors of tourist satisfaction: lodging, dining, shopping, attractions, activities, events, environment and accessibility. Do Valle et al. [8] established that factors of beaches, hospitality, landscape, restaurants, food, and lodging and the competence and kindness of the locals are more positively perceived by tourists. Mendes et al. [13] verified the higher levels of satisfaction on items such as restaurants, lodging, competence, courtesy, landscape, hospitality and beaches. Yoon and Uysal [19] identified that cleanness, shopping, reliable weather, safety; different culture and water activities are the significant factors for tourism destination in Cyprus. Emir [9] found that front office, food and beverage, housekeeping, customer relations, health and hygiene, physical facilities and employees are the most important factors of customer satisfaction with hotel services in Antalya (Turkey) for Russian tourists. Marcussen [11] established that location, price, information, service, facilities and friendly people are the most important factors of overall accommodation satisfaction. Wan and Chan [17] noted that good weather affected tourists' moods and satisfaction levels. In summary, as seen in the discussion above, we can suppose that factors related to attractions, activities, accommodation, food, hospitality of local residents, courtesy personnel and prices are the most significant factors of tourist satisfaction. These attributes were incorporated in the questionnaire survey about tourists' satisfaction of vacation in the Czech Republic.

The data for this study were collected by a self-administered questionnaire method in Prague, Brno and in the large cities of Russia during November 2014 to April 2015. The target population of this study involves Russian tourists who visited the Czech Republic. From target population, the sample population was selected using a quota sampling method. The number of tourists which is included in each quote was defined proportionally to the age of tourist in the target population. According to the investigation of "CzechTourism" agency, 40% of Russian tourists arrive to the Czech Republic at the age of 30-44 years, 45-59 years (27%), 15-29 years (26%) and the group of tourists over 60 years is (7%) [1]. The survey questionnaire consisted of the questions that measured the tourists' attribute satisfaction, overall satisfaction and questions about tourists' demographic information and travel behaviour in the Russian language.

Using the satisfaction attributes from the previously conducted in-depth interviews and focus groups with Russian tourists, tourist experts and previous studies, an attribute list consisting of 55 items was established. The destination attributes encompassed five areas of tourism activities: transport, accommodation, food, attractions and activities, price. Using seven-point Likert-scales, tourists were asked to evaluate their satisfaction with each tourist attracting attribute (1= Very dissatisfied and 7=Very satisfied).

A total of 343 usable questionnaires were collected. The data were analysed using the program "STATISTICA" 12. Descriptive analysis was chosen as the method for analysing respondent socio-demographic profiles, travel characteristics and their satisfaction with destination attributes in the Czech Republic. The exploratory factor analysis (EFA) was performed for determination the most significant attributive satisfaction factors for Russian tourist in the Czech Republic. Using Varimax rotation, the latent root criterion of 1.0 was used for factor inclusion, and a factor loading of 0.70 was used as the benchmark to include items in a factor.

3 Results and discussion

A total of 343 Russian tourists were attended in questionnaire survey. There were more female respondents (76.3%) than male ones (23.7%) and there was a higher proportion of younger respondents (more than 63 %

were under 30 years old) and respondents from city with more than 1 000 000 residents (44.7%). A high proportion of respondents possessed a higher education (64.6%). There was a wide occupational spread ranging from an employee of a company (48%), an entrepreneur (8.6%), a student (37%), a senior (1.8%) and an unemployed (4.6%). The most respondents had a monthly family income between 20 001 - 60 000 roubles (47.7%).

The most respondents were travelling to the Czech Republic for recreation (34.3%), new experiences (23.5%), study (14.3%) and treatment in the Czech spas (10.7%). A high proportion of the tourists organized their trips independently (56.5%) while 43.5% bought package tours to the Czech Republic. Most of the respondents (46%) used information that had been provided by their relatives or friends. The duration of travel was predominantly 8-14 days (44.2%). Most of the respondents were travelling to the Czech Republic by airplane (64.7%) and were staying in the hotel 3* (35.7%). With regard to party composition, 27.1% of respondents were travelling with their families, while 26.3% with their friends, 21% alone, 13.4% with a spouse and 12.2% with a partner.

The mean scores for respondent satisfaction were calculated. First of all, respondents were asked to rate their overall satisfaction with travelling experience with „1“ signifying least satisfied, and „5“ very satisfied on the scale of 1–5. The mean of tourist overall satisfaction is 4.61, which means that Russian tourists are the highly satisfied with travel experience in the Czech Republic. Also respondents were asked to indicate their levels of satisfaction with 55 attributes, with „1“ signifying least satisfied, and „7“ very satisfied on the scale of 1–7. Attributes were divided into five dimensions: transport, accommodation, food, attractions and activities, price and depicted in Table 1.

Tourism dimensions	The number of variables	Satisfaction (Mean)
Transport	8	5.65
Accommodation	10	5.48
Food	10	5.74
Attractions and activities	19	5.66
Price	8	5.36
Overall satisfaction (scale of 1-5)	1	4.61

Table 1. Satisfaction evaluations in different tourism dimensions

The analysis of the means allows the identification of the attribute recognized as most satisfactory by the respondents. Attributes related to food were evaluated as most satisfactory by the respondents (mean =5.74) while attributes related to price had the lowest scores (mean =5.36). The variables which scored the highest of the fifty five attributes for satisfaction were historical and cultural sights, architecture, nature, traditional Czech food, impressions with the Czech Republic, the variety of restaurants and cafes, traditional Czech alcoholic beverages, the quality and freshness of food, public transport, atmosphere and the safety of accommodation.

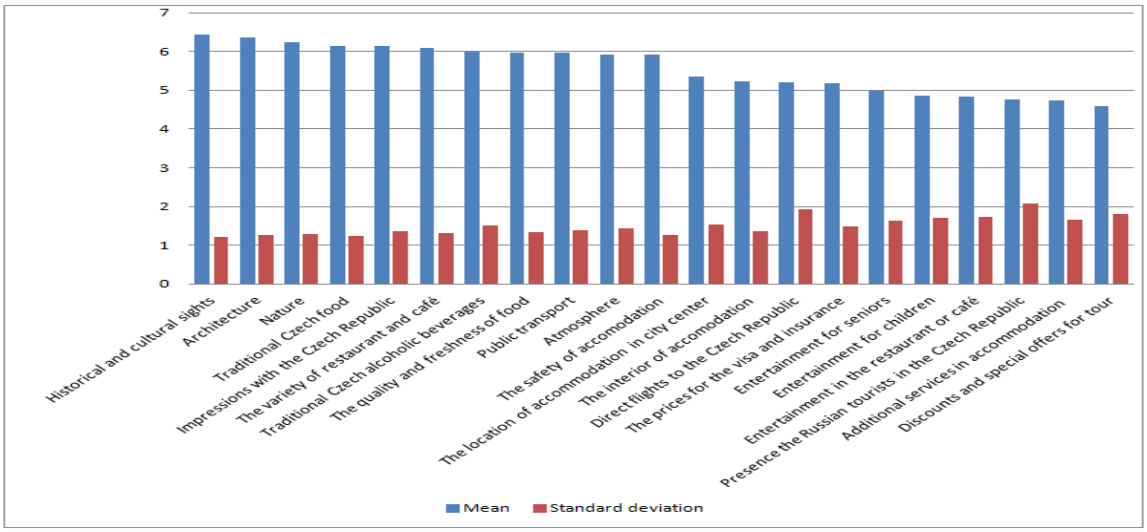


Figure 1. The most important satisfaction and dissatisfaction attributes for Russian tourists in the Czech Republic

However, the attributes which scored the lowest for satisfaction were the location of the accommodation in the city centre, the interior of accommodation, direct flights to the Czech Republic, the prices for visa and insurance, entertainment for seniors, children and in the restaurant or the cafe, presence the Russian tourists in the Czech Republic, additional services in accommodation and discounts and special offers by when purchasing the tour. The satisfaction attributes and their means and standard deviation are listed in Figure 1.

In order to simplify the analysing of a huge number of variables, EFA was employed to identify the underlying dimensions or factors in the data. EFA was used to reduce the 55 satisfaction attributes to a more manageable number of underlying constructs and to determinate the most significant satisfaction factors for Russian tourist in the Czech Republic. The EFA was performed using the principal component and the Varimax rotation methods, with a cut-off eigenvalue of 1.0. and the factor loading of 0.70. Six factors have arisen (Table 2).

The first accounts for 51.12% of the total data variance and is associated with the observed variables – the traditional Czech alcoholic beverages, the intercity transport, helpful personnel of the transport, the interior of restaurant and cafe, the comfort of the public transport and public transport. This factor is concerned essentially with transport and food variables and was designated "transport and food". The second factor was named "entertainment" and explains 6.13% of the total data variance. This factor includes the following items: entertainment for children, seniors and the young people. The third factor, named "price", is related to the price for accommodation, transport, tour to the Czech Republic, food, visa and insurance representing 4.50% of the total data variance. The fourth accounts for 3.77% of the total data variance and is associated with the variables service in the accommodation, polite personnel of the accommodation and easy system of the hotel reservation. This factor was called "service in the accommodation". The fifth factor was named "uniqueness of the Czech Republic" and explains 3.67% of the total data variance; it is related to the historical and cultural sights, architecture, and the authenticity of the Czech Republic, nature and unforgettable impressions with country. The sixth factor, named "service in restaurant and cafe", and representing 2.65% of the total data variance. Thus, these six factors explain approximately 72% of the total data variance. The results from EFA suggest that these six factors can be used as a summary measure of the indicators that make up each factor.

Variables	Transport and food	Entertainment	Price	Service in the accommodation	Uniqueness of the Czech Republic	Service in restaurant and cafe
The traditional Czech alcoholic beverages	0.755					
The intercity transport	0.740					
Helpful personnel of the transport	0.731					
The interior of the restaurant and cafe	0.714					
The comfort of the public transport	0.708					
The public transport	0.700					
Entertainment for children		0.811				
Entertainment for seniors		0.793				
Entertainment for young people		0.714				
Price for accommodation			0.829			
Price for transport			0.792			
Affordable price for tour to the Czech Republic			0.757			
Price for food			0.743			
Price for visa and insurance			0.729			
Service in the accommodation				0.742		

Polite personnel of the accommodation				0.725		
Easy system of the hotel reservation				0.717		
Historical and cultural sights					0.761	
Architecture					0.756	
The authenticity of the country					0.741	
Nature					0.719	
The unforgettable impressions with country					0.717	
Service in restaurant and cafe						0.704
Explained variance	51.12	6.13	4.5	3.77	3.67	2.67

*Extraction method: principal component analysis. Rotation method: Varimax.

Table 2. The results of EFA

The findings support previous studies of tourists' satisfaction factors [2][3][14][15][16]. The results of research conducted by Meng et al. [14] revealed that friendly/quality services and lodging performance are significant factors in determining the overall satisfaction, whereas food and location are of significant importance in the satisfaction evaluation. Campo-Martinez and Garau-Vadel [3] found the factors that received the highest satisfaction levels were the characteristics of the destination (the scenery and beaches) and the hospitality of local residents, while the lowest ratings corresponded to the price of leisure activities. Song and Cheung [16] established that the venue, service quality and ticket price/value for money are important factors to the level of tourist satisfaction. Pulpanova and Simova [15] identified the five main factors for Czech tourists influencing customer satisfaction by purchasing vacation tour - personnel, transportation, price, information and communication. Aktas et al. [2] verified the dimension of accommodation services also was the strongest predictor of the Russian tourists' overall holiday satisfaction, followed by destination facilities and incoming travel agency services. A lot of previous studies [3][8][13] which are concentrated on the tourists satisfaction mentioned that natural characteristics (beaches, landscape etc.) are the most important factors of tourists satisfaction. While, according to our research, we prove that Russian tourists in the Czech Republic are more interested in city infrastructure and sights than other satisfaction factors. That can be explained by different type of tourist destination.

4 Conclusion

As the Czech inbound market is dominated by Russian visitors, there is an urgent necessity for the Czech tourism industry to increase visitor arrivals from Russia. Although satisfaction factors of Russian tourists have been the subject of numerous studies, Russian tourists' satisfaction with their travel experience in the Czech Republic is virtually non-existent. The purpose of this paper was to determine the significantly correlated satisfaction factors for Russian tourists in the Czech Republic. Using an exploratory factor analysis technique, this study has found the most important factors of tourist satisfaction: "transport and food", "entertainment", "price", "service in the accommodation", "uniqueness of the Czech Republic", "service in restaurant and cafe".

The major findings of this study have significant managerial implications for the Czech Republic. First of all, the exploratory factor analysis revealed that tourist satisfaction had six underlying factors. These results could help Czech destination marketers better understand the factors contributing to tourist satisfaction of Russian tourists better that they are able to carefully deliver appropriate products and services that accommodate tourists' needs and wants. Especially, managers should pay the attention to the attributes of factor "entertainment" – entertainment for children and seniors, which Russian tourists are dissatisfied with. Also the problems with absence of the direct flights to the Czech Republic, high prices for visa and insurance and the absence of special incentives by purchasing the tour were identified. Thus, it is suggested that destination marketers consider the practical implications of these satisfaction variables, because they can be fundamental factors in increasing satisfaction of Russian tourist with destination services and products in the Czech Republic. Consequently, destination managers should establish a higher tourist satisfaction level to create positive post-purchase Russian tourist behaviour in order to improve and sustain destination competitiveness of the Czech Republic. Also the gained factors will be used for measuring tourist satisfaction in the destinations which specify in cultural and heritage tourism. The results presented in this study need to be qualified with limitation. The population of this

study was limited by Russian tourists. Therefore, the results from the study may not be generalized beyond this population

Acknowledgements

This paper was based on research realized in the scope of the project IGA n. 21/2015; GACR n. 15-21179S.

References

[1] Agency Czech Tourism (2012). How Russians perceive the Czech Republic. [Online]. Available at: http://vyzkumy.czechtourism.cz/#;q=jak%20nas;orderby=novinky
[2] Aktas, A., Cevirgen, A., Toker, B.: Assessing holiday satisfaction of German and Russian tourists visiting Alanya. *Tourism and Hospitality Management* **15** (2008), 1-12.
[3] Campo-Martinez, S., Garau-Vadell, J. B.: The generation of tourism destination satisfaction. *Tourism Economics* **16** (2010), 461-475.
[4] Chi, C. G-Q., Qu, H.: Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tourism Management* **29** (2008), 342-636.
[5] Czech Statistical Office. (2015). Cestovní ruch 1. čtvrtletí 2015. [Online]. Available at: https://www.czso.cz/csu/czso/cri/cestovni-ruch-1-ctvrtleti-2015
[6] Czech Statistical Office. (2014). Návštěvnost hromadných ubytovacích zařízení podle zemí 2012-2013. [Online]. Available at: http://vdb.czso.cz/vdbvo/tabparam.jsp?voa=tabulka&cislotab=CRU0031UU_KR&&kapitola_id=654
[7] Dogan, H., Unguren, E., Barutcu, S.: Russian and German tourist potential and satisfaction in sustainable competitive advantage of Alanya destination. *An International Multidisciplinary Journal of Tourism* **7** (2012), 381-399.
[8] do Valle, P., Mendes, J., Guerreiro, M., Silva, J.: Can welcoming residents increase tourist satisfaction? *Anatolia: An International Journal of Tourism & Hospitality Research* **22** (2011), 260-277.
[9] Emir, O.: A multi-national satisfaction analysis: An application on tourists in Antalya. *Tourism* **61** (2013), 347-359.
[10] Eusebio, C., Vieira, A. L.: Destination attributes' evaluation, satisfaction and behavioural intentions: a structural modelling approach. *International Journal of Tourism Research* **15** (2013), 66-80.
[11] Marcussen, C. H.: Determinants of tourist satisfaction and intention to return. *Tourism* **59** (2011), 203-221.
[12] Mavlikaeva, L., Netopil.: Russian tourists as a perspective resource market for Czech tourism PEFnet2013:European Scientific Conference of doctoral students (2013), 1-9.
[13] Mendes, J. C., Valle, P.O., Guerreiro, M. M., Silva, J. A.: The tourist experience: Exploring the relationship between tourist satisfaction and destination loyalty. *Tourism* **58** (2010), 111-126.
[14] Meng. F., Tepanon, Y., Uysal M.: Measuring tourist satisfaction by attribute and motivation: The case of a nature-based resort. *Journal of Vacation Marketing* **14** (2008), 41-56.
[15] Pulpanova, L., Simova, J.: Factors of customer satisfaction in tourism. *Economics and Management* (2012), 160-170.
[16] Song, H., Cheung, C.: Attributes affecting the level of tourist satisfaction with and loyalty towards theatrical performance in China: Evidence from a qualitative study. *International Journal of Tourism Research* **12** (2010), 665–679.
[17] Wan, Y. K. P., Chan, S. H. J. C.: Factors that affect the levels of tourists' satisfaction and loyalty towards food festivals: a case study of Macau. *International Journal of Tourism Research* **15** (2013), 226–240.
[18] Williams, P., Soutar, G. N.: Value, satisfaction and behavioural intentions in an adventure tourism context. *Annals of Tourism Research* **36** (2009), 413–438.
[19] Yoon, Y., Uysal, M.: An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management*, **26** (2005), 45–56.

On Consistency and Uncertainty of Interval Pairwise Comparison Matrices

Jiří Mazurek[1]

Abstract: In the analytic hierarchy process (AHP) it is assumed that pairwise comparisons of objects with regard to a superior element of a hierarchy are expressed by crisp numbers from Saaty's fundamental scale $\{1/9,1/8,...,1,...,8,9\}$.

But real-world decision making almost always involves some degree of uncertainty. In this paper uncertainty is modeled by interval numbers, and pairwise comparisons of all objects from the same hierarchy level form interval pairwise comparison matrices (IPCMs). The aim of the paper is to examine consistency of IPCMs in the AHP framework, and to propose several measures of IPCMs' degree of inconsistency, namely average random inconsistency, mid-point inconsistency and maximal inconsistency. Also, a measure of information uncertainty (entropy) of IPCMs is introduced as well. These measures provide useful information on quality of decision makers' judgments, and can be used, for instance, to estimate decision makers' (a posteriori) weights in a group decision making framework. The proposed approach is demonstrated with examples.

Keywords: consistency, decision making, interval AHP, pairwise comparisons, uncertainty.

JEL Classification: C44, C61, D81
AMS Classification: 15B99, 68T37

1. Introduction

In the analytic hierarchy process (AHP) pairwise comparisons of objects with regard to a superior element of a hierarchy are expressed by crisp numbers from Saaty's fundamental scale $\{1/9,1/8,...,1,...,8,9\}$, see [6] or [7]. However, real-world decision-making problems often involve uncertainty related to lack of time of decision makers, insufficient or imprecise information, etc. In this paper uncertainty in pairwise comparison of object is modeled with the use of interval numbers. Hence, when a decision maker is not sure whether an object A is better than an object B three or four times, he/she can assign the value $[3,4]$ to this pair instead of 3 or 4.

In the recent decades many methods, mainly of linear programming, for solving interval AHP problems were proposed, see e.g. [1], [2], [4] or [9]. However, a problem of consistency in interval AHP was not addressed so thoroughly, as there is not a single measure of degree of inconsistency (an analogue to consistency index or consistency ratio as in standard AHP) of interval comparison matrices known to author yet. Also, information content of different interval comparison matrices is different, as pairwise comparisons with longer intervals provide less specific information than those with shorter ones.

Therefore, the aim of the paper is to examine consistency of interval pairwise comparisons matrices (IPCMs) in the analytic hierarchy process framework, and to propose several measures of IPCMs' degree of inconsistency. Furthermore, a measure of information uncertainty of an IPCM is proposed as well. The proposed approach is demonstrated by examples.

The paper is organized as follows: section 2 provides a brief review of consistency in standard AHP, section 3 deals with inconsistency of IPCMs, in section 4 a measure of information uncertainty is proposed and section 5 provides illustrative examples. Conclusions close the article.

[1] School of Business Administration in Karviná, Silesian University in Opava, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40, Czech Republic, mazurek@opf.slu.cz

2. Consistency of crisp pairwise comparison matrices

In the analytic hierarchy process (AHP), proposed by Saaty in 1980, all elements (alternatives, criteria, etc.) are pairwise compared with regard to a superior element of a hierarchy, and a weight (priority vector) of alternatives and criteria is usually obtained by an eigenvalue method or row geometric mean method.

Pairwise comparisons of a decision maker should be consistent (should preserve the transitive property). Consistency in AHP is introduced as follows: let n alternatives compared pairwise, and let $s_{ij} \in \{1/9, 1/8, ..., 1, ..., 8, 9\}$ expresses the relative importance of an alternative i over an alternative j (with regard to a given criterion). Furthermore, s_{ij} is considered (multiplicative) reciprocal, that is $s_{ij} = 1/s_{ji}$ is satisfied for all i. Then the square and reciprocal matrix S (s_{ij}) is called pairwise comparison matrix (PCM).

The matrix S is said to be consistent if and only if $s_{ij} \cdot s_{jk} = s_{ik}, \forall i, j, k$.

However, real-world judgments of decision makers are seldom consistent. To assess inconsistency in AHP Saaty proposed two measures: consistency index (C.I.) and consistency ratio (C.R.), where $\lambda_{\max}$ is the highest (positive) eigenvalue of an interval comparison matrix of the order n, and R.I. is the random index representing consistency of randomly generated PCMs of the order n:

$$C.I. = \frac{\lambda_{\max} - n}{n-1} \tag{1}$$

$$C.R. = \frac{C.I.}{R.I.} \tag{2}$$

According to Saaty, pairwise comparisons matrices with *C.R.* lower than 0.1 can be considered tolerable, as DMs are usually not fully consistent in their judgments.

3. Consistency of interval comparison matrices

In interval AHP a decision maker expresses a relative importance of criteria and alternatives in a form of real intervals (in this paper it is assumed that these intervals conform to the Saaty's fundamental scale), which form the interval pairwise comparison matrix (IPCM), see Figure 1.

$$\begin{bmatrix} 1 & [l_{12}, u_{12}] & [l_{13}, u_{13}] & ... \\ [l_{21}, u_{21}] & 1 & [l_{23}, u_{23}] & ... \\ [l_{31}, u_{31}] & ... & 1 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

Figure 1 A general form of IPCM

Elements of IPCM satisfy (multiplicative reciprocity) relations: $l_{ij} = \frac{1}{u_{ji}}, u_{ij} = \frac{1}{l_{ji}}, \forall i, j$. Moreover, in this paper it is assumed tha all l_{ij} and u_{ij} are from Saaty's fundamental scale, but they can be easily extended to more general scales.

Definition 1. IPCM is consistent if and only if there exist $a_{ij} = [l_{ij}, u_{ij}], \forall i, j$, such that crisp matrix $A(a_{ij})$ is consistent.

But, of course, for some other values $a'_{ij} \in [l_{ij}, u_{ij}]$ IPCMs are inconsistent. Wang et al. [8] provide simple consistency check of IPCMs based on inspection of interval bounds:

Theorem 1 ([8]). $A(a_{ij})$ is a consistent interval comparison matrix if and only if it satisfies the following inequality: $\max_k (l_{ik} l_{kj}) \le \min_k (u_{ik} u_{kj})$ for all i, j, k.

However, there is no measure of degree of IPCM inconsistency known to author. Therefore, in the following definitions three different measures of IPCM inconsistency (even for IPCMs consistent according to Definition 1) are proposed.

Definition 2. Let *A* be a IPCM with elements $a_{ij} = \left[l_{ij}, u_{ij} \right], \forall i, j$, and let $m_{ij} = \sqrt{l_{ij} \cdot u_{ij}}$ ($\forall i, j$) be elements of a crisp mid-point matrix *M*. Then *midpoint consistency* (*MPC*) of the matrix *A* is defined as Saaty's consistency ratio of the matrix *M*:

$$MPC(A) = C.R.(M) = \frac{C.I.(M)}{R.I.} = \frac{\lambda_{\max} - n}{(n-1) \cdot R.I.} \tag{3}$$

Definition 3. Let *A* be an IPCM with elements $a_{ij} = \left[l_{ij}, u_{ij} \right], \forall i, j$. Then the *average random consistency* (*ARC*) of the matrix *A* is defined as the arithmetic mean of consistency ratio *C.R.* of a large number (at least 1000) of randomly generated matrices $P(p_{ij})$, $p_{ij} \in \left\lfloor l_{ij}, u_{ij} \right\rfloor, \forall i, j$.

Definition 4. Let *A* be an IPCM with elements $a_{ij} = \left[l_{ij}, u_{ij} \right], \forall i, j$. Let $B(b_{ij})$, $b_{ij} = l_{ij} \vee u_{ij}, \forall i, j$ be all (crisp) matrices with elements equal to the boundaries of intervals of *A*. Then the *maximal inconsistency* (*MIC*) of *A* is defined as:

$$MIC(A) = \max_{B} \{C.R.(B)\} \tag{4}$$

These three measures can be useful in judging (in)consistency of IPCMs. Small inconsistencies might be tolerated in the same way as in the standard AHP. Rationale behind the use of the geometric mean rather than the arithmetic mean for assessing midpoint consistency in Definition 2 rests in the fact that an IPCM in the latter case would not be reciprocal in general, and therefore the largest (positive) eigenvalue might not exist (see Perron-Frobenius theorem). For the estimation of average random consistency Monte Carlo simulations (with at least 1000 cases) can be performed. Nevertheless, midpoint consistency and maximal inconsistency might be more suitable for practical use, as they are less time demanding.

In a case of high inconsistency IPCM have to be rebuilt, see e.g. [8].

4. Uncertainty of interval pairwise comparison matrices

Information content of various IPCMs might be very different. Consider two IPCMs from Figure 2. As can be seen, IPCM in Figure 2a) provides more specific information than IPCM in Figure 2b). Information comprised in the latter matrix is more uncertain, because its intervals are larger. Therefore, some *measure of information uncertainty* (*MIU*) of IPCMs is needed. In [3] a non-probabilistic entropy was defined as a measure of uncertainty, which can be modified in a case of interval AHP as follows:

Definition 5. Let *A* be an IPCM with elements $a_{ij} = \left[l_{ij}, u_{ij} \right], \forall i, j$. Then a *measure of information uncertainty* (*MIU*) of IPCM satisifes the following relations:

i) $MIU(A) = 0 \Leftrightarrow u_{ij} = l_{ij}, \forall i, j.$ (*MIU* of a crisp case is 0),

ii) $MIU(A) = 1 \Leftrightarrow \left[l_{ij}, u_{ij} \right] = [1/9, 9], \forall i, j, i \neq j$ and $\left[l_{ij}, u_{ij} \right] = [1,1] = 1, \forall i = j$. (*MIU* of a maximally uncertain case is 1),

iii) $MIU(A) \geq MIU(B) \Leftrightarrow \left| u_{ij}^{A} - l_{ij}^{A} \right| \geq \left| u_{ij}^{B} - l_{ij}^{B} \right|, \forall i, j.$ (monotonicity).

Proposition 1. Let *A* be an IPCM with elements $a_{ij} = \left[l_{ij}, u_{ij} \right], \forall i, j$. Let $L^{A}(l_{ij})$ and $U^{A}(u_{ij})$ be (crisp) matrices of lower and upper bounds of *A*(a_{ij}). Let $A^{*}(a_{ij}^{*}), a_{ij}^{*} = [1/9, 9], \forall i \neq j$ and $a_{ij}^{*} = [1,1] = 1, \forall i = j$, be a maximally uncertain matrix. Then the following measure given as follows:

$$MIU(A) = \frac{\left\| U^A - L^A \right\|}{\left\| U^{A^*} - L^{A^*} \right\|} \tag{5}$$

is a measure of information uncertainty satisfying relations from Definition 5.

Proof: trivial.

In (5) the symbol $\| \ \|$ denotes arbitrary (induced or entry-wise) matrix norm. Matrices $L^A(l_{ij})$ and $U^A(u_{ij})$ are not reciprocal, albeit reciprocity is not required in the evaluation of information uncertainty.

The measure of information uncertainty can be useful for example in a group decision making, where a set of decision makers provide their IPCMs. In such a setting weights of DMs can be based on the degree of information uncertainty of IPCM provided by an individual DM, so that the higher is the degree, the lower is the weight.

It should be noted that the use of Shannon entropy to measure an uncertainty of interval matrices was proposed too, see e.g. [4].

$$\begin{pmatrix} 1 & [2,3] & [5,6] \\ \left[\frac{1}{3},\frac{1}{2}\right] & 1 & [3,4] \\ \left[\frac{1}{6},\frac{1}{5}\right] & \left[\frac{1}{4},\frac{1}{3}\right] & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & [2,5] & [3,8] \\ \left[\frac{1}{5},\frac{1}{2}\right] & 1 & [2,6] \\ \left[\frac{1}{8},\frac{1}{3}\right] & \left[\frac{1}{6},\frac{1}{2}\right] & 1 \end{pmatrix}$$

a) b)

Figure 2. Two interval pairwise comparison matrices.

5. Numerical examples

In this section two numerical examples are provided. For a matrix norm in relation (5) entry-wise linear (p = 1) norm was selected.

Example 1. Let A be IPCM shown in Figure 3. Evaluate its *MPC*, *ARC* and *MIC* (in)consistency. Also, estimate the measure of information uncertainty *MIU* by relation (5).

Solution:

The matrix A is inconsistent, because it fails consistency check from Theorem 1. Hence, we can evaluate the degree of inconsistency.

Midpoint consistency: Midpoint comparison matrix is provided in Figure 4, $\lambda_{\max} = 3.069$ and with the use of (3) we obtain: $MPC(A) = C.R.(M) = \frac{C.I.(M)}{R.I.} = \frac{0.069}{0.52} = 0.13$.

Average random consistency : $ARC(A) = 0.088$

Maximum inconsistency: there are different 8 matrices with elements equal to interval bounds of the matrix A. The most inconsistent matrix A' is shown in Figure 5. The maximal eigenvalue $\lambda_{\max} = 3.136$, and $MIC(A) = C.R.(A') = \frac{0.136}{0.52} = 0.26$.

Information uncertainty: Upper and lower matrices of A, $L^A(l_{ij})$ and $U^A(u_{ij})$, are shown in Figure 6. Let $\|A\|_p = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}$ be an entry-wise matrix norm. Let's set p = 1 and with the use of (5) we obtain: $MIU(A) = \dfrac{\|U^A - L^A\|}{\|U^{A^*} - L^{A^*}\|} = \dfrac{4.383}{53.33} = 0.082$.

Example 2. Let A and B be IPCM shown in Figure 2a) and 2b) respectively. Evaluate their information uncertainty.

Solution:

The matrix A: $MIU(A) = \dfrac{\|U^A - L^A\|}{\|U^{A^*} - L^{A^*}\|} = \dfrac{3.283}{53.33} = 0.062$.

The matrix B: $MIU(B) = \dfrac{\|U^B - L^B\|}{\|U^{B^*} - L^{B^*}\|} = \dfrac{12.542}{53.33} = 0.235$.

So, indeed, the matrix B contains more uncertainty than the matrix A.

$$A = \begin{pmatrix} 1 & [3,4] & [4,5] \\ \left[\frac{1}{4},\frac{1}{3}\right] & 1 & [2,4] \\ \left[\frac{1}{5},\frac{1}{4}\right] & \left[\frac{1}{4},\frac{1}{2}\right] & 1 \end{pmatrix}$$

Figure 3 Interval pairwise comparison matrix A from Example 1

$$M = \begin{pmatrix} 1 & 3.464 & 4.472 \\ 0.289 & 1 & 2.828 \\ 0.224 & 0.354 & 1 \end{pmatrix}$$

Figure 4 Mid point pairwise comparison matrix M from Example 1

$$A' = \begin{pmatrix} 1 & 4 & 4 \\ 0.25 & 1 & 3 \\ 0.25 & 0.33 & 1 \end{pmatrix}$$

Figure 5 The most inconsistent matrix from A interval bounds

$$L^A = \begin{pmatrix} 1 & 3 & 4 \\ 0.25 & 1 & 2 \\ 0.20 & 0.25 & 1 \end{pmatrix}, \; U^A = \begin{pmatrix} 1 & 4 & 5 \\ 0.33 & 1 & 4 \\ 0.25 & 0.5 & 1 \end{pmatrix}$$

Figure 6 Matrices $L^A(l_{ij})$ and $U^A(u_{ij})$ from the left to the right

Conclusions

The aim of this paper was to examine consistency and uncertainty of interval pairwise comparison matrices. Also, three measures of inconsistency and one measure of information uncertainty were proposed. These measures provide useful information on quality of decision makers' judgments, and can be used to estimate decision makers' (a posteriori) weights in a group decision making framework. Further research may focus on interval fuzzy or interval intuitionistic fuzzy pairwise comparison matrices as well.

Acknowledgement

The paper was supported by the grant of the Czech Grant Agency no. 14-02424S.

References

[1] Arbel, A.: A linear programming approach for processing approximate articulation of preference, in: P. Korhonen, A. Lewandowski, J. Wallenius (Eds.), *Multiple Criteria Decision Support, Lecture Notes in Economics and Mathematical Systems*, vol. 356, Spring-Verlag, Berlin, 1991, 79–86.

[2] Chandran, B., Golden, B, Wasil, E.: Linear programming models for estimating weights in the analytic hierarchy process, *Computers & OperationsResearch* 32 (9), 2005, 2235-2254.

[3] De Luca, A.,Termini, S.: A definition of a non-probabilistic entropy in the setting of fuzzy sets theory, *Inform. and Control* 20, 1972, 301-312.

[4] Islam, R., Biswal, M. P., Alam, S. S.: Preference programming and inconsistent interval judgments, *European Journal of Operational Research* 97(1), 1997, 53-62.

[5] Lotfi, F. H., Fallahnejad, R.: Imprecise Shannon's Entropy and Multi Attribute Decision Making. *Entropy* 2010, 12, 53-62.

[6] Saaty, T. L.: *The Analytic Hierarchy Process*. McGraw Hill, New York, 1980.

[7] Saaty, T. L.: Decision making with the analytic hierarchy process. *Int. J. Services Sciences* **1**, 2008, 83–98.

[8] Wang, Y. M. , Yang, J. B., Xu, D. L.: Intervalweight generation approaches based on consistency test and interval comparison matrices, *Applied Mathematics and Computation*, 167(1), 2005, 252-273.

[9] Jing Rung Yu, Yu-Wen Hsiao, and Her-Jiun Sheu: A Multiplicative Approach to Derive Weights in the Interval Analytic Hierarchy Process, *International Journal of Fuzzy Systems*, 13(3), 2011, 225-231.

Weighting Scheme for Measuring the Composite Index of EU Regional Resilience

Lukáš Melecký[1], Michaela Staníčková[2]

Abstract. Composite indices have received substantial attention in recent years and various methodologies have been developed to handle different aspects of economic and business issues. General scheme of building composite indicators includes several steps that have to be made and corresponding methods have to be chosen. Primarily, selection of sub-indicators, normalizing methods, weighting schemes and aggregation formulas are fundamental. This paper deals with the selection of appropriate tool for weighting scheme, i.e. the issue of weights. When used in benchmarking framework, weights can have a significant effect on the overall composite indicator and the evaluated units' rankings. Therefore an objective approach of determining weighting scheme in composite weighted index for measuring EU regional resilience is suggested. Several dimensions of regional resilience as sub-indicators exist, therefore it is necessary to use different weighting for (1) each dimension of regional resilience, (2) each EU NUTS 2 region and (3) each year of reference period. For this purpose, the entropy method is used to determine objective weighting scheme based on information available in the data set.

Keywords: Entropy method, composite index, regional resilience, weighting scheme.

JEL Classification: C43, C82, R11, R12
AMS Classification: 28D20, 62H25, 90B50

1 Topic of Economic Resilience

Economic shocks occur periodically in economies, though the effect of these shocks have varies from region to region as well as region's adjustment and its recovery. We are particularly concerned with regional economic resilience: why are some regional economies that are adversely affected by shocks able to recover in a relatively short period of time while others are not? Economic resilience is a concept that is frequently used but rarely well defined. Conceptually, there are two separate, though not necessarily unrelated, concepts. The first is based on 'equilibrium analysis' in which resilience is the ability to return to a pre-existing state in a single equilibrium system. The second defines resilience in terms of complex adaptive systems and relates to the ability of a system to adapt and change in response to stresses and strains [8]. For regional economic analysis, perhaps the most natural conceptual meaning of economic resilience, is the ability of a regional economy to maintain or return to a pre-existing state (typically assumed to be an equilibrium state) in the presence of some type of exogenous shock. Resilience thinking thus constitutes an alternative approach. 'Planning for resilience' can find a home in planning theory as an analysis of the external dynamics that accelerate economic, social and spatial vulnerability and as an approach that helps to link social and economic processes with ecological processes, calling for a reconsideration of the 'substance' of planning so as to enhance capacity to deal with slow and sudden changes of different forms. This can occur within a process that focuses on 'building a self-organization capacity' alongside a change in the value system that can overcome the unequal power relations [11].

Opinions vary to the definition of resilience, and there is no mainstream approach for measurement and expression of resilience and therefore no uniform strategies for strengthening resilience of economies. Research studies below show that indicators and subsequently designed factors of regional resilience are considered as crucial for purposes of this paper. We do not want to just use these previously defined factors, but to find out relevant indicators which could be part of factors of regional resilience crucial for constructing the composite weighted index of European Union (EU) regional resilience. But what are the main characteristics for regional resilience? The first group of factors is suggested by Martin [5] and among the key factors of regional resilience ranks: dynamic growth of region, structure of the economy, export orientation and specialization of region, human capital, innovation rate, business and corporate culture, localization of region, and institutional arrangement in region. The

[1] VŠB–TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, lukas.melecky@vsb.cz.
[2] VŠB–TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, michaela.stanickova@vsb.cz.

second group of factors is defined in [1] and among the key factors of regional resilience suggests: regional economic capacity, socio-demographic capacity of region and regional community capacity. In the Czech Republic, Koutský et al [4] engage issues of regional resilience determinants and define following factors: the main macroeconomic indicators, labor market indicators and additional ones.

Based on sets of regional resilience factors mentioned above, it is possible to define, with certain degree of generalization, a set of indicators of regional resilience in the field of competitiveness. In the paper, we link the concepts of resilience with competitiveness. It is very important to understand the extent to which areas (territories/localities, in our case regions) compete with each other, where this competition comes from, and what factors determine a territorial economic attractiveness. Taking the competitiveness concept a step further, understanding territorial resilience challenges allows us to not only think about wealth generation of our territories, but also ensure the wellbeing of all citizens, enable sustainable economic development, and how to manage economic shocks and decline into our territorial strategies [12]. The authors point out that regional resilience is certainly influenced at least by the nature of state economic policy, export-orientation of region, business and corporate culture as well as institutional arrangement of region. However, the above mentioned indicators are considered as the basic indicators for initial research in this area.

2 Weighting Scheme for Construction of Composite Index

Construction of composite index of EU regional resilience is based on procedure in Table 1. The first step on analysis is to find out relevant indicators for resilience measuring, in the second step factor analysis is applied for factors of resilience defining, in the third step entropy method is applied for using different weighting scheme for each resilience dimension, and in the last step – construction of composite index is provided.

Input data analysis
Pre-processing phase » Collection of indicators » Groups of indicators for resilience
Factor analysis
Correlation » Standardization (Z-score) » Principal component analysis » Varimax with Kaiser Normalization » Resilience factors » Set of new composite indicators » Factors description
Entropy method
Weighting scheme for factors » Set of weights for each resilience dimension
Composite index construction
Combination of factor analysis and entropy method results » Construction of composite weighted index

Table 1 Scheme of Analysis

Based on the information mentioned above, basic approach for choosing relevant indicators is the EU Regional Competitiveness Index 2013 (RCI). Used indicators in the framework of 'flagship' areas of resilience with link to competitiveness are as follows: (1) *institutional dimension*: government effectiveness (GE), corruption (C), rule of law (RL); (2) *infrastructure dimension*: motorway potential accessibility (MPA), railway potential accessibility (RPA); (3) *health dimension*: healthy life expectancy (HLE), cancer disease death rate (CDDR), heart disease death rate (HDDR); (4) *education dimension*: population 25-64 with higher education (PE), lifelong learning (LL), accessibility to universities (AU); (5) *labor market dimension*: employment rate (ER), long-term unemployment (LTUR), labor productivity (LP); (6*) market size dimension*: disposable income (DI), gross domestic product (GDP); (7) *business sophistication*: employment in sophisticated (K-N) sectors (ESS), gross valued added of sophisticated (K-N) sectors (GVA); (8) *innovation dimension*: total patent applications (TPA), core creative class employment (CCCE), gross expenditure on research and development (GERD), human resources in science and technology (HRST), high-tech patents (HTP), ICT patents (ICT). The reference period for calculating Composite Weighted Index of EU Regional Resilience (CI) for 273 EU28 NUTS 2 regions is the year 2013 as last actual year for data availability for this territorial unit.

Why carry out factor analysis (FA)? If we can summarize a multitude of measurements with a smaller number of factors without losing too much information, we have achieved some economy of description, which is one of the goals of scientific investigation. It is also possible that FA will allow us to test theories involving variables which are hard to measure directly. Finally, at a more prosaic level, FA can help us establish that sets of questionnaire items (observed variables) are in fact all measuring the same underlying factor (perhaps with varying reliability) and so can be combined to form a more reliable measure of that factor. There are a number of different varieties of FA [10]. FA is calculated via IBM SPSS Statistics 22 and are applied following FA features: Principal Component Analysis as extraction method; Varimax with Kaiser Normalization as rotation method.

Indicators are pieces of information that summarize the characteristics of a multiple-criteria decision-making system or highlight what is happening in this system. A mathematical combination (or aggregation) of a set of indicators is most often called an 'index' or a 'composite indicator'. The definition of composite indicators used in

this paper is adopted by the European Commission, i.e. *composite indicators are based on sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting these sub-indicators* [9, p.5]. Weighting and aggregation systems have a crucial effect on outcome of composite index. There is not only one proper method. Although various functional forms for the underlying aggregation rules of a composite indicator have been developed in the literature, e.g. [6] or [7], in the standard practice, a composite indicator can be considered a *weighted linear aggregation rule* applied to a set of variables. In this framework, a crucial role is played by the concept of weight. The evaluation of the criteria weights may be subjective, objective and integrated. List of the most common method is summarized, for instance, in [2] or [7]. All quantitative approaches of criteria weighting are based on the matrix $\mathbf{R} = \|r_{ij}\|\ (i = 1,...,p;\ j = 1,...,k)$ of the criteria significances $R_1,...,R_k$, characterizing the compared alternatives $A_1,...,A_p$. These significances r_{ij} may be statistical data or the estimates provided by experts. *Subjective methods* of weight determination are based on expert evaluation. His/her experience and knowledge allows for providing the most valuable information about the compared objects. There are numerous techniques for subjective determining the criteria weights (significances), including e.g. ranking or pairwise comparison, see e.g. [2]. This type of criteria weighting includes common practice in attaching weights where greater weight should be given to criteria (components) which are considered to be more significant in the context of multiple-criteria decision-making process or the particular composite indicator. The *objective approaches* to calculating the criteria weights evaluate the structure of matrix **R** representing the values r_{ij}, while the values of the weights may change together with the values themselves. In this paper we used the *entropy method* to determine the weight of evaluating factors, and applied it in resilience evaluation in the case of EU28 NUTS 2 regions and Member States. The entropy method based on information on alternatives can be used only in case of a finite number of alternatives. This method requires knowledge of the values of all the criteria for all variants in the matrix **R**. In the theory of information the entropy is the criterion of uncertainty posed by a discrete probability distribution p_i. This degree of uncertainty is expressed by [3] in the formula:

$$S(p_1, p_2,..., p_n) = -c.\sum_{i=1}^{n} p_i.ln\ p_i, \tag{1}$$

where c is a positive constant. Equation (1) express entropy in statistical concept, therefore entropy can be found as probability distribution p_i and thus terms of entropy and probability are considered as synonyms. Suppose all p_i equal, then for given i, $p_i = \frac{1}{n}$ reaches $S(p_1, p_2,..., p_n)$ maximum value. From matrix **R** we can determine share of the i-th variant on the sum of the j-th criteria for all criteria p_{ij} from the formula:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^{p} r_{ij}}, i = 1,2,..., p, j = 1,2,...,k. \tag{2}$$

For the j-th criterion entropy (s_j) can be determined by formula:

$$s_j = -c.\sum_{i=1}^{p} p_{ij}.ln\ p_{ij}, j = 1,2,...,k. \tag{3}$$

If suppose $c = \frac{1}{\ln p}$, then $0 \le s_j \le 1$ is guaranteed. Non normalized entropy weight of j-th criteria (d_j) can be found in formula:

$$d_j = 1 - s_j, j = 1,2,...,k, \tag{4}$$

while respective normalized weights w_i are obtained from the formula:

$$w_j = \frac{d_j}{\sum_{j=1}^{k} d_j}, j = 1,2,...,k. \tag{5}$$

Finally, calculation of Composite Weighted Index of Regional Resilience (CWIRR) is based on weighted linear aggregation:

$$CWIRR_r = \sum_{i=1}^{n} w_{i_F}.F_{i,r}, \tag{6}$$

where w_{i_F} is a weight attached to each factor F of regional resilience normalized between zero and one, and $F_{i,r}$ presents factor score of each extracted factor and EU NUTS 2 region.

3 Composite Weighted Index of EU Regional Resilience Evaluation

FA was applied for finding relevant factors of resilience based on the used data set – indicators were divided into factors that are crucial for EU regional resilience and competiveness as well. In the paper, five dominating factors: *community links* (CL), *human capital and socio-demographic structure* (HC-SDS), *labor market* (LM), *economic performance* (EP), *innovation, science and research* ISR). These factors explained 84,368 % of total variability in reference period 2013 (see Table 2), what can be considered as very satisfactory result. Table 2 also shows indicators and their belonging to relevant factors, which are also classified with respect to their importance to resilience, i.e. weights for each dimension are mentioned. Based on FA results it is clear, that indicators associated within each factor are relevant for its dimension of resilience; also number (No.) of indicators is balanced across factors. Based on Entropy method results it is evident, that values of weights are also balanced across factors. The greatest impact on the overall regional resilience has *human capital and socio-demographic structure* dimension, what is logical considering the importance of human capital and its manifestations in all economic areas.

Factors	**CL**	**HC-SDS**	**LM**	**EP**	**ISR**	Sum of Normalized Weights
Indicators	GE, C, RL, MPA, RPA	HLE, CDDR, HDDR, PE, LL, AU	ER, LTUR, ESS, CCCE	LP, GVA, DI, GDP	TPA, GERD, HRST, HTP, ICT	
No. of indicators	6	6	4	4	5	
Weights	0,205	0,223	0,195	0,194	0,182	1,000

Table 2 Results of Factor Analysis and Entropy Method

In Figure 1, results of composite index for 273 EU 28 NUTS 2 regions are illustrated. Curve closer to value 0, region is less resistant to crises, and conversely, higher value and curve has more distant from center, region is more resistant to crises and more competitive. There are obvious differences between traditionally developed and known less developed regions what means that results of composite index are conclusive and relevant to topic.

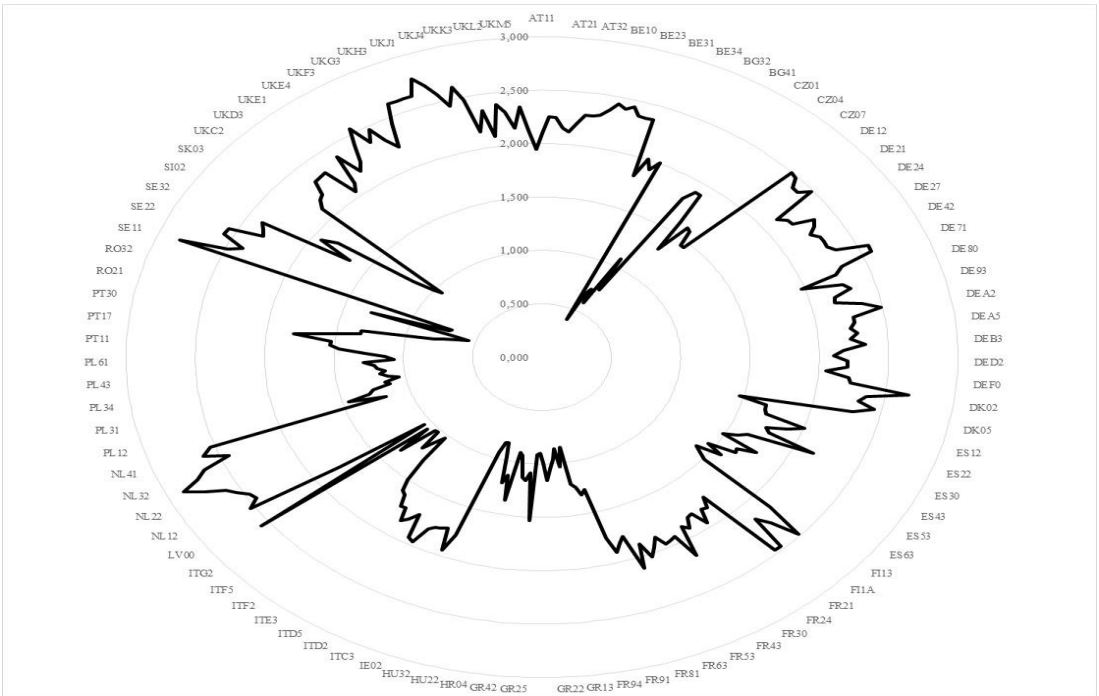


Figure 1 Composite Index of Regional Resilience for EU28 NUTS 2 Regions – Regional Level

In Figure 2, results of composite index for national level of analysis, i.e. for 28 EU Member States, are illustrated. Composite index for countries was calculated as median of composite index values for regions within each country. There is the same logic in results interpretation as in the case of regional level in Figure 2 and confirmed thus differences not only among regions, but also at national level between group of 'old' and 'new' EU countries. Exact values of composite index for all EU NUTS 2 regions NUTS 0 countries are shown in Annex 1 where (based on traffic light method) dark-grey color means higher values of composite index, i.e. greater resilience, and conversely (white color means percentile 50).

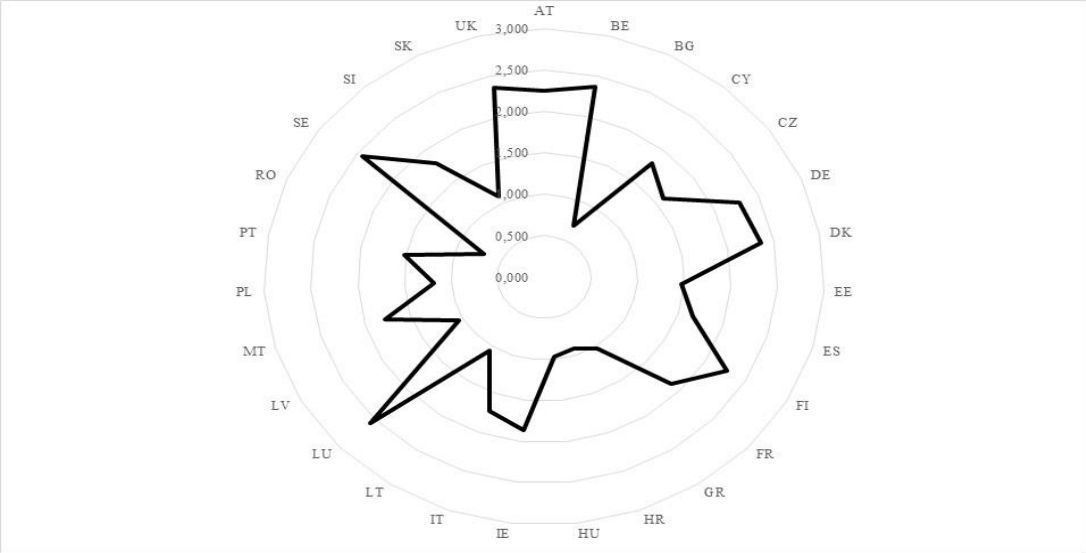


Figure 2 Composite Index of Regional Resilience for EU28 Countries (NUTS 0) – National Level

4 Conclusion

Regional resilience is much broader concept beyond the economic dimension. It is also reasonable to assume that application of similar indices at lower territorial level will require adaptation to national conditions and specifics. The index designed in this paper represents the initial concept needs to be developed in further research. Because it takes a long time to change the regional characteristics that affect resilience-related outcomes, policies and strategies that are put in place after a region has experienced an economic shock are challenging activities, what is our future research orientation. In the framework of results, the entropy method used within the context of decision making can be further used to define wider background or remove red tape so as to achieve no-regret conditions in the long term period.

Acknowledgements

Supported by the SGS project (SP2015/106) of Faculty of Economics, VŠB-Technical University of Ostrava and the Operational Programme Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Foster, K.A.: A Case Study Approach to Understanding Regional Resilience. *Working paper 2007-08*, Institution of Rural and Regional Development, Berkeley, 2007.
[2] Ginevičius, R., Podvezko, V. Objective and Subjective Approaches to Determining the Criterion Weight in Multicriteria Models. In: *Proceedings of International Conference RelStat 04*, Riga, Latvia, 2004, 133–139.
[3] Karmeshu: *Entropy Measures, Maximum Entropy Principles and Emerging Applications*. Springer-Verlag, Berlin Heidelberg, 2003.
[4] Koutský, J. et al: *Profilace měkkých faktorů regionálního rozvoje jako nástroj posilování regionální odolnosti a adaptability*. Certifikovaná metodika Ministerstva pro místní rozvoj, Praha, 2012.
[5] Martin, R.: Regional Economic Resilience, Hysteresis and Recessionary Shocks. *Journal of Economic Geography*, **12/1** (2012), 1–32.
[6] Munda, G., Nardo, M.: Constructing Consistent Composite Indicators: the Issue of Weights. *Working paper EUR 21834 EN*, EC – Joint Research Centre, Luxembourg, 2005.
[7] OECD: *Handbook on Constructing Composite Indicators*. *Methodology and User Guide*. OECD, Paris, 2008.
[8] Pindus, N., Weir, M., Wial, H., Wolman, H.: *Building Resilient Regions: Urban and Regional Policy and Its Effects*. Brookings Institution Press, Washington, 2012.
[9] Saisana, M., Tarantola, S.: *State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development*. EC – Joint Research Centre, 2002.
[10] Stevens, J.: *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, New Jersey. 1986.
[11] Svoboda, O.: Economic Performance and Research Activity as an Important Factor of Economic Resilience of Region. In: *Proceedings of 17th International Colloquium on Regional Sciences.* Masaryk University, Brno, 2014, 179–185.
[12] Tamásy, Ch., Diez, J.R.: *Regional Resilience, Economy and Society*. Ashgate, Farnham, 2013.

Annex 1: Values of Composite Index of Regional Resilience for EU28 NUTS 2 Regions (in alphabetical order)

NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 2	CI	NUTS 0	CI
AT11	2,093	DE13	2,404	DK01	2,665	FR30	1,925	HU23	0,962	NL31	2,876	SE11	2,831	UKJ2	2,677	AT	2,250
AT12	2,250	DE14	2,403	DK02	2,365	FR41	1,881	HU31	0,837	NL32	2,739	SE12	2,485	UKJ3	2,574	BE	2,365
AT13	2,250	DE21	2,485	DK03	2,320	FR42	2,168	HU32	0,843	NL33	2,651	SE21	2,387	UKJ4	2,447	BG	0,696
AT21	2,149	DE22	2,128	DK04	2,444	FR43	1,931	HU33	0,943	NL34	2,440	SE22	2,564	UKK1	2,610	CY	1,802
AT22	2,120	DE23	2,184	DK05	2,297	FR51	1,951	IE01	1,777	NL41	2,613	SE23	2,555	UKK2	2,477	CZ	1,592
AT31	2,211	DE24	2,200	EE00	1,472	FR52	1,966	IE02	1,949	NL42	2,541	SE31	2,307	UKK3	2,156	DE	2,276
AT32	2,285	DE25	2,349	ES11	1,677	FR53	1,873	ITC1	1,737	PL11	1,186	SE32	2,334	UKK4	2,346	DK	2,365
AT33	2,289	DE26	2,311	ES12	1,680	FR61	1,951	ITC2	1,756	PL12	1,458	SE33	2,375	UKL1	2,094	EE	1,472
AT34	2,307	DE27	2,253	ES13	1,700	FR62	2,033	ITC3	1,769	PL21	1,291	SI01	1,657	UKL2	2,388	ES	1,658
BE10	2,365	DE30	2,302	ES21	2,003	FR63	1,885	ITC4	1,804	PL22	1,258	SI02	1,935	UKM2	2,311	FI	2,266
BE21	2,435	DE41	2,302	ES22	1,892	FR71	2,108	ITD1	1,964	PL31	1,123	SK01	1,816	UKM3	2,158	FR	1,877
BE22	2,402	DE42	2,302	ES23	1,745	FR72	1,777	ITD2	1,939	PL32	1,152	SK02	1,164	UKM5	2,352	GR	1,038
BE23	2,440	DE50	2,346	ES24	1,783	FR81	1,828	ITD3	1,720	PL33	1,047	SK03	1,008	UKM6	2,125	HR	0,919
BE24	2,365	DE60	2,575	ES30	2,152	FR82	1,907	ITD4	1,797	PL34	1,124	SK04	0,944	UKN0	1,950	HU	0,962
BE25	2,362	DE71	2,571	ES41	1,652	FR83	1,755	ITD5	1,841	PL41	1,176	UKC1	2,103			IE	1,863
BE31	2,365	DE72	2,318	ES42	1,592	FR91	1,274	ITE1	1,717	PL42	1,141	UKC2	2,174			IT	1,717
BE32	1,825	DE73	2,253	ES43	1,485	FR92	1,321	ITE2	1,727	PL43	1,197	UKD1	2,202			LT	1,068
BE33	2,010	DE80	1,975	ES51	1,782	FR93	1,243	ITE3	1,625	PL51	1,211	UKD2	2,320			LU	2,566
BE34	1,919	DE91	2,276	ES52	1,643	FR94	1,211	ITE4	1,602	PL52	1,287	UKD3	2,326			LV	1,059
BE35	2,010	DE92	2,318	ES53	1,658	GR11	0,852	ITF1	1,503	PL61	1,069	UKD4	2,329			MT	1,782
BG31	0,396	DE93	2,183	ES61	1,441	GR12	1,038	ITF2	1,285	PL62	1,130	UKD5	2,049			NL	2,572
BG32	0,660	DE94	2,174	ES62	1,584	GR13	0,859	ITF3	1,033	PL63	1,274	UKE1	2,117			PL	1,181
BG33	0,732	DEA1	2,365	ES63	1,385	GR14	0,959	ITF4	1,121	PT11	1,470	UKE2	2,392			PT	1,525
BG34	0,589	DEA2	2,489	ES64	1,434	GR21	1,047	ITF5	1,210	PT15	1,535	UKE3	2,177			RO	0,706
BG41	1,085	DEA3	2,276	ES70	1,508	GR22	1,154	ITF6	1,026	PT16	1,525	UKE4	2,250			SE	2,436
BG42	0,752	DEA4	2,273	FI13	2,161	GR23	0,995	ITG1	1,034	PT17	1,805	UKF1	2,401			SI	1,796
CY00	1,802	DEA5	2,243	FI18	2,485	GR24	0,898	ITG2	1,341	PT18	1,531	UKF2	2,551			SK	1,086
CZ01	1,895	DEB1	2,285	FI19	2,266	GR25	0,917	LT00	1,068	PT20	1,331	UKF3	2,360				
CZ02	1,895	DEB2	2,229	FI1A	2,161	GR30	1,529	LU00	2,566	PT30	1,325	UKG1	2,472				
CZ03	1,617	DEB3	2,336	FI20	2,470	GR41	1,084	LV00	1,059	RO11	0,792	UKG2	2,327				
CZ04	1,315	DEC0	2,180	FR10	2,465	GR42	1,158	MT00	1,782	RO12	0,720	UKG3	2,226				
CZ05	1,598	DED1	2,102	FR21	1,749	GR43	1,139	NL11	2,534	RO21	0,623	UKH1	2,442				
CZ06	1,585	DED2	2,204	FR22	1,848	HR03	0,932	NL12	2,447	RO22	0,549	UKH2	2,617				
CZ07	1,449	DED3	2,209	FR23	1,828	HR04	0,905	NL13	2,470	RO31	0,669	UKH3	2,617				
CZ08	1,445	DEE0	2,052	FR24	1,953	HU10	1,362	NL21	2,543	RO32	1,304	UKI1	2,617				
DE11	2,496	DEF0	2,222	FR25	1,821	HU21	1,129	NL22	2,602	RO41	0,717	UKI2	2,617				
DE12	2,486	DEG0	2,240	FR26	1,830	HU22	1,211	NL23	2,739	RO42	0,696	UKJ1	2,773				

Note: CI = Composite Index; NUTS 0 = Country/National Level; NUTS 2 = Regional Level

Real Options Pricing by the Finite Element Method with Hermite Cubic Elements

Kateřina Mičudová[1], Ladislav Lukáš[2]

Abstract. The paper is focused on numerical method that allows finding numerical solution of real option valuation problem which is based on the known Black-Scholes framework. Opposite to the lattice type methods, or binomial option pricing method, the solution is discretized by properly defined finite elements. The importance of valuation of real options stem from the fact it provides more flexible tool for assessment of opportunities within investment projects. Hence, real options represent interesting managerial instrument for investment project evaluation along with traditional procedures based upon cash flows scenes as discounted cash flow methods. The paper discusses in detail a way of space domain discretization made by cubic finite elements, which possess node interpolation conditions of Hermitean type. The general form of such approximation depending upon selected node unknowns is described in detail. Such type of node unknowns allows direct calculation of nodal values of real option and also some of their Greeks, i.e. the sensitivities of real option values upon marginal changes of underlying asset. All numerical calculations and symbolic manipulations are performed in sw Mathematica.

Keywords: option pricing, real options, Black-Scholes model, shape functions, finite element method.

JEL Classification: C63, G13
AMS Classification: 91G80

1 Introduction

At present, real option analysis (ROA) seems to be a promising concept which links capital budgeting decisions and corporate valuation approach under a common framework. So, in theory, ROA looks like a perfect tool for managers to use since it is based on financial option pricing theory being deeply elaborated now. However, financial world is less complex than the real world of corporations with investments, projects, productions, and all kinds of business. Well-known book [4] brings thorough overview of investment valuation methods based upon discounted cash flows and real options, as well. In compact form, real options and investment under uncertainty is described in [10].

The traditional approach to valuing investment projects, based on net present value (NPV), essentially involves discounting the expected net cash flows from a project at a discount rate that reflects the risk of those cash flows, i.e. risk-adjusted discount rate. Such approach points the adjustment for risk to the discount rate. On the contrary, an alternative approach is to make proper adjustment for risk to the cash flows and consequently to discount the resulting certainty-equivalent cash flows, instead of the expected cash flows, at the risk-free rate of interest. Generally, certainty-equivalent cash flows are defined as the certain financial amounts which would have the same value as the uncertain cash flows. Usual popularity of NPV approach in practice is based simply on speculative idea that estimation of risk-adjusted discount rate is easier than the certainty-equivalent cash flows. However, in certain cases, such as in the case of commodities or products for which futures contracts exist, the contrary is evidently true, as the market-traded future prices provide framework for determination of certainty-equivalent cash flows. Once the adjustment for risk has been appropriately made to the cash flows, the relevant discount rate is the risk-free rate of interest.

The critical advantage of working in that risk-neutral environment in which the relevant discount rate is the risk-free rate of interest is that it is an appropriate and convenient for option pricing.

[1] University of West Bohemia/Faculty of Economics, Dept. of Economics and Quantitative Methods, Husova 11, 306 14 Pilsen, pitrovak@kem.zcu.cz.
[2] University of West Bohemia/Faculty of Economics, Dept. of Economics and Quantitative Methods, Husova 11, 306 14 Pilsen, lukasl@kem.zcu.cz

2 Real option pricing model

We follow [1] and use also [2] for building real option pricing model. Generally, there are two important ways in financial option pricing that allow one to price real option in the investment projects considering the underlying asset as a real asset. First, it is a continuous time model based usually on Black-Scholes model (B-S) and its various generalizations. For example, Glasserman [5] gives a family of various continuous time models and appropriate simulation solving procedures. At second, it is a discrete time model, which is represented by binomial option pricing model in particular, and further generalizations. Both of them are based on the assumptions of riskless arbitrage opportunities absence and on risk-neutral valuation.

The B-S model assumes that the rate of return, defined by $\mathrm{d}S_t/S_t$ as usual, of the underlying or subjacent real asset, denoted S_t traditionally, follows a generalized Wiener process. It means that the real asset S_t obeys a lognormal distribution. In this case, the real option price $V_t = F(S_t, t)$, $0 \le t \le T$, being an unknown function of S_t and t obeys the well-known B-S partial differential equation

$$\frac{\partial F}{\partial t} + rS\frac{\partial F}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 F}{\partial S^2} - rF = 0, \tag{1}$$

where we omit index t in S_t and both arguments in $F(S_t, t)$ thus to elevate readability of (1), assuming real asset value $0 \le S_t < +\infty$, r is the risk-free rate, t denotes the time since the real option was issued, T represents the expiration date or maturity date of the real option, and σ is the real asset volatility.

In finance naturally, the eq. (1) is a backward moving equation, i.e. it is solved on the domain $\Omega_{S,t} = [0,+\infty[\times [0,T]$ from the future $t = T$ with final condition set to the present time $t = 0$ providing all other boundary conditions are given, which depend upon the type of real option.

If the real option corresponds in type with the European call option then the final condition at $t = T$ is defined as follows

$$F(S_T, T) = \max(S_T - K, 0), \quad 0 \le S_T < +\infty, \tag{2}$$

where K being given is exercise or strike price of the option. We know in general, a call/put option being issued on time $t = 0$ gives the owner the right to buy/sell the underlying real asset for the price K on an exercise date, which coincides with expiration date T in case of the European option, or it can be any time between 0 and T in case of the American option.

Boundary conditions for $S_t = 0$ and $S_t \to +\infty$ are following

$$F(0, t) = 0, \quad \lim \partial F(S_t, t)/\partial S_t = 1,\ S_t \to +\infty, \quad 0 \le t \le T, \tag{3}$$

which mean that the call option has no value if value of the underlying real asset falls to zero, and if $S_t \to +\infty$ then the option value V_t is asymptotically equivalent to S_t. The first boundary condition in (3) is called the Dirichlet boundary condition, while the second one the Neumann boundary condition, within theory of partial differential equations.

In similar way, if the real option corresponds in type with the European put option then the final condition at $t = T$ is defined by

$$F(S_T, T) = \max(K - S_T, 0), \quad 0 \le S_T < +\infty, \tag{4}$$

where K being given is exercise or strike price of the option. We know in general, a call/put option being issued on time $t = 0$ gives the owner the right to buy/sell the underlying real asset for the price K on an exercise date, which coincides with expiration date T in case of the European option, or it can be any time between 0 and T in case of the American option.

Boundary conditions for $S_t = 0$ and $S_t \to +\infty$ are following

$$F(0, t) = K\exp(-r(T - t)), \quad \lim \partial F(S_t, t)/\partial S_t = 0,\ S_t \to +\infty, \quad 0 \le t \le T, \tag{5}$$

which mean that the put option gets value of strike price discounted by risk-less rate r during time period $T - t$, and if $S_t \to +\infty$ then the option value V_t is asymptotically negligible to S_t.

There are several other approaches for numerical solution of option pricing problems, e.g. [3], and [9]. However, we still follow [1] whereas the next step of solving procedure is transformation of boundary value problems (1) – (3), and (1), (4) – (5) into corresponding BV problems for the well-known one-dimensional heat conduction

equation, which takes a canonic form (6) thus representing the simplest partial differential equation of parabolic type. Moreover, the final condition (4) at $t = T$ will be transformed into an initial one which is more advantageous from numerical point of view.

$$\frac{\partial u}{\partial \tau} - \frac{\partial^2 u}{\partial X^2} = 0, \tag{6}$$

where new variables X_τ, τ, and $u(X_\tau, \tau)$ are introduced to replace S_t, t, and $F(S_t, t)$ as follows

$$X_\tau = \ln(S_t/K), \quad \tau = \sigma^2(T-t)/2, \quad u(X_\tau, \tau) = (1/K)F(S_t, t)\exp((1/2)(\gamma-1)X_\tau + (1/4)(\gamma+1)^2\tau), \ \gamma = 2r/\sigma^2. \tag{7}$$

Original ranges $0 \le S_t < +\infty$, and $0 \le t \le T$ are mapped onto $-\infty < X_\tau < +\infty$, and $0 \le \tau \le \sigma^2 T/2$ by transformations (7). However, we need to recast the boundary conditions (2) – (5), too. Following [1], we may write

$$(2) \ \sim> \ u(X_0, 0) = \max(\exp((1/2)(\gamma+1)X_0) - \exp((1/2)(\gamma-1)X_0)), 0), \ -\infty < X_0 < +\infty, \tag{8}$$

$$\begin{gathered}(3) \ \sim> \ \lim u(X_\tau, \tau) = 0, \quad X_\tau \to -\infty, \qquad 0 \le \tau \le \sigma^2 T/2, \\ \lim \partial u(X_\tau, \tau)/\partial X_\tau = (1/2)((\gamma+1)\exp(X_\tau)) - (\gamma-1)\exp(-\gamma\tau))\exp((1/2)(\gamma-1)X_\tau + (1/4)(\gamma+1)^2\tau), \quad X_\tau \to +\infty,\end{gathered} \tag{9}$$

and also

$$(4) \ \sim> \ u(X_0, 0) = \max(\exp((1/2)(\gamma-1)X_0) - \exp((1/2)(\gamma+1)X_0)), 0), \ -\infty < X_0 < +\infty, \tag{10}$$

$$\begin{gathered}(5) \ \sim> \ \lim u(X_\tau, \tau) = \exp((1/2)(\gamma-1)X_\tau + (1/4)(\gamma+1)^2\tau), \quad X_\tau \to -\infty, \\ \lim \partial u(X_\tau, \tau)/\partial X_\tau = 0, \quad X_\tau \to +\infty, \qquad 0 \le \tau \le \sigma^2 T/2.\end{gathered} \tag{11}$$

So, instead of final BV problems for backward moving eq. (1) on the domain $\Omega_{S,t}$ we have got equivalent initial BV problems for forward moving eq. (6), in particular (6) – (9), and (6), (8) – (9), respectively, being defined on the domain $\Omega_{X,\tau} = \{(X_\tau, \tau) \mid -\infty < X_\tau < +\infty, \ 0 \le \tau \le \sigma^2 T/2\}$.

3 Domain discretization and Hermitean finite elements

There is well-known that wide variety of methods exists for solving parabolic equations numerically, see e.g. [7]. However, basic classes are the following:

- Full discretizations:
  - Application of finite difference methods to the classical initial BV problem.
  - Application of so called space-time finite element methods to weak formulation of the initial BV problem.
- Semidiscretizations (methods of lines):
  - The method of vertical lines, i.e. discretization starts with respect to spatial variable and yields a system of ordinary differential equations which is further to be solved.
  - The method of horizontal lines (Rothe's method), i.e. discretization starts with respect to time variable.

First, for numerical solving we need to shrink the domain $\Omega_{X,\tau}$ as the space variable X_τ can take infinite values. Hence, instead of $\Omega_{X,\tau}$ we accept a subdomain $Q_{X,\tau} = \{(X_\tau, \tau) \mid -X_b < X_\tau < X_d, \ \tau_0 \le \tau \le \tau_f\}$, introducing given finite bounds X_b, $X_d > 0$, and denoting $\tau_0 = 0$, and $\tau_f = \sigma^2 T/2$, simply.

Now, let us consider discretization of the space variable with a step ΔX, and integer m given, as follows

$$\Delta X = (X_b + X_d)/m, \quad X_i = -X_b + (i-1)\Delta X, \quad i = 1, \ldots, m+1. \tag{12}$$

Such discretization is adopted in [1], too, and one-dimensional linear and quadratic Lagrange finite elements are utilized therein. Both have the length ΔX, they are formed by intervals $[X_i, X_{i+1}]$ geometrically, and carry either linear shape functions, or quadratic ones. Generic linear shape functions are defined by unit value at one node and zero value at the other one, so there are two different linear shape functions, in general. While, generic quadratic functions are defined in similar way, i.e. by unit value at one node, and zero values at two residual

ones, so there are three different quadratic shape functions, in general, provided the three nodes are defined by points X_i, $(X_i + X_{i+1})/2$, X_{i+1}, respectively. Such finite elements serve well for method of vertical lines.

At present, there is huge number of books and papers dealing with finite element method theory and applications, as well. As examples, see books [6] – [8].The majority of sources is focused upon numerical solution of problems stemming from continuum mechanics. However, applications in finance are increasing rapidly.

Our goal in this paper is to construct both one-dimensional and two-dimensional Hermitean finite elements implementing symbolic power of sw Mathematica. Such elements are suitable for either method of vertical lines, or full-discretization method realized by space-time rectangular elements. Their importance and advantage consists in nodal parameters. In general, finite elements constructed by Lagrange interpolation procedure consider nodal function values only, whereas elements constructed by Hermite interpolation procedure provide except nodal values also values of derivatives at nodes. In option pricing computation, it means that together with nodal values of function u, and F, subsequently, we obtain also nodal values of some Greeks directly, which play important role within hedging analysis, in general.

3.1 One-dimensional cubic and quintic Hermitean elements

Unit cubic Hermitean element is defined on unit interval [0,1] having two nodes at $\xi_0 = 0$, and $\xi_1 = 1$, respectively, and providing four different shape functions, in general. An interval $[X_i, X_{i+1}]$ is mapped from [0,1] by simple linear mapping (13), which is invertible one

$$x = X_i + (X_{i+1} - X_i)\xi, \quad 0 \le \xi \le 1. \tag{13}$$

Now, we assume general cubic polynomial $v(\xi)$ with its coefficients a_k, $k = 0,\ldots,3$ to be determined by Hermitean interpolation conditions, with four values u_l, v_l, $l = 0,1$, given, as follows

$$v(\xi) = \sum_{k=0}^{3} a_k \xi^k, \quad v(\xi_l) = u_l, \quad \mathrm{d}v(\xi_l)/\mathrm{d}\xi = v_l, \; l = 0,1. \tag{14}$$

Determination of a_0, and a_1 is trivial, as it holds $a_0 = u_0$, and $a_1 = v_0$, respectively, whilst a_2, and a_3 are determined by Mathematica function Solve as follows

```
Solve[a2+a3==u1–u0–v0 && 2a2+3a3==v1–v0,{a2,a3}]//FullSimplify
```

Generator of cubic shape functions defined on [0,1] is constructed by Mathematica pure function, where # stands for an argument and u0,u1,v0,v1 represents nodal parameters u_l, v_l, $l = 0,1$ in (14)

```
u3ρ:=u0+v0*# + (3(u1–u0) – (2v0+v1))*#^2 + (–2(u1–u0)+v0+v1)*#^3;
```

In **Fig. 1**, there are plotted cubic shape functions generated by pure function u3ρ, and denoted $\varphi_1(\xi\,;\, u_0{=}1)$ ~ full line, $\varphi_2(\xi\,;\, u_1{=}1)$ ~ dotted, $\varphi_3(\xi\,;\, v_0{=}1)$ ~ dashed, and $\varphi_4(\xi\,;\, v_1{=}1)$ ~ dash-dotted, where in parenthesis appears argument ξ being separated by semicolon from the only one unit interpolation parameter accepted while all other ones are zero.

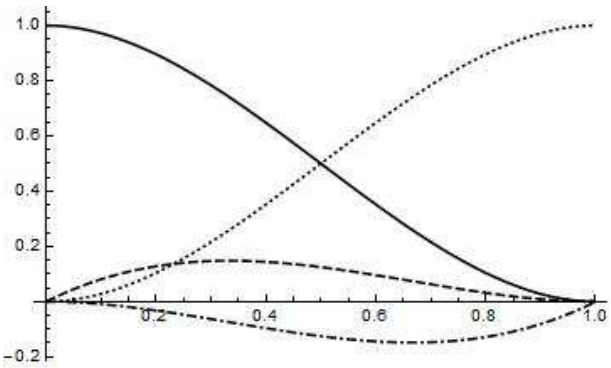


Figure 1 Cubic shape functions $\varphi_n(\xi\,;\,\cdot)$, n=1,…,4, where · stands for unit interpolation parameter

Unit quintic Hernitean finite element is defined on the unit interval [0,1], too, thus having two nodes ξ_0, and ξ_1, as well. However, instead of four it provides six different shape functions. We assume general quintic polynomial $v(\xi)$ with its coefficients a_k, $k = 0,\ldots,5$ to be determined by Hermitean interpolation conditions, with six values u_l, v_l, w_l, $l = 0,1$, as follows

$$v(\xi) = \sum_{k=0}^{5} a_k \xi^k, \quad v(\xi_l) = u_l, \quad \mathrm{d}v(\xi_l)/\mathrm{d}\xi = v_l, \quad \mathrm{d}^2v(\xi_l)/\mathrm{d}\xi^2 = w_l, \quad l = 0,1. \tag{15}$$

Determination of a_0, a_1, and a_2 is trivial, again, as it holds $a_0 = u_0$, $a_1 = v_0$, and $a_2 = w_0/2$, respectively, whilst a_3, a_4, and a_5 are determined by Mathematica function Solve, again

```
Solve[a3+a4+a5==u1–u0–v0–w0/2 && 3a3+4a4+5a5==v1–v0–w0 &&
6a3+12a4+20a5==w1–w0,{a3,a4,a5}]//FullSimplify
```

Generator of quintic shape functions defined on [0,1] is constructed by Mathematica pure function, again

```
u5ρ:=u0+v0*# + .5w0*#^2 + .5(20(u1–u0) –12v0 – 8v1 – 3w0+w1)*#^3 + (–15(u1–u0)+8v0+7v1+3w0/2 – w1)*#^4
+ .5(12(u1–u0) – 6(v0+v1) – w0 + w1)*#^5;
```

In **Fig. 2**, there are plotted quintic shape functions generated by pure function u5ρ, and denoted $\psi_1(\xi\,;\,u_0{=}1) \sim$ full(left), $\psi_2(\xi\,;\,u_1{=}1) \sim$ full(right), $\psi_3(\xi\,;\,v_0{=}1) \sim$ dashed(left), and $\psi_4(\xi\,;\,v_1{=}1) \sim$ dashed(right), $\psi_5(\xi\,;\,w_0{=}1) \sim$ dotted(left), and $\psi_6(\xi\,;\,w_1{=}1) \sim$ dotted(right).

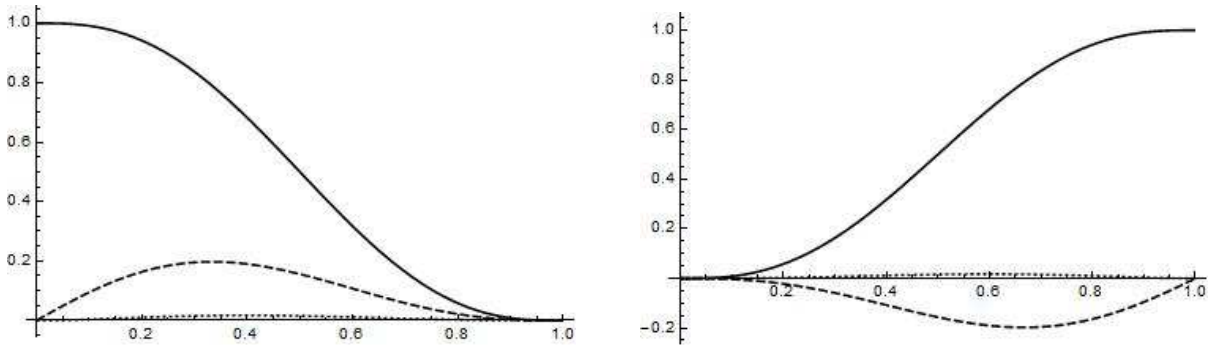


Figure 2 Quintic shape functions: $\psi_n(\xi\,;\,\cdot)$, n=1,3,5, on the left; $\psi_n(\xi\,;\,\cdot)$, n=2,4,6, on the right

3.2 Two-dimensional bi-cubic Hermitean rectangular element

Unit cubic Hermitean 2D element is defined on unit square [0,1]x[0,1] having four nodes at (0,0), (1,0), (0,1), and (1,1). However, we need to make regular discretization, first, by covering domain $Q_{X,\tau}$ by m x n rectangles, each one with sides ΔX and $\Delta\tau$, where ΔX is determined by (12), and $\Delta\tau$ by (16) in similar way

$$\Delta\tau = (\tau_f - \tau_0)/n, \quad \tau_j = \tau_0 + (j-1)\Delta\tau, \quad j = 1,\ldots,n{+}1. \tag{16}$$

Covering $Q_{X,\tau}$ by these rectangles generates $(m{+}1)$x$(n{+}1)$ regular mesh of nodes on $Q_{X,\tau}$. An arbitrary (i,j)-th rectangle element is mapped from the unit square by invertible linear mapping (17) in following way

$$x = X_i + (X_{i+1} - X_i)\xi, \quad \tau = \tau_j + (\tau_{j+1} - \tau_j)\eta, \quad 0 \le \xi \le 1,\ 0 \le \eta \le 1. \tag{17}$$

Now, we assume general bi-cubic polynomial $w(\xi, \eta)$ with its coefficients a_k, k = 0,…,15 to be determined by Hermitean interpolation conditions, with sixteen values in total, i.e. four ones at each corner A(0,0), B(1,0), C(0,1), and D(1,1) given, as follows

$$w(\xi, \eta) = \sum_{r=0}^{3} \alpha_r \xi^r \ \sum_{s=0}^{3} \beta_s \eta^s = \sum_{s=0}^{3} \sum_{r=0}^{3} a_k \xi^r \eta^s, \quad k = r{+}4s, \ r,s = 0,\ldots,3, \tag{18}$$

$$\{{}_{\mathrm{P}}w_{0,0} = w(\mathrm{P}),\ {}_{\mathrm{P}}w_{1,0} = \partial w(\mathrm{P})/\partial\xi,\ \ {}_{\mathrm{P}}w_{0,1} = \partial w(\mathrm{P})/\partial\eta,\ \ {}_{\mathrm{P}}w_{1,1} = \partial^2 w(\mathrm{P})/\partial\xi\partial\eta\,\}, \quad \mathrm{P} = \mathrm{A, B, C, D}.$$

Determination of the coefficients a_k, k = 0,…,15 from Hermitean interpolation conditions is performed symbolically by Mathematica, again. Since the code is much longer than in case of one-dimensional elements, we give just a short code snippet here.

First, the bi-cubic polynomial is defined by pure Mathematica function with two dummy arguments #1, #2 and coefficients a_k being represented by variables c00, …, c33, as follows

```
uξ3η3:=c00+c01*#2+c02*#2^2+c03*#2^3 + (c10+c11*#2+c12*#2^2+c13*#2^3)*#1 +
(c20+c21*#2+c22*#2^2+c23*#2^3)*#1^2 + (c30+c31*#2+c32*#2^2+c33*#2^3)*#1^3;
```

Now, we express all derivatives we need in symbolic way, and compose Solve command which starts as follows

```
Solve[uξ3η3[0,0] ==uA00 && uξ3η3[1,0] ==uB00 && uξ3η3[0,1] ==uC00 && uξ3η3[1,1] ==uD00 && …;
```

It yields a_k in symbolic form, which being plunged into bi-cubic polynomial forms the shape functions generator.

For illustration, **Fig. 3** shows difference between bi-cubic and bi-linear shape functions having only ${}_{A}w_{0,0} = 1$, and all others parameters set zero. In total, there is sixteen different shape functions available for a rectangular bi-cubic polynomial with Hermitean interpolation conditions at its vertices.

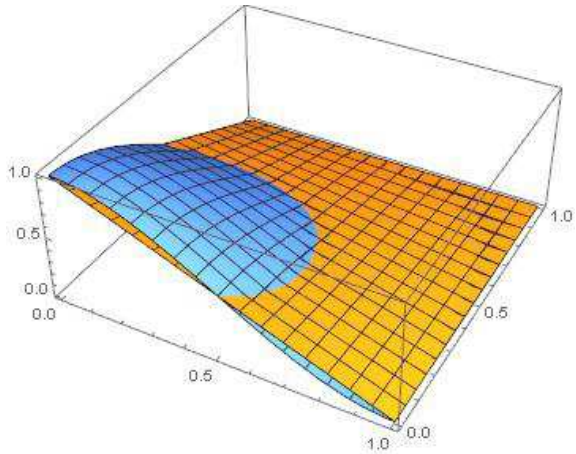


Figure 3 Comparison of bi-cubic (blue) and bi-linear (orange) shape functions with ${}_{A}w_{0,0} = 1$

4 Conclusions

Within the paper we discussed following topics:

- Formulation of real option pricing problem using traditional Black-Scholes model.
- Recasting final BV problems for B-S equation being recognized as natural formulation framework within option pricing theory into equivalent initial BV problems for simple parabolic partial differential equation which seem to be more attractive from computational point of view.
- Construction of basic shape functions for cubic and quintic 1-D finite elements, and bi-cubic rectangular element with Hermitean interpolation conditions using symbolic computation procedure in Mathematica.
- Near future work will be focused upon algorithmic realization of element assembly procedure using Mathematica, and running numerical experiments of solving some real option problems with finite element method.

Acknowledgements

The research project was supported by the grant no. 15-20405S of the Grant Agency, Prague, Czech Republic, and the grant SGS-2014-047 of the Faculty of Economics, University of West Bohemia in Pilsen.

References

[1] Andalaft-Chacur, A., Ali, M.M., Salazar, J.G.: Real options pricing by the finite element method. *Computers and Mathematics with Applications* **61** (2011), 2863–2873.
[2] Cipra, T.: *Financial and insurance formulae* (in Czech: *Finanční a pojistné vzorce*). Grada Publishing, Praha, 2006.
[3] Cortazar, G., Gravet, M., Urzua, J.: The valuation of multidimensional American real option using LSM simulation method. *Computers & Operations Research* **35** (2008), 113–129.
[4] Damodaran, A.: *Damodaran on Valuation – Security Analysis for Investment and Corporate Finance*. 2-nd ed., John Wiley and Sons, New York, 2006.
[5] Glasserman, P.: *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
[6] Johnson, C.: *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge Univ. Press, Cambridge, 1988.
[7] Knabner, P., and Angermann, L: *Numerical methods for elliptic and partial differential equations*. Springer, New York, 2003.
[8] Kolář, V., Kratochvíl, J., Leitner, F., Ženíšek, A.: *Computation of plane and space structures by finite element method* (in Czech: *Výpočet plošných a prostorových konstrukcí metodou konečných prvků*). SNTL, Praha, 1972.
[9] Rambeerich, N., et al.: High-order computational methods for option valuation under multifactor models. *EJOR* **224** (2013), 219–226.
[10] Schwartz, E.S., Trigeorgis, L.: Real Options and Investment under Uncertainty: An Overview. Online: http://mitpress.mit.edu/sites/default/files/titles/content/9780262693189_sch_0001.pdf.

Shapley Values of Cooperative Games with I-Fuzzy Expectations

Elena Mielcová[1]

Abstract. In the cooperative game theory players of a game are cooperating in order to increase a mutual profit. The cooperative game theory considers the question of a profit distribution, and provides several solution concepts; one of them is the concept of a Shapley value. In reality, the expected pay-offs are not always exactly given – they can be only expected with some precision. One of the possible approaches that include an uncertainty into cooperative games is the fuzzy-sets approach. This approach expect that players knows a degree of membership of specific expected payoffs. However, the real situation should be better covered, when also the part of indecisiveness about the future pay-offs is considered, as in I-fuzzy sets, described originally as Atanassov's intuitionistic fuzzy sets.

The main aim of this article is to discuss the construction of the Shapley value of transferable utility cooperative games when pay-offs are vague, in this case expressed as I-fuzzy numbers, and compare properties of this Shapley value with properties of the original crisp Shapley value.

Keywords: Shapley values, cooperative games, I-fuzzy sets.

JEL classification: C71
AMS classification: 91A12

1 Introduction

In the cooperative game theory players of a game are cooperating in order to increase a mutual profit. Considering that the profit can be distributed among players with respect to some coalitional agreement, these games are called also transferable utility games. Cooperative game theory considers the question of profit distribution, and provides several solution concepts; one of them is the concept of Shapley value.

In reality, the expected pay-offs are not always exactly given they can be only expected with some precision. One of the possible approaches that include uncertainty into cooperative games is the fuzzy approach, described for example in [9]. This approach expect that players knows degree of membership (and implicitly the member of non-membership) of specific expected payoffs. However, the real situation should be better covered, when also the part of indecisiveness about the future pay-offs is considered, as in I-fuzzy sets, described originally by Atanassov [2] as intuitionistic fuzzy sets. In general, the theory of I-fuzzy sets is considered to be an extension of fuzzy set theory [6].

The main aim of this article is to discuss the construction Shapley value of the transferable utility cooperative game when pay-offs are vague in this case expressed as I-fuzzy numbers, and compare properties of this Shapley value with properties of the original crisp Shapley value.

2 Preliminaries

2.1 Cooperative games

Let $N = \{1, 2 \dots n\}$ be a finite set of n players, let 2^N be the collection of all subsets of N. Any nonempty subset of N is called a coalition. A cooperative game is a pair (N, v) where $N = \{1, 2 \dots n\}$ is a set of n

[1]Silesian University in Opava/School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, mielcova@opf.slu.cz

players $v : 2^N \to R$ is a mapping defined on subsets of N with the property $v(\emptyset) = 0$ The function v is called a characteristic function of a game. The characteristic function v connects each coalition $K \subset N$ with a real number $v(K) \subset R$ representing total profit of coalition K; we assume $v(\emptyset) = 0$. A cooperative game can by denoted as (N, v), or simply only by characteristic function v.

For any pair of disjoint coalitions $K, L \subset N$, $K \cap L = \emptyset$ a coalition game (N, v) is called superadditive if $v(K \cup L) \geq v(K)+v(L)$; subadditive if $v(K \cup L) \leq v(K)+v(L)$; and additive if $v(K \cup L) = v(K)+v(L)$. A game (N, v) is additive if and only if it is superadditive and subadditive. We say that a TU cooperative game (N, v) is convex if for every pair of coalitions $K, L \subset N$ there hold $v(K \cup L)+v(K \cap L) \geq v(K)+v(L)$. In a cooperative game theory, games are usually expected to be superadditive.

The Shapley value was first introduced in 1953 when L. S. Shapley published his study "A value for n-person games" [15]. He defined three axioms (symmetry, efficiency and law of aggregation - additivity) that the value of a game should fulfill. He had shown that there exists one unique value fulfilling all three axioms, now called Shapley value. I will follow the Shapley's definitions:

Definition 1. Let U denote the universe of players. A carrier of v is any set $N \subset U$ with $v(S) = v(S \cap N)$ for all $S \subset N$. Let $\Pi(N)$ denote the set of permutations of N. If $\pi \in \Pi(N)$, then writing πS for the image of S under π , we may define function πv by $\pi v(\pi S) = v(S)$ for all $S \subset N$.

Definition 2. By the value $\phi(v)$ of the game v we shall mean a function that associates with each $i \in U$ a real number $\phi_i(v)$ satisfying the conditions of the following axioms:

Axiom 1 ("symmetry"): For all $\pi \in \Pi(N)$ there is $\phi_{\pi i}(\pi(v)) = \phi_i(v)$.

Axiom 2 ("efficiency"): For each carrier S of v, $\sum_S \phi_i(v) = v(S)$.

Axiom 3 ("additivity"): For all u, v such that $(u + v)(S) = u(S) + v(S)$, the value $\phi(u + v) = \phi(u) + \phi(v)$.

Theorem 1 (Shapley [15])**.** *The axioms 1, 2, 3 are sufficient to determine a unique ϕ for all games.*

In the proof of the theorem, Shapley has shown that there is a unique value obeying all three axioms defined on the space of all games. He expressed this unique value ϕ explicitly:

$$\phi_i(v) = \sum_{i \in S \subset N} \frac{(s-1)!(n-s)!}{n!} \big(v(S) - v(S \backslash i)\big) \tag{1}$$

where N is any finite carrier of U . This value ϕ is called Shapley value.

The equivalent form of the Shapley value was introduced in [8] and is based on considering marginal contributions in respective subgames:

$$\phi_j(N, v) = \frac{1}{n}[v(N) - v(N \backslash j)] + \frac{1}{n} \sum_{i \in N \backslash j} \phi_j(N \backslash i, v) \tag{2}$$

Shapley value as a solution concept is the one that assigns fair share of players' contributions to their payoffs. The next theorem was proven using theory from [12] and discussed in [14]. It states that the solution concept is a Shapley value if and only if for any player j "the sum of of the contributions of player j to the other players is equal to the sum of contributions of the other players to player j" [14]:

Theorem 2 (Perez-Castrillo, Wettstein [13])**.** *A value ϕ is the Shapley value iff for each game (N, v) and for all $i, j \in N$ it satisfies:*

$$\sum_{i \in N \backslash j} (\phi_i(N, v) - \phi_i(N \backslash j, v)) = \sum_{i \in N \backslash j} (\phi_j(N, v) - \phi_j(N \backslash i, v)) \tag{3}$$

Shapley value is expected to be a solution concept assigning fair distribution of mutual profit to players. According to Perez-Castrillo and Wettstein [14], the calculation of Shapley value can be based on the contribution of player j to the Shapley value of player i. Let's refer to the expression $\phi_i(N, v) - \phi_i(N \backslash j, v)$ as a concession, what player j concedes to player i, and denote it by c_i^j. The next Remark is direct consequence of Theorem 2.

Remark 1 (Perez-Castrillo, Wettstein [14]). A value ϕ is the Shapley value if and only if for each game (N, v) there exists a matrix of concessions $\left(c_i^j(N,v)\right)_{i,j}$ with $c_i^j(N,v) \in R$ for all $i, j \in N,\ i \neq j$, such that:

(1) $\phi_i(N,v) = \phi_i(N\backslash j, v) + c_i^j(N,v)$ for all $i, j \in N,\ i \neq j$;

(2) $\sum_{i \in N\backslash j} c_i^j(N,v) = \sum_{i \in N\backslash j} c_j^i(N,v)$

The part (1) can be viewed as a consistency property of the Shapley value, while the part (2) can be interpreted as a fairness requirement (the sum of concessions one player makes to the others equals the sum of concessions the others make to him).

I-fuzzy sets

The main aim of this article is to discuss the construction Shapley value expressed via intuitionistic fuzzy sets, described originally by Atanassov [2]. In general, the theory of intuitionistic fuzzy sets is considered to be an extension of fuzzy set theory. Considering the discussion about terminological difficulties concerning intuitionistic fuzzy sets, see [6], throughout this paper terms "I-fuzzy set", or, "IF set" are used instead of "intutitionistic fuzzy set".

Let a set $X = \{x_1, x_2, \ldots, x_n\}$ be fixed. Then an I-fuzzy set is defined as a set of triples $A = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle; x_i \in X\}$ where functions $\mu_A : X \to L$ and $\nu_A : X \to L$ for $L = [0,1]$ define the degree of membership and the degree of nonmembership of the element $x_i \in X$ to $A \subset X$, respectively. For an I-fuzzy set, the condition $0 \leq \mu_A(x_i) + \nu_A(x_i) \leq 1$ holds for all $x_i \in X$.

A definition of I-fuzzy quantity is based on the basic definition of I-fuzzy set, extended with respect to infinite set, in this case with respect to real numbers. In general, any I-fuzzy subset A of the set of real numbers R is called an I-fuzzy quantity with membership function $\mu_A : R \to [0,1]$, and nonmembership function $\nu_A : R \to [0,1]$ iff: (a) $\exists x_0 \in R$ such that $\mu_A(x_0) = 1$ and $\nu_A(x_0) = 0$; and (b) $\exists x_1, x_2 \in R, x_1 < x_2$ such that for all $x \notin [x_1, x_2]$ there is $\mu_A(x) = 0$ and $\nu_A(x) = 1$. Special subset of I-fuzzy quantities cover all I-fuzzy numbers, defined for example in [1], [4], and [7]. An I-fuzzy subset of the real line $A = \{\langle x, \mu_A(x), \nu_A(x)\ \rangle; x \in R\}$ is called an I-fuzzy number if:

(a) A is if-normal (there exist at least two points $x_0, x_1 \in X$ such that $\mu_A(x_0) = 1$ and $\nu_A(x_1) = 1$)

(b) A is if-convex (its membership function μ_A is fuzzy convex and its non-membership function ν_A is fuzzy concave);

(c) μ_A is upper semi-continuous and ν_A is lower semi-continuous;

(d) $A = \{x \in X; \nu_A(x) < 1\}$ is bounded.

Elementary arithmetic operations over I-fuzzy sets are derived with respect to so-called extension principle with some level of generality. Extension principle over I-fuzzy sets was introduced by Çoker [5] in his article discussing a topology of I-fuzzy sets. The idea of an extension principle was applied and discussed for example in [11] and [16].

The addition can be defined by an extension principle: Let a, b are two I-fuzzy quantities. Then $a \oplus b$ is also an I-fuzzy quantity with a membership and non-membership function:

$$\mu_{a\oplus b}(x) = \sup_{y \in R}[\min(\mu_a(y), \mu_b(x-y))]; \tag{4}$$

$$\nu_{a\oplus b}(x) = \inf_{y \in R}[\max(\nu_a(y), \nu_b(x-y))]. \tag{5}$$

Subtraction can be defined as an addition with the opposite element $a \oplus (-b)$. The membership and nonmembership functions of an opposition element are:

$$\mu_{-b}(x) = \mu_b(-x); \tag{6}$$

$$\nu_{-b}(x) = \nu_b(-x);. \tag{7}$$

In order to compare two I-fuzzy quantities, an I-fuzzy ordering relation should give the membership and nonmembership degree of given ordering relation. This type of ordering relation can be derived from the extension principle:

Definition 3. Let $\mathbb{R}$ be the set of all intuitionistic fuzzy quantities. For $a, b \in \mathbb{R}$ we define fuzzy relations $a \preccurlyeq$ and $a \approx b$ with membership functions $\mu_{\preccurlyeq} : \mathbb{R} \times \mathbb{R} \to [0,1]$, $\mu_{\approx} : \mathbb{R} \times \mathbb{R} \to [0,1]$ and nonmembership functions $\nu_{\preccurlyeq} : \mathbb{R} \times \mathbb{R} \to [0,1]$, $\nu_{\approx} : \mathbb{R} \times \mathbb{R} \to [0,1]$ such that:

$$\mu_{\preccurlyeq}(a,b) = \sup_{x \leq y}(\min(\mu_a(x), \mu_b(y))); \tag{8}$$

$$\nu_{\preccurlyeq}(a,b) = \inf_{x \leq y}(\max(\nu_a(x), \nu_b(y))); \tag{9}$$

$$\mu_{\approx}(a,b) = \sup_{x \in R}(\min(\mu_a(x), \mu_b(x))); \tag{10}$$

$$\nu_{\approx}(a,b) = \inf_{x \in R}(\max(\nu_a(x), \nu_b(x))). \tag{11}$$

Two fuzzy numbers a and b are identical, we write $a = b$, if for all $x \in R$ membership and nonmembership degrees of both I-fuzzy numbers are equal: $\mu_a(x) = \mu_b(x)$ and $\nu_a(x) = \nu_b(x)$.

3 Cooperative games with I-fuzzy payoffs

In general, taking into account cooperative games with vague payoffs, the uncertainty of possible payoffs implies also shifts in characteristics and in solution concepts of such games. The definition of I-fuzzy extension of a cooperative game is based on the fuzzy extension of cooperative game defined in [9]:

Definition 4. Let (N, v) be a cooperative game and let $w(K) = \langle x, \mu_K(x), \nu_K(x) \rangle$ be an I-fuzzy number with a membership function $\mu_K : R \to [0,1]$, and a non-membership function $\nu_K : R \to [0,1]$ fulfilling condition $0 \leq \mu_K(x) + \nu_K(x) \leq 1$; such that

- $\mu_K(v(K)) = 1$, and hence $\nu_K(v(K)) = 0$;
- $\mu_K(x)$ is non-decreasing for $x < v(K)$ and non-increasing for $x > v(K)$;
- $\nu_K(x)$ is non-increasing for $x < v(K)$ and non-decreasing for $x > v(K)$;
- $\mu_\emptyset(x) = 1$ iff $x = 0$, and $\nu_\emptyset(x) = 1$ for all $x \neq 0$.

Then w is the I-fuzzy characteristic function, and the pair (N, w) is called an I-fuzzy extension of the cooperative game (N, v).

3.1 Shapley value of cooperative games with I-fuzzy payoffs

The expression of the Shapley value for the cooperative games with crisp pay-offs (1) can be transformed into the expression the Shapley value for extensions of cooperative games with I-fuzzy pay-offs using notation for I-fuzzy addition the same way as was derived in fuzzy case [9], [3]:

$$\Phi_i(w) = \sum_{i \in S \subset N}^{\oplus} \frac{(s-1)!(n-s)!}{n!} \big(w(S) \oplus (-w(S \backslash i))\big) \tag{12}$$

where summation symbol $\sum^{\oplus}$ denotes addition of I-fuzzy quantities. $\Phi_i(w)(x)$ is an I-fuzzy number, determined by its membership and non-membership function defined for all $x \in R$. The I-fuzzy Shapley value $\Phi_i(w)$ fulfill properties according to Shapley Axioms 1-3 in I-fuzzy sense similarly as in the fuzzy case [9], [3]:

- **Symmetry:** Values $\Phi_i(w)(x)$ are not dependent on order of $i \in N$.

- **Efficiency:** Clearly summation of Shapley values of original crisp game (N, v), denoted $\sum_i \phi_i$ gives modal values (values with membership function equal 1, and non-membership function equal 0) of the summation of Shapley values of I-fuzzy extension of this game (N, w), denoted $\sum_i^{\oplus} \Phi_i$. According to (10) and (11) this imply that the membership function of relation operator $\mu_{\approx}(\sum_i^{\oplus} \Phi_i, w(N)) = 1$, and non-membership function of the operator $\nu_{\approx}(\sum_i^{\oplus} \Phi_i, w(N)) = 0$.
- **Additivity:** Let (N, v_1) and (N, v_2) are two games such that $(v_1 + v_2)(S) = v_1(S) + v_2(S)$, and let (N, w_1) and (N, w_2) are I-fuzzy extensions of respective games. Shapley values of crisp games are denoted $\phi_i^{(1)}$, and $\phi_i^{(2)}$, and of respective I-fuzzy extensions are denoted $\Phi_i^{(1)}$, and $\Phi_i^{(2)}$. Defining the characteristic function of a game game $(N, v_1 + v_2)$ to be $(v_1 + v_2)(K) = v_1(K) + v_2(K)$, with its I-fuzzy extension $(N, w_1 \oplus w_2)$ with characteristic function $(w_1 \oplus w_2)(K) = w_1(K) \oplus w_2(K)$ and I-fuzzy Shapley value Φ_i^*, then $\phi_i^{(1)} + \phi_i^{(2)}$ are modal values of Φ_i^*. Moreover, $\phi_i^{(1)}$, and $\phi_i^{(2)}$ are modal values of $\Phi_i^{(1)}$, and $\Phi_i^{(2)}$, respectively. Taking into account crisp game efficiency relation $\phi(v_1) + \phi(v_2) = \phi(v_1 + v_2)$, then $\mu_{\approx}(\Phi(w_1) \oplus \Phi(w_2), \Phi^*) = 1$, and $\nu_{\approx}(\Phi(w_1) \oplus \Phi(w_2), \Phi^*) = 0$.

Similarly, the expression of the equivalent form of Shapley value can be transformed into the expression the Shapley value for extensions of cooperative games with I-fuzzy pay-offs using notation for I-fuzzy addition relations:

$$\Phi_j(N, w) = \frac{1}{n}[w(N) \oplus (-w(N\backslash j))] \oplus \frac{1}{n} \sum_{i \in N\backslash j}^{\oplus} \Phi_j(N\backslash i, w) \tag{13}$$

After rearranging:

$$n\Phi_j(N, w) = w(N) \oplus (-w(N\backslash j)) \oplus \sum_{i \in N\backslash j}^{\oplus} \Phi_j(N\backslash i, w)$$

Taking into account efficiency condition:

$$w(N) = \sum_{i \in N\backslash j}^{\oplus} \Phi_i(N, w) \oplus \Phi_j(N, w) \text{ and } w(N\backslash j) = \sum_{i \in N\backslash j}^{\oplus} \Phi_i(N\backslash j, w) \oplus \Phi_j(N\backslash j, w)$$

this expression can be written as

$$n\Phi_j(N, w) = \sum_{i \in N\backslash j}^{\oplus} \Phi_i(N, w) \oplus \Phi_j(N, w) \oplus \Big(- \sum_{i \in N\backslash j}^{\oplus} \Phi_i(N\backslash j, w)\Big) \oplus (-\Phi_j(N\backslash j, w)) \oplus \sum_{i \in N\backslash j}^{\oplus} \Phi_j(N\backslash i, w).$$

Because $(n-1)\Phi_j(N, w)$ can be expressed as $\sum_{i \in N\backslash j}^{\oplus} \Phi_j(N, w)$, and $\Phi_j(N\backslash j, w)$ has nonmembership function equal to 1 at R, this expression can be rewritten as I-fuzzy equivalent form of Equation (3):

$$\sum_{i \in N\backslash j}^{\oplus} \Big(\Phi_i(N, w) \oplus (-\Phi_i(N\backslash j, w))\Big) = \sum_{i \in N\backslash j}^{\oplus} \Big(\Phi_j(N, w) \oplus (-\Phi_j(N\backslash i, w))\Big). \tag{14}$$

Taking into account approach introduced in [14], the expression $\Phi_i(N, w) \oplus (-\Phi_i(N\backslash j, w))$ is an I-fuzzy concession, what player j concedes to player i, and the expression $\Phi_j(N, w) \oplus (-\Phi_j(N\backslash i, w))$ is an I-fuzzy concession, what player i concedes to player j (concessions denoted as C_i^j and C_j^i, respectively). This implies that the fairness requirement as referred in Remark 1 condition (2) is fulfilled in I-fuzzy way:

$$\sum_{i \in N\backslash j}^{\oplus} C_i^j(N, w) = \sum_{i \in N\backslash j}^{\oplus} C_j^i(N, w). \tag{15}$$

4 Conclusions

The main aim of this article was to discuss the construction of the Shapley value of transferable utility cooperative games when pay-offs expressed as I-fuzzy numbers. Presented approach represents an extension of the theory of fuzzy games with fuzzy payoffs considered in [9], [3], or [10] into the class of cooperative games with I-fuzzy payoffs. The Shapley value based on presented class of I-fuzzy extension

of cooperative games fulfills basic properties put on Shapley value — symmetry, additivity, and efficiency — in I-fuzzy sense, which is dependent on definition of I-fuzzy number ([1], [4], and [7]), as well as on the type of I-fuzzy extension of the crisp game. The Shapley value of this class of games fulfills a fairnes requirement represented by players' concessions.

Acknowledgements

This research was supported by the GAČR project No. 14-02424S.

References

[1] Anzilli, L., Facchinetti, G., Mastroleo, G.: Evaluation and Ranking of Intuitionistic Fuzzy Quantities. In: *WILF 2013, LNAI 8256* (Masulli, F., Pasi, G., Yager R., eds.), Springer International Publishing, Switzerland, 2013, 139–149.

[2] Atanassov, K. T.: Intuitionistic Fuzzy Sets, *Fuzzy Sets and Systems* **20**(1) (1986), 87–96.

[3] Borkotokey, S., Mesiar, R.: The Shapley Value of Cooperative Games under Fuzzy Settings: A Survey. *International Journal of General Systems* **43**(1) (2014), 75–95.

[4] Burillo, P., Bustince, H., Mohedano, V.: Some Definitions of Intuitionistic Fuzzy Number. First properties. In: *Proceedings of the 1st Workshop on Fuzzy Based Expert Systems*, 1994, 53–55.

[5] Çoker, D.: An Introduction to Intuitionistic Fuzzy Topological Spaces. *Fuzzy Sets and Systems* **88**(1) (1997), 81–89.

[6] Dubois, D., Gottwald, S., Hajek, P., Kacprzyk, J., Prade, H.: Terminological Difficulties in Fuzzy Set Theory – The Case of "Intuitionistic Fuzzy Sets", *Fuzzy Sets and Systems* **156**(3) (2005), 485–491.

[7] Grzegorzewski, P.: Distances and orderings in a family of intuitionistic fuzzy numbers. In: *Proceedings of EUSFLAT Conference*, (Wagenknecht, M., Hampel R., eds.), 2003, 223–227.

[8] Hart, S., Mas-Colell A.: Potential Value and Consistency, *Econometrica* **57**(3), (1989), 589–614.

[9] Mareš, M.: *Fuzzy Cooperative Games*. Physica-Verlag, Heidelberg, 2001.

[10] Mareš, M., Vlach, M.: Fuzzy Classes of Cooperative Games with Transferable Utility. *Scientiae Mathematicae Japonicae* **60**(2) (2014), pp. 269–278.

[11] Marinov, E.: On Extension Principle for Intuitionistic Fuzzy Sets. *Notes on Intuitionistic Fuzzy Sets* **20**(3) (2014), pp. 34–41.

[12] Peleg, B.: Voting by Count and Account. In: *Rational Interaction: Essays in Honor of John C. Harsanyi*, (R. Selten, ed.), Springer-Verlag, New York, 1992, pp. 41-51

[13] Perez-Castrillo, D., Wettstein, D. : An Ordinal Shapley Value for Economic Environments (revised version), *Universidad Autonoma de Barcelona* Discussion Paper 04-12, 2004.

[14] Perez-Castrillo, D., Wettstein, D. : An Ordinal Shapley Value for Economic Environments, *Journal of Economic Theory* **127**(1) (2006), 296–308.

[15] Shapley, L. S.: A value for n-Person Games. In: *Contribution to the Theory of Games* (Kuhn, H. W., and Tucker A. W., eds.), **2**, Princeton University Press, Princeton, New Jersey, 1953, 307–317.

[16] Stoeva, M.: Intuitionistic Fuzzy Extension Principle. *Notes on Intuitionistic Fuzzy Sets*, **5**(3) (1999), 29–30

The omega function for continuous distribution

Ewa Michalska[1], Donata Kopańska-Bródka[2]

Abstract. The omega function (proposed in 2002 by Keating and Shadwick) is used as a performance measure of an investment asset or portfolio. According to the authors, omega function was developed as a measure which takes into account all the information about the return distribution. In this article the most important properties of omega function are presented as well as the derivation of its form for some continuous distribution.

Keywords: performance measure, omega function, continuous distribution.

JEL Classification: D81
AMS Classification: 91B06

1 Introduction

The inspiration for Keating and Shadwick (authors of the omega function) was the observation that the moments of the distribution (mean and variance) do not fully describe the distribution of random rate of return. Seeking for alternative methods of investments evaluation of that manage to capture the asymmetry and fat tails of distribution, resulted in the measure that takes into account higher moments of the distribution rate of return [6]. The form of the omega function also gives the investor the opportunity to take into account the preferences expressed by the threshold value (e.g. acceptable rate of return) in respect of which the investment results are judged as desired (values higher than the threshold) and undesired (values lower than the threshold). Omega function for a fixed value of argument (threshold level) is the performance measure of investment in particular investment fund. It is the ratio of the expectations of gains above the threshold level (e.g. acceptable rate of return) to the expectations of the losses below the threshold.

In recent years we can observe enormous interest in using alternative ways to evaluate investments taking into account the distribution of random rate of return [10]. There are many publications devoted to this subject. Among them there are articles whose authors explore properties and look for new areas of application of the omega function proposed in the literature [1], [9], [7]. Several authors present modified measures or create new measures showing relationships with existing ones [2], [4], [5], [3], [8]. Regardless of the direction of research related to the omega function, it must be emphasized that knowledge of analytical form of omega function (for a given distribution of random variable) and its properties increases the usefulness of omega function for making investment decisions. In this article we present the main properties of the omega function as well as the derivation of its form for some continuous distribution (uniform distribution). We also present an elementary example which illustrates the application of the omega function for capital allocation in risky and risk-free assets.

2 The definition and some properties of the omega function

The search for the performance measure which includes all information about distribution of the random variable led to the omega function. In this section the definition and the most important properties of the omega function are presented.

Definition 1. *If F is a cumulative distribution with non-trivial domain (a,b) and if F has finite mean μ then the omega function is defined as*

$$\Omega(L) = \frac{I_2(L)}{I_1(L)} = \frac{\int_L^b (1 - F(t))dt}{\int_a^L F(t)dt} \tag{1}$$

The parameter L is a threshold (benchmark) selected by the investor and $L \in (a,b)$. Figure 1 presents the geometric interpretation of the omega function as a ratio of area $I_2(L)$ to area $I_1(L)$. The omega function is defined as the ratio of expected gains to expected losses (gains and losses are relative to a threshold L).

[1] University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, ewa.michalska@ue.katowice.pl.
[2] University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, broda@ue.katowice.pl.

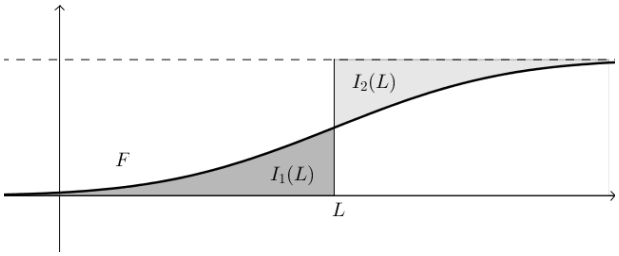


Figure 1 The geometric interpretation of the omega function

Suppose that the integrals I_1 and I_2 exist for all $L \in (a,b)$. The omega function exhibits following properties:

Property 1. Ω is continuous on (a,b)

Property 2. Ω is decreasing on (a,b)

Property 3. $\lim_{L\to a^+} \Omega(\mathrm{L}) = +\infty$ and $\lim_{L\to b^-} \Omega(\mathrm{L}) = 0$

Property 4. $\mathrm{Im}(\Omega) = (0,+\infty)$

Property 5. $I_2(L) - I_1(L) = \mu - L$ for $L \in (a,b)$

Property 6. $\Omega(\mu) = 1$

Property 7. $\Omega(L) = 1 + \dfrac{\mu - L}{I_1(L)}$ for $L \in (a,b)$

Property 8. Ω is differentiable and $\dfrac{d\Omega(L)}{dL} = \dfrac{F(L)[I_1(L) - I_2(L)]}{I_1^2(L)} - \dfrac{1}{I_1(L)}$

Property 9. *F* is symmetric about μ if and only if $\Omega(\mu - L) = \Omega(\mu + L)^{-1}$ for all $L \in (a,b)$

Property 10. $\Omega_X = \Omega_Y$ if and only if $F_X = F_Y$ (F_X and F_Y are cumulative distributions of continuous random variable X , Y).

The proofs of properties 1-7 are given in [9] and 8-10 in [1]. In this article, we will prove that the omega function is a quasi-concave function. It is important for optimization algorithm which is used in portfolio selection when short selling is not allowed. At first we present definition for a quasi-concave function.

Definition 2. *A function f: S→ R defined on an interval S is a quasi-concave function if it satisfies*

$$f(\lambda x + (1-\lambda)y) \geq \min(f(x), f(y)) \quad \forall x, y \in S, \forall \lambda \in [0,1] \tag{2}$$

In other words, for any two points in S, the value of the quasi-concave function at the convex combination of selected points is always not lower than the minimum of the function at these points.

Property 11. *Omega function is a quasi-concave function..*

Proof. As we know, the omega function is a decreasing function (property 2), then for two points $x, y \in S$ and $x \leq \lambda x + (1-\lambda)y \leq y$, $\lambda \in [0,1]$ we get the following result:

$$\Omega(y) \leq \Omega(\lambda x + (1-\lambda)y) \leq \Omega(x) \tag{3}$$

and hence

$$\Omega(\lambda x + (1-\lambda)y) \geq \Omega(y) = \min\{\Omega(x), \Omega(y)\} \tag{4}$$

Whereas the omega function is a quasi-concave function, we can find the global maximum if the domain is restricted to the non-negative region.

3 The omega function for a uniform distribution

For a random variable X with the uniform distribution $U(a,b)$ the probability density function is given as

$$f_X(x)=\begin{cases}\dfrac{1}{b-a}, & \text{for } x\in[a,b]\\ 0, & \text{for } x\notin[a,b]\end{cases} \tag{5}$$

and the cumulative distribution function is

$$F_X(x)=\begin{cases}0, & \text{for } x<a\\ \dfrac{x-a}{b-a}, & \text{for } a\le x<b\\ 1, & \text{for } x\ge b\end{cases} \tag{6}$$

The omega function of a random uniformly distributed variable X can be taken in the following analytical form:

$$\Omega_X(L)=\left(\frac{L-b}{a-L}\right)^2, \quad \text{for } L\in(a,b) \tag{7}$$

Example 1. Suppose that X is a random variable with the uniform distribution $U(0,4)$, the omega function is

$$\Omega_X(L)=\left(\frac{L-4}{L}\right)^2=\frac{L^2-8L+16}{L^2}, \quad \text{for } L\in(0,4) \tag{8}$$

The figure 2 shows the graph of the omega function, if the reference point L equals the expected value $E(X)$ of random variable X, then the omega function equals one ($\Omega(E(X))=1$).

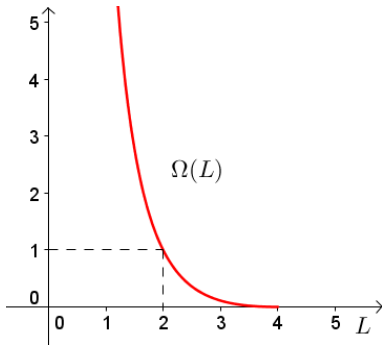


Figure 2 The omega function for the uniform distribution U(0,4)

Example 2. Let us consider two investments with the random returns R1 and R2, uniformly distributed $U(a_1,b_1)$ and $U(a_2,b_2)$, respectively. Using the omega measure the investment R1 is preferred to R2 if $a_1=a_2$ and $b_2<b_1$ for any threshold level $L\in(a_1,b_2)$ and the investment R2 is preferred to R1 if $b_1=b_2$ and $a_1<a_2$ for any threshold level $L\in(a_2,b_1)$. These preferences are coherent with the mean-variance criterion.

For a random variable $Y=cX+d$ (where X is a random variable with a uniform distribution $U(a,b)$), we derived the following formula for the omega function for Y

$$\Omega_Y(L)=\Omega_X(L)\cdot\left(\frac{c+\dfrac{d-L(1-c)}{b-L}}{c+\dfrac{L(1-c)-d}{L-a}}\right)^2 \quad \text{for } L\in(ca+d,cb+d) \tag{9}$$

For this special case of continuous distribution, the omega function (9) for the linear function of X depends on the omega function for X (which is constant for a given threshold level L) and the factor expressed by parameters c, d.

4 Capital allocation between a risky and risk-free assets

Most investment professionals consider asset allocation as the most important part of investment strategy. Here we assume that the investor must decide how to invest all of his wealth and has only two options: a risk-free asset (e.g. treasury bills) and a risky asset (e.g. stocks, portfolio of stocks or a mutual fund). Since all of his wealth must be invested then we denote the fraction of his wealth which he invests in the risky asset as w and the fraction of his wealth which he invests in risk-free asset as $1-w$ (we assume no short sale so $w \geq 0$). In this way, we get the portfolio for which the base of rate of return is the risk-free rate. Moreover, the portfolio is expected to earn a risk premium that depends on the risk premium of the risky asset. Denote the risky rate of return by R and the rate of return of risk-free asset by R_f. The rate of return on the complete portfolio R_P is as

$$R_P = wR + (1-w)R_f \tag{10}$$

Assuming that R is a random variable with the uniform distribution $U(a,b)$, the omega function for random variable R_P (based on the relationship (9)) has the following form

$$\Omega_{R_P}(L) = \Omega_R(L) \cdot \left(\frac{w + \dfrac{(R_f - L)(1-w)}{b-L}}{w + \dfrac{(L-R_f)(1-w)}{L-a}} \right)^2 \quad \text{for } L \in (wa + (1-w)R_f, wb + (1-w)R_f) \tag{11}$$

For a given threshold level L (level of return), using the very simple rule of "prefer more to less", we should always prefer a portfolio with a higher value of omega in comparison with one with a lower omega. The portfolio with the higher omega has greater probability of delivering returns which match or exceed the L threshold. We can change the portfolio performance by changing the share of the risky asset in the portfolio. The uniform distribution of the random rate of return can be assumed when the data are gathered from the short period of time.

For any threshold level L there exists a value w for which the function (11) has a minimum. Thus we can determine the worst portfolio for the investor. Next we will consider the least desirable portfolios. In Example 3 we will create the portfolio consisting of a risky asset represented by WIG20 (index of 20 Polish stocks listed on the Warsaw Stock Exchange) and a risk-free asset represented by bonds with fixed annual return.

Example 3. We assume that a random variable $R = R_{WIG20}$ (denoting rate of return of WIG20 index) has a uniform distribution $U(a,b)$. For WIG20 index we consider data from period 10.04.2015-30.04.2015 (15 rates of return) and three different values of annual return of risk-free asset: 5%, 10%, 15%. The rate of return of the complete portfolio can be written as

$$R_P = wR_{WIG20} + (1-w)R_f \tag{12}$$

The value of the omega function (11) for the portfolio (12) depends on the value of the omega function for a random variable R_{WIG20} (positive value for a given level L) and the factor expressed by the fraction of risky asset in portfolio. Portfolio with the lowest value of the omega function is the least desirable.

Table 1 presents the fractions of risky asset in portfolio for various threshold level L. With the increase of the L level the share of risky asset in the portfolio also increases. As we can see, in the worst portfolio the share of risky asset does not increase with the increase of risk-free rate of return for a given threshold level.

L	R_f (5%)	R_f (10%)	R_f (15%)
[%]	*w*	*w*	*w*
100.1	0.05	0.04	0.04
100.2	0.11	0.10	0.10
100.3	0.17	0.16	0.16
100.4	0.23	0.23	0.22
100.5	0.29	0.29	0.28
100.6	0.35	0.35	0.34
100.7	0.41	0.41	0.40
100.8	0.47	0.47	0.46
100.9	0.53	0.53	0.52
101.0	0.59	0.59	0.59
101.1	0.65	0.65	0.65
101.2	0.71	0.71	0.71
101.3	0.77	0.77	0.77
101.4	0.83	0.83	0.83
101.5	0.89	0.89	0.89
101.6	0.95	0.95	0.95
101.7	1	1	1

Table 1 The fraction of risky asset in portfolio for various threshold level L

The dependence between the share of risky asset in the worst portfolio and the changing threshold level is presented in Figure 3.

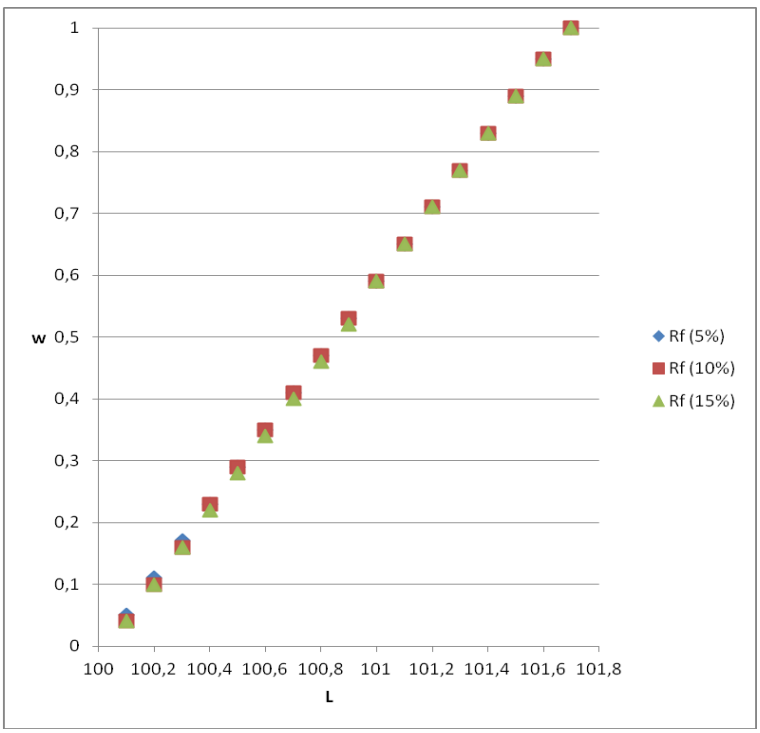


Figure 3 Dependence between the threshold level L and the fraction for risky asset in portfolio

It is easy to observe almost steady increase of fraction for risky asset in our portfolio.

5 Conclusion

In recent years there are many publications devoted to the evaluation of investments. Researchers are especially interested in the omega function which takes into account an investor's preferences as well as full knowledge of the distribution of random rate of return. In our article we presented main properties of the omega function and we also proved its quasi-convexity. We appointed the analytical form of the omega function for the case of uniform distribution. Then we used it in the valuation of performance of portfolio with one risky asset and one risk-free asset. In further research we are going to derive the omega function for portfolios with two or more risky assets assuming other continuous distributions.

References

[1] Cascon, A., Keating, C., Shadwick, W.: *The Omega Function*. The Finance Development Centre, 2003. https://faculty.fuqua.duke.edu.

[2] Farinelli, S., Tibilettti, L.: *Sharpe thinking with asymmetrical preferences*. Social Science Research, 2002 http://www.edge-fund.com/FaTi02.pdf

[3] Kaplan, P.: *A unified approach to risk-adjusted performance*, 2005 http://corporate.morningstar.com

[4] Kaplan, P., Knowels, J.: *Kappa: a generalized downside risk-adjusted performance measure*, 2004 http://corporate.morningstar.com

[5] Kazemi, H., Schneeweis, T., Gupta, R.: Omega as a performance measure. *Journal of Performance Measurement* **8** (2004), 16-25.

[6] Keating, C., Shadwick, W.: A universal performance measure. *Journal of Performance Measurement* **6** (2001), 59-84.

[7] Michalska, E.: Zastosowanie wskaźnika omega w podejmowaniu decyzji przy niepełnej informacji liniowej (Application of the omega ratio in decision making with partial information). In: *Metody i zastosowania badań operacyjnych*'13, Uniwersytet Łódzki (in print).

[8] Michalska, E., Dudzińska-Baryła, R.: Związek funkcji omega z dominacją stochastyczną. (Relationship between omega function and stochastic dominance). In: *Metody i zastosowania badań operacyjnych*'14, Uniwersytet Ekonomiczny w Katowicach (in print).

[9] Nguyen, V., N.: *Omega function: a theoretical introduction*. 2009 http://scholarspace.manoa.hawaii.edu/handle/10125/25918

[10] Xiong, J. X., Idzorek, T., M.: The impact of skewness and fat tails on the asset allocation decision. *Financial Analysts Journal* **67** (2011), 23-35.

Estimation of regression parameters using simulation methods

Michał Miłek[1]

Abstract. Statistical analysis in economics are often based on explaining the phenomena using appropriate econometric models. It is therefore necessary to estimate the unknown parameters of those models. In the case of the estimate parameters of time series models, we can use a large number of estimation methods, both parametric (if we know the class of distribution) and non-parametric. A special group are the solutions based on the median. The unquestionable advantage of the median methods is their resistance to the presence of the outliers, which can negatively affect the final form of the estimated model. In this paper a simulation method based on median was proposed to estimate the unknown parameter. Also different methods which can be used to estimate the parameters of the regression function were presented. In particular a modification of the Theil's method where the bootstrap and the jackknife simulations were used to estimate the appropriate median. The simulation analysis was conducted based on the data generated according to four scenarios, next, the results ware compared.

Keywords: Median, time series, bootstrap, jackknife, simulation.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In the time series analysis it is very important to form a suitable choice of the model. This is possible only if the parameters of the model are estimated with reasonable accuracy. There are numerous publications describing different methods of parameter estimation for the regression function, both parametric (when we know the distribution of the class which the distribution of the test belongs to) and non-parametric, which special type are the methods based on the median [3]. The unquestionable advantage of the median methods is their resistance to the presence of the outliers, which can greatly influence the final form of the estimated model. The results received by comparing the efficiency of the parametric and the non-parametric estimators [3] confirm that fact. In addition, they indicate that, in case of the disorders of the outliers, none of the estimators can be used interchangeably, neither the classical estimators can be regarded as optimal [3]. However, the application of the median methods to estimate the model parameters entails some risk associated with the presence of the median bias [8]. In this paper a simulation method based on median was proposed to estimate the unknown parameter. Moreover this article presents a comparison of different methods of estimating parameters of the linear time series model - in particular, the modification of the Theil's method which uses the bootstrap [4] and the Jackknife [11] simulation to estimate the appropriate median.

2 Methods of the estimation of regression parameters

In order to calculate the parameters of the linear regression equations, we can use numerous methods based on the median. They can also be used to estimate the parameters of the linear time series model. They consist in dividing the sample into several areas (usually 2 or 3), with respect to one of the variables. The median is also used when considering the residuals, on the basis of which the parameters are subsequently calculated [3]. The most popular median method is the Wald's method, which divides the variation area into two, and the modification of the method proposed by Brown and Mood [2]. Theil's method [13] is based on the median of indicators, which are tangents of the angle for two-element pairs (x_i, y_i). Yet another method is the so-called scavenging method proposed by Andrews [1]. It is considered from different perspectives and has many modifications: the classic approach, method of the repeated median [12], the Rousseuw's method of least median of squares [9], the method of Mosteller and Tukey [7]. In the following chapter of this paper, we will present Theil's method as example and proposal of method in order to estimate the parameter of the linear time series model. In addition to the classical methods of statistical analysis, the simulation methods are becoming more and more popular. The range of their application is very wide. They can be used to estimate the variances of estimators or verify the hypotheses in a situation where the distribution of the test statistics is not known. Calculations can also be per-

[1] University of Economics in Katowice, Department of Statistics, Katowice, Bogucicka 4, michal.milek@ue.katowice.pl.

formed, due to the complex nature of economic phenomena, are not feasible using analytical methods. For this purpose, the most commonly used methods are the bootstrap and the jackknife, which have a lot of advantages. Among other things, the sample does not have to come from a normally distributed population, the jackknife method allows to reduce the bias of the estimator, those methods can be used in small samples, which do not allow the use of limit theorems. In the literature, one can find a comprehensive description of different modifications of those methods, to extend the scope of their applications, among other things, allowing their use in time series analysis [6].

In further analysis, we will employ model in the following form:

$$Y_t = \alpha + \beta X_t + \varepsilon_t \tag{1}$$

where:
$t = 1, 2, \ldots, n$,
α, β – unknown parameters,
ε – random component with a constant variance and expected value of 0.

2.1 Theil's method

Pairs of values:

$$(x_1; y_1), (x_2; y_2), \ldots, (x_n; y_n) \tag{2}$$

are n-element sample P, ordered with respect to the variable X:

$$P = \{(x_t; y_t) : x_1 < x_2 < \ldots < x_n\} \tag{3}$$

The estimator of the β parameter would be estimated on the basis of the sample of $N = \binom{n}{2}$ coefficients, calculated as follows:

$$S_{ij} = \frac{y_j - y_i}{x_j - x_i} \quad \text{for } i<j \tag{4}$$

In the further step, the median of the coefficients should be calculated, which will be the evaluation of the unknown parameter β.

$$\hat{\beta} = med\{S_{ij}\} \tag{5}$$

It can be said that the β-parameter will be estimated on the basis of the S_{ij} coefficients, representing tangents angle of the slope related to the x-axis. The method of operation is illustrated in Figure 1.

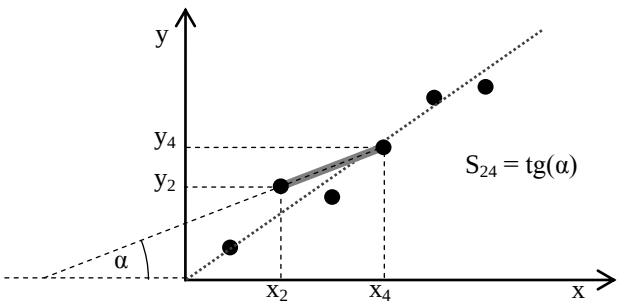


Figure 1. Method of the calculation of the S_{ij} coefficients in Theil's method

This method, after minor modifications, can also be used if the inequalities in the expression (3) are not sharp, then the value of the variable X can be repeated. The solution to this problem is proposed by Sen [10].

2.2 Bootstrap

One of the most popular methods for simulation is the bootstrap method. Bootstrap is a variation of the Monte Carlo method. It is assumed that the population is examined with respect to the random variable *X* [8]. A simple sample comes from the population:

$$P_{boot} = \{X_t : X_1, X_2, ..., X_n\} \tag{6}$$

The implementation of that sample can be marked with symbols: $x_1, x_2, ..., x_n$.

In the case of the parameter estimation using the bootstrap method, the procedure is as follows - from the input data series *B*-fold *n*-element bootstrap sample with replacement is drawn.

$$P^*_{boot,b} = \{X^*_{t,b} : X^*_{1,b}, X^*_{2,b}, ..., X^*_{n,b}\}, \quad b = 1, 2, ..., B \tag{7}$$

On the basis of each sample. the evaluation of the parameter estimated is determined by the use of estimator $T^*_{n,b}$:

$$T^*_{n,b} = T_n\left(X^*_{1,b}, X^*_{2,b}, ..., X^*_{n,b}\right) \tag{8}$$

From those values, the average is calculated, resulting in bootstrap estimate being obtained:

$$\overline{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_{n-1,i} \tag{9}$$

The variance estimator is given by:

$$v_{boot}(T_n) = \frac{n-1}{n}\sum_{i=1}^{n}\left(T_{n-1,i} - \overline{T}_n\right)^2 \tag{10}$$

2.3 Jackknife

This method, similar to the bootstrap, is widely used because of the property of enabling the reduction of the bias of the estimator. As previously, a population with respect to random variable *X* is being considered [8]. A simple sample derived from this population will be marked as P_{jack}.

$$P_{jack} = \{X_t : X_1, X_2, ..., X_n\} \tag{11}$$

The implementation of this test will be marked as follows: $x_1, x_2, ..., x_n$. The estimator Θ of the parameter of the *X*-distribution, will be denoted by a symbol T_n. The idea behind the jackknife method is to estimate the value of the estimator on the basis of a subsample formed by omitting one of the observations, the estimator can be determined as following:

$$T_{n-1} = T_{n-1}\left(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n\right), \quad i = 1, 2, ..., n \tag{12}$$

Jackknife estimator of the unknown parameter will be expressed by the following formula [11]:

$$T_{JACK} = T_n - b_{JACK} = nT_n - (n-1)\overline{T}_n \tag{13}$$

where:

$\overline{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_{n-1,i}$,

$b_{JACK} = (n-1)\left(\overline{T}_n - T_n\right)$.

The variance of the estimator is given by:

$$v_{JACK}(T_n) = \frac{n-1}{n}\sum_{i=1}^{n}\left(T_{n-1,i} - \bar{T}_n\right)^2 \tag{14}$$

2.4 Proposed method

In this paper a method to calculate the unknown β parameter was proposed. It base on simulation method what have unquestionable advantages that allow us to omit crucial assumptions about analyzed phenomena. We consider pairs of values as in formula 2. The procedure of proposed method was as follows:

1. Estimate the parameters of classical linear trend.
2. Calculate the residuals from the model.
3. Generate *B*-time series by changing the order of residuals.
4. For each time series the β^* value was calculated as follows:

$$\beta_b^* = \frac{y_B - y_A}{x_B - x_A} \tag{15}$$

Where *A* and *B* are random numbers of observations and *B*>*A*.

5. Estimator of β parameter is calculated as follows:

$$\hat{\beta} = med\left\{\beta_b^*\right\} \tag{16}$$

Beside of the advantages connected with using median to estimate the unknown parameter, this method allows also to reduce the number of computation needed to estimate the parameter.

3 Simulation analysis

In the analysis, a comparison of the results is obtained using different methods of determining the median of the formula (5). The classical approach proposed by Theil (17) was considered, as well as the median-determined (T_1) by formula (18) [8], and using the method of Bootstrap (T_2), Jackknife (T_3) and proposed solution (T_4).

$$med\{S\} = \begin{cases} S_{k+1} & n = 2k+1 \\ \dfrac{S_k + S_{k+1}}{2} & n = 2k \end{cases} \tag{17}$$

$$med\{S\} = \{S_{k+1} \qquad k = \lceil n/2 \rceil \tag{18}$$

$$P_{boot} = \{X_t : X_1, X_2, \ldots, X_n\} \tag{19}$$

The analysis was conducted on the basis of the simulated data. All considered time series were generated according to model (20):

$$Y_t = \alpha + \beta X_t + \varepsilon_t \tag{20}$$

Where α=0 and β=1. Random component ε has different distribution according to four scenarios. First was generated from a normal distribution $N(0, 1)$. The second was introduced disorder - 10% came from the observation distribution $N(4, 1)$, while in the third rate was 20%. The last one assumes that random component was generated from log-normal distribution. Sample histograms of these populations are shown in Figure 2.

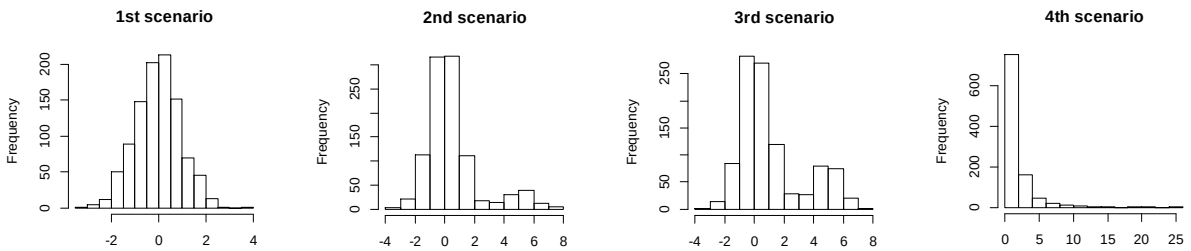


Figure 2. Histograms of the random component ε according to four scenarios

The simulation was repeated 1,000 times, each time the value of the unknown parameter, estimator bias, its standard deviation and Root Mean Squared Error (RMSE) was calculated. In the case of the bootstrap method and proposed solution, in each case B=1000 bootstrap samples were determined. Figure 3 show the results of the simulation in relation to the size of the sample for the second scenario. Shown is the Root Mean Squared Error of the β-parameter estimator. The results for all scenarios are included in Table 1.

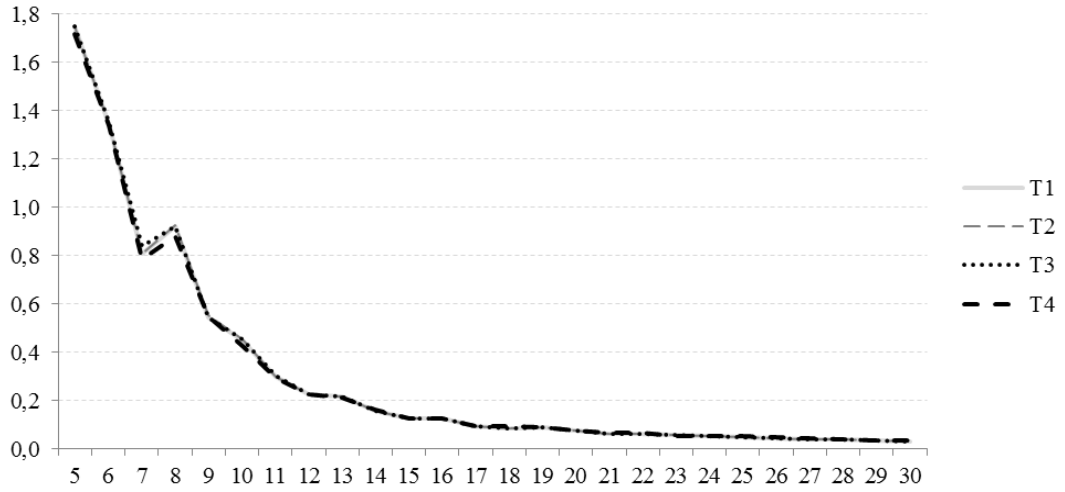


Figure 3. The values of Root Mean Squared Error of the β-parameter estimator for 2nd scenario

	I scenario				II scenario				III scenario				IV scenario			
n	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4	T_1	T_2	T_3	T_4
3	0,73	0,73	0,73	0,73	26,5	26,5	26,5	26,5	35,6	35,6	35,6	35,6	69,5	69,5	69,5	69,5
4	0,50	0,50	0,50	0,51	11,2	11,2	11,2	13,0	8,54	8,54	8,54	9,14	8,66	8,66	8,66	10,4
5	0,34	0,34	0,34	0,35	1,75	1,75	1,75	1,72	3,22	3,79	3,22	3,05	1,97	1,97	1,97	2,04
6	0,26	0,26	0,26	0,26	1,34	1,34	1,36	1,34	1,22	1,22	1,17	1,17	1,83	1,83	1,73	1,83
7	0,19	0,19	0,18	0,19	0,81	0,81	0,84	0,78	1,23	1,23	1,25	1,23	0,84	0,84	0,86	0,84
8	0,16	0,16	0,16	0,16	0,92	0,92	0,92	0,88	0,79	0,77	0,79	0,71	0,56	0,56	0,56	0,56
9	0,14	0,14	0,14	0,14	0,54	0,55	0,54	0,55	0,45	0,45	0,45	0,47	0,66	0,66	0,66	0,68
10	0,16	0,16	0,16	0,16	0,45	0,45	0,45	0,43	0,34	0,34	0,35	0,34	0,41	0,41	0,42	0,40
11	0,12	0,12	0,12	0,12	0,30	0,30	0,30	0,29	0,28	0,28	0,28	0,28	0,28	0,28	0,28	0,28
12	0,11	0,11	0,11	0,11	0,23	0,23	0,23	0,23	0,25	0,25	0,25	0,24	0,22	0,22	0,22	0,22
13	0,09	0,09	0,09	0,10	0,22	0,22	0,22	0,21	0,19	0,19	0,19	0,19	0,20	0,20	0,20	0,21
14	0,08	0,08	0,08	0,08	0,16	0,16	0,16	0,16	0,20	0,20	0,20	0,20	0,16	0,16	0,16	0,16
15	0,08	0,07	0,07	0,08	0,12	0,12	0,12	0,13	0,13	0,13	0,12	0,13	0,14	0,14	0,14	0,14

Table 1. The values of Root Mean Squared Error of the β-parameter estimator

Analyzing the graph it is visible that all the estimators quite good evaluate the average value of the estimated parameter. Theil, Bootstrap and Jackknife estimators behave in almost the same way and give similar results. Proposed solution gives some advantages when the random component ε has different distribution including some disturbances. For small sample sizes 5-8 elements Root Mean Squared Error (RMSE) is lower them in other estimators. Furthermore there are some benefits connected with computation time in proposed solution.

4 Conclusions

A suitable method for estimating the parameters of the econometric model can be crucial in the analysis of the economic phenomena. This article proposes a method that allow us to estimate of the slope of the linear time series model. The proposed method was compared with the results obtained in determining the parameter β in the classical manner proposed by Theil and its modification proposed by author, based on Bootstrap and Jackknife method. The results indicate that this modification, in specific cases, provides better results than the original method, in particular to reduce bias.

References

[1] Andrews D. F. *A Robust Method for Multiple Linear Regression*, Technometrics 16, 1974, 523-531.

[2] Brown B. M., Mood A. M., *On Median for Linear Hypotheses,* Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951, 159-166.

[3] Domański C., Pruska K., Wagner W., *Wnioskowanie statystyczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1998, 99-108, 201-215.

[4] Efron B., *Bootstrap methods: Another look at the jackknife*, Annals Statistics 7, 1979, 1–26

[5] Efron B., Tibshirani R.: *An Introduction to the Bootstrap*, Science Business Media Inc, 1993.

[6] Kozarski R., *Metody bootstrapowe w analizie szeregów czasowych*, Roczniki Kolegium Analiz Ekonomicznych – Szkoła Główna Handlowa, 2004.

[7] Mosteller F. C., Tukey J., Data Analysis and Regression, Adison-Wesley Reading, Massachusetts, 1997.

[8] Pruska K., *Estimation of bias and variance of sample median by Jacknife and Bootstrap Methods*, Acta Universitatis Lodziensis, Folia Oeconomica 206, 2007, 67-78.

[9] Rousseuw P. J., *Least Median of Squares Regression*, Jurnal of the American Statistical Analysis 79, 1984, 871-880.

[10] Sen P. K. Estimates of the Regression Coefficient Based on Kendall's tau, journal of the American Statistical Association 63, 1968, 1379-1389.

[11] Shao J., Tu D., *The Jackknife and Bootstrap,* Springer-Verlag, 1996, New York.

[12] Siegel A. F., *Robust Regression sing Repeated Medians*, Biometrica 69, 1982, 242-244.

[13] Theil H., *A Rank-invariant Method of Linear and Polynomial Regression Analysis, I, II, III,* Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, 1950, 386-392, 521-525, 1937-1412.

Computable Equilibrium Model Parameters Calibration vs. Estimation

Veronika Miťková[1]

Abstract. The paper deals with two approaches to obtaining the parameters of the computable general equilibrium models. We concern to the parameters of the production function in the model of closed economy for the Slovak republic. The differences among the calibration and econometric estimation results are shown. The dataset consists of quarterly data for Slovakia from 1995q1 to 2014q4, the sector focused is the public sector - public administration and defense, compulsory social security, education, human health and social work activities.

Keywords: CGE model, calibration, econometric estimation, Slovakia

JEL Classification: C68
AMS Classification: 62P20

1 The econometric estimation versus the calibration procedure

This paper deals with the parameter values of the computable general equilibrium model. The model consists of two types of equations: identity relationships and behavioral relationships. Identity relationships, given by accounting identities are usually non-controversial. Behavioral relationships are unobservable and are given by model parameters. They may be either calibrated or estimated by econometric methods. Some authors state the third method – imposing the values by the model user [1].

The calibration uses the empirical evidence, some parameters are given arbitrarily and the rest of the parameters are set such that the model replicates the benchmark year dataset. The calibration procedure is mentioned in [7]. This approach is under the critics of i.e. Jorgensen, Lau, Diewert and Lawrence and McKitrick [3].

The econometric estimation is usually limited by non-observable data, short term time series and it also requires simple functional forms used in the model. The detailed work dealing with the estimation and calibration can be found in our previous research [6].

2 The underlying dataset and model structure

We deal with the dataset used either for the econometric estimation and calibration and either with the model structure – the simple computable general equilibrium model without the intermediate consumption relations in this chapter [4].

2.1 The model structure

The model was inspired by the work of Hans Löfgren [2], for the beginning of our research in the field of the CGE econometrics it was chosen the model with a simple structure. It is discussed in some previous works, i.e. Miťková [4]. The model consists of three equilibrium conditions and seven behavioral relationships. For the main goal of this paper we deal only with the production function, which has the following form

$$QA_{\alpha} = ad_{\alpha} \cdot \prod_{f \in F} QF_{f\alpha}^{a_{f\alpha}} \quad \alpha \in A \tag{1}$$

where:
- ad_{α} efficiency parameter in the production function for activity a,
- $\alpha_{f\alpha}$ share of value-added for factor f in activity a,
- QA_{α} level of activity a,
- $QF_{f\alpha}$ demand for factor f from activity a.

[1] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 821 05 Bratislava, Slovakia, veronika.mitkova@fses.uniba.sk

Notations in the model follow the IFPRI[2] notation: for activities $\alpha \in A = \{PS - A, R - A\}$, for factors $f \in F = \{lab, cap\}$. The *PS - A* means Public Sector activities, it includes sectors according to NACE Rev. 2 (Statistical Classification of Economic Activities in the European Community) O, P and Q: Public administration, defense, education, human health and social work activities. *R - A* means rest of the sectors activities in the national economy, *lab* means labor and *cap* means capital. In the model, there is a disaggregation of households to urban and rural.

2.2 The dataset

The dataset consists of quarterly data for Slovakia from 1995q1 to 2014q4. The data are from the Slovak Statistical Office and Eurostat databases. For the purpose of the model were used the following aggregates: P.1 Output, P.2 Intermediate consumption, P.3 Final consumption expenditure, P.51 Gross fixed capital formation, P.52 Changes in inventories, D.1 Compensation of employees, D.2-D.3 Taxes less subsidies on product and export, B.1g Gross domestic product, B.2g Gross operating surplus, B.3g Gross mixed income and B.8g Gross national saving. The data are in current prices, millions of euro, not seasonally adjusted.

For the econometric estimation all the 80 quarterly observations were used and for the calibration procedure the average value of each aggregate was put to the social accounting matrix (SAM). The SAM for the model has the following structure shown by Table 1 in which one can find the three methods of the gross domestic product value calculation.

	activities	**commodities**	**factors**	**households**
activities		P.1 – P.2 + (D.2-D.3)		
commodities				Expenditure approach P.4 + P.51 + P.52 + B.8g
factors	Production approach B.1g + (D.2-D.3)			
households			Income approach D.1 + B.2g + B.3g + (D.2-D.3)	

Table 1: The social accounting matrix structure

In the intersection of the activities row and commodities column is the value of the level of activity α: QA_{α}, later disaggregated to the public sector activities and the rest of the activities volume. The intersection factors and activities cell shows the demand for factor *f* from activity α*:* $QF_{f\alpha}$, later divided to the demand for labor in the public sector, demand for capital in the public sector and correspondingly for the rest of the sectors in the national economy. The commodities and households cell contains the consumption of commodity *c* by household *h*, the urban and rural households were modeled. The households and factors cell contains the income of household *h* from factor *f*, again disaggregated to the urban household demand for labor and so on.

3 The calibration

The calibration procedure uses the average data from the 80 quarterly observations and the disaggregated Social accounting matrix has the structure and values as shown in the Table 2.

averages in mil. €		**activities**		**commodities**		**factors**		**households**	
		PS	R	PS	R	lab	cap	urban	rural
act.	PS			8 010					
	R				4 138				
com.	PS							3 976	4 034
	R							2 072	2 066
fact.	lab	2 637	1 413						
	cap	5 374	2 724						
hhlds.	urban					2 024	4 024		
	rural					2 026	4 074		

Table 2: The social accounting matrix for average data 1995q1 – 2014q4

[2] International Food Policy Research Institute

For the disaggregation was used the procedure developed in Excel, see [6]. The aggregated average value of the GDP was 12 148 mil. €, the initial information is the share of production the PS sector: 8.3%, the share of intermediate consumption in the PS sector: 4,5%, the households income from the labor production factor 30%, the households income from the capital production factor 70% - the average values for the time series of 80 observations.

The parameters of all the values of the model were calibrated; here are introduced only the production functions parameters values. The values of the parameters ensure the replication of the benchmark dataset.

$ad_{PS} = 1.884$ $\quad ad_R = 1.900$

$\alpha_{lab,PS} = 0.329$ $\quad \alpha_{lab,R} = 0.342$

$\alpha_{cap,PS} = 0.671$ $\quad \alpha_{cap,R} = 0.658$

The values of the parameters ensure the replication of the benchmark dataset. The production function takes the following final form:

$$QA_{PS} = 1.884 \cdot QF_{lab,PS}^{0.329} \cdot QF_{cap,PS}^{0.671} \tag{2}$$

$$QA_R = 1.900 \cdot QF_{lab,R}^{0.342} \cdot QF_{cap,R}^{0.658} \tag{3}$$

The technological parameter for the public sector is 1.88 and for the rest of the other sectors 1.9. If the amount of the labor force increases by 1%, the activity level increases by 0.33% in the public sector and by 0.34% in the rest of the other sectors. For the production factor capital holds true that if it increases by 1%, the activity level increases by 0.67% in the public sector and by 0.66% in the rest of the other sectors.

4 The econometric estimation

The production function models for two modeled sectors are as follows:

$$QA_{PS,t} = ad_{PS} \cdot QF_{lab,PS,t}^{\alpha_{lab,PS}} \cdot QF_{cap,PS,t}^{\alpha_{cap,PS}} \cdot e^{u_t} \tag{4}$$

$$QA_{R,t} = ad_R \cdot QF_{lab,R,t}^{\alpha_{lab,R}} \cdot QF_{cap,R,t}^{\alpha_{cap,R}} \cdot e^{u_t} \tag{5}$$

the logarithmic form of the model will be used for the estimation with the ordinary least squares method.

$$\ln QA_{PS,t} = \ln ad_{PS} + \alpha_{lab,PS} \cdot \ln QF_{lab,PS,t} + \alpha_{cap,PS} \cdot QF_{cap,PS,t} + u_t \tag{6}$$

$$\ln QA_{R,t} = \ln ad_R + \alpha_{lab,R} \cdot \ln QF_{lab,R,t} + \alpha_{cap,R} \cdot QF_{cap,R,t} + u_t \tag{7}$$

The estimation of the production function with the ordinary least squares method gives the values of the parameters as shown in the Tables 3 and 4.

For the production function in the public sector:

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.557319	0.0172293	32.3472	<0.00001	***
l_QFlab_ps	0.679013	0.00560381	121.1699	<0.00001	***
l_QFcap_ps	0.323877	0.00454837	71.2071	<0.00001	***

Mean dependent var	7.305466	S.D. dependent var	0.452608
Sum squared resid	0.005480	S.E. of regression	0.008437
R-squared	0.999661	Adjusted R-squared	0.999653
F(2, 77)	113649.4	P-value(F)	2.5e-134
Log-likelihood	270.0286	Akaike criterion	-534.0571
Schwarz criterion	-526.9111	Hannan-Quinn	-531.1921
rho	-0.092062	Durbin-Watson	2.178626

Table 3: The initial estimation for the equation (4)

For the production function in the rest of the other sectors:

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.457148	0.0422347	10.8240	<0.00001	***
l_QFlab_r	0.331313	0.0178152	18.5972	<0.00001	***
l_QFcap_r	0.668441	0.0153966	43.4149	<0.00001	***

Mean dependent var	9.169055	S.D. dependent var	0.446486
Sum squared resid	0.024100	S.E. of regression	0.017691
R-squared	0.998470	Adjusted R-squared	0.998430
F(2, 77)	25120.44	P-value(F)	4.1e-109
Log-likelihood	210.7883	Akaike criterion	-415.5767
Schwarz criterion	-408.4306	Hannan-Quinn	-412.7116
rho	0.122586	Durbin-Watson	1.752874

Table 4: The initial estimation for the equation (5)

$\ln ad_{PS} = 0.5573 \Rightarrow ad_{PS} = 1.746$ $\quad$ $\ln ad_{R} = 0.4571 \Rightarrow ad_{R} = 1.5795$

$\alpha_{lab,PS} = 0.679$ $\quad$ $\alpha_{lab,R} = 0.3313$

$\alpha_{cap,PS} = 0.3239$ $\quad$ $\alpha_{cap,R} = 0.6684$

The final form of the model (4 and 5) is:

$$\overline{QA_{PS,t}} = 1.746 \cdot QF_{lab,PS,t}^{0.679} \cdot QF_{cap,PS,t}^{0,3239} \quad (6)^3$$

$$\overline{QA_{R,t}} = 1.5759 \cdot QF_{lab,R,t}^{0.3313} \cdot QF_{cap,R,t}^{0.6684} \quad (7)$$

The variability of the activity level is explained by the econometric models on 99.97% for the public sector and on 99.85% for the rest of the sectors. All the explanatory variables are highly statistically significant. The technological parameter for the public sector is 1.75 and for the rest of the other sectors 1.58. If the amount of the labor force increases by 1%, the activity level increases by 0.68% in the public sector and by 0.33% in the rest of the other sectors. For the production factor capital holds true that if it increases by 1%, the activity level increases by 0.32% in the public sector and by 0.67% in the rest of the other sectors. The public sector is more sensitive to the changes in the labor force, in contrary to the rest of the other sectors which are more sensitive to the capital changes.

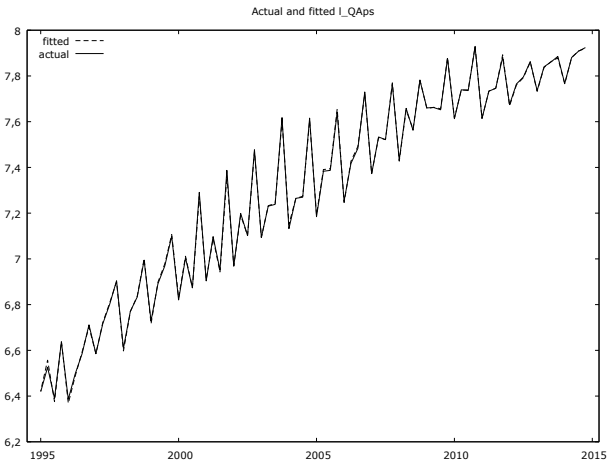


Figure 1: Actual vs. fitted for the public sector production function

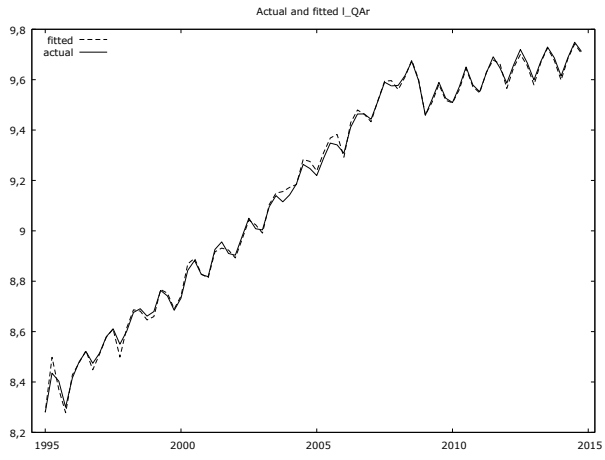


Figure 2: Actual vs. fitted for the rest of other sectors production function

Testing for autocorrelation with the Durbin-Watson test gives us DW_1=2.1786 for the first model and DW_2=1.7529 for the second model. The table values for 80 observations and two explanatory variables are d_L=1.5859 and d_U=1.6882. Since $d_L < DW_1 < d_U$ and $d_L < DW_2 < d_U$ the null hypotheses about no autocorrelation

[3] the lines above $QA_{PS,t}$ and $QA_{PS,t}$ are instead of hat

in both models (4) and (5) is accepted. The seasonal dummy variables added to both models are statistically not significant.

The residuals are normally distributed. The Jacques-Bera test gives $\chi^2_{2;0,05} = 14.348$ with p-value 0.0008 for the model (4) and $\chi^2_{2;0,05} = 10.195$ with p-value 0.0061 for the model (5).

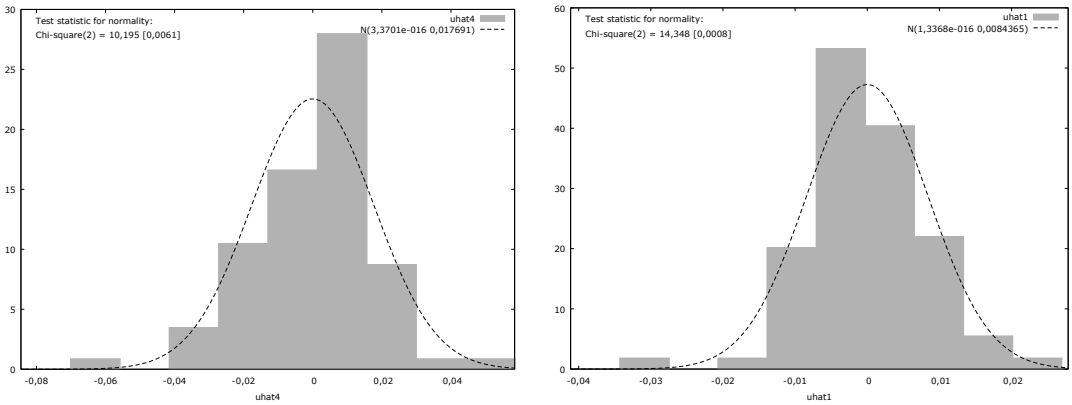


Figure 3: Normality distribution for the models (4 and 5)

The Variance Inflation Factor (VIF) value in the model (4) is 6.385 which us under 10, meaning no significant collinearity in explanatory variables. The VIF in the model (5) has value of 13.279 suspecting a multicollinearity problem. Since there are only two explanatory variables in each model we accept the multicollinearity in the case of the rest of other sectors.

5 Conclusion

The values of the parameters of the production function of the computable general equilibrium model obtained by the procedure of calibration and by the econometric approach are quite different for the public sector. It may be explained by the sensitivity of the CGE model to the initial set of labor – capital shares in the production function, which is the same for both sectors. In the econometric estimation these shares are different for the public sector and the rest of the sectors.

The further research will be focused on the econometric estimation of the commodity sector and households behavior. Later the model will be enlarged by the government and the rest of the world sector.

Acknowledgements

This paper research was supported by the Slovak Ministry of Education, Science and Sport, VEGA 1/0197/13: Ekonometrický odhad parametrov modelu všeobecnej ekonomickej rovnováhy s dezagregáciou verejného sektora.

References

[1] Lee, S., Pollitt, H., and Ueta, K.: A Model- Based Econometric Assessment of Japanese Environmental Tax Reform, *The Scientific World Journal* (2012), 1-9.
[2] Löfgren, H.: Exercises in General Equilibrium Modeling Using GAMS. *Microcomputers in Policy Research 4a*. International Food Policy Research Institute. 2003.
[3] McKitrick, R.R.: The econometric critique of computable general equilibrium modeling: the role of functional forms. *Economic Modelling* **15** (1998), 543-573.
[4] Miťková, V.: General equilibrium model of closed economy: case of Slovakia. Proceedings of the international workshop *Micro and Macro Models in Economics*. Bratislava, 2006, 102-110.
[5] Miťková, V.: Modely všeobecnej ekonomickej rovnováhy. Dissertation theses. University of Economics, Bratislava, 2007.
[6] Miťková, V.: The Computable General Equilibrium Model Parameters Econometrics Estimation versus their Calibration. In: *Proceedings of the Využitie kvantitatívnych metód vo vedecko-výskumnej činnosti a v praxi X*, Bratislava, 2013.
[7] Shoven, J. B. and Whalley, J.: Applying General Equilibrium. Cambridge: Cambridge University Press, 1992

Should Politeness Help in Optimization Metaheuristics?

Matej Mojzeš[1], Jaromír Kukal[2], Quang Van Tran[3]

Abstract. Discrete optimization metaheuristics typically start in random initial points. In our work we focus on random point generators: uniform random, greedy, and a polite one. Uniform randomness is the standard option typically used when no better concept is available. Greedy approach is based on component-wise sequential point construction. First coordinate is chosen randomly and the remaining ones are set to have the best values related to the previous coordinates. While this approach can help in some cases, it generates points with low diversity and can be trapped in local extremes. Novelty of this paper consists in the politeness of random point generation process. First coordinate is chosen randomly and the process is sequential as well. As opposed to greedy approach we are not concerned only with the best choice, but we select k-th best variant using Binomial distribution with constant mean value of the politeness parameter. Such approach may be considered a trade off between greedy and uniform random approach. This is demonstrated via numerical experiments on Traveling Salesman Problem using Random Shooting, Iterated Local Search and Fast Simulated Annealing.

Keywords: Discrete optimization, Metaheuristics, Randomness, Greediness, Politeness, Travelling Salesman Problem, Random Shooting, Iterated Local Search, Fast Simulated Annealing

JEL classification: C44
AMS classification: 90C15

1 Introduction

When considering general-purpose metaheuristics, there exist reputable and relatively sophisticated ones, e.g. Fast Simulated Annealing [1], and also some rather straightforward, but often very efficient ones, like Iterated Local Search [2] or even pure Random Search [3]. All of these algorithms have one feature in common - they always start their search by generating a random initial point.

We chose the Traveling Salesman Problem (TSP) as our objective function with the only reason being its high computational complexity and therefore suitability as a benchmark problem. For practical TSP applications, instead of general-purpose heuristics, one should choose e.g. the Concorde TSP Solver written by David Applegate et al. [4].

2 Random Point Generation

2.1 Integer Task Optimization Description

First of all, the following sections are relevant to metaheuristic integer task optimization, i.e. let $n \in \mathbb{N}$ be the problem dimension, $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^d$ be bounds satisfying $\mathbf{a} \leq \mathbf{b}$. The domain of optimization can be defined as $\mathbb{D} = \{\mathbf{x} \in \mathbb{Z}^d \,|\, \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$. Having an objective function g: $\mathbb{D} \to \mathbb{R}$ the optimization task consist in search for the optimal objective function value $g_{\text{opt}} = \min_{\mathbf{x} \in \mathbb{D}} \mathrm{g}(\mathbf{x})$ as well as optimal solution of the problem $\mathbf{x}_{\text{opt}} \in \arg\min_{\mathbf{x} \in \mathbb{D}} \mathrm{g}(\mathbf{x})$.

[1]FNSPE CTU Prague, Applied Software Engineering, Trojanova 13, Prague, Czech Republic, mojzemat@fjfi.cvut.cz
[2]FNSPE CTU Prague, Applied Software Engineering, Trojanova 13, Prague, Czech Republic, jaromir.kukal@fjfi.cvut.cz
[3]FNSPE CTU Prague, Applied Software Engineering, Trojanova 13, Prague, Czech Republic, tran@vse.cz

In the TSP case let n be the number of cities in a given TSP instance and $\mathbf{C} = \{c_{i,j}\} \in \mathbb{R}_+^{n \times n}$ a symmetric anti-reflexive matrix of travel costs between the cities. The goal then is to find a tour, i.e. sequence of visited cities, which includes all of the cities and has the lowest possible travel costs.

2.2 Data representation

A trivial tour visiting all cities can be represented as $\mathbf{y} \in \{1, \ldots, n\}^n$ satisfying $y_i = y_j \Leftrightarrow i = j$. Tour cost, i.e. our objective function, to minimize can be defined as

$$\mathrm{g}(\mathbf{y}) = C_{y_n, y_1} + \sum_{k=1}^{n-1} C_{y_k, y_{k+1}} \ . \tag{1}$$

For practical reasons we recommend following tour encoding: for given $y_1 \in \{1, \ldots, n\}$ use vector $\mathbf{x} \in \mathbb{H}$ where $\mathbb{H} = \{\mathbf{x} \in \mathbb{N}^{n-1} \,|\, \forall k : x_k \leq n - k\}$. Here, x_k represents city rank in the set of not visited cities which will be visited in the k-th step.

Therefore, we employ transformation $\mathbf{y} = \mathrm{h}([y_1, \mathbf{x}]) = \mathrm{h}(\boldsymbol{\xi})$ to reformulate TSP as minimization of

$$\mathrm{f}(\boldsymbol{\xi}) = \mathrm{f}([y_1, \mathbf{x}]) = \mathrm{g}(\mathrm{h}([y_1, \mathbf{x}])) \tag{2}$$

for $\boldsymbol{\xi} \in \mathbb{D} = \{1, \ldots, n\} \times \mathbb{H}$. Using this formulation we can introduce three approaches to random point generation.

2.3 Traditional approaches

The first two approaches are rather traditional ways of generating new random points. The first one, uniform approach, is the standard option typically used when we have no better idea. Another extreme is the second one, greedy approach, which can help generate much better point, but it does generate points with low diversity and often from local extremes.

Uniform approach

We can construct the tour by selecting cities with uniform probabilities, i.e. $\boldsymbol{\xi} = \mathrm{U}(\mathbb{D})$.

Greedy approach

We select only the first city with uniform probability, i.e. $y_1 \sim \mathrm{U}(\{1, \ldots, n\})$ and then, using recurrence, we choose next city using the greedy way:

$$y_{j+1} \in \underset{i \notin \{y_1, \ldots, y_j\}}{\arg\min} \ C_{y_j, i} \ . \tag{3}$$

The resulting $\boldsymbol{\xi}$ is easily obtained via decoding as $\boldsymbol{\xi} = \mathrm{h}^{-1}(\mathbf{y})$.

2.4 Polite approach

In this case we select the first city with uniform probability $y_1 \sim \mathrm{U}(\{1, \ldots, n\})$ as well. Let $\lambda > 0$ be politeness parameter and Bi be Binomial distribution. Then y_{j+1} is recursively generated as follows:

- we collect next tour step costs as $r_i = C_{y_j, i}$ for $i \notin \{y_1, \ldots, y_j\}$,
- we sort them in ascending order as $r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(n-j)}$,
- we generate random index

$$ind \sim \mathrm{Bi}(n - j - 1, \frac{\lambda}{n - j - 1}) \ , \tag{4}$$

or set directly $ind = 0$ if $\lambda > n - j - 1$,

- we set $y_{j+1} = r_{ind+1}$,

and again $\boldsymbol{\xi} = \mathrm{h}^{-1}(\mathbf{y})$. There are following particular cases of politeness:

- $\lambda = 0$ would be pure greedy approach,
- $0 < \lambda < 1$: only the very last step selection is greedy (which is enforced anyway),
- $1 \leq \lambda < n - 1$: would be greedy selection in last $\lfloor \lambda \rfloor + 1$ steps.

Hence, in the following analysis we will focus on the $0 < \lambda < 1$ interval.

3 Experimental Part

3.1 Metaheuristics

Random Search (RS)

The RS algorithm [3] is naive and very straightforward - it just randomly generates and evaluates points from $\mathbb{D}$ until a solution is found, or maximum number of evaluations $maxeval \in \mathbb{N}$ is exhausted. Nonetheless it may be useful for many ill-structured global optimization problems.

Iterated Local Search (ILS)

The ILS algorithm can be regarded as RS which after each random point $\mathbf{x} \in \mathbb{D}$ is generated performs local optimization on neighborhood $\mathrm{N}(\mathbf{x})$ of this point. The local optimization procedure is:

1. Evaluate $\mathrm{g}(\mathbf{z}_i)$ for all $\mathbf{z}_i \in \mathrm{N}(\mathbf{x})$
2. If $\min \mathrm{g}(\mathbf{z}_i) < \mathrm{g}(\mathbf{x})$ set $\mathbf{x} \leftarrow \arg\min \mathrm{g}(\mathbf{z}_i)$ and continue with the first step
3. Else quit local optimization procedure

Success of this technique lies in the reasonable selection of the neighborhood function. We have used

$$\mathrm{N}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{D} \mid ||\mathbf{y} - \mathbf{x}||_2 = 1\} \ . \tag{5}$$

Also in this case, search is being performed until *maxeval* is exhausted or optimal solution found.

Fast Simulated Annealing (FSA)

FSA begins with $k = 0$, temperature $T_k > 0$ and an initial random point. Then, using $\boldsymbol{\psi} \sim \mathrm{N}(\mathbf{0}, \mathbb{I})$ and $\sigma > 0$ we generate basic point in $\mathbb{Z}^n$

$$\mathbf{x}_{\mathrm{trial}} = [\mathbf{x}_k + \sigma \cdot \boldsymbol{\psi}] \ , \tag{6}$$

which is used to create feasible point inside $\mathbb{Z}^n$

$$\mathbf{y}_{\mathrm{k}} = \mathrm{perturb}(\mathbf{x}_{\mathrm{trial}}, \mathbb{T}) \tag{7}$$

and conditionally accepted

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{y}_k & \mathrm{f}(\mathbf{y}_k) < \mathrm{f}(\mathbf{x}_k) + T_k \tan(\frac{\pi \eta_k}{2}) \\ \mathbf{x}_k & \mathrm{f}(\mathbf{y}_k) \geq \mathrm{f}(\mathbf{x}_k) + T_k \tan(\frac{\pi \eta_k}{2}) \end{cases} \tag{8}$$

where $\eta_k \sim \mathrm{U}([-1, +1])$. We repeat this process until a solution from the goal set is found or the number of objective function evaluations $maxeval \in \mathbb{N}$ is exhausted. The cooling strategy used was

$$T_k = \frac{T_0}{1 + k/n_0} \ , \tag{9}$$

with $n_0 = maxeval/100$. The two main parameters influencing FSA performance: initial temperature and mutation scale were set to $T_0 = 0.1$ and $\sigma = 1$.

Hybrid Fast Simulated Annealing (HFSA)

Inspired by its performance on other tasks [5], we have implemented also a hybrid version of FSA using a local optimization technique as in the case of ILS in every iteration of the standard FSA algorithm. Thus we were able to set a higher initial temperature of $T_0 = 100$.

3.2 TSP Instances

We have chosen three Symmetric TSP instances from the well-known TSPLIB library. For each instance we have decided to set $maxeval = 1000\,n$, i.e. (only) a thousand objective function evaluations for each city. Instances with their number of cities n, minimal tour cost g_{opt}, probability of its successful random guess p_{guess} logarithm and maximum number of evaluations are listed in Table 1.

Table 1 Selected symmetric TSP instances

Instance	n	g_{opt}	$\log_{10} p_{\text{guess}}$	$maxeval$
FRI26	26	937	-25.2	26 000
BAYS29	29	2020	-29.5	29 000
GR48	48	5046	-59.4	48 000

3.3 Results

Every metaheuristic was run 100 times on each TSP instance using all three random point generation strategies. Results are included in Tables 2 - 4 for every TSP instance. For each metaheuristic and random point generating method the best objective function value found g_{best}, relative distance to optimum as $dist_{\text{opt}} = (g_{\text{best}} - g_{\text{opt}})/g_{\text{opt}}$. Also, if optimum was reached, we include basic metaheuristic performance measures [6]:

- REL as *reliability*, $REL = m/q$, where m is the number of successful runs (when optimal solution was found) and q total number of runs ($q = 100$),
- MNE as *mean number of evaluations*, $MNE = \frac{1}{m}\sum_{i=1}^{m} NE_i$ where NE_i is the number of objective function evaluations until optimal solution was found,
- SNE as *standard deviation of the number of evaluations*, $SNE = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(NE_i - MNE)^2}$

in the form of $MNE \pm SNE$ ($100\,REL\,\%$) or just MNE ($100\,REL\,\%$) for $m = 1$.

Parameter λ was empirically optimized over values $0.05, 0.10, 0.15, \ldots, 0.95$ and the best choices are documented in Tables 2 - 4 under the polite random point generators.

Even though the FRI26 instance with 26 cities could be regarded as simple task, for general-purpose metaheuristics it is a very challenging problem. Only the best metaheuristics turned out to find optimal solution. Comparing distinct algorithms, the ILS was the most successful and quick of the tested. In this regard, RS could make the second best use of polite random point generator (proving its usefulness), while FSA and its hybrid version were more consistent than RS, although less successful in general. However, the most important outcomes come from the comparison of different random point generators over all metaheuristics. Polite generator is undoubtedly the best choice, closely followed by the greedy one and finally the uniform approach is of very little use.

As far as BAYS29 instance is considered, results are very similar. Although, it has to be said that we are able to compare random point generator performance only based on the best objective function value found, apart from one exception in the case of RS and polite generator.

In the case of GR48 instance, every heuristic failed to reach optimum solution. Nevertheless, there is noticeable improvement achieved by polite random generator, again.

Table 2 Results for FRI26 instance.

Metaheuristic	Generator	g_{best}	$dist_{\text{opt}} \cdot 100[\%]$	Performance
RS	Uniform	1751	86.9	
RS	Greedy	965	3.0	
RS	Polite, $\lambda = 0.15$	937	0.0	11325 ± 6951 (66%)
ILS	Uniform	1344	43.4	
ILS	Greedy	955	1.9	
ILS	Polite, $\lambda = 0.05$	937	0.0	9763 ± 7755 (71%)
FSA	Uniform	1509	61.0	
FSA	Greedy	987	5.3	
FSA	Polite, $\lambda = 0.05$	1085	15.8	
HFSA	Uniform	1190	27.0	
HFSA	Greedy	961	2.6	
HFSA	Polite, $\lambda = 0.10$	937	0.0	37537 ± 56496 (4%)

Table 3 Results for BAYS29 instance.

Metaheuristic	Generator	g_{best}	$dist_{\text{opt}} \cdot 100[\%]$	Performance
RS	Uniform	3750	85.6	
RS	Greedy	2134	5.6	
RS	Polite, $\lambda = 0.10$	2020	0.0	5933 (1%)
ILS	Uniform	2789	38.1	
ILS	Greedy	2091	3.5	
ILS	Polite, $\lambda = 0.10$	2028	0.4	
FSA	Uniform	3775	86.9	
FSA	Greedy	2288	13.3	
FSA	Polite, $\lambda = 0.10$	2146	6.2	
HFSA	Uniform	2934	45.2	
HFSA	Greedy	2118	4.9	
HFSA	Polite, $\lambda = 0.05$	2084	3.2	

Table 4 Results for GR48 instance.

Metaheuristic	Generator	g_{best}	$dist_{\text{opt}} \cdot 100[\%]$	Performance
RS	Uniform	14676	190.8	
RS	Greedy	5840	15.7	
RS	Polite, $\lambda = 0.15$	5303	5.1	
ILS	Uniform	9951	97.2	
ILS	Greedy	5818	15.3	
ILS	Polite, $\lambda = 0.15$	5367	6.4	
FSA	Uniform	14439	186.1	
FSA	Greedy	6014	19.2	
FSA	Polite, $\lambda = 0.10$	5847	15.9	
HFSA	Uniform	9043	79.2	
HFSA	Greedy	5851	16.0	
HFSA	Polite, $\lambda = 0.10$	5786	14.7	

4 Conclusions

As our experimental results show, no matter what general-purpose metaheuristic is chosen, there is substantial difference in performance based on what type of random point generator is used. Unsurprisingly, the default uniform random generator is of little use when it comes to larger than trivial instances of TSP. Greedy approach can help the metaheuristic to get very close to optimal objective function value, however, the solution is very probably trapped in a local minimum and different generated points tend to be the same.

In this context, the polite random point generator can be considered as a reasonable trade-off in that it generates almost greedy random points in a stochastic manner and thus generates high-quality candidate points but still with some amount of randomness which then help the metaheuristic to fine-tune them using its own logic. At least in our experiments, this has been the best approach for every combination of metaheuristic and problem instance. Nonetheless, it is very important to select the right politeness parameter. From these experiments the suggested value is around $\lambda = 0.10$.

Despite studying the random point generation theory closely related to TSP, it could be useful even more generally for problems where initial solution can be generated sequentially, selecting the next coordinate from a finite set where each option yields certain objective function value.

Acknowledgements

The paper was created with the support of CTU in Prague, Grant SGS14/208/OHK4/3T/14.

References

[1] Cerny, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. In: *Journal of optimization theory and applications*, pp. 41–51, Springer (1985)

[2] Akl, S. G.: Discrete steepest descent in real time. *Parallel and Distributed Computing Practices* 4, 301–317 (2001)

[3] Zabinsky, Z. B.: *Random Search Algorithms*. Wiley Encyclopedia of Operations Research and Management Science (2010)

[4] Applegate, D. L., Bixby, R. E., Chvatal, V., Cook, W. J.: *The Traveling Salesman Problem: A Computational Study*. Princeton university press (2011)

[5] Mojzeš, M., Kukal, J., Lauschmann, H.: Sub-model Testing in Fractographic Analysis. In: *Proc. of Mendel 2012 Soft Computing Conference*, pp. 350–355, Brno Univ Technology Press (2012)

[6] Mojzeš, M., Kukal, J., Tran, V.Q., Jablonský, J.: Performance Comparison of Heuristic Algorithms via Multi-criteria Decision Analysis. In: *Proc. of Mendel 2011 Soft Computing Conference*, pp. 244–251, Brno Univ Technology Press (2011)

Robustness of Monge fuzzy matrices with inexact data

M. Molnárová[1]

Abstract. Robustness of interval Monge matrices over binary max-min algebra (fuzzy matrices) is studied. The max-min algebra is an extremal algebra with operations maximum and minimum. An interval matrix A over fuzzy algebra is a set of matrices given by a lower bound matrix and an upper bound matrix. There are two possibilities to define the robustness of an interval matrix. Namely, the possible robustness, if there is at least one robust matrix in A or universal robustness, if all matrices from A are robust. Equivalent conditions for possible robustness and universal robustness of interval Monge matrices in binary case, were proved. Polynomial algorithms for checking the necessary and sufficient conditions for interval Monge matrices are introduced.

Keywords: (max, min) algebra, robustness, Monge matrix

JEL classification: C02
AMS classification: 08A72, 90B35, 90C47

1 Introduction

The max-min algebra (with operations maximum and minimum), known also as fuzzy algebra is one of the so-called extremal algebras. The operation maximum creates no new element (as well as the operation minimum). This crucial property of an extremal algebra predestinates max-min algebra to model applications in many divers areas as discrete dynamic systems (DDS), graph theory, knowledge engineering or description of technical devices. Properties of fuzzy matrices were described in [3]. The Monge matrices and their applications were studied in [1], [4]. Robust matrices over fuzzy algebra were investigated in [9]. Robustness of Monge fuzzy matrices in binary case was presented in [6]. Sufficient and necessary conditions for robustness of Monge fuzzy matrices were proved in [7]. Robustness of interval fuzzy matrices was studied in [8].

The aim of this paper is to find equivalent conditions for possible robustness and universal robustness of interval Monge fuzzy matrices in binary case and to introduce polynomial algorithms for verifying the possible robustness and universal robustness as well.

2 Preliminaries

The fuzzy algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus$, $\otimes$. The least element in B will be denoted by O, the greatest one by I. By $\mathbb{N}$ we denote the set of all natural numbers. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$, the least common multiple of the set S is denoted by $\operatorname{lcm} S$. For a given natural $n \in \mathbb{N}$, we use the notation N for the set of all smaller or equal positive natural numbers, i.e., $N = \{1, 2, \ldots, n\}$.

For any $m, n \in \mathbb{N}$, $B(m, n)$ denotes the set of all matrices of type $m \times n$ and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. The matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes$) as matrix operations over any field. The rth power of a matrix $A \in B(n, n)$ is denoted by A^r, with elements a_{ij}^r. For $A, C \in B(n, n)$ we write $A \leq C$ if $a_{ij} \leq c_{ij}$ holds for all $i, j \in N$.

[1]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, e-mail: Monika.Molnarova@tuke.sk

A *digraph* is a pair $G = (V, E)$, where V, the so-called vertex set, is a finite set, and E, the so-called edge set, is a subset of $V \times V$. A digraph $G' = (V', E')$ is a subdigraph of the digraph G (for brevity $G' \subseteq G$), if $V' \subseteq V$ and $E' \subseteq E$. A path in the digraph $G = (V, E)$ is a sequence of vertices $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1, \ldots, k$. The number k is the length of the path p and is denoted by $\ell(p)$. If $i_1 = i_{k+1}$, then p is called a cycle. For a given matrix $A \in B(n, n)$ the symbol $G(A) = (N, E)$ stands for the complete, edge–weighted digraph associated with A, i.e. the vertex set of $G(A)$ is N, and the capacity of any edge $(i, j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A, h)$ is the digraph $G = (N, E')$ with the vertex set N and the edge set $E' = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$.

The following lemma describes the relation between matrices and corresponding threshold digraphs.

Lemma 1. [8] Let A, $C \in B(n, n)$. Let h, h_1, $h_2 \in B$.

(i) If $A \leq C$ then $G(A, h) \subseteq G(C, h)$,

(ii) if $h_1 < h_2$ then $G(A, h_2) \subseteq G(A, h_1)$.

By a *strongly connected component* of a digraph $G(A, h) = (N, E)$ we mean a subdigraph $\mathcal{K} = (N_\mathcal{K}, E_\mathcal{K})$ generated by a non-empty subset $N_\mathcal{K} \subseteq N$ such that any two distinct vertices $i, j \in N_\mathcal{K}$ are contained in a common cycle, $E_\mathcal{K} = E \cap (N_\mathcal{K} \times N_\mathcal{K})$ and $N_\mathcal{K}$ is the maximal subset with this property. A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. For any non-trivial strongly connected component $\mathcal{K}$ is the *period* of $\mathcal{K}$ defined as $\operatorname{per}\mathcal{K} = \gcd\{\,\ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0\,\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per}\mathcal{K} = 1$. By $\mathrm{SCC}^\star(G)$ we denote the set of all non-trivial strongly connected components of G.

Let $A \in B(n, n)$ and $x \in B(n)$. The sequence $O(A, x) = \{x^{(0)}, x^{(1)}, x^{(2)}, \ldots, x^{(n)}, \ldots\}$ is the orbit of $x = x^{(0)}$ generated by A, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.
For a given matrix $A \in B(n, n)$, the number $\lambda \in B$ and the n–tuple $x \in B(n)$ are the so-called *eigenvalue* of A and *eigenvector* of A, respectively, if they are the solution of the *eigenproblem* for matrix A, i.e. they satisfy the equation $A \otimes x = \lambda \otimes x$. The corresponding *eigenspace* $V(A, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e. $V(A, \lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.
Let $\lambda \in B$. A matrix $A \in B(n, n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \geq R$. The smallest natural number p with above property is called the period of A, denoted by $\operatorname{per}(A, \lambda)$. In case $\lambda = I$ we denote $\operatorname{per}(A, I)$ by abbreviation $\operatorname{per} A$.

Definition 1. Let $A = (a_{ij}) \in B(n, n)$, $\lambda \in B$. Let $T(A, \lambda) = \{x \in B(n);\ O(A, x) \cap V(A, \lambda) \neq \emptyset\}$. A is called *λ–robust* if $T(A, \lambda) = B(n)$. A λ–robust matrix with $\lambda = I$ is called *a robust matrix.*

In our considerations we will use the following result (adapted for $\lambda = I$) proved in [9] to study robustness of a matrix.

Lemma 2. [9] Let $A = (a_{ij}) \in B(n, n)$. Then A is robust if and only if $\operatorname{per} A = 1$.

Definition 2. We say, that a matrix $A = (a_{ij}) \in B(m, n)$ is a convex Monge matrix (concave Monge matrix) if and only if

$$a_{ij} \otimes a_{kl} \leq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l$$
$$\big(\, a_{ij} \otimes a_{kl} \geq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l \,\big).$$

In this paper, we assume that the considered matrices are convex.

Under restriction to matrices which satisfy the condition $A \geq I_{ad}$ (only the weight of arcs $(1, n)$, $(2, n-1)$, $\ldots$, $(n, 1)$ is equal to 1 in I_{ad}) the following necessary and sufficient condition for a binary fuzzy matrix was proved.

Theorem 1. [6] *Let $A = (a_{ij}) \in B(n, n)$ be a Monge matrix with $A \geq I_{ad}$. Then A is robust if and only if $G(A, 1)$ is strongly connected and contains a loop.*

We recall a necessary and sufficient condition for a fuzzy Monge matrix to be robust modified for binary case.

Theorem 2. [7] *Let $A = (a_{ij}) \in B(n, n)$ be a Monge matrix. Then A is robust if and only if $G(A, 1)$ contains at most one non-trivial strongly connected component and this has a loop.*

3 Robustness of interval Monge matrices

In this section we shall deal with matrices with interval elements. Similarly to [2], [5], we define an interval matrix $\mathbf{A}$.

Definition 3. Let $\underline{A}, \overline{A} \in B(n,n)$, $\underline{A} \leq \overline{A}$. An interval matrix $\mathbf{A}$ with bounds $\underline{A}$ and $\overline{A}$ is defined as follows

$$\mathbf{A} = [\underline{A}, \overline{A}] = \left\{ A \in B(n,n);\ \underline{A} \leq A \leq \overline{A} \right\}.$$

We can set following questions investigating robustness of an interval matrix $\mathbf{A}$. Is A robust for some $A \in \mathbf{A}$ or for all $A \in \mathbf{A}$?

Definition 4. An interval matrix $\mathbf{A}$ is called

- *possibly robust* if there exists a matrix $A \in \mathbf{A}$ such that A is robust,
- *universally robust* if each matrix $A \in \mathbf{A}$ is robust.

Definition 5. An interval matrix $\mathbf{A}^M = [\underline{A}, \overline{A}]$ is called interval Monge, if $\underline{A}$, $\overline{A} \in B(n,n)$ are Monge matrices and $\mathbf{A}^M = \{A \in \mathbf{A};\ A \text{ is Monge}\}$.

Since $\underline{A}$, $\overline{A} \in \mathbf{A}^M$, the set $\mathbf{A}^M$ is non-empty.

Possible robustness

A necessary and sufficient condition for an interval matrix to be possibly robust was proved in [8]. An $O(n^5)$ algorithm for checking the possible robustness and finding a robust matrix $A^\star \in \mathbf{A}$ was introduced. Let $H = \{\overline{a}_{ij};\ i,j \in N\} \cup \{\underline{a}_{ij};\ i,j \in N\}$.

Theorem 3. [8] *An interval matrix* $\mathbf{A}$ *is possibly robust if and only if for each* $h \in H$ *and for each* $\mathcal{K} \in \mathrm{SCC}^\star(G(\overline{A}, h))$ *such that* $\operatorname{per} \mathcal{K} \neq 1$ *the digraph* $G(\underline{A}, h)/N_{\mathcal{K}}$ *is acyclic.*

Applying the algorithm introduced in [8] for binary case of max-min algebra we can define the matrix $A^\star$ and describe the corresponding algorithm. Denote by $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ the non-trivial strongly connected components of $G(\overline{A}, 1)$ with period equal to one (if there exist).

We define the matrix $A^\star$ by following equation

$$a^\star_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \bigcup_{s=1}^{m} E_{\mathcal{K}_s} \text{ and } \underline{a}_{ij} < 1, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases} \tag{1}$$

However, the resulting matrix $A^\star$ need not to have the Monge property although the input matrices $\underline{A}$ and $\overline{A}$ do as illustrated on the following example.

Example 1. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5,5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The digraph $G(\overline{A}, 1)$ consists of two non-trivial strongly connected components (see Figure 1). The period of the component $\mathcal{K}$ generated by the node set $K = \{1, 5\}$ is 2, but $G(\underline{A}, 1)/N_{\mathcal{K}}$ is acyclic (see Figure 1). The only non-trivial strongly connected component of $G(\overline{A}, 1)$ with period equal to one is $\mathcal{K}_1$ generated by the node set $K_1 = \{2, 3, 4\}$. Hence we can compute the robust matrix $A^\star$ by (1). The corresponding digraph $G(A^\star, h)$ for threshold $h = 1$ is depicted on Figure 2. However, the Monge property does not hold, since $a^\star_{13} \otimes a^\star_{24} > a^\star_{14} \otimes a^\star_{23}$. Hence $A^\star \notin \mathbf{A}^M$.

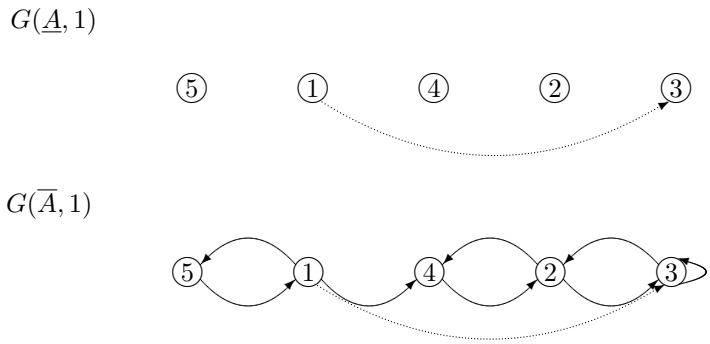


Figure 1: Possible robustness in binary case

$$A^\star = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

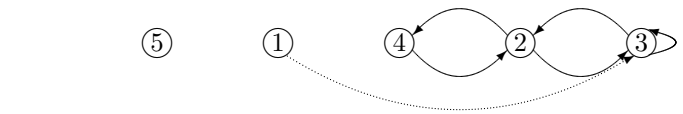


Figure 2: Robust matrix without Monge property

Restricting ourselves to interval matrices which satisfy the condition $A \geq I_{ad}$, we can prove a necessary and sufficient condition for possible robustness of interval Monge matrices using Theorem 1.

Theorem 4. *An interval Monge matrix* $\mathbf{A}^M$ *with* $\underline{A} \geq I_{ad}$ *is possibly robust if and only if* $\overline{A}$ *is robust.*

Proof. Let $\mathbf{A}^M$ with $\underline{A} \geq I_{ad}$ be possibly robust. There exists a robust matrix $A \in \mathbf{A}^M$. By Theorem 1 is the digraph $G(A,1)$ strongly connected with a loop. Since $A \leq \overline{A}$ then by Lemma 1 $G(A,1) \subseteq G(\overline{A},1)$. Thus the digraph $G(\overline{A},1)$ is strongly connected with a loop, too. Hence $\overline{A}$ is robust.
The converse implication is trivial. □

Algorithm **Possible Robustness for binary case**

Input. $\mathbf{A}^M = [\underline{A}, \overline{A}]$, I_{ad}.
Output. 'non-Monge matrix' in variable *prbin* if $\mathbf{A}^M$ is not an interval Monge matrix; 'non-proper matrix' in variable *prbin* if $\mathbf{A}^M$ does not satisfied the condition $\underline{A} \geq I_{ad}$; 'yes' in variable *prbin* if $\mathbf{A}^M$ is possibly robust; 'no' in *binpr* if $\mathbf{A}^M$ is not possibly robust.
begin

(i) **If** $\underline{A}$ or $\overline{A}$ is not Monge **then** *prbin* :='non-Monge matrix'; **go to** end;

(ii) **If** the condition $\underline{A} \geq I_{ad}$ is not satisfied **then** *prbin* :='non-proper matrix'; **go to** end;

(iii) **If** the digraph $G(\overline{A},1)$ is not strongly connected **then** *prbin* :='no'; **go to** end;

(iv) **If** $G(\overline{A},1)$ contains no loop **then** *prbin* :='no', else *prbin* :='yes';

end

Theorem 5. *The algorithm* **Possible Robustness for binary case** *correctly decides in* $O(n^3)$ *time for an interval matrix* $\mathbf{A}$ *whether* $\mathbf{A}^M$ *is an interval Monge matrix and verifies the condition* $\underline{A} \geq I_{ad}$ *and the possible robustness in positive case.*

Proof. To verify the Monge property for matrices $\underline{A}$ and $\overline{A}$ takes $O(n^3)$ time. To check the condition $\underline{A} \geq I_{ad}$ takes $O(n)$ time. To verify the strong connectivity of $G(\overline{A},1)$ requires $O(n^2)$ operations. To check the existence of a loop takes $O(n)$ time. Thus the total complexity of the algorithm is $O(n^3)$. □

Remark 1. The robustness in fuzzy algebra can reflect the following economic background. Suppose there is some commodity on the market, let's say, houses with properties p_1, p_2, ..., p_n such as location, number of floors, number of bedrooms, number of bathrooms, etc. The customer fills in the fuzzy matrix $A = (a_{ij}) \in B(n,n)$ his preferences, i.e. $a_{ij} = 1$ if the property p_i has a preference to property p_j, otherwise $a_{ij} = 0$. In the same manner fills in the vector $x \in B(n)$ his initial preferences. If the matrix A is robust, then the system reaches an eigenvector $y = A^k \otimes x$, for some natural k, regardless of the choice of the vector x, i.e. it reaches the maximum stable vector y of the matrix A, which represents the final vector of preferences. Replacing the fixed matrix A by an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ allows to find at least one robust matrix within minimum $\underline{A}$ and maximum $\overline{A}$ restrictions.

Universal robustness

In contrast to possible robustness there is no polynomial algorithm for checking the universal robustness of interval matrices in fuzzy algebra. We were able to prove the necessary and sufficient condition for an interval Monge matrix under condition $\underline{A} \geq I_{ad}$ to be universally robust in binary case. Moreover, we can check the universal robustness in $O(n^3)$ time.

Theorem 6. *An interval Monge matrix $\mathbf{A}^M$ with $\underline{A} \geq I_{ad}$ is universally robust if and only if $\underline{A}$ is robust.*

Proof. Let $\underline{A}$ be robust. By Theorem 1 is $G(\underline{A}, 1)$ strongly connected and contains a loop. Let $A \in \mathbf{A}^M$. Using the fact that $\underline{A} \leq A$ we get by Lemma 1 $G(\underline{A}, 1) \subseteq G(A, 1)$. Hence $G(A, 1)$ is strongly connected and contains a loop. Consequently A is robust.
The converse implication is trivial. □

Theorem 7. *There is an $O(n^3)$ algorithm which decides for an interval matrix $\mathbf{A}$ whether $\mathbf{A}^M$ is an interval Monge matrix and verifies the condition $\underline{A} \geq I_{ad}$ and the universal robustness in positive case.*

Proof. To check the Monge property of matrices $\underline{A}$ and $\overline{A}$ takes $O(n^3)$ time and dominates the total complexity of the complete algorithm. To verify the condition $A \geq I_{ad}$ takes $O(n)$ time, to check the existence of a loop takes $O(n)$ operations as well and, finally, to verify the strong connectivity of $G(\underline{A}, 1)$ takes $O(n^2)$ time. □

Example 2. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5,5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\underline{A}$ and $\overline{A}$ are Monge matrices the corresponding interval matrix $\mathbf{A}^M$ is an interval Monge matrix. The digraph $G(\underline{A}, h)$ for threshold $h = 1$ is strongly connected and contains a loop (see Figure 3). Consequently the considered interval Monge matrix $\mathbf{A}^M$ is universally robust.

$G(\underline{A}, 1)$

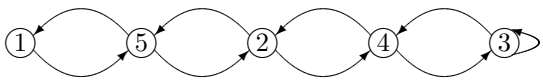


Figure 3: Universal robustness

It is not sufficient that matrices $\underline{A}$ and $\overline{A}$ are robust to guarantee universal robustness without restriction $\underline{A} \geq I_{ad}$ as we can see in the next example.

Example 3. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5,5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The Monge property of both matrices $\underline{A}$ and $\overline{A}$ guarantees that the corresponding interval matrix $\mathbf{A}^M$ is an interval Monge matrix. Each of the digraphs $G(\underline{A}, h)$ and $G(\overline{A}, h)$ for threshold $h = 1$ contains one non-trivial strongly connected component with a loop. Hence the matrices $\underline{A}$ and $\overline{A}$ are by Theorem 2 robust. In spite of this fact there exists a matrix $A^\star \in \mathbf{A}^M$ which is not robust (see Figure 4).

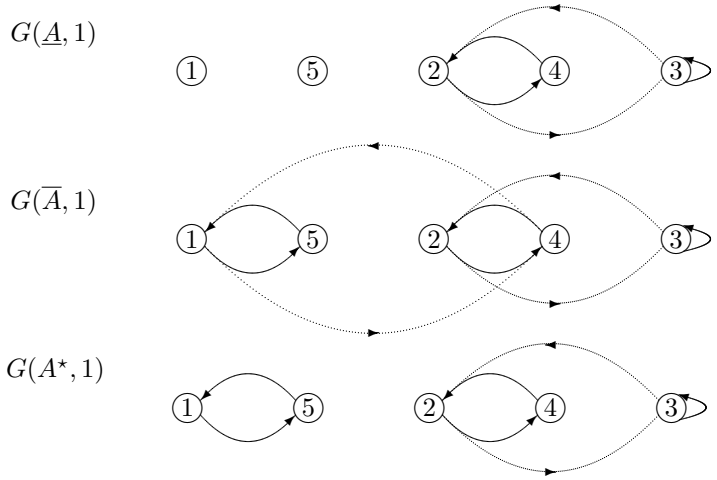


Figure 4: Universal robustness in non-robust case

4 Conclusion

Problems related to robustness of interval fuzzy matrices in general were introduced in [8]. Sufficient and necessary conditions for possible as well as universal robustness of interval fuzzy matrices were proved. However the suggested polynomial algorithm for checking the possible robustness has the computational complexity $O(n^5)$ and the computational complexity of the algorithm for checking the universal robustness can be even exponentially large. The aim of this paper is to present more effective algorithms for special class of interval fuzzy matrices, namely Monge matrices in binary case. Based on sufficient and necessary conditions we have proved, we can introduce polynomial algorithms, both with computational complexity $O(n^3)$, for verifying the possible or universal robustness.

Acknowledgements

This work was supported by the Slovak Cultural and Education Agency No. 032 TUKE- 4/2013.

References

[1] Burkard, R. E., Klinz, B. and Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.

[2] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J. and Zimmermann, K.: Linear Optimization Problems with Inexact Data. Springer–Verlag, Berlin 2006.

[3] Gavalec, M.: *Periodicity in extremal algebras.* GAUDEAMUS, Hradec Králové, 2004.

[4] Gavalec, M. and Plavka, J.: An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max algebra, *DAM*, Volume **127** (2003), 651–656.

[5] Gavalec, M. and Zimmermann, K.: Classification of solutions to systems of two-sided equations with interval coefficients, *Inter. J. of Pure and Applied Math.* Volume **45** (2008), 533–542.

[6] Molnárová, M.: Robustness of (0–1) Monge fuzzy matrices, In: *Proceedings of* 31^{st} *Int. Conference Mathematical Methods in Economics 2013*, Jihlava, 2013, 636–641.

[7] Molnárová, M.: Robustness of Monge matrices in fuzzy algebra, In: *Proceedings of* 32^{nd} *Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.

[8] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[9] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue 5 (2011), 381–388.

Fuzzy Matrix Equations

Helena Myšková[1]

Abstract.

Fuzzy algebra is the triple $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I} = [O, I]$ is a linear ordered set with the least element O and the greatest element I and $\oplus$, $\otimes$ are binary operations defined by $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

In this paper, we shall deal with the solvability of fuzzy matrix equations of the form $A \otimes X \otimes C = B$, where A, B, and C are given matrices of suitable sizes and X is an unknown matrix. We derive the necessary and sufficient conditions for the solvability of an fuzzy matrix equation, which can be checked in $O(n^4)$ time.

Keywords: fuzzy algebra, matrix equation, tensor product

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Motivation

Fuzzy equations have found a broad area of applications in causal models which emphasize relationships between input and output variables. They are used in diagnosis models [1], [6], [7], [8] or models of nondeterministic systems [9]. Diagnostic models are particularly important because they cope with the uncertainty on the many real-life situations either concerning medical diagnosis or diagnosis of technical devices. In the simplest formulation we are faced with a space of symptoms and a space of faults. Elements of faults are related with elements of symptoms by means of a fuzzy relation. As usual, the higher the value of the relationship for a certain pair of arguments, the stronger is the appropriate relationship between the symptom and fault. The solution of the fuzzy relational equation of the form $A \otimes x = b$, where A is a matrix, b and x are vectors of suitable dimensions and classical addition and multiplication operations are replaced by maximum and minimum, provides a maximal set of symptoms that produce the given fault. The solvability of the systems of fuzzy linear equations is well reviewed. In this paper, we shall deal with the solvability of fuzzy matrix equations of the form $A \otimes X \otimes C = B$, where A, B, and C are given matrices of suitable sizes and X is an unknown matrix. In the following example we will show one of possible applications.

Example 1. Let us consider a situation, in which passenger from places P_1, P_2, P_3, P_4 want to transfer to holiday destinations D_1, D_2, D_3. Different transportation means provide transporting goods from places P_1, P_2, P_3, P_4 to airport terminals T_1, T_2 (See Figure 1). We assume that the connection between P_i and T_l is possible only via one of the check points Q_1, Q_2, Q_3. There is the arrow $(P_i\, Q_j)$ on Figure 1 if there is a road from P_i to Q_j and there is the arrow $(T_l\, D_k)$ if terminal T_l handles passengers traveling to destination D_k ($i = 1, 2, 3, 4$, $j = 1, 2, 3$, $k = 1, 2, 3$, $l = 1, 2$). The symbols along the arrows represent the capacities of the corresponding connections.

Suppose that the number of passengers travelling from place P_i to destination D_k is denoted by b_{ik}. If place Q_j is linked with T_l by a road with a capacity x_{jl}, then the capacity of the connection between P_i and D_k via Q_j and using terminal T_l is equal to $\min\{a_{ij}, x_{jl}, c_{lk}\}$.

To ensure the transportation for all passengers from P_1 to their destinations the following equations

[1]Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, e-mail: helena.myskova@tuke.sk

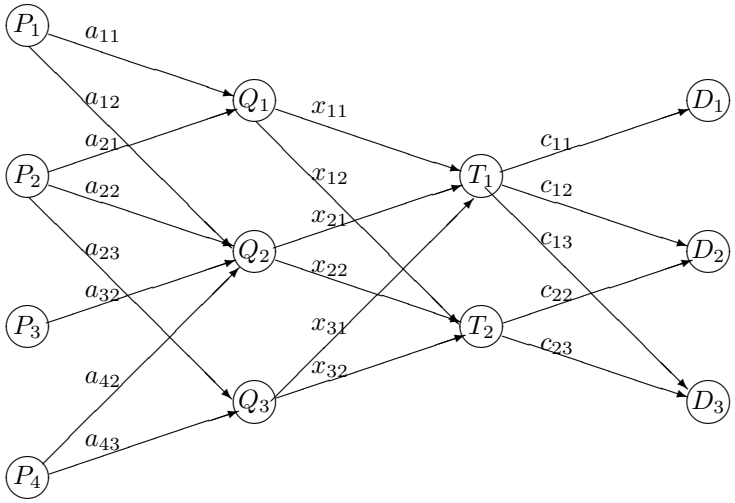


Figure 1: Transportation system

must be satisfied:

$$\begin{aligned}
&\max\Big\{\min\{a_{11},x_{11},c_{11}\},\min\{a_{12},x_{21},c_{11}\}\Big\}=b_{11},\\
&\max\Big\{\min\{a_{11},x_{11},c_{12}\},\min\{a_{12},x_{21},c_{12}\},\min\{a_{12},x_{22},c_{22}\}\Big\}=b_{12},\\
&\max\Big\{\min\{a_{11},x_{12},c_{23}\},\min\{a_{11},x_{11},c_{13}\},\min\{a_{12},x_{21},c_{13}\},\min\{a_{12},x_{22},c_{23}\}\Big\}=b_{13}.
\end{aligned}\tag{1}$$

The similar equations must be satisfied to ensure transportation for all passengers from P_2, P_3 and P_4 to their destinations.

In general, suppose that there are m places $P_1, P_2, \ldots, P_m$, n transfer points $Q_1, Q_2, \ldots, Q_n$, s terminals $T_1, T_2, \ldots, T_s$ and r destinations $D_1, D_2, \ldots, D_r$. Denote by a_{ij} (c_{lk}) the capacity of the road from P_i to Q_j (from T_l to D_k). If there is no road from P_i to Q_j (from T_l to D_k), we put $a_{ij} = O$ ($c_{lk} = O$). Our task is to choose the appropriate capacities $x_{jl},\ j \in N = \{1,\ 2, \ldots,\ n\}, l \in S = \{1, 2, \ldots, s\}$ such that the maximum capacity of the road from P_i to D_k is equal to a given number b_{ik} for all $i \in M = \{1, 2, \ldots, m\}$ and for all $k \in R = \{1, 2, \ldots, r\}$, i.e.,

$$\max_{j\in N,\, l\in S} \min\{a_{ij}, x_{jl}, c_{lk}\} = \ b_{ik} \tag{2}$$

for each $i \in M$.

2 Preliminaries

Fuzzy algebra is the triple $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I} = [O, I]$ is a linear ordered set with the least element O, the greatest element I, and two binary operations $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

Denote by M, N, R, and S the index sets $\{1, 2, \ldots, m\}$, $\{1, 2, \ldots, n\}$, $\{1, 2, \ldots, r\}$, and $\{1, 2, \ldots, s\}$, respectively. The set of all $m \times n$ *matrices* over $\mathcal{I}$ is denoted by $\mathcal{I}(m, n)$ and the set of all column *n-vectors* over $\mathcal{I}$ by $\mathcal{I}(n)$.
Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in the classical algebra. We will consider the *ordering* $\leq$ on the sets $\mathcal{I}(m, n)$ and $\mathcal{I}(n)$ defined as follows:

- for $A, C \in \mathcal{I}(m, n):\ A \leq C$ if $a_{ij} \leq c_{ij}$ for all $i \in M,\ j \in N$,
- for $x, y \in \mathcal{I}(n):\ x \leq y$ if $x_j \leq y_j$ for all $j \in N$.

We will use the *monotonicity of* $\otimes$, which means that for each $A, C \in \mathcal{I}(m, n)$ and for each $B,\ D \in \mathcal{I}(n, s)$ the implication

$$\text{if } A \leq C \text{ and } B \leq D \text{ then } A \otimes B \leq C \otimes D$$

holds true.

Let $A \in \mathcal{I}(m,n)$ and $b \in \mathcal{I}(m)$. In fuzzy algebra we can write the system of equations in the matrix form

$$A \otimes x = b. \tag{3}$$

The crucial role for the solvability of system (3) in fuzzy algebra is played by the *principal solution* of system (3), defined by

$$x_j^*(A,b) = \min_{i \in M}\{b_i : a_{ij} > b_i\} \tag{4}$$

for each $j \in N$, where $\min \emptyset = I$.

The following theorem describes the importance of the principal solution for the solvability of (3).

Theorem 1. [4, 10] *Let $A \in \mathcal{I}(m,n)$ and $b \in \mathcal{I}(m)$ be given.*

i) If $A \otimes x = b$ for $x \in \mathcal{I}(n)$, then $x \leq x^*(A,b)$.
ii) $A \otimes x^*(A,b) \leq b$.
iii) The system $A \otimes x = b$ is solvable if and only if $x^*(A,b)$ is its solution.

The properties of a principal solution are expressed in the following assertions.

Lemma 2. [3] Let $A \in \mathcal{I}(m,n)$ and $b, d \in \mathcal{I}(m)$ be such that $b \leq d$. Then $x^*(A,b) \leq x^*(A,d)$.

Lemma 3. [5] Let $b \in \mathcal{I}(m)$ and $C, D \in \mathcal{I}(m,n)$ be such that $D \leq C$. Then $x^*(C,b) \leq x^*(D,b)$.

Lemma 4. Let $A \in \mathcal{I}(m,n)$, $b \in \mathcal{I}(m)$ and c, $d \in \mathcal{I}$, $c \leq d$. Then $x^*(A \otimes d, b) \leq x^*(A \otimes c, b)$.

Proof. According to (4) for each $j \in N$ we have

$$x_j^*(A \otimes c, b) = \min\{b_i : c \otimes a_{ij} > b_i\} = \min\{b_i : \min\{c, a_{ij}\} > b_i\}.$$

Similarly, $x_j^*(A \otimes d, b) = \min\{b_i : \min\{d, a_{ij}\} > b_i\}$.

Since $\{b_i : \min\{c, a_{ij}\} > b_i\} \subseteq \{b_i : \min\{d, a_{ij}\} > b_i\}$, we have

$$x_j^*(A \otimes c, b) = \min\{b_i : \min\{c, a_{ij}\} \ > \ b_i\} \geq \min\{b_i : \min\{d, a_{ij}\} > b_i\} = x_j^*(A \otimes d, b).$$

□

3 Matrix Equations and Tensor Product

Let $A \in \mathcal{I}(m,n)$, $B \in \mathcal{I}(m,r)$, $X \in \mathcal{I}(n,s)$ and $C \in \mathcal{I}(s,r)$ are given matrices. It is easy to see that $[A \otimes X \otimes C]_{ik} = \max\limits_{j \in N,\, l \in S} \min\{a_{ij}, x_{jl}, c_{lk}\}$. Hence, we can (2) write in the form

$$A \otimes X \otimes C = B. \tag{5}$$

In the following, we shall deal with the solvability of (5). We shall use the notion of tensor product.

Definition 1. Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be an $r \times s$ matrix. The *tensor product* of A and B is the following $mr \times ns$ matrix:

$$A \boxtimes B = \begin{pmatrix} A \otimes b_{11} & A \otimes b_{12} & \dots & A \otimes b_{1s} \\ A \otimes b_{21} & A \otimes b_{22} & \dots & A \otimes b_{2s} \\ \dots & \dots & \dots & \dots \\ A \otimes b_{r1} & A \otimes b_{r2} & \dots & A \otimes b_{rs} \end{pmatrix}.$$

Let $X \in B(n,s)$. Denote by $\mathrm{vec}\,(X)$ the vector $(X_1, X_2, \dots, X_s)^\top$, where X_l is l-th column of matrix X. Similarly we define $\mathrm{vec}\,(B)$.

Theorem 5. [2] *Matrix equation*

$$(A_1 \otimes X \otimes C_1) \oplus (A_2 \otimes X \otimes C_2) \oplus \cdots \oplus (A_r \otimes X \otimes C_r) = B, \tag{6}$$

where A_i, C_i and B are of compatible sizes, is equivalent to the vector-matrix system

$$(A_1 \boxtimes C_1^\top \oplus A_2 \boxtimes C_2^\top \oplus \dots A_r \boxtimes C_r^\top) \otimes \mathrm{vec}\,(X) = \mathrm{vec}\,(B). \tag{7}$$

Proof. The proof in max-plus algebra is given in [2]. It is sufficient to prove the statement for $r = 1$. We have

$$\mathrm{vec}\,(A \otimes X \otimes C) = \mathrm{vec}\,((A \otimes X_1, \dots, A \otimes X_s) \otimes C)$$

$$= \begin{pmatrix} A \otimes X_1 \otimes c_{11} \oplus A \otimes X_2 \otimes c_{21} \oplus \cdots \oplus A \otimes X_s \otimes c_{s1} \\ A \otimes X_1 \otimes c_{12} \oplus A \otimes X_2 \otimes c_{22} \oplus \cdots \oplus A \otimes X_s \otimes c_{s2} \\ \dots \\ A \otimes X_1 \otimes c_{1r} \oplus A \otimes X_2 \otimes c_{2r} \oplus \cdots \oplus A \otimes X_s \otimes c_{sr} \end{pmatrix}$$

$$= \begin{pmatrix} A \otimes c_{11} & A \otimes c_{21} & \dots & A \otimes c_{s1} \\ A \otimes c_{12} & A \otimes c_{22} & \dots & A \otimes c_{s2} \\ \dots & \dots & \dots & \dots \\ A \otimes c_{1r} & A \otimes c_{2r} & \dots & A \otimes c_{sr} \end{pmatrix} \otimes \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_s \end{pmatrix}$$

$$= (A \boxtimes C^\top) \otimes \mathrm{vec}\,(X)$$

□

For $r = 1$, matrix equation the matrix equation in the form (6) takes form $A \otimes X \otimes C = B$.

Denote by $X^*(A, C, B) = (x^*_{jl}(A, C, B))$ the matrix defined as follows

$$x^*_{jl}(A, C, B) = \min_{k \in R}\{x^*_j(A \otimes c_{lk}, B_k)\}. \tag{8}$$

We shall call the matrix $X^*(A, C, B)$ a *principal matrix solution* of (5). The following theorem expresses the properties of $X^*(A, C, B)$ and gives the necessary and sufficient condition for the solvability of (5).

Theorem 6. *Let $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(m, n)$.*

i) If $A \otimes X \otimes C = B$ for $X \in \mathcal{I}(n, s)$, then $X \leq X^*(A, C, B)$.
ii) $A \otimes X^*(A, C, B) \otimes C \leq B$.
iii) The matrix equation $A \otimes X \otimes C = B$ is solvable if and only if $X^*(A, C, B)$ is its solution.

Proof. The consequence of Theorem 5 is that the solvability of (5) is equivalent to the solvability of

$$(A \boxtimes C^\top) \otimes \mathrm{vec}\,(X) = \mathrm{vec}\,(B). \tag{9}$$

By Theorem 1 iii) the solvability of (9) is equivalent to

$$(A \boxtimes C^\top) \otimes x^*(A \boxtimes C^\top, \mathrm{vec}\,(B)) = \mathrm{vec}\,(B).$$

We will prove that $x^*(A \boxtimes C^\top, \mathrm{vec}\,(B)) = \mathrm{vec}\,(X^*(A, C, B))$. We rewrite (9):

$$\begin{pmatrix} A \otimes c_{11} & A \otimes c_{21} & \dots & A \otimes c_{s1} \\ A \otimes c_{12} & A \otimes c_{22} & \dots & A \otimes c_{s2} \\ \dots & \dots & \dots & \dots \\ A \otimes c_{1r} & A \otimes c_{2r} & \dots & A \otimes c_{sr} \end{pmatrix} \otimes \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_s \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_r. \end{pmatrix}$$

By (4) we get

$$x^*_{jl} = \min\Big\{\min_{i\in M}\{b_{i1} : a_{ij}\otimes c_{l1} > b_{i1}\}, \min_{i\in M}\{b_{i2} : a_{ij}\otimes c_{l2} > b_{i2}\}, \ldots, \min_{i\in M}\{b_{ir} : a_{ij}\otimes c_{lr} > b_{ir}\}\Big\} =$$

$$\min\Big\{x^*_j(A\otimes c_{l1}, B_1), x^*_j(A\otimes c_{l2}, B_2), \ldots, x^*_j(A\otimes c_{lr}, B_r)\Big\} = \min_{k\in R} x^*_j(A\otimes c_{lk}, B_k).$$

Hence the proof of parts i), ii) and iii) follows directly from Theorem 1. □

Remark 1. Equality (8) is equivalent to

$$X^*(A,C,B) = (X^*_1(A,C,B), X^*_2(A,C,B), \ldots, X^*_s(A,C,B)),$$

where

$$X^*_l(A,C,B) = \min_{k\in R} x^*(A\otimes c_{lk}, B_k). \tag{10}$$

The following theorem deals with the complexity of checking the solvability of fuzzy matrix equation. For the sake of simplicity, we will suppose that $m = r = s = n$, i.e., all matrices are square of order n.

Theorem 7. *There is an algorithm which decide whether the given fuzzy matrix equation is solvable in $O(n^4)$ steps.*

Proof. Checking the solvability is based on verification of the equality $A\otimes X^*(A,C,B)\otimes C = B$. Since computing $x^*(A\otimes c_{lk}, B_k)$ requires $O(n^2)$ arithmetic operations, computing $X^*_l(A,C,B)$ by (10) for fixed l and fixed requires $n\cdot O(n^2) = O(n^3)$ arithmetic operation. Hence, computing the matrix $X^*(A,C,B)$ requires $n\cdot O(n^3) = O(n^4)$ operations. Matrix multiplications need $O(n^3)$ arithmetic operations and checking the matrix equality requires $O(n^2)$ arithmetic operations.

Hence the total complexity of the algorithm for checking the solvability of (5) is $O(n^4) + O(n^3) + O(n^2) = O(n^4)$. □

Example 2. Let $\mathcal{I} = [0, 10]$ and let

$$A = \begin{pmatrix} 3 & 8 & 5 \\ 2 & 6 & 4 \\ 7 & 3 & 6 \end{pmatrix}, \qquad C = \begin{pmatrix} 6 & 7 \\ 3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 5 & 5 \\ 5 & 5 \\ 6 & 6 \end{pmatrix}.$$

We check whether the matrix equation $A\otimes X\otimes C = B$ is solvable.

Solution:
We have

$$A\otimes c_{11} = \begin{pmatrix} 3 & 6 & 5 \\ 2 & 6 & 4 \\ 6 & 3 & 6 \end{pmatrix}, \; A\otimes c_{12} = \begin{pmatrix} 3 & 7 & 5 \\ 2 & 6 & 4 \\ 7 & 3 & 6 \end{pmatrix}, \; A\otimes c_{21} = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}, \; A\otimes c_{22} = \begin{pmatrix} 3 & 4 & 4 \\ 2 & 4 & 4 \\ 4 & 3 & 4 \end{pmatrix}.$$

We compute the principal matrix solution by (10):

$$X^*(A,C,B) = \begin{pmatrix} \min\left\{\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 6 \\ 5 \\ 10 \end{pmatrix}\right\}, \min\left\{\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}\right\} \end{pmatrix} = \begin{pmatrix} 6 & 10 \\ 5 & 10 \\ 10 & 10 \end{pmatrix}.$$

Since $A\otimes X^*(A,C,B)\otimes C = B$, the given matrix equation is solvable and $X^*(A,C,B)$ is the greatest solution. √

4 Conclusion

In this paper, we dealt with the solvability of matrix equations in fuzzy algebra. Fuzzy algebra is a useful tool for describing real situation in the economy and industry. In Example 1, the values a_{ij}, x_{jl}, and c_{lk} represent the capacities of corresponding connections. In economics, those values can represent for example the financial costs for the production or transporting of some products. In another example, a_{ij} represents a measure of the preference of the property P_i of some object before the property Q_j, similarly x_{jl} (c_{lk}) represent a measure of the preference of the property Q_j before the property T_l (the property T_l before the property D_k).

In practice, the values a_{ij} and c_{lk} may depend on external conditions, so they are from intervals of possible values. Due to this fact, our main objective for the future will be studying fuzzy matrix equations with inexact data.

References

[1] A. Asse, P. Mangin, D. Witlaeys: Assisted diagnosis using fuzzy information, NAFIPS 2 Congress (Schenectudy, NY,1983).

[2] P. Butkovič, M. Fiedler, Tropical tensor product and beyond, `http://web.mat.bham.ac.uk/P.Butkovic/Pubs.html`.

[3] K. Cechlárová, Solutions of interval systems in fuzzy algebra, Proc.of SOR 2001, Eds. V. Rupnik, L.Zadnik-Stirn, S. Drobne, Preddvor, Slovenia, 321-326.

[4] R. A. Cuninghame-Green, Minimax Algebra, Lecture notes in Economics and Mathematical systems 1966, Springer, Berlin, 1979.

[5] H. Myšková, Interval systems of max-separable linear equations, Linear Alg. Appl. 403 (2005), 263-272.

[6] A. Di Nola, S. Salvatore, W. Pedrycz, E. Sanchez, Fuzzy Relation Equations and Their Applications to Knowledge Engineering, Kluwer Academic Publishers, Dordrecht, 1989.

[7] E. Sanchez, Medical diagnosis and composite relations, In: M. M. Gupta, R. K. Ragade, R. R. Yager, Eds, Advances in Fuzzy Set Theory and Applications (North-Holland, Amsterdam-New York, 1979), 437-444.

[8] T. Terano, Y. Tsukamoto, Failure diagnosis by using fuzzy logic. In: Proc. IEEE Conference on Decision Control (New Orleans, LA, 1977), 1390-1395.

[9] L. A. Zadeh: Toward a theory of fuzzy systems, In: R. E. Kalman, N. De Claris, Eds., Aspects of Network and Systems Theory (Hold, Rinehart and Winston, New York, 1971), 209-245.

[10] K. Zimmermann, Extremální algebra, Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha, 1976.

How to improve happiness in the EU countries?

Kateřina Myšková[1], Ladislava Issever Grochová[2]

Abstract. Contemporary economic and political goals, originally aimed at pure economic targets, have biased towards new benchmarks as promoting sustainable development, well-being and happiness for all, enhancing environmental protection as well as improving human rights, suppressing poverty or supporting good governance. In this paper we focus on the first categories of nowadays political priorities. In particular, we study how the EU countries do well in supporting happiness for their citizens. The question of our interest is what can be done to improve happiness in the EU states. The indicators of happiness that we use in our analysis are the Happy Planet Index and experienced well-being. Data envelopment analysis is used with the aim to identify the areas of possible improvement in areas contributing to happiness of population and to assess what exactly should be done to make EU citizens happier.

Keywords: happy planet index, experienced well-being, better life index, data envelopment analysis, EU countries.

JEL Classification: Q56, I3, C61
AMS Classification: 90C08

1 Introduction

Nowadays, the stress on purely economic aspects in society shifts towards new values that have become relevant. For inhabitants in richer countries ecological and social issues complement the economic ones creating so a new goal of becoming happy [5, 6, 7, 8]. This provides a new insight into economics when assessing welfare combining the techniques typically used by economists with those more commonly used by psychologists. Individuals do not consume just because they want to consume but because consumption has value as it contributes to human happiness. Furthermore, even if income per capita has risen remarkably over time, the level of happiness is much more stable in many countries. This implies that well-being cannot be explained only with economic variables but other non-financial factors must be included. Economists so expand the notion of utility compared to conventional economics, highlighting the role of non-income aspects that affect well-being.

Besides economic factors such as income, wealth, inflation or unemployment also other non-financial factors contribute to individual's happiness [4]. These are demographic and personality factors including nationality, education and health; political factors such as political regimes, governance quality or rights enforcement; and finally, environmental factors being represented with pollution or resource depletion.

Growing interest in well-being, happiness or life satisfaction requires measures of this phenomenon. Organizations like OECD or the New Economics Foundation (NEF) developed actually widely used both one- or multi-dimensional measures: the Happy Planet Index (HPI), Experienced Well-Being (EWB) and the Better Life Index (BLI). The possibility of numerical capturing of happiness facilitates formulating policy recommendations on the ways how to improve well-being in a particular country based on quantitative evidence.

The BLI focuses on a wide range of dimensions, starting from traditional measures such as income and jobs, health, education and the local environment, to personal safety and overall satisfaction with life, essential to well-being and happiness. This variety enables to identify relative strengths and weaknesses in the national well-being [9]. In other words, we can say that Better Life Index dimensions contribute to national happiness or well-being. This implies that the level of happiness can be improved enhancing the individual dimensions. In this paper, we are interested in how individual dimensions contribute to happiness in the EU countries. The aim of this paper is then to assess how individual European countries are efficient in supporting happiness for their inhabitants. For this purpose we employ data envelopment analysis already adopted by [3]. They study happiness from the individual perspective, in particular, how efficiently individuals transform their available resources into happiness. This approach necessarily lacks any suggestions whether and how the happiness can be enhanced

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, myskova@mendelu.cz
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, 613 00 Brno, Czech Republic, ladislava.grochova@mendelu.cz

even by the state authorities. The main contribution of our research is the shift from the individual level to a national level which enables us to formulate overall policy recommendations.

The paper is organized as follows: First, the concept of happiness and its importance is introduced. The second section provides the overview of data and methods used in this paper. Next, results on how states are efficient in enhancing happiness are discussed. The last section concludes.

2 Data and Methods

To assess how individual European countries are efficient in supporting national happiness we adopt data envelopment analysis using selected BLI dimensions as inputs and the HPI and EWB, respectively, as outputs. As our scope is also to provide political recommendations we focus only on the BLI aspects that can be influenced by authorities. The data cover the EU (22 countries provide relevant data) in the year 2013.

In particular, the list of inputs that influence the level of happiness includes: Household net adjusted disposable income (US dollars at current PPPs per capita, *income*) directly influences the amount of goods and services that households can afford without reducing their wealth. Long-term unemployment (% of labour force, *unempl*) even if causes more leisure time, it frustrates and deprives people and so reduces their life-satisfaction, while Working very long hours (% of dependent employed, *workhour*). Educational attainment (% of adults holding at least upper secondary degree, *educ*) supports employability which influences consumption. Rule of law (World Bank governance matters index, *rol*) is small the enforcement of human rights is poor which creates insecurity in people's lives. Last, Air pollution (annual concentrations of particulate matters in micrograms per cubic meter, *air*) determines directly the quality of life but also health status.

	Inputs						Outputs	
Country	**Income**	**Unempl**	**Workhour**	**Educ**	**Rol**	**Air**	**EWB**	**HPI**
AT	28.85	1.07	8.76	82	1.80	27	7.35	47.09
BE	26.87	3.45	4.43	70	1.40	21	6.85	37.09
CZ	16.96	2.80	7.58	92	1.00	17	6.15	39.35
DK	24.68	1.85	1.97	76	1.90	16	7.77	36.61
EE	12.80	7.06	4.10	89	1.20	9	5.14	34.95
FI	25.74	1.75	3.89	83	1.90	15	7.39	42.69
FR	28.31	3.83	8.96	71	1.40	12	6.80	46.52
DE	28.80	2.84	5.41	86	1.60	16	6.72	47.20
GR	20.44	8.75	5.23	65	0.40	31	5.84	40.53
HU	13.86	5.36	3.10	81	0.60	15	4.73	37.40
IE	24.10	8.52	3.94	73	1.70	12	7.26	42.40
IT	24.22	4.36	4.07	55	0.40	21	6.35	46.35
LU	35.52	1.41	2.62	78	1.80	13	7.10	28.99
NL	25.49	1.49	0.66	73	1.80	30	7.50	43.09
NO	31.46	0.38	2.83	81	2.00	15	7.63	51.43
PL	15.37	3.05	7.24	89	0.80	34	5.78	42.58
PT	19.37	6.14	8.50	32	1.00	20	4.87	38.68
SK	16.68	8.65	6.38	91	0.40	12	6.05	40.13
SI	19.12	3.61	5.55	83	1.00	26	6.08	40.17
ES	22.85	8.99	6.34	53	1.00	25	6.19	44.06
SE	26.24	1.29	1.23	87	2.00	10	7.50	46.17
UK	26.90	2.62	12.06	75	1.70	13	7.03	47.93

Table 1 Actual state of happiness and its determinants

We compute two versions of DEA with one output, the EWB and HPI, respectively. The EWB (*ewb*) reflects the overall satisfaction with one's live and ranges from 0-10. A weakness can be considered the fact that respondents need not necessarily comprise all dimensions of lives in their responses. The HPI is an index of human well-being and environmental impact that was introduced by NEF. The HPI (*hpi*) measures the extent to which countries enable long, happy, sustainable lives to the people living there [1]. The data analyzed are shown in Table 1. The inputs that should be maximized are consequently multiplied by (-1) to maintain the logics of minimization process.

As was mentioned above we use data envelopment analysis, now we briefly describe this method. In case that a decision making unit (DMU) employs *m* inputs and *r* outputs, the model for *q*th DMU can be formally written as:

$$\min\ z = \theta_q - \varepsilon(\mathbf{1}^{\mathbf{T}}\boldsymbol{s}^{+} + \mathbf{1}^{\mathbf{T}}\boldsymbol{s}^{-}) \tag{1}$$

$$\text{s.t.: } \boldsymbol{X}\boldsymbol{\lambda} + \boldsymbol{s}^{-} = \theta_q \boldsymbol{x}_q,$$

$$\boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{s}^{+} = \boldsymbol{y}_q,$$

$$\mathbf{e}^{\mathbf{T}}\boldsymbol{\lambda} = \mathbf{1},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{s}^{+} \geq \mathbf{0}, \boldsymbol{s}^{-} \geq \mathbf{0}, \tag{2}$$

where n is the number of DMUs, $\boldsymbol{X}=\{x_{ij},\ i=1, \ldots, m;\ j=1, \ldots, n\}$ is matrix of inputs and $\boldsymbol{Y}=\{y_{ij},\ i=1, \ldots, r;\ j=1, \ldots, n\}$ is matrix of outputs, θ_q is relative efficiency of qth DMU, ε is an infinitesimal, $\boldsymbol{s}+$ and $\boldsymbol{s}^{-}$ are input and output slacks, respectively, $\boldsymbol{\lambda}$ is a vector of weights, $\boldsymbol{x}_q$ is the qth column of matrix $\boldsymbol{X}$ and $\boldsymbol{y}_q$ is the qth column of matrix $\boldsymbol{Y}$ representing inputs and outputs of qth DMU, respectively, and $\mathbf{1}^{\mathbf{T}}$ is a row vector with all elements equal to 1, for more details see for example [2, 10].

3 Results and Discussion

In Table 2 and 3 we report results of two versions of DEA. Regarding less comprehensive indicator of life satisfaction – experienced well-being, the most satisfied are Danes, Norwegians, Swedes and Dutch and in their countries relative efficiency can be seen (see Table 2).

Denmark, Estonia, Luxembourg, Netherlands, Norway, Sweden result to be relatively efficient in supporting national well-being. With the exception of Estonia, these countries are typical of high EWB and low values of inputs. Estonia with medium EWB exhibits extremely low air pollution, which is the reason why DEA classifies this country as efficient.

The second group of countries – Ireland and France – reaches at least 80% of the efficiency level of the best performers. To reach the frontier both countries should motivate to education so that additional 15 p. p. of population hold at least upper secondary degree. Moreover, Ireland should also reduce long-term unemployment by 6.65 units and France should reduce very long working hours by 7.04 units. In case of France inhabitants would be satisfied in the same manner even if their income slightly decreased.

Next cluster of countries has efficiency between 0.7 and 0.8. Countries that belong to this group are United Kingdom and Slovakia. To achieve efficiencies the United Kingdom must reduce very long working hours by 10.48 units and enhance educational attainment by 12.24 p. p. Lower income would not affect experienced well-being. Slovakia must improve all indicators except for educational attainment that could be even lower, the most interest should be in improving rule of law and long-term unemployment. Air pollution is the factor which moves Slovakia relatively close towards frontier.

Countries which have relative efficiency between 0.6 and 0.7 are Finland, Germany but also Hungary. All these countries have to improve air pollution by about 5 units. To reach current level of well-being Germans do not need such a high income. In addition to improvements of air quality Hungary should stimulate educational attainment by 6.54 units and output per capita by 8.73 thousands US dollars per year. Nevertheless, even these improvements would not lead to efficiency, there should be incentives from other areas that would increase well-being by 2.13 units (EWB slacks may explain the unexpected inclusion of Hungary into the group with Finland and Germany).

Country	Theta	Income	Unempl	Workhour	Educ	Rol	Air	EWB
AT	0.520	1.60	-0.51	-6.24	0.16	0.20	-12.97	0.26
BE	0.473	-1.43	-1.82	-3.03	17.12	0.55	-11.06	0.50
CZ	0.585	8.48	-1.16	-6.18	-4.88	0.95	-7.06	1.20
DK	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EE	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FI	0.694	0.93	-0.54	-2.53	3.50	0.10	-4.59	0.11
FR	0.813	-5.28	-1.16	-7.04	16.48	0.41	-2.24	0.13
DE	0.620	-3.65	-1.08	-3.95	1.16	0.33	-6.08	0.58
GR	0.318	3.78	-6.59	-3.57	22.30	1.48	-21.15	1.30
HU	0.649	8.73	-2.50	-1.09	6.54	1.18	-5.27	2.13
IE	0.825	0.78	-6.65	-2.42	14.20	0.22	-2.10	0.00
IT	0.470	0.26	-2.31	-2.46	32.26	1.49	-11.13	0.83
LU	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NL	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NO	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PL	0.337	12.38	-2.02	-5.55	-3.74	1.20	-22.55	1.76
PT	0.485	2.94	-3.16	-6.43	55.59	0.77	-10.29	1.93
SK	0.782	1.33	-3.83	-3.39	-2.78	1.11	-2.61	0.00
SI	0.384	6.90	-2.22	-4.27	4.03	0.99	-16.02	1.37
ES	0.385	-1.66	-5.53	-4.03	34.75	0.70	-15.38	0.42
SE	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UK	0.760	-1.44	-0.69	-3.91	-1.73	0.20	-9.73	0.03

Table 2 Recommended changes – version *EWB*

The Czech Republic and Austria has the efficiency between 0.5 and 0.6. Both countries must diminish excessively long working hours for about 6 p. p. of labour force. Air pollution should be reduced in both countries, namely by 13 and 7 units in Austria and the Czech Republic, respectively. The latter also needs to increase income per capita by 8.48 thousands US dollars per year. In contrast, educational attainment could be even lower and yet Czech population would reach the same well-being.

Countries with an efficiency between 0.4 and 0.5 are Portugal, Italy and Belgium. All these states should improve air pollution by about 11 units and educational attainment by 55 p. p. in Portugal, 32 p. p. in Italy and 17 p. p. in Belgium. Belgium can achieve the same well-being even with lower income. Italians should also enhance rule of law. Additionally, Portugal should reduce excessive working hours for 6.43 p. p. of labour force together with adoption of tools enhancing EWB by 1.93 (which stands behind the inclusion of Portugal in the group).

The last group are the least efficient countries, their efficiency is less than 0.4. These are Greece, Poland, Slovenia, and Spain. All these countries could achieve relative efficiencies when reducing air pollution (Greece and Poland by about 22 units, Slovenia and Spain by about 16 units) and very long working hours for 4 p. p. of labour force (in Poland even for 5.55 p. p.). Spain and Greece must also improve long-term unemployment by about 6 p. p. and educational attainment by more than 22 p. p. Poland should also work on raising income per capita by about 12 thousands US dollars per year, but they could even reduce educational attainment by 3.74 p. p. Slovenia should focus more on improving output per head by 6.9 thousands US dollars per year.

DEA for the version with HPI gives very similar results, both in terms of efficiency and input slacks. The only two changes can be seen in Denmark and the United Kingdom. The United Kingdom becomes efficient when analyzing HPI. This is due to the fact that HPI reflects more aspects of happiness which results in better position of this country. In contrast, the position of Denmark worsens in terms of HPI while in the other measure this country is the best performer. From the efficiency point of view Denmark belongs to the cluster with Finland and Germany, and likewise these countries Danes must improve air pollution by about 5 units and motivate to

higher educational level for about 10.35 p. p. of labour force to achieve efficiency. Moreover, they should also improve other areas of HPI not included in this study by 9.82 units.

Country	Theta	Income	Unempl	Workhour	Educ	Rol	Air	HPI
AT	0.520	1.60	-0.51	-6.24	0.16	0.20	-12.97	3.33
BE	0.473	-1.43	-1.82	-3.03	17.12	0.55	-11.06	8.41
CZ	0.585	8.48	-1.16	-6.18	-4.88	0.95	-7.06	6.14
DK	0.669	1.87	-0.61	-0.65	10.35	0.10	-5.30	9.82
EE	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FI	0.694	0.93	-0.54	-2.53	3.50	0.10	-4.59	3.92
FR	0.861	-1.72	-2.60	-7.62	15.60	0.60	-1.67	0.00
DE	0.686	-1.54	-1.73	-3.87	-0.17	0.40	-5.02	0.00
GR	0.318	3.78	-6.59	-3.57	22.30	1.48	-21.15	3.96
HU	0.649	8.73	-2.50	-1.09	6.54	1.18	-5.27	5.72
IE	0.805	-2.38	-5.29	-1.75	14.67	0.03	-2.34	0.00
IT	0.484	2.20	-3.10	-2.79	31.80	1.60	-10.83	0.00
LU	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NL	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NO	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PL	0.337	12.38	-2.02	-5.55	-3.74	1.20	-22.55	5.11
PT	0.485	2.94	-3.16	-6.43	55.59	0.77	-10.29	4.21
SK	0.789	2.33	-4.26	-3.61	-2.92	1.17	-2.54	0.00
SI	0.384	6.90	-2.22	-4.27	4.03	0.99	-16.02	5.81
ES	0.392	0.87	-6.62	-4.57	34.38	0.85	-15.19	0.00
SE	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UK	0.898	0.32	-1.00	-3.37	-3.76	0.20	-8.04	0.00

Table 3 Recommended changes – version HPI

4 Conclusions

Growing interest in more comprehensive approach towards standard of living has been shifting the interest of economists from purely economic also to non-financial aspects of well-being. The goal of this paper is then to assess how the EU countries are efficient in supporting happiness of their citizens. For this purpose DEA is used to identify the most efficient countries and the areas of possible improvements for those that are inefficient.

Countries that need to work on improving long-term unemployment by more than 6 p. p. are Greece and Spain, by more than 5 p. p. is Ireland and more than 4 p. p. is the Slovak Republic. Reduction of air pollution by more than 21 units is necessary in Greece and Poland, by more than 15 units in Slovakia and Spain. The quality of air is not optimal even in Austria that should abate pollution by almost 13 units. The problem with working for very long hours is evident in the United Kingdom where the reduction of 10.3 p. p. is recommended. The UK is followed by France with the desired reduction of about 7.62 p. p.. Also in Austria, the Czech Republic and Portugal this variable is recommended to decrease by more than 6 p. p. Improvements in institutional environment are needed particularly in Italy (by 1.6), in Greece (by 1.48), but also in Poland and Hungary (by about 1.2 units). In terms of economic activity, there are countries that have unnecessarily high incomes, namely Ireland, France, Germany and Belgium. On the other hand, countries which should significantly increase their incomes are Poland (by 12.38), Hungary (by 8.73), the Czech Republic (by 8.48) and Slovenia (by 6.9 US dollars per capita and year). Similarly, as for education there are the countries in which raising average level of education does not help to achieve higher life satisfaction. These are the Czech Republic, Poland, Slovak Republic and also Germany. In contrast, increasing the average education would elevate happiness of citizens in Portugal (by 55.59), Spain (by 34.38), Italy (by 31.8) and Greece (by 22.3 units). Countries that need to improve

well-being or happiness by other than mentioned means are Denmark, Belgium, the Czech Republic, as well as Slovenia, Hungary and Poland.

To sum up, national well-being or happiness efficiency is closely associated with air pollution. We also support actual debate that for satisfaction not only economic aspects are important (derived from more relevant variables for EWB and HPI and also negative values in corresponding slacks). Likewise, raising the level of education may not be the key to higher happiness. The happiest population in terms of the HPI is in Norway, the United Kingdom, Germany, Austria, but only Norway is efficient, while the most satisfied in terms of WB are Denmark, Norway, Netherlands, Sweden, and all are efficient.

Acknowledgements

The support of EUREKA (Project E! 6752 DETECTGAME) is appreciated.

References

[1] Abdallah, S., Michaelson, J., Shah, S., Stoll, L., and Marks, N.: The Happy Planet Index: 2012 Report - A Global Index of Sustainable Well-being, 2012. Available at http://www.happyplanetindex.org/assets/happy-planet-index- report.pdf.

[2] Banker, R., Charnes, A. and Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* **30** (1984), 1078–1092.

[3] Cordero Ferrera, J. M., Javier Salinas-Jiménez, J., and Salinas-Jiménez, M.M.: Assessing the level of happiness across countries: A robust frontier approach. MPRA Paper No. 57784, 2014.

[4] Frey, B. S., and Stutzer, A.: *The Economics of Happiness*. Princeton University Press, New Jersey, 2001.

[5] Graham, C.: The Economics of Happiness - Insights on globalization from a novel approach. *World Economics* **6** (2005), 41–55.

[6] Issever Grochová, L.: Can Economic Growth Be Acceptable if Environmental Issues are Considered? Economic Consequences in the EU Countries. In: *Proceeding of 12th International Conference of Numerical Analysis and Applied Mathematics 2014* (Simos, T. E. eds.). American Institute of Physics, Melville, New York, 2015, 720007-1 – 720007-4.

[7] Issever Grochová, L., and Myšková, K.: Ecology versus Economic Growth in Europe: How to Improve Relative Efficiency? In: *Proceeding of 12th International Conference of Numerical Analysis and Applied Mathematics 2014* (Simos, T. E. eds.). American Institute of Physics, Melville, New York, 2015, 720015-1 – 720015-4.

[8] Kabát, L., Hampel, D., Issever Grochová, L., Janová, J. and Střelec, L.: Alternative approaches for assessing the European countries economic and social results. *Procedia Economics and Finance* **12** (2014), 273-282.

[9] OECD: *How's life? 2013: Measuring well-being.* OECD Publishing, 2013.

[10] Tone, K.: A Slack-Based Measure of Efficiency in Data Envelopment Analysis. *European Journal of Operational Research* **130** (2001), 498–509.

The concept of an enterprise's financial situation assessment fuzzy model

Tomasz Leszek Nawrocki[1]

Abstract. The main aim of the paper is to present the concept and practical application of an enterprise financial situation assessment fuzzy model, as an alternative to dominant in the literature econometric models for predicting bankruptcy of enterprises. The proposed model is an attempt to combine the fuzzy methodology with financial indicators and in contrast to the current methodology in this area, it includes all, not only selected, dimensions of companies liquidity and solvency, i.e. financial liquidity in static, income and structural form, as well as debt level and ability to service the debt. Such a comprehensive approach to enterprise financial situation assessment makes it more resistant to changes of business conditions, and thus should be more objective in comparison with econometric models basing on several, often limited in terms of information, financial ratios. In addition, it should be noted, that the use of fuzzy logic allows for significant flexibility in both – the model-creating step and during its later use. As a result, the proposed solution is relatively easy to adapt to a particular specificity of enterprises and can be further modified.

Keywords: fuzzy logic, financial liquidity, solvency, financial situation assessment.

JEL Classification: C69, G33
AMS Classification: 94D05, 26E50

1 Introduction

Reliable information on the financial situation of companies is an object of interest to many market players, in particular, suppliers, creditors, customers, and finally the owners themselves. Due to the significant effort of full and detailed company's financial condition analysis over decades there has been much effort into developing different methods to make a quick and relatively certain diagnosis concerning company's financial situation, based at the same time on the smallest possible number of parameters. Although the first quantitative assessment models of creditworthiness date back to the 1930s (P.J. Fritz Patrick), a significant development acceleration in this field occurred in 1960s, along with the wider use of statistics and econometrics (W.H. Beaver, E.I. Altman), especially discriminant analysis [2]. Now, with the increasingly common use of computers and perceiving certain drawbacks and limitations of statistical and econometric methods [3], [6], it can be seen further develop broadly defined methods of enterprises financial condition assessment, which focus on the neural networks or fuzzy modeling use [3]. Therefore, the main purpose of this article is to present a comprehensive conception and practical application of the financial situation assessment model based on fuzzy set theory.

2 Research methodology

In contrast to the approach used in the statistical and econometric methods [6] assessment criteria selection in proposed solution was carried out arbitrarily, on the basis of author's knowledge and practical experience in the field of financial analysis. Efforts were made at the same time to take into account, with a limited number of criteria, all the dimensions of enterprises financial condition assessment, that are crucial from the viewpoint of the bankruptcy risk, or creditworthiness.

Therefore, in the proposed model it was assumed that the financial situation of an enterprises can be analysed and assessed in two basic dimensions considered in the literature on financial analysis [5]: its financial liquidity and indebtedness. In the assessment of financial liquidity there were three basic dimensions included: static, income (cash flows) and structural (net working capital), and in the case of indebtedness two: debt level and ability to service the debt.

The structure of the suggested enterprise's financial situation assessment model, along with the most detailed assessment criteria within the particular modules, is presented in Figure 1.

[1] Silesian University of Technology, Faculty of Organization and Management, Institute of Economics and Informatics, Roosevelta 26-28, 41-800 Zabrze, Poland, e-mail: tomasz.nawrocki@polsl.pl

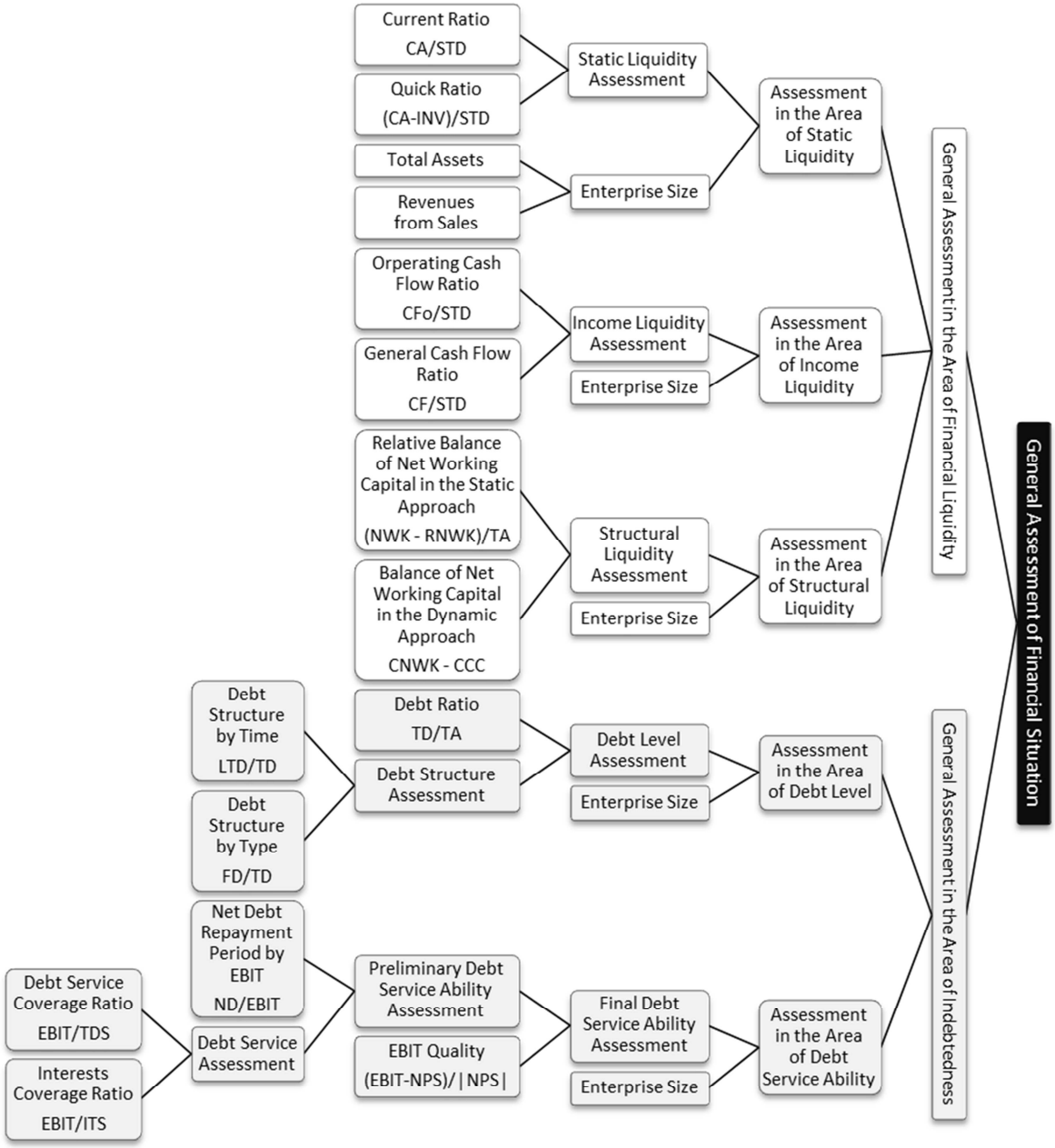


Figure 1 General structure of enterprise's financial situation assessment model

CA – current assets, STD – short-term debt, INV – inventories, CFo – cash flows from operations, CF – general cash flow (net increase in cash), NWK – net working capital, RWK – request for net working capital (non-financial current assets – non-financial short-term liabilities), CWKn – net working capital cycle, CCC – cash conversion cycle, TD – total debt (liabilities and reserves), TA – total assets, LTD – long-term debt, FD – financial debt, ND – net debt (FD – cash), TDS – total debt service (principal, interests and financial leasing payments), ITS – interests, NPS – net profit from sales.

In the proposed model firstly it is intended to obtain partial assessments within the distinguished basic assessment criteria of financial situation. These assessments will result from the ratios calculated on the basis of data from financial statements. Next aggregated assessment results may be obtained in the areas of financial liquidity in static, income and structural dimension, debt level and ability to service the debt. Furthermore, these results constitute foundations for calculating general situation measures in the areas of financial liquidity and indebtedness, so that in the final stage, on their basis, it is possible to achieve overall financial situation assessment for the analysed enterprise.

The calculation tool in the suggested solution is based on the fuzzy set theory, which is one of the approximate reasoning methods [7], [4].

Detailed assumptions of fuzzy model construction

In relation to the construction of proposed fuzzy model, based on the Mamdani approach [4], the following assumptions were made:

- for all input variables of the model, the same dictionary of linguistic values was used, and their value space was divided into three fuzzy sets named {low, medium, high};
- for output variables of the model, in order to obtain more accurate intermediate assessments, the space of linguistic values was divided into five fuzzy sets named {low, mid-low, medium, mid-high, high};
- in case of all membership functions to the particular fuzzy sets, a triangular shape was decided for them (Figure 2 and Figure 3);
- the values of fuzzy sets characteristic points (x_1, x_2, x_3) for the particular input variables of the model were determined partly basing on the literature on enterprises financial analysis and partly arbitrarily, basing on the distribution of analysed variables values and on the author years of experience in the area of companies financial situation analysis (Table 1);
- for fuzzification of input variables, the method of *simple linear interpolation* was used [1];
- fuzzy reasoning in the particular knowledge bases of the model was conducted using *PROD* operator (fuzzy implication) and *SUM* operator (final accumulation of the conclusion functions received within the particular rule bases into one output set for each base) [4];
- for defuzzification of fuzzy reasoning results within the particular rule bases simplified *Center of Sums* method was used [4].

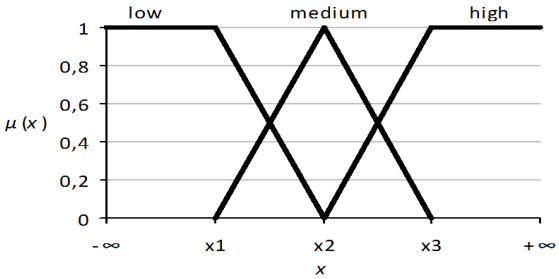


Figure 2 The general form of input variables membership function to distinguished fuzzy sets.

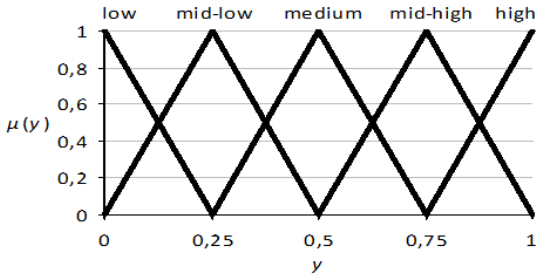


Figure 3 The output variables membership function to distinguished fuzzy sets.

	x1 $\mu(x)$ = 1/low	x2 $\mu(x)$ = 1/medium	x3 $\mu(x)$ = 1/high
Current Ratio (-)	0	1	2
Quick Ratio (-)	0	0,6	1,2
Operating Cash Flow Ratio (-)	0	0,5	1
General Cash Flow Ratio (-)	0	0,5	1
Relative Balance of Net Working Capital in the Static Approach (-)	-0,05	0,025	0,1
Balance of Net Working Capital in the Dynamic Approach (days)	-15	7,5	30
Debt Structure by Time (-)	0	0,5	1
Debt Structure by Type (-)	0	0,25	0,5
Debt Ratio (-)	0	0,35	0,7
Debt Service Coverage Ratio (-)	0	1	2
Interests Coverage Ratio(-)	0	2	4
Net Debt Repayment Period by EBIT (years)	0	3	6
EBIT Quality (-)	0	1	2
Total Assets (‘000 EUR)	0	172.000	344.000
Revenues from Sales (‘000 EUR)	0	200.000	400.000

Table 1 The values of fuzzy sets characteristic points for particular input variables of the enterprise’s financial situation assessment fuzzy model

Next, taking into consideration the general structure of the financial situation assessment model presented in Figure 1, author, basing on his knowledge and experience in the area of analysed issue, designed 17 rules bases in the form of „IF – THEN” (16 bases with 9 rules and 1 base with 27 rules), achieving this way a „ready to use” form of the financial situation assessment fuzzy model. Due to large volume of all knowledge (rules) bases included in the proposed model, below are presented only the last three of them.

Rules base for assessment in the area of indebtedness – *AI* (*ADL* – assessment in the area of debt level, *ADSA* – assessment in the area of debt service ability):

R1: IF *ADL* is low AND *ADSA* is low THEN *AI* is low
R2: IF *ADL* is medium AND *ADSA* is low THEN *AI* is low
R3: IF *ADL* is low AND *ADSA* is medium THEN *AI* is mid-low
R4: IF *ADL* is high AND *ADSA* is low THEN *AI* is mid-low
R5: IF *ADL* is medium AND *ADSA* is medium THEN *AI* is medium
R6: IF *ADL* is low AND *ADSA* is high THEN *AI* is mid-high
R7: IF *ADL* is medium AND *ADSA* is high THEN *AI* is mid-high
R8: IF *ADL* is high AND *ADSA* is medium THEN *AI* is mid-high
R9: IF *ADL* is high AND *ADSA* is high THEN *AI* is high

Rules base for assessment in the area of financial liquidity – *AFL* (*AStaL* – Assessment in the Area of Static Liquidity, *AIncL* – Assessment in the Area of Income Liquidity, *AStrL* – Assessment in the Area of Structural Liquidity):

R1: IF *AStaL* is low AND *AIncL* is low AND *AStrL* is low THEN *AFL* is low
R2: IF *AStaL* is medium AND *AIncL* is low AND *AStrL* is low THEN *AFL* is low
R3: IF *AStaL* is low AND *AIncL* is low AND *AStrL* is medium THEN *AFL* is mid-low
R4: IF *AStaL* is low AND *AIncL* is medium AND *AStrL* is low THEN *AFL* is mid-low
R5: IF *AStaL* is medium AND *AIncL* is medium AND *AStrL* is low THEN *AFL* is mid-low
R6: IF *AStaL* is medium AND *AIncL* is low AND *AStrL* is medium THEN *AFL* is mid-low
R7: IF *AStaL* is low AND *AIncL* is high AND *AStrL* is low THEN *AFL* is mid -low
R8: IF *AStaL* is high AND *AIncL* is low AND *AStrL* is low THEN *AFL* is mid -low
R9: IF *AStaL* is medium AND *AIncL* is medium AND *AStrL* is medium THEN *AFL* is medium
R10: IF *AStaL* is low AND *AIncL* is low AND *AStrL* is high THEN *AFL* is medium
R11: IF *AStaL* is low AND *AIncL* is medium AND *AStrL* is medium THEN *AFL* is medium
R12: IF *AStaL* is high AND *AIncL* is high AND *AStrL* is low THEN *AFL* is medium
R13: IF *AStaL* is medium AND *AIncL* is low AND *AStrL* is high THEN *AFL* is medium
R14: IF *AStaL* is medium AND *AIncL* is high AND *AStrL* is low THEN *AFL* is medium
R15: IF *AStaL* is high AND *AIncL* is medium AND *AStrL* is low THEN *AFL* is medium
R16: IF *AStaL* is high AND *AIncL* is low AND *AStrL* is medium THEN *AFL* is medium
R17: IF *AStaL* is high AND *AIncL* is low AND *AStrL* is high THEN *AFL* is mid -high
R18: IF *AStaL* is medium AND *AIncL* is medium AND *AStrL* is high THEN *AFL* is mid-high
R19: IF *AStaL* is medium AND *AIncL* is high AND *AStrL* is medium THEN *AFL* is mid-high
R20: IF *AStaL* is high AND *AIncL* is medium AND *AStrL* is medium THEN *AFL* is mid-high
R21: IF *AStaL* is low AND *AIncL* is medium AND *AStrL* is high THEN *AFL* is mid-high
R22: IF *AStaL* is low AND *AIncL* is high AND *AStrL* is medium THEN *AFL* is mid-high
R23: IF *AStaL* is low AND *AIncL* is high AND *AStrL* is high THEN *AFL* is high
R24: IF *AStaL* is medium AND *AIncL* is high AND *AStrL* is high THEN *AFL* is high
R25: IF *AStaL* is high AND *AIncL* is high AND *AStrL* is medium THEN *AFL* is high
R26: IF *AStaL* is high AND *AIncL* is medium AND *AStrL* is high THEN *AFL* is high
R27: IF *AStaL* is high AND *AIncL* is high AND *AStrL* is high THEN *AFL* is high

Rules base for general assessment of financial situation – *AFS*:

R1: IF *AFL* is low AND *AI* is low THEN *AFS* is low
R2: IF *AFL* is low AND *AI* is medium THEN *AFS* is low
R3: IF *AFL* is medium AND *AI* is low THEN *AFS* is low
R4: IF *AFL* is low AND *AI* is high THEN *AFS* is mid-low
R5: IF *AFL* is medium AND *AI* is medium THEN *AFS* is medium
R6: IF *AFL* is high AND *AI* is low THEN *AFS* is medium
R7: IF *AFL* is medium AND *AI* is high THEN *AFS* is mid-high
R8: IF *AFL* is high AND *AI* is medium THEN *AFS* is mid-high
R9: IF *AFL* is high AND *AI* is high THEN *AFS* is high

The intermediate and final assessments generated by the model take values in the range between 0 and 1, where from the viewpoint of analysed issue, values closer to 1 mean a very favourable result (better financial situation, lower risk of financial problems and bankruptcy), while values closer to 0 indicate a result less favourable (worse financial situation, higher risk of financial problems and bankruptcy).

All calculations related to the presented fuzzy model were based on self-developed structure of formulas in MS Excel.

3 Research results

In order to verify the proposed model the financial situation assessment was conducted for three enterprises from the retail industry involved in managing clothing brands, which shares are listed on the Warsaw Stock Exchange – LPP, Monnari Trade (Monnari) and Redan. According to the adopted methodology, the basis for the financial situation assessment of mentioned above entities were data acquired from the annual reports published by these companies in the years 2007-2015. The results obtained during the research are presented in Figures 4-8.

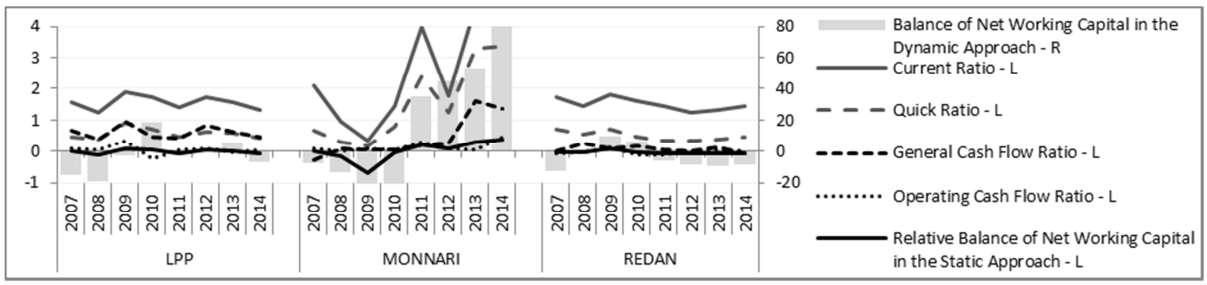


Figure 4 Financial liquidity ratios in analyzed companies

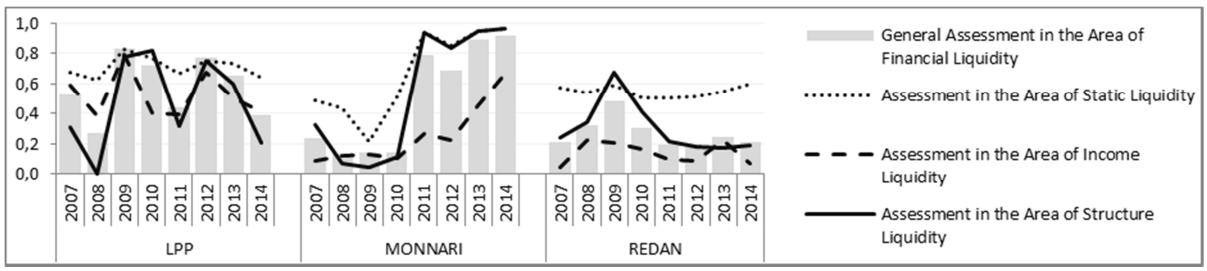


Figure 5 Financial liquidity assessment of analyzed companies

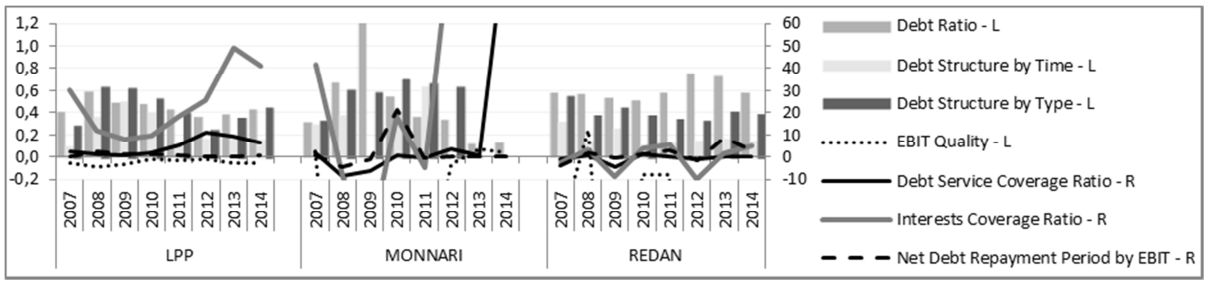


Figure 6 Indebtedness ratios in analyzed companies

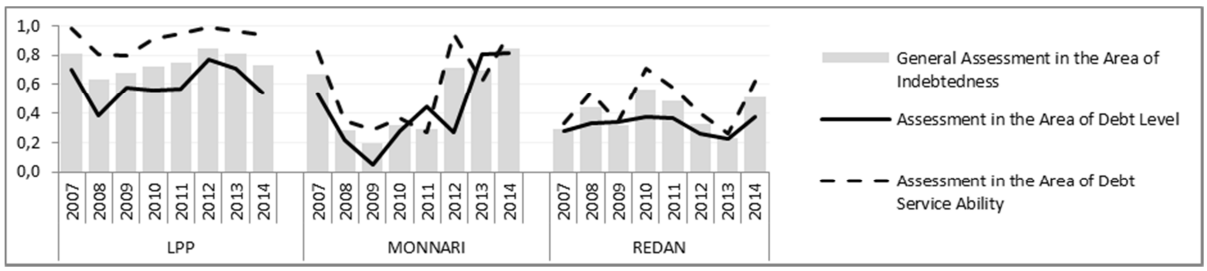


Figure 7 Indebtedness assessment of analyzed companies

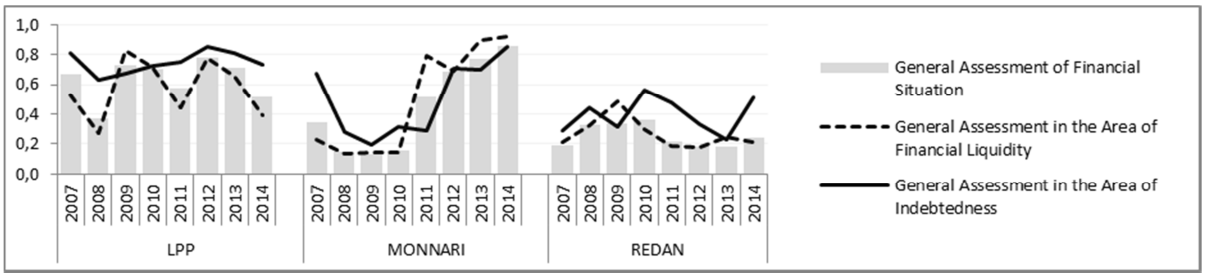


Figure 8 Financial situation assessment of analyzed companies

Taking into account carried out research it should be noted, that financial situation each of the three considered companies developed over analyzed period differently. In case of the LPP it can be seen generally maintaining (with some variations in plus and in minus) assessment at average level in the range between 0.5-0.7. Also relatively stable, although much lower than in case of LPP and pointing to some financial problems, maintained Redan assessment – interval 0.2-0.3. The biggest financial situation changes in the considered period occurred in Monnari, where initially in the years 2007-2010 this company was facing serious financial problems (score below 0.2), and then in subsequent years, due to restructuring, significantly improved its financial standing (in the end of 2014 an overall assessment of financial situation was at level above 0.8).

In the context of fuzzy logic use to analyze enterprises financial situation it should be noted, that thereby assessments in individual thematic areas with different levels of aggregation were obtained, which on the one hand are consistent with the interpretation of analyzed companies situation on the basis of partial liquidity and indebtedness criteria, and on the other hand gave the possibility of better discernment in comparison with the processing of each criterion separately. It should be also noted, that proposed model has a much greater completeness of considered issues coverage in comparison with statistical models, making its results more objective.

4 Conclusions

The model of enterprise's financial situation assessment, developed using the methodology of fuzzy sets, is characterized by the following advantages, that partially enable reducing the defects of methods previously applied: (i) the use of the full range, and not just selected, financial indicators to provide greater objectivity of final results and less susceptible to falsify them on grounds of industry; (ii) combination of analytic and synthetic assessment, achieved through the use of sub-criteria, that are later aggregated into a more general financial situation assessments; (iii) combination of quantitative and qualitative approach to financial situation assessment, manifested in a qualitative dimension of ratios used for its measurement and qualitative method of its diversification determination; (iv) using for the assessment process data included in generally accessible business periodical reports.

The concept of an enterprise's financial situation assessment fuzzy model presented in this article has preliminary character, that during further research will be subject to certain modifications. They will cover in particular the issue of input variables value space division, so that the final results were less vulnerable to issues of analyzed companies industry membership, and thus the possibility of model application become more universal compared to alternative to him statistical models.

Acknowledgements

This research is financed from BK-221/ROZ1/2015.

References

[1] Bartkiewicz, W.: *Zbiory rozmyte* [in:] *Inteligentne systemy w zarządzaniu*, Zieliński, J.S., (Ed.), PWN, Warszawa, 2000, p. 72-140.
[2] Jagiełło, R.: *Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw*, Materiały i Studia NBP, nr 286/2013. Available from: http://www.nbp.pl/publikacje/materialy_i_studia/ms286.pdf.
[3] Korol, T.: *Fuzzy Logic in Financial Management* [in:] *Fuzzy Logic – Emerging Technologies and Applications*, Dadios E. (Ed.), InTech, p. 259-286. Available from: http://www.intechopen.com/books/fuzzy-logic-emerging-technologies-and-applications/fuzzy-logic-in-financial-management.
[4] Piegat, A.: *Modelowanie i sterowanie rozmyte*. EXIT, Warszawa, 2003.
[5] Wędzki, D.: *Analiza wskaźnikowa sprawozdania finansowego* t.2. Wolters Kluwer, Kraków, 2009.
[6] Wiatr, M.S.: *Zarządzanie indywidualnym ryzykiem kredytowym*, SGH, Warszawa 2011.
[7] Zadeh, L.A.: Fuzzy sets, *Information and Control* 8 (1965), p. 338-353.

Parametric asymptotic portfolio decisions

S. Ortobelli[1], T. Tichy[2], T. Lando[3], F. Petronio[4]

Abstract. In this paper, we deal with portfolio selection decisions considering the asymptotic behavior of the portfolio returns. Thus, using the stable Paretian portfolio parameterization, we examine some dominance rules to determine the optimal choices of non-satiable risk averse investors. In particular, we first preselect a subclass of assets which are not dominated by the point of view of non-satiable and risk-averse investors. Secondly we optimize a multi-parametric portfolio optimization problem that takes into account the asymptotic stochastic dominance rule. Finally, we compare the ex-post wealth obtained by optimal portfolios with different levels of asymptotic skewness and stability index.

Keywords: Portfolio optimization, asymptotic stochastic dominance rules, large-scale portfolio problem.

JEL Classification: C16, C44, G11
AMS Classification: 60E15, 91GXX, 91G10

1 Introduction

In financial literature, the Gaussian distribution has been largely used, due to its important role in statistical inference and to the Central Limit Theorem for random variable with finite variance. However it is well known that asset returns are not normally distributed, as several studies by Mandelbrot (see [6], [7], [8], [9]) and Fama (see [1], [2], [3]) recognized an excess of kurtosis and non-zero skewness in the empirical distributions of financial assets, which often lead to the rejection of the assumption of normality and proposition of the stable Paretian distribution as an alternative model for asset returns. The Fama and Mandelbrot's conjecture was supported by numerous empirical investigations (see [14] and [15]).

In view of the several financial applications of the stable Paretian distribution, the aim of this paper is to (stochastically) order stable distributed random variables. Stochastic dominance rules quantifies the concept of one random variable being "preferable" to another, by establishing a partial order in the space of distribution functions. For instance, in a financial context, stochastic orderings are used to establish an order of preferences for investors whose utility functions share certain characteristics [17]. Indeed, it is well known that stochastic dominance rules are generally aimed at addressing investors and institutions towards the best choices in terms of expected gain and risk (see, among others, [5], [12], [13]). It is well known in ordering literature that we can obtain the second order stochastic dominance between stable distributions by a mean-dispersion comparison (similar to the Gaussian case), but only when the stable distributions present the same skewness parameter and index of stability ([11], [14]). In this paper, we recall some recent general results [10], by considering two fundamental aspects of the stable Paretian distributions, namely: the tail behavior and the asymmetry. Thus, using these asymptotic stochastic dominance rules, we propose to optimize the portfolio by the point of view of non-satiable risk averse investors in the US stock market. In this framework, we have to solve two computational problems. First, we need sufficiently robust approximations of the stable Paretian parameters, in a reasonable computational time; secondly, since we deal with large scale portfolio problems, we have to reduce their dimensionality. Moreover, since these portfolio optimization problems admit different local optima, we have to use a heuristic argument to approximate the optimal choices. Finally, we compare the ex-post final wealth obtained for different levels of skewness and index of stability.

In Section 2, we discuss the use of stable Paretian distributions and their orderings. In Section 3, we examine the multi-parametric portfolio selection problem for different non satiable risk averse investors. Finally, in section 4 we briefly summarize the paper results.

[1] University of Bergamo, via dei Caniana, 2, Bergamo, Italy; and VŠB -TU Ostrava, Sokolská třída 33, Ostrava, Czech Republic; e-mail: sergio.ortobelli@unibg.it.
[2] VŠB - Technical University of Ostrava, Sokolská třída 33, Ostrava, Czech Republic; e-mail: tomas.tichy@vsb.cz.
[3] University of Bergamo, Via dei Caniana, 2, Bergamo, Italy; and VŠB -TU Ostrava, Sokolská třída 33, Ostrava, Czech Republic; e-mail: tommaso.lando@unibg.it.
[4] VŠB-Technical University Ostrava, Sokolská třída 33, Ostrava, Czech Republic;
e-mail: filomena.petronio@vsb.cz.

2 Asymptotic stochastic dominance rules

In this section, we first recall the definitions of some classical stochastic orders and their possible applications to non-satiable risk averse investors.

Recall that any non-satiable investor prefers a portfolio *X* to another one *Y* if and only if *X* dominates *Y* with respect to the first stochastic dominance order (in symbols *X FSD Y*) or, equivalently *X FSD Y* if and only if $\mathrm{E}(g(X)) \geq \mathrm{E}(g(Y))$ for any non-decreasing function g. Similarly, we recall that any non-satiable risk averse investor prefers a portfolio *X* to another one *Y* if and only if *X* dominates *Y* with respect to the second stochastic dominance order (in symbols *X SSD Y*) or, equivalently *X SSD Y* if and only if $\mathrm{E}(g(X)) \geq \mathrm{E}(g(Y))$ for any non-decreasing and concave function g. Finally, we say that *X* dominates *Y* with respect to the Rothschild-Stiglitz order (in symbols *X RS Y*) when *X SSD Y* and $\mathrm{E}(X) = \mathrm{E}(Y)$ (or $Y \geq_{icx} X$ and $\mathrm{E}(X) = \mathrm{E}(Y)$) or equivalently *X RS Y* if and only if $\mathrm{E}(g(X)) \geq \mathrm{E}(g(Y))$ for any concave function g. For more details about the use of non-decreasing function in this framework see [4].

In order to deal with ordering of stable Paretian distributions, let us briefly summarize some important characteristics of the stable distribution. As discussed in the introduction, the stable Paretian law is especially appropriate for approximating the distribution of a random variable X whose tails are significantly heavier than the Gaussian law, that is, for large x

$$P(|X| > x) \approx x^{-\alpha} L(x), \tag{1}$$

where $0 < \alpha < 2$ and $L(x)$ is a slowly varying function at infinity. This tail condition implies that the random variable X is in the domain of attraction of a stable law. Thus, assume X be a random variable in the domain of attraction of a stable law. Let $\{X_i\}_{i\in\mathbb{N}}$ be independent and identically distributed observations of X. Under this hypothesis, we know that there exists a sequence of positive real values $\{d_{X,i}\}_{i\in\mathbb{N}}$ and a sequence of real values $\{a_{X,i}\}_{i\in\mathbb{N}}$, such that, as $n\to+\infty$:

$$\frac{1}{d_{X,n}} \sum_{i=1}^{n} X_i + a_{X,n} \xrightarrow{d} X', \tag{2}$$

where $X' \sim S_{\alpha_X}(\sigma_X, \beta_X, m_X)$ is an α-stable Paretian random variable, where $0 < \alpha_X \leq 2$ is the so-called stability index, which specifies the asymptotic behavior of the tails, $\sigma_X > 0$ is the dispersion parameter, $\beta_X \in [-1,1]$ is the skewness parameter and $m_X \in \mathbb{R}$ is the location parameter. This convergence result is a consequence of the Stable Central Limit Theorem (SCLT) for normalized sums of i.i.d. random variables (see [14], [16]) and it is the main justification for the use of stable distribution in many areas of study, such as finance and econometrics. In particular, the SCLT makes it possible to characterize the skewness and kurtosis of a wide range of phenomena in a statistically proper way. We recall that, if $X \sim S_\alpha(\sigma, \beta, \mu)$, and $\alpha < 2$, then $E(|X|^p) < \infty$ for any $p < \alpha$ and $E(|X|^p) = \infty$ for any $p \geq \alpha$. Therefore, stable distributions do not generally have finite variance, which happens only when $\alpha = 2$ (i.e. Gaussian distribution, $E(|X|^p) < \infty$ for any p). Unfortunately, except in few cases, we do not have a closed form expression for the density of stable Paretian distribution, which is identified by its characteristic function, given by:

$$E(\exp\{itX\}) = \begin{cases} \exp\left\{it\mu - |t\sigma|^\alpha \left(1 - i\beta \operatorname{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right)\right\} & \alpha \neq 1 \\ \exp\{it(\mu + 2\beta\sigma \ln(\sigma)/\pi) & \\ -|t\sigma|(1 + 2i\beta \ln|t\sigma| \operatorname{sign}(t)/\pi)\} & \alpha = 1 \end{cases} \tag{3}$$

It is worth noting that, since the density and distribution functions of the stable Paretian distribution cannot be expressed with elementary functions, it is not possible to verify the integral conditions for the RS and SSD orders.

In particular, from [10], we know that even if RS order is not verifiable, it is possible to analyze a weaker order of "risk" by studying the distributions of the absolute values. Let Z be a random variable and define:

$$\varphi_Z(p) = \operatorname{sign}(p) E(|Z|^p),\ p \in \mathbb{R}. \tag{4}$$

We are now able to define a new stochastic order of risk, expressed in terms of the absolute centered moments of order p.

Definition 1 *Let X and Y be random variables belonging to* $L^q = \{X |\ E(|X|^q) < \infty\ \}$, *for* $q \geq 1$. *We say that X dominates Y with respect to the moment dispersion (MD) order* (*in symbols* $X \geq_{md} Y$) *if and only if*

$$\varphi_{X-E(X)}(p) \leq \varphi_{Y-E(Y)}(p),\ \forall p \geq 1.$$

Observe that in [10] is proved that if X RS Y, then $X \geq_{md} Y$. Therefore (MD) order is consistent with the choices of risk averse investors. Moreover given two stable distributed random variables $X_1 \sim S_{\alpha_1}(\sigma_1, \beta_1, \mu_1)$ and $X_2 \sim S_{\alpha_2}(\sigma_2, \beta_2, \mu_2)$ then if $\alpha_1 \geq \alpha_2 > 1$, $|\beta_1| < |\beta_2|, \sigma_1 \leq \sigma_2$, and $\hat{\mu}_1 \geq \hat{\mu}_2$ then there exists a stable random

variable X_3 with parameters $\alpha_2, \sigma_2, \beta_1, \mu_2$ such that $X_1 SSD\ X_3 \geq_{md} X_2$. Thus, if we maximize the ratio between the location parameter and the scalar parameter as suggested by [11] and [14], varying properly different level of beta and alpha we are able to approximate the optimal choices of different non satiable risk averse investors.

3 An empirical ex post comparison among portfolio strategies

In this section, we examine different portfolio strategies based on the stochastic dominance rule discuss in Section 2. We use the daily observations of the S&P 500 components from January 1, 1999 to July19, 2014. We assume that no short sales are allowed. We point out with $x = [x_1, \dots, x_{500}]'$ the vector of percentages invested in each asset and with $r = [r_1, \dots, r_{500}]'$ the vector of returns. Therefore, starting from 1 January 2000 and using 1 year of historical observations, we optimize monthly a portfolio selection rule that takes into account the asymptotic behavior of the series. Since this large scale portfolio problem requires the estimation and optimization of the stable parameters $\alpha_{x'r}, \sigma_{x'r}, \beta_{x'r}, \mu_{x'r}$ of each portfolio *x'r*, we clearly need to reduce the dimensionality of the problem. Thus we preselect the first 100 assets with the highest ratio between the stable location parameter and the scalar parameter. On this preselected assets we optimize the choice for non satiable risk averse investors. Finally, we examine the ex post wealth obtained with this procedure. In particular, we optimize the following portfolio problem

$$\begin{gathered} max_x \frac{\mu_{x,r}}{\sigma_{x,r}} \\ s.t.\ \ \alpha_{x,r} > \alpha^*; \sum_{i=1}^{n} x_i = 1 \\ |\beta_{x,r}| \leq \beta^*;\ \ x_i \geq 0 \end{gathered} \tag{5}$$

Where we used $\alpha^* = 1.2, 1.35, 1.5, 1.7, 1.8$ and $\beta^* = 0.9, 0.7, 0.5, 0.3, 0.01$ for a total of 25 strategies.

The results are reported in Tables 1, 2, and Figures 1, 2. Table 1 reports three risk measures (the standard deviation, the conditional value at risk $CVaR_p(x'r)$ at p=1%, 5%, where $CVaR_p(X) = \frac{-1}{p}\int_0^p F_X^{-1}(u)du$) of the ex-post returns obtained for the 25 strategies. While Table 2 reports three performance measures (the ex-post final wealth, the Sharpe ratio – i.e. (mean excess return)/(standard deviation)- and the STARR ratio 1% -i.e. (mean excess return)/($CVaR_{1\%}(x'r)$)) of the ex-post returns obtained for the 25 strategies. Figures 1 and 2 show the ex post wealth evolution for the strategies with skewness bounds respectively equal to 0.01 and 0.9 (i.e. $\beta^* = 0.01$, 0.9). It appears evident from Table 1 that the risk of the strategy decreases (except in few cases) when the skewness parameter constraints is decreasing and the index of stability is increasing.

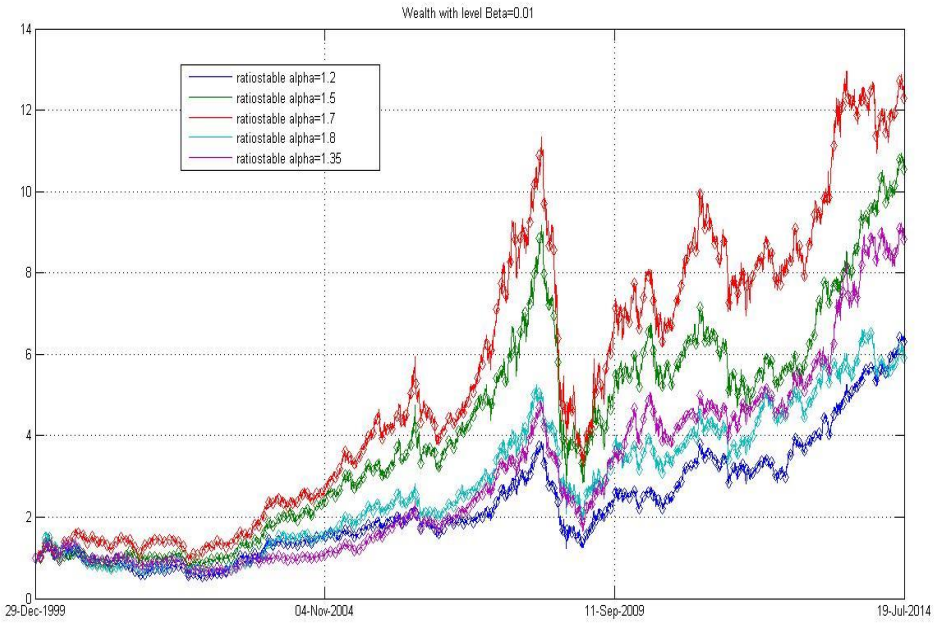


Figure 1: Ex post wealth obtained solving monthly problem (5) for $\beta^* = 0.01$.

Standard Dev		Beta* 0.9	0.7	0.5	0.3	0.01
	1.2	0,0244	0,0238	0,0230	0,0224	0,0208
	1.35	0,0241	0,0243	0,0239	0,0232	0,0206
Alpha*	**1.5**	0,0240	0,0236	0,0224	0,0221	0,0211
	1.7	0,0215	0,0217	0,0206	0,0216	0,0204
	1.8	0,0211	0,0211	0,0214	0,0226	0,0215
$CVaR_{5\%}$		**Beta* 0.9**	**0.7**	**0.5**	**0.3**	**0.01**
	1.2	0,058143	0,05541	0,055155	0,052179	0,049439
	1.35	0,05695	0,057372	0,057762	0,056064	0,048745
Alpha*	**1.5**	0,055919	0,057109	0,051988	0,052665	0,050221
	1.7	0,050627	0,051465	0,049121	0,05122	0,047954
	1.8	0,051765	0,050891	0,05048	0,053219	0,050427
$CVaR_{1\%}$		**Beta* 0.9**	**0.7**	**0.5**	**0.3**	**0.01**
	1.2	0,093583	0,090162	0,088361	0,082761	0,074736
	1.35	0,090851	0,095369	0,095655	0,091553	0,074297
Alpha*	**1.5**	0,091336	0,096127	0,081373	0,085744	0,079054
	1.7	0,082296	0,081849	0,077993	0,081596	0,078773
	1.8	0,08072	0,080082	0,082058	0,087038	0,077444

Table 1: Standard deviation, $CVaR_{1\%}$, and $CVaR_{5\%}$ of the ex-post returns obtained for the 25 strategies

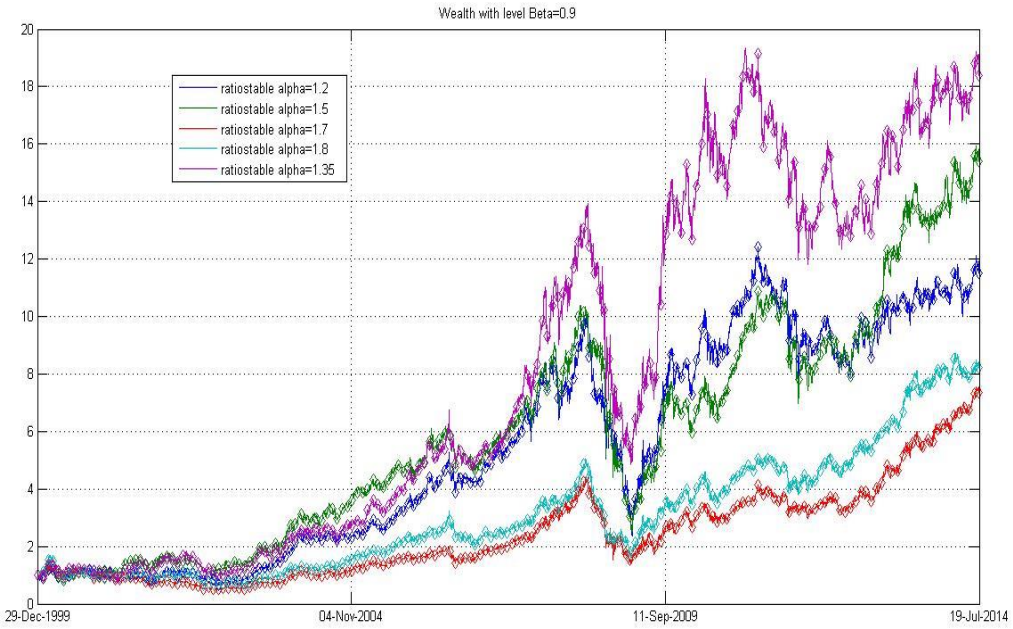


Figure 2: Ex post wealth obtained solving monthly problem (5) for $\beta^* = 0.9$.

While from Table 2 we deduce that there is a strategy (among the 25) corresponding to constraints $\beta^* = 0.5$. and $\alpha^* = 1.5$ of problem (5) that presents the best ex-post performance. Moreover we do not observe a trend in the performance with respect the index of stability and the skewness parameter. Figures 1 and 2 confirm the high

variability of the strategies during the 15 years of ex-post observations, in particular during the sub-prime crisis period (2008-2009) and the country credit risk period 2011-2012.

				Beta*		
Final Wealth		**0.9**	**0.7**	**0.5**	**0.3**	**0.01**
	1.2	11,50443	15,81598	15,20934	10,2482	6,307824
	1.35	18,41026	11,27399	6,140071	3,943075	8,821332
Alpha*	**1.5**	15,39988	12,62042	45,42906	6,889942	10,5512
	1.7	7,339395	5,546076	2,711941	9,49929	12,28207
	1.8	8,228558	5,683344	8,090717	17,99372	5,903721
				Beta*		
Sharpe ratio		**0.9**	**0.7**	**0.5**	**0.3**	**0.01**
	1.2	0,027357	0,031679	0,032361	0,028378	0,024264
	1.35	0,033099	0,027336	0,020784	0,016209	0,03212
Alpha*	**1.5**	0,031192	0,029391	0,046528	0,023858	0,032878
	1.7	0,025382	0,021627	0,013276	0,028529	0,022947
	1.8	0,027279	0,02254	0,02669	0,034986	0,022152
				Beta*		
$STARR_{1\%}$		**0.9**	**0.7**	**0.5**	**0.3**	**0.01**
	1.2	0,007143	0,00838	0,00843	0,007696	0,006746
	1.35	0,008773	0,006952	0,005194	0,004104	0,008923
Alpha*	**1.5**	0,008236	0,007275	0,010915	0,005771	0,009338
	1.7	0,005974	0,004879	0,003359	0,007186	0,005932
	1.8	0,007147	0,00594	0,006974	0,009086	0,006142

Table 2: Ex-post final wealth, Sharpe ratio and $STARR_{1\%}$ ratio of the ex-post returns obtained for the 25 strategies.

4 Concluding remarks

This paper examines the ex-post wealth of some portfolio strategies based on the optimization of an asymptotic stochastic dominance rule. In particular, we observe that the asymptotic skewness and tail behavior could have a crucial effect on the choices and the future ex-post wealth. Moreover, we also observe that the risk of the ex-post wealth generally confirms the asymptotic valuation of the ex-ante risk. This simple observation suggests that the asymptotic behavior should be considered in portfolio strategies at least for the risk valuation.

Acknowledgements

This work was supported by the Italian funds ex MURST 60% 2014 and MIUR PRIN MISURA Project, 2013–2015, and ITALY project (Italian Talented Young researchers). The research was also supported through the Czech Science Foundation (GACR) under project 15-23699S and through SP2015/15, an SGS research project of VSB-TU Ostrava, and furthermore by the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296.

References

[1] Fama, E.: Mandelbrot and the stable Paretian hypothesis, *Journal of Business* **36**, (1963), 420-429
[2] Fama, E.: The behavior of stock market prices, *Journal of Business*, **38** (1965a), 34-105.
[3] Fama, E.: Portfolio analysis in a stable Paretian market, *Management Science*, **11** (1965b), 404-419.
[4] Kopa, M., Post, T.: A general test for SSD portfolio efficiency, OR Spectrum, **37**(3) (2015), 703-734.
[5] Lando, T., and Bertoli-Barsotti, L.: Statistical Functionals Consistent with a Weak Relative Majorization Ordering: Applications to the Mimimum Divergence Estimation, *WSEAS Transactions on Mathematics* **13** (2014) 666-675.
[6] Mandelbrot, B.: New methods in statistical economics, *Journal of Political Economy* **71** (1963a), 421-440.
[7] Mandelbrot, B.: The variation of certain speculative prices, *Journal of Business* **26** (1963b), 394-419.
[8] Mandelbrot, B.: The variation of some other speculative prices, *Journal of Business* **40** (1967a), 393-413.

[9] Mandelbrot, B., and Taylor, M.: On the distribution of stock price differences, *Operations Research* **15** (1967b), 1057-1062.
[10] Ortobelli, S., Lando, T., Petronio, F and Tichy, T.: Asymptotic stochastic dominance rules for sums of i.i.d. random variables, submitted to *Journal of Computational and Applied Mathematics.*
[11] Ortobelli, S., and Rachev, S.T.: Safety-First Analysis and Stable Paretian Approach to Portfolio Choice Theory, *Mathematical and Computer Modelling* **34** (2001), 1037-1072.
[12] Ortobelli, S., Rachev, S., Shalit, H., and Fabozzi, F.: Orderings and Risk Probability Functionals in Portfolio Theory, *Probability and Mathematical Statistics* **28**(2) (2008), 203-234.
[13] Ortobelli, S., Tichy, T., and Petronio, F.: Dominance among financial markets, *WSEAS transactions on business and economics* **11**, (2014), 707-717.
[14] Rachev, S.T., and Mittnik, S.: *Stable Paretian models in finance*, Wiley, New York, 2000.
[15] Rachev, S.T., Mittnik, S., Fabozzi, F.J., Focardi, S., and Jasic, T.: *Financial econometrics*, John Wiley & Sons, Inc., 2007.
[16] Samorodnitsky, G. and Taqqu, M. S.: *Stable Non Gaussian Random Processes*, Chapman & Hall, 1994.
[17] Shaked, M., and Shanthikumar, G.: *Stochastic orders*, Springer Series in Statistics, 2007.

Review of possibilities for speed-up of distance matrix calculation

Stanislav Palúch[1], Tomáš Majer[2]

Abstract. A distance matrix in a graph can be calculated row by row by a "point to all" algorithm. During the course of this algorithm for row i not only i-j distances for all j are the result but many other distance matrix entries can be easily derived. This paper investigates the possibilities how to make use of these values for speedup of distance matrix calculation.

Keywords: graph, digraph, shortest path, distance.

JEL classification: C61
AMS classification: 05C12

1 Introduction

Themes related to the shortest path problems or distance matrix calculation seem to be closed. However, papers concerning those problems still appear – e.g. paper [1] studying methods of distance matrix calculation, [4] proposing an shortest path algorithm for bipartite graph or [2] designing a new k-shortest path algorithm. This paper will be concerned with distance matrix calculation in general graph or digraph.

There are several methods for calculation a distance matrix for an arc weighted digraph or for an edge weighted graph. Some of them can calculate whole distance matrix all at once e.g. using Floyd algorithm. Another method of distance matrix computation is to compute it row by row using a point to all shortest path algorithm like Ford algorithm, Dijkstra algorithm or their implementation called label correct or label set implementation.

2 Known algorithms for distance matrix

2.1 Floyd algorithm

Let us have a graph or digraph $G = (V, E, c)$ with vertex set V, edge (or arc) set E and edge cost $c : E \to R$, where R is the set of real numbers and c is a nonnegative real function. Let us denote by $n = |V|$ the number of vertices and by $m = |E|$ number of edges of the graph G. Let us suppose form simplicity that $V = \{1, 2., \ldots, n\}$. Finally denote by $c(i, j)$ the cost of the edge $(i, j) \in E$.

Floyd algorithm is as follows

Define two squared matrices $\mathbb{D} = \{d(i, j)\}_{i=1,2,\ldots,nj}$, $\mathbb{X} = \{x(i, j)\}$ both of the type $n \times n$. Set

$$d(i,j) = \begin{cases} 0 & \text{if } i = j \\ c(i,j) & \text{if } (i,j) \in E \\ \infty & \text{if } i \neq j \text{ and } (i,j) \notin E \end{cases} \tag{1}$$

$$x(i,j) = \begin{cases} 0 & \text{if } i = j \\ j & \text{if } (i,j) \in E \\ \infty & \text{if } i \neq j \text{ and } (i,j) \notin E \end{cases} \tag{2}$$

[1]University of Žilina /Faculty of Management Science and Informatics, Department of Mathematical Methods and Operation Analysis, Univerzitná 8215/1, 010 26 Žilina, Slovakia, stanislav.paluch@fri.uniza.sk
[2]University of Žilina /Faculty of Management Science and Informatics, Department of Mathematical Methods and Operation Analysis, Univerzitná 8215/1, 010 26 Žilina, Slovakia, tomas.majer@fri.uniza.sk

```
for(k=1; k<=n, k++){
    for(i=1; i<=n, i++){
          if(d[i][k] < infty then{
                for(j=1; j<=n, j++){
                      if(d[i][j]<d[i][k]+d[k][j]) then{
                            d[i][j]=d[i][k]+d[k][j];
                            x[i][j]=x[k][j];}
                            }}}}
```

After finishing Floyd algorithm matrix $\mathbb{D}$ becomes the distance matrix and matrix $\mathbb{X}$ contains in entry $x(i,j)$ last but one vertex of the shortest i-j path. The complexity of Floyd algorithm is $O(n^3)$ since it contains three nested loops every one of them is executed (almost) n times. Floyd algorithm computes all distance matrix in one computational proves, therefore it is sometimes referred to as a all to all algorithm.

2.2 Label set implementation of Dijkstra algorithm and label correct implementation of Ford algorithm

Ford shortest path algorithm is designated to compute all i-j shortest paths from fixed vertex i to all vertices $j \in V$. Such algorithms are referred to as point to all algorithms. It is not suitable for searching a single i-j shortest paths since it does not stop until all i-j shortest paths from i to all j are determined.

Dijkstra shortest path algorithm is designated to compute a single i-j shortest path, therefore it is sometimes referred to as a point to point algorithm. This algorithm can detect during computation that it has just found i-j earlier that it should compute any longer i-k path. However, Dijkstra algorithm can be easily used as a point to all algorithm, if we let it stop after computing all i-j distances.

Both Dijkstra and Ford algorithms have very implementations called label set and label correct implementations, therefore we will not present here original algorithms but only their mentioned implementations.

Algorithm for finding all shortest u–v paths in edge weighted graph or digraph $G = (V, E, c)$ with non-negative edge weights.

- **Step 1.** Initialization. For every vertex $i \in V$ assign two label $t(i)$ a $x(i)$. Label $t(i)$ represents till now the best u-i path, label $x(i)$ is the last but one of this u-i path,

  Set $t(u) := 0$, $t(i) := \infty$ for $i \in V,\ i \neq u$ a $x(i) := 0$ for every $i \in V$.

  Create the set $\mathcal{F} := \{u\}$. The set $\mathcal{F}$ is the set of such vertices for which there is a hope to find a shorter u-j path by examining their neighborhoods. $\mathcal{F}$ contains initially only starting vertex u.

- **Step 2.** Remove a vertex r called pivot from $\mathcal{F}$, set $\mathcal{F} := \mathcal{F} - \{r\}$.

  For all edges of the form $(r, j) \in F$ do:

  If $t(j) > t(r) + c(r, j)$, then $t(j) := t(r) + c(r, j)$, $x(j) := r$ and insert j into $\mathcal{F}$ – set $\mathcal{F} := \mathcal{F} \cup \{j\}$.

- **Step 3.** If $\mathcal{F} \neq \emptyset$ GOTO Step 2. Otherwise $t(i)$ is the length of the shortest u-i path and $x(i)$ is its last but one vertex for every $i \in V$.

Selection of pivot r – element of $\mathcal{F}$ which is removed in Step 2. of just presented algorithm is not specified within this algorithm. If we choose pivot r at random without any further requirements we speak about *label correct algorithm*. If we choose as a pivot r the element of $\mathcal{F}$ with least label $t(r)$, we speak about *label set algorithm*.

Label set algorithm has that advantage that if we extract from $\mathcal{F}$ pivot r we are sure that the value $t(r)$ represents the length of the shortest u-r path. Therefore we can stop algorithm if we need only this concrete path. Every vertex can appear as a pivot only once. Therefore another advantage of label set algorithm is that it examines every edge in Step 2. at most once in comparison with label correct algorithm which can use the same edge in Step 2. more times. On the other hand the selection of pivot r with the least label $t(r)$ requires additional overhead computation.

It is known that the complexity of label set algorithm is

$$O(m + n.\log_2 n) \quad \text{where } n = |V| \text{ and } m = |E|. \tag{3}$$

3 An attempt to speed-up distance matrix calculation

Distance matrix can be calculated row by row using a point to all shortest paths algorithm. We have tried three point to all algorithms – label correct algorithm with set $\mathcal{F}$ organized as LIFO, label correct algorithm with set $\mathcal{F}$ organized as FIFO and label set with the set $\mathcal{F}$ organized as a binary heap.

Once we have the shortest u-v path in the form of alternating sequence of vertices and edges

$$u \equiv v_1, (v_1, v_2), v_2, (v_2, v_3), v_3, \ldots, (v_{k-1}, v_k), v_k \equiv v, \tag{4}$$

then every subsequence of (4) for $1 \leq i < j \leq k$ in the form

$$v_i, (v_i, v_{i+1}), v_{i+1}, \ldots, v_{j-1}, (v_{j-1}, v_j), v_j \tag{5}$$

is the shortest v_i-v_j path.

Array resp. vector $x(i)$ for $i = 1, 2, \ldots, n$ is the vector of pointers to last but one vertices of the shortest u-i paths. Digraph $T = (V, X, f)$ with vertex set V and directed edge set

$$\{(i, x(i)) \mid i \in V\} \tag{6}$$

and edge weight

$$f(i, x(i)) = c(x(i), i) \tag{7}$$

is a rooted tree with the root u in which every i-u path represents in reverse order the shortest u-i path in original graph G. Moreover every directed path in digraph T is, in reverse order a shortest path in original graph G. We will say that T is a *short distance tree*.

Suppose we are given a graph G modeling a road network with one vertex of degree 2, other vertices with degree 3. (Degree of a vertex is the number of incident edges). The tree T with minimum different paths (regardless of their length) is a tree in the form of binary heap having k complete levels where $k = \lfloor \log_2 n \rfloor$, one vertex on level 0, two vertices on level 1, 2^i vertices on level i, 2^{k-1} vertices on level $k-1$. There are

$$1 + 2^1 + 2^2 + \cdots + 2^{k-1} = \frac{1 - 2^k}{1 - 2} = 2^k - 1 \tag{8}$$

vertices on all complete levels.

Every vertex on level i is starting vertex of i different paths. So vertices on k complete levels are starting vertices of

$$1.2 + 2.2^2 + 3.2^3 + \cdots + k.2^{k-1} = 2 + (k-2)2^k \tag{9}$$

different paths. There are $n - (1 + 2 + 2^2 + \cdots + 2^{k-1}) = n - 2^k + 1$ vertices on incomplete level k and they are starting vertices of $k(n - 2^k + 1)$ different paths.

Number of additional paths different from all u-i paths is

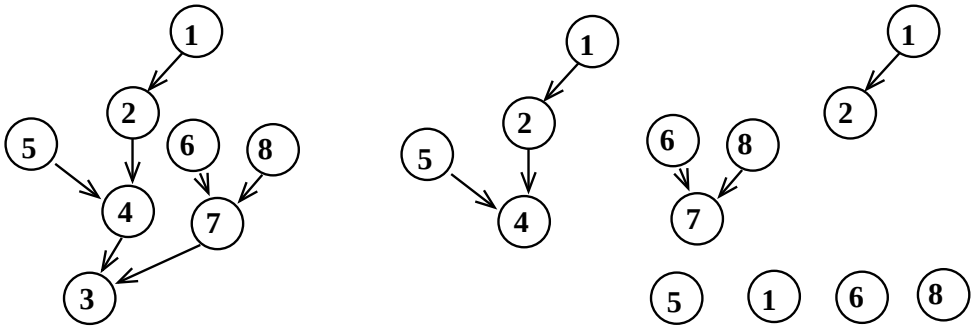
$$\begin{aligned} 2 + (k-2)2^k + k(n - 2^k + 1) - n = 2 + k.2^k - 2.2^k + kn - k.2^k + k - n > \\ > -2.2^k + kn - n \geq -2n + kn - n = (k-3)n \approx (\log_2 n - 3)n \end{aligned} \tag{10}$$

The last number can be considered as a very gross assessment of additional distances obtained while calculating a row of a distance matrix using a point to all algorithm. The essential conception for speed-up is to try to make use or those additional data while calculating other rows.

The idea how to make it is as follows. We know only trivial paths of the type u-u for every $u \in V$ in the beginning of calculation. Therefore corresponding matrix $\mathbb{T}$ has zeros on main diagonal and ∞ on other entries. Matrix $\mathbb{X}$ contains only trivial trees for every $u \in V$, therefore it contains zeros on main diagonal and ∞ on other places.

After calculating line u we will have complete short distance tree rooted in u. We can derive from this tree some part of new short distance tree rooted in i for every vertex i belonging to the same component as u together with corresponding labels $t(j)$ for all j of this new tree. An example of just mentioned fact is displayed in fig 1.

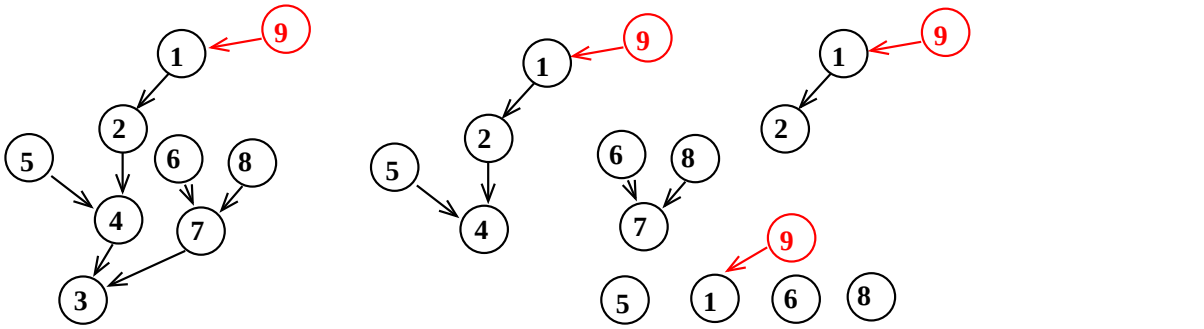
When calculating a row with partial a tree, we will make use of this information as follows. The set $\mathcal{F}$ will contain initially all such neighbors j of all vertices contained in partial tree for which label $t(j)$



a) Short distance tree obtained when calculating row 3

b) Partial trees obtained as additional data when calculating row 3

Figure 1



a) A change of short distance tree during computing row 3 after extracting pivot 9

b) Changes in partial trees as a consequence of extracting vertex 9 from $\mathcal{F}$ when calculating row 3

Figure 2

can be reduced. Further computation is the same as in label set algorithm with only difference that if some r is selected as the pivot directed edge $(r, x(r))$ is introduced into all partial trees – see fig. 2.

Here is the proposed algorithm:

- **Step 1.** Initialization. Let us have matrices $\mathbb{T}$ and $\mathbb{X}$ initialized as follows

$$t(i,j) = \begin{cases} 0 & \text{if } i = j \\ \infty & \text{if } i \neq j \end{cases} \qquad x(i,j) = \begin{cases} 0 & \text{if } i = j \\ \infty & \text{if } i \neq j \end{cases} \tag{11}$$

- **Step 2.** Calculation of subsequent row u.

  Here we suppose that we have in every row i several precomputed entries $t(i,j)$ and $x(i,j)$ representing distances from i to j and where $x(i,j)$ is the last but one vertex of shortest i-j path. Let us emphasize that if $t(i,j) = \infty$ no shortest i-j path was discovered.

  Set $\mathcal{F} := \emptyset$.

  Set $\mathcal{R} = \{r \mid t(r,j) < \infty\}$

  For all edges of the form (r,j) where $r \in \mathcal{R}$ do:

  If $t(u,j) > t(u,r) + c(r,j)$, then

  $t(u,j) := t(u,r) + c(r,j)$, $x(j) := r$ and insert j into $\mathcal{F}$ – set $\mathcal{F} := \mathcal{F} \cup \{j\}$.

- **Step 3.** Extract pivot r with least label $t(u,r)$ from $\mathcal{F}$. Set $\mathcal{F} := \mathcal{F} - \{r\}$.

  Now we are sure that distance from u to r is $t(u,r)$ and the last but one vertex of shortest u-r path is $x(u,r)$. We are able also easily determine distances from i to u for all vertices

$$i \in \{x(u,r), x(u,x(u,r), x(u,x(u,(u,r))), \dots\}$$

  Modification of partial trees:

Set $i := r$

While $x(u,i) > 0$ do
{ Set $i := x(u,i)$
$x(i,r) := x(u,r), \quad t(i,r) = t(i,x(i,r) + c(x(i,r),r)$
}

Exploring the neighborhood of pivot r

For all edges of the form $(r,j) \in E$ do
If $t(u,j) > t(u,r) + c(r,j)$ then
$t(u,j) := t(u,r) + c(r,j), \quad x(u,j) := r, \quad \mathcal{F} := \mathcal{F} \cup \{j\}$.

- **Step 4.** If $\mathcal{F} \neq \emptyset$ GOTO Step 3.

  If $\mathcal{F} = \emptyset$ and there exists a not calculated row of matrix $\mathbb{T}$ GOTO Step 2.

4 Computing experiments

We have made several computational tests on Fujitsu server computer with Intel(R) Xeon(R) CPU E3-1246 v3 3.50GHz, 8GB RAM and OS Debian Linux 8.0 (jessie) amd64 installed.

We have implemented Floyd all to all algorithm and following four row by row algorithms: Label Set, Label Correct with LIFO queue, Label Correct with FIFO queue and our proposed algorithm. Corresponding computer implementations were written in C language. We have applied GNU-C compiler (gcc) version 4.9.2 to compile and GNU-make to build executable application.

Test instances were made as subsets of real Slovak road network with 10, 20, 50, 100, 200, 500, 1000, 2000, 5000 and 10000 vertices (junctions).

Every instance was created by choosing a random vertex and subsequently choosing desired number of closest vertices in the same order as the succession in which they appear as pivots in Dijkstra algorithm.

Results are shown on following tables and graphs.

Network	Floyd's	LabelSet	LabCorr FIFO	LabCorr LIFO	Our alg.
SK10	5 μs	11 μs	6 μs	5 μs	12 μs
SK20	23 μs	40 μs	20 μs	21 μs	51 μs
SK50	217 μs	272 μs	124 μs	243 μs	291 μs
SK100	1.23 ms	0.973 ms	0.408 ms	0.783 ms	1.13 ms
SK200	8.79 ms	4.14 ms	1.52 ms	3.84 ms	4.443 ms
SK500	114 ms	32.7 ms	19.0 ms	261 ms	39.1 ms
SK1000	509 ms	136 ms	106 ms	3 202 ms	175 ms
SK2000	2.30 s	0.585 s	0.669 s	56.6 s	0.918 s
SK5000	37.5 s	4.03 s	6.45 s	1 610 s	14.3 s
SK10000	364 s	17.1 s	46.7 s	$>$ 5 h	102 s

Table 1 Computation times for tested algorithms and instances.

Best results for all instance were achieved by Label Set algorithm with the set $\mathcal{F}$ implemented as a binary heap. Label Correct algorithm with the set $\mathcal{F}$ implemented as FIFO queue. Our algorithm was three time faster than Floyd. The worst algorithm was Label Correct with the set $\mathcal{F}$ implemented as LIFO stack.

Authors feel a great disappointment about effectiveness of proposed algorithm. Assessment that during computation of a row of distance matrix we can get $(\log_2 n - 3)n$ additional matrix entries promised a hope to design a faster algorithm. Unfortunately, administration of additional matrix entries and the way in which proposed algorithm makes use of precomputed data are more time consuming that Dijkstra used row by row. Another cause of low effectiveness can be in multiplicity of additional data. Although the computation for one row gives a rich set of additional data, the one for other row may give also

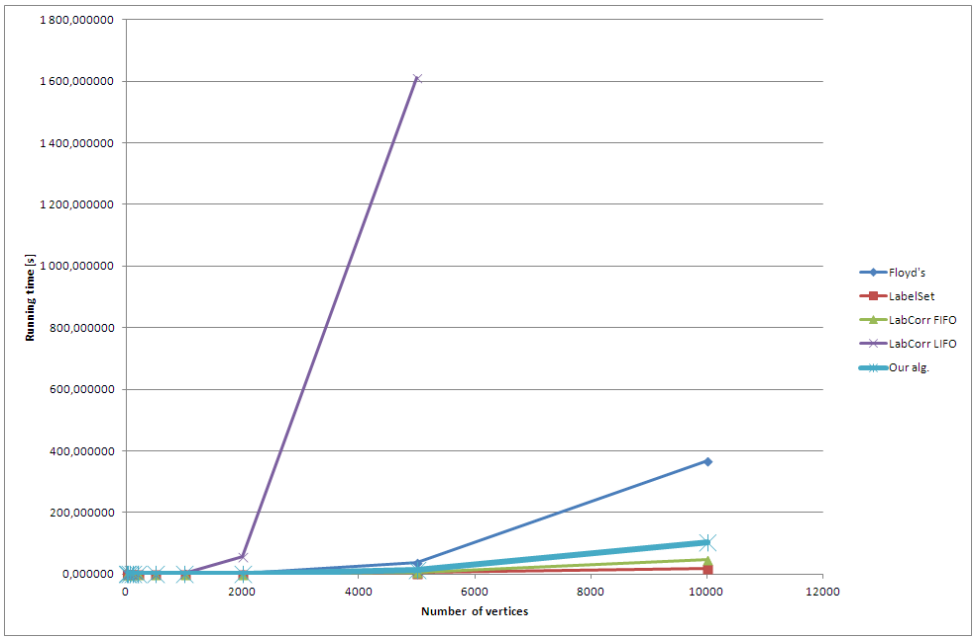


Figure 3 Comparison of computational times of tested algorithms

rich but similar set which contains entries discovered by preceding row. However, the idea of exploiting mentioned additional data is so challenging, that we cannot give-up it.

Open question. Is there any way how to make use of mentioned additional data efficiently?

5 Conclusion

Our experiments show that the most efficient way of distance matrix calculation for network similar to road network is row by row calculation using Dijkstra algorithm with the set $\mathcal{F}$ implemented as a binary heap.

Acknowledgements

The research was supported by the Slovak research grants VEGA 1/0158/15 "Resilient rescue systems with uncertain accessibility of service" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks."

References

[1] Dachuan Wei: An Optimized Floyd Algorithm for the Shortest Path Problem. *Journal of Networks* Vol.**18**, No.**12**, (December 2010)

[2] Palúch, S.: A multi label algorithm for k shortest paths problem, *Communications: Scientific Letters of the University of Žilina* Vol.**11**, No.**3**, 2009, 11-14.

[3] Plesník, J.: *Grafové algoritmy*, Bratislava, 1985

[4] Torgassin, S., Zimmermann, K.,H.,: An All Pairs Shortest Path Algorithm for Bipartite Graphs. *Central European Journal of Computer Science* **3(4)**, (2013), 149-157.

GDP Targeting. Econometric Approach.

Václava Pánková[1]

Abstract. Nominal GDP (nGDP) targeting is an alternative framework for a monetary policy; in contrast to strict inflation targeting it may result in more stable real economic activity as it was un-lately demonstrated in case of advanced economies. Attention to middle income countries in the nGDP targeting is scant but required. Such countries are more likely to export commodities and to be price takers on world markets. So they are more exposed to trade and supply shocks; also productivity shocks are more likely to occur. A general knowledge is that the more common are supply shocks, the more convenient is nGDP targeting.
The discretionary policy is formalised as an optimising problem to choose aggregate demand so as to minimise a given loss function. Afterwards, an econometric model is derived, estimated and relevant conclusions drawn.
The principles on which the theoretical models are developed may differ for advanced and for middle income economies. The differences will be discussed with the goal to show a potential application to the economy of the Czech Republic.

Keywords: nominal GDP targeting, loss function, simultaneous equations.

JEL Classification: C3, C60, E1, E6
AMS Classification: 90C30, 91G70

1 Introduction

In many economies some sort of macroeconomic targeting is practiced with the goal to earn macroeconomic benefits. A nominal target can affect expectations in two ways. First, it influences markets' expectations; second, a nominal target should encourage firms and workers to behave in a way which seems to be optimal according to their own criteria.

After 1990, inflation targeting became very popular due to its transparency and accountability. It was preferred to the ten years older concept of nominal GDP (nGDP) targeting. After 2008 – 9 financial crisis, central bankers captured its un-sufficiency and nGDP targeting occurred in a focus of economists again as a more useful alternative. First proposals to prefer nGDP targeting occurred in largest advanced economies. Frankel in [3] shows that in fact, such a targeting is more useful for middle – income countries and gives some examples.

In the Czech Republic, inflation targeting was performed by the Czech National Bank with a surprising result: in 2014 CNB started to weaken the exchange rate of Czech crown because of a disproportionately low inflation rate. That is why an eventual applicability of nGDP targeting is studied in this article using Czech economic data.

A loss function of the economy, once formulated is than minimised by central bankers. Criteria concerning inflation rule and nGDP rule are formulated and compared.

2 Theoretical background

The open economy loss function to be minimised is

$$F = \alpha p^2 + (y - \hat{y})^2 + \gamma s^2 \tag{1}$$

where α is a weight to the inflation objective, $\hat{y}$ is desired level of output. The aggregate supply y is given as

$$y = \bar{y} + \beta(p - p^e) + u \tag{2}$$

where y a real output, $\bar{y}$ is a potential output, p is the price index, p^e expected price index, u is a supply disturbance. Further, s in (1) is the spot exchange rate, γ is the weight placed on exchange rate stability. Bhandari and Frankel in [1] propose

$$s = m - y + \varepsilon \tag{3}$$

[1] University of Economics, Department of Econometrics, 130 69 Praha 3, nám. Winstona Churchilla 4, pankova@vse.cz

with money supply *m* and error term ε. They argue that money supply variable will allow authorities the possibility of affecting the exchange rate. Adding the money market equilibrium condition

$$m = p + y - v \tag{4}$$

where *v* are velocity shocks and inserting (2), (3) and (4) to (1) and including an assumption that *v* is not correlated with *u* we have

$$F = \alpha p^2 + \left[\bar{y} - \hat{y} + \beta(p - p^e) + u\right]^2 + \gamma(p - v - \varepsilon)^2 . \tag{5}$$

The loss function (5) is minimised by applying a chosen macroeconomic rule.

(i) inflation targeting rule

In such a case the inflation variables are set to zero and the expected loss function is

$$EF = \left[\bar{y} - \hat{y}\right]^2 + Var(u) + \gamma Var(\varepsilon) + \gamma Var(v) \ . \tag{6}$$

Disturbance terms, as e.g. *u*, follow the rule $Var(u) = (u - E(u))^2 = u^2$ when fulfilling the Gaussian – Markovian assumption $E(u) = 0$.

(ii) nGDP targeting rule

Minimizing *F* in (5) with respect to *p* we get the first order condition $dF / dp = 0$ what enables to express *p* as a function of other variables. Further, $p^e = E(p)$. Substituting both this findings into equation (5) (for details see [1]) the expected value of loss function is

$$EF = \left[\bar{y} - \hat{y}\right]^2 + \left[\frac{1 + \alpha + \gamma}{1 + \beta^2}\right] Var(u) + \gamma Var(\varepsilon) \tag{7}$$

Under an assumption that authorities vary money supply *m* to keep *p+y* constant the velocity shocks drop out. Comparing inflation rule (6) with the nGDT rule (7) we can see that nGDP rule dominates the inflation one if

$$\frac{1 + \alpha + \gamma}{1 + \beta^2} < 1$$

what may be rearranged as

$$\alpha + \gamma < \beta^2 . \tag{8}$$

As central bankers usually prefer smaller exchange rate fluctuations to price level fluctuations, $\gamma < \alpha$ may be supposed as a direct consequence. Higher α represents a higher penalisation of the inflation term in the loss function. Instead of (8) we than may expect $2\alpha < \beta^2$, hence

$$\alpha < \frac{\beta^2}{2} \tag{9}$$

as a sufficient condition of nGDP rule dominancy.

Parameter α is the weight on price stability from the Taylor rule, originally introduced in [6]. As it is argued in [1], the weights to output and price stability are widely supposed to be equal what implies $\alpha = 1$. Instead of (9) we than get the condition

$$\beta > \sqrt{2} \ . \tag{10}$$

Of course, the problem how to find β remains. The answer is to be found in the context of aggregate supply (AS) and aggregate demand (AD) curves. From a theory (e.g. [5]) it is

AS: $y_t = \beta p_t - u_t$ with u_t being an adverse supply shock (11)

AD: $p_t + \delta y_t = m_t + v_t$ with v_t being positive demand shock (12)

All variables are in logs so that AS and AD curves become straight lines. Though we are interested in the slope of AS curve only, we have to pay attention to the potential dependency of p_t on u_t. Treating (11) and (12) a s a simultaneous equations system we apply necessary instrumental variables technique by using 2SLS to the AS equation which is exactly identified (details e.g. in [7]). So, the technical treatment of (11) is very straightforward; all variables are observable and the estimation process is used only once. The weak point is the evidence that (11) is not the same as (2). Hence, the economic interpretation of the result is questionable.

Taking

AS: $y = \bar{y} + \beta(p - p^e) + u$ (11A)

we have to find unobservable variables potential output and expected price first.

As for the potential output, [2] is a representable survey of relevant methods divided into four groups: (i) trend, (ii) univariate filters, (iii) multivariate filters, (iv) production function approaches. The techniques as well as (dis)edvantages of their choice are discussed. For the following computations, the Hodrik - Prescott filter was used to create $\bar{y}$. The difference $y - \bar{y}$ than is an output gap which in fact is the cyclical component of the real output.

The expectations also are not represented by a unique method only. As for the econometric approach, a basic choice is between adaptive or rational hypotheses (REH) expectations. Both techniques are explained e.g. in [4]. For the further treatment of (11A), the computation was performed according to the formula

$$p^e = \zeta_0 + \zeta_1 p_{-1} + \varepsilon \tag{13}$$

what is a REH application.

Formally, (11A) and (12) still represent a simultaneous equations system but it is hardly to believe that the 2SLS with the same instrument as it was used for (11) and (12) should work. The reason is that the output gap as well as the $(p - p^e)$ part are differences. That is why the 2SLS as well as the simultaneous system were dropped here and a general instrumental variables (IV) technique with a wider choice of instruments was applied to (11A).

3 Application to the Czech economy

Having the data concerning Czech economy the period from 2002Q1 to 2014Q4 is covered by quarterly data. GDP measured in current prices, unit million euro and HICP (2005 = 100) are received from Eurostat, money supply M2 information is published by CZSO, the transformation from millions of Czech crown to millions of Euro was done by the author. The variables are in the logarithmic form. Referring to the simultaneous system, *y* and *p* are endogenous.

The results of the estimates are summarised in Table 1. Equation (11) offers reliable results (2SLS applied to seasonally adjusted quarterly data) but it does not meet the economic rules. Equation (11A) was estimated in three different ways. First, as a part of simultaneous system (column 2); according to the expectation, the β parameter is not significant from the statistical point of view. Second (column 3), using OLS the potential correlation of the exogenous variable and the disturbance term is ignored. Third (column 4), general IV approach used combining available variables as possible instruments. In this case, all necessary operations were performed. But having in mind the variability of choices of (i) methods of calculating potential output, (ii) methods of calculating expected

prices and (iii) instruments as well, the number of different combinations and therefore also different results is a little bit devaluating.

equations	(11), (12)	(11A), (12)	(11A)	(11A)
method	2SLS	2SLS	OLS	IV
β parameter	2.218	2.609	3.160	9.160
prob. of t-stat. 5%	0.000	0.305	0.000	0.002
instruments	m	$m,\ p^e,\ \bar{y}$	none	m, p, y

Table 1. Estimation of parameter β, source: own computation

The estimated value of parameter β is $\hat{\beta} = 9.16$. Evidently, $\hat{\beta} > \sqrt{2}$ what means that nGDP rule in the conditions of the Czech economy dominates the inflation targeting.

4 Conclusions

After an evident failure of inflation targeting in most part of economies a nominal GDP targeting became a possible alternative. Its appropriateness is discussed in latest literature distinguishing between most advanced and middle – income economies. The letter variant is briefly summarised including quantitative criteria to make a decision about eventual dominancy of nGDP rule to the inflation one.

Using quarterly data covering macroeconomic processes in Czech Republic during 2002 to 2014, a predominant evidence is given that nGDP rule in the conditions of Czech economy apparently dominates the inflation targeting.

Acknowledgement

Financial support of VŠE IGA F4/24/2014 is gratefully announced.

References

[1] Bhandari, P., Frankel, J. (2015): Nominal GDP Targeting for Developing Countries, *NBER Working Paper* No. 20898

[2] Cotis, J.P., Elmeskov, J., Mourougane, A. (2015): Estimates of potential output: benefits and pitfalls from a policy perspective, OECD Economics Department, http://www.oecd.org/economy/growth/23527966.pdf

[3] Frankel J. (2014): Nominal GDP Targeting for Middle – Income Countries, Harvaard Kennedy School, *Faculty Researche Working Paper Series*, RWP 14-033

[4] Ghosh, S.K. (1991): *Econometrics*, Prentice Hall, ISBN 0-13-224668-6

[5] Mankiw, N. G. (2012) *Macroeconomics*, Worth Publishers, Inc., ISBN-13: 978-1429240024

[6] Taylor, J. B. (1993): Discrertion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy 39*, Horth Holland, pp. 195 - 214

[7] Wooldridge, J.M. (2009) *Introductory Econometrics*, South – Western CENGAGE Learning, ISBN-10: 1-111-53104-8

Labor Market Frictions in the Czech Republic and Hungary

Adam Pápai[1], Daniel Němec[2]

Abstract. The goal of this paper is to investigate and compare the structural and dynamical characteristics of the Czech and Hungarian economy. The focus lies mainly on the examination of the development of key labor market variables. We also want to capture the changes that occurred due to the Great Recession in 2008. We estimate a DSGE model with search and matching frictions, price and wage rigidities and hiring costs. The monetary authority sets the nominal interest rates according to a Taylor-type rule. The wages setting mechanism and hours worked are the result of the Nash bargaining process. This model is estimated for the quarterly data of the Czech Republic and Hungary for the period 2001Q2 – 2014Q4. The results show that the reactions of variables to monetary shock are larger in the Czech Republic. This suggests that the monetary policy is less efficient in Hungary during the examined period. The bargaining power of workers is stronger in the Czech economy. This coefficient is smaller in Hungary, which is in line with the low trade union participation of workers. The model shows the preference, foreign and disutility from work shocks to be the main cause of the Great Recession in both countries.

Keywords: DSGE, small open economy, labor market, search and matching frictions, Great Recession

JEL classification: E32, J60
AMS classification: 91B40

1 Introduction

The aim of this paper is to examine the structural characteristics and dynamism of the Czech and Hungarian economy in the past 14 years. We focus mainly on the behavior of the labor market variables and the interactions between the labor market and the rest of the economy. The whole examined time period (2001Q2 – 2014Q4) is divided into two sections – *before* and *after* crisis. In total, we perform four separate estimations – two for each country – to investigate the differences between the economies and to identify the structural changes caused by the Great Recession. This paper is a follow up of our previous research [5], where we investigated the behavior of the Hungarian economy. Now we extend our field of interest also to the Czech Republic and provide a more detailed look at the impact of the Great Recession.

There are several other papers which examine the development of the labor market. Lubik [3], for example, estimates a model for the United States containing vacancy posting cost developed by Rotemberg [6]. This is one way to implement frictions to the model: the firms are less willing to hire and fire workers, because they cannot create vacancies for free. Lubik states, that most of the structural parameters of the model are not dependent on the model specification. However, specific parameters, like the search costs or worker benefit, can vary widely across specifications, and thus are likely not identified in either an econometric or economic sense.

Thomas [7] decided to examine the role of the monetary policy in models with labor market. He investigated the optimal monetary policy under different levels of wage rigidities. Under completely flexible wages, he finds the ideal inflation to be zero. However under wage frictions, the monetary

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, papai@mail.muni.cz

[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, nemecd@econ.muni.cz

authority should secure a positive inflation to decrease the volatility of unemployment. Therefore, the monetary policy holds a tool which can actively help in the labor market stabilization.

The role of monetary policy for labor market dynamics is discussed also by Christoffel et al. [2]. They use a DSGE model augmented with search and matching frictions and wage rigidities to determine the factors important for the monetary policy's effectiveness. They find flexible wages to increase the efficiency of the central bank. However, apart from the wage rigidities, other labor market characteristics have limited impact on the transmission of the monetary policy. They also suggest the bargaining power of workers to be an important information carrier for the monetary policy makers, and they emphasize the need for labor market modeling for the central banks.

1.1 Economic background of the Czech Republic and Hungary

The historical development of output is similar in both countries. It shows a relatively steady growth until the financial crisis, when there was a substantial drop in the GDP. On the other hand, the labor market variables show some significant differences between the two countries. The unemployment rate in the Czech Republic declined from 8% in 2001 to below 5% in 2008 with the average around 7%. During the same time period, the unemployment rate in Hungary had almost the same mean value (6.6%), but increased from 6% to 8%. The increase of this variable in Hungary to above 10% due to the recession was followed by a four-year period of high unemployment. After this, it decreased to its current level around 7.5%. The unemployment in the Czech Republic was also affected by the crisis and peaked temporarily at 7.8%. Its decrease to its current value (6.7%) was interrupted by a second decrease in the output in 2012. The movements in the vacancy rate are the opposite to the development of the unemployment rate in the Czech Republic. The vacancy rate reached its maximum value in 2008, when the unemployment was the lowest. This negative relationship also holds for the Hungarian economy, however not to such extent. The average working hours slightly increased during the whole examined time period in both countries. This increase was accompanied by rising real wages in the Czech Republic. On the other hand, the real wage shows a slight decrease in Hungary.

2 Model characteristics and structure

The selected model was introduced by Albertini et al. [1]. They used it to estimate the behavior of the New Zealand economy. This small open economy model consists of a monetary authority and homogenous firms and households. The monetary authority is present in the form of a Taylor rule. The sole input in the production function is the labor and there is no government sector present in the model. The model incorporates a detailed description of the movements on the non-Walrasian labor market. These movements are captured by varying hours worked (intensive margin) and also choosing whether or not the members of the households want to participate in the labor market at all (extensive margin). The Nash bargaining process sets the hours worked and wages and determines the allocation of the surplus from the production between the firms and workers based on their negotiating power. The model includes a matching function, initially introduced by Mortensen and Pissarides [4]. To capture the behavior of a real economy, four kinds of frictions are implemented into the models structure. First, due to the matching function, a certain part of the unemployed population fails to get paired with a vacant job position. Second, creating this vacant job position increases the firm's costs, therefore the firms cannot increase the employment freely. Next, the firms face a price adjustment cost, so they might be reluctant to change their prices. Finally, the change in the wage of the employees also increases the firms costs.

The most important element of the model is the labor market. This is the place, where the unemployed search for vacant job positions and firms seek to fill their vacancies. However, searching for workers and job positions is a costly and time consuming process. Being unemployed or vacant is a unproductive state. Only employees combined with filled vacancies create output. The number of these matches is the result of the following Cobb-Douglas matching function

$$M_t = \varepsilon_t^m S_t^\nu V_t^{1-\nu}.$$

This function gives us the amount of seeker-vacancy pairs that were created in each time period. On the other hand, there are several jobs that are destroyed each period at an exogenous and constant job destruction rate. Thus, the number of employed in each time period is given by the number of those,

who kept their job and those, who managed to find a new one. This model assumes constant labor force over time and differentiates between the number of unemployed and job seekers. Everyone is looking for jobs except those, who kept their job from the previous period. The number of unemployed in period t is given by the whole labor force without the number of employed in this period t. This setting allows workers who lost their job in one period to find a new one in the same time period.

The examined model consists of the foreign sector and three domestic agents: households, firms and monetary authority. Furthermore, there are three kinds of firms in the model: domestic producers and retailers, and foreign good importers. The foreign sector (output, inflation and interest rate) is exogenous and is represented by independent autoregressive processes – AR(1).

The representative household maximizes its expected intertemporal utility function. This function depends positively on the difference of consumption and consumption habit, and negatively on the supplied labor:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \varepsilon_t^c \left[\log(C_t - \vartheta \bar{C}_{t-1}) - N_t \kappa_t^h \frac{h_t^{1+\phi_h}}{1+\phi_h} \right].$$

Because there is no capital in the model, the income of the household – wage from work and interests from bonds and assets – are spent solely on the consumption and domestic and foreign assets. The consumption consist of bundles of domestic and foreign goods.

The domestic intermediate good producers are the only type of firms which hire workers and set their wages. These firms maximize their intertemporal profit function

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} \left[mc_t Y_t - \frac{W_t}{P_{H,t}} h_t N_t - \Gamma(V_t) - \Upsilon(W_t) N_t \right], \tag{1}$$

while facing hiring ($\Gamma(V_t)$) and quadratic wage adjustment ($\Upsilon(W_t)$) costs as in Rotemberg [6]. The good producers negotiate the wages and the worked hours by a Nash bargaining process. They sell their products on a perfectly competitive market to the second kind of firms, the domestic retailers. The retailers combine the differentiated intermediate goods to a final good and sell it to the representative household on a monopolistically competitive market. This gives them the opportunity to adjust their prices. However, each price change induces a increase of costs. The retailers have to choose whether to change their prices despite the increased costs or keep them at their old levels. The last kind of firms, the importers function similarly as the retailers. They import goods and set their domestic prices on a monopolistically competitive market. The importers also face price adjustment costs together with the fluctuations of the exchange rates.

The final agent in the model is the monetary authority, which is modeled using the following Taylor-rule

$$i_t = i_{t-1}^{\rho_r} \left[\frac{1}{\beta} \left(\frac{E_t \pi_{t+1}}{\pi} \right)^{\rho_\pi} \left(\frac{Y_t}{Y} \right)^{\rho_Y} \left(\frac{Y_t}{Y_{t-1}} \right)^{\rho_{\Delta Y}} \left(\frac{e_t}{e_{t-1}} \right)^{\rho_e} \right]^{1-\rho_r} \varepsilon_t^m.$$

The central bank sets the current nominal interest rate based on its previous value, deviation of inflation from its target, output from its steady state and the differences of output and exchange rate from their previous values.

3 Data and calibration

The time period is selected to cover the whole inflation targeting era in Hungary, 2001Q2 – 2014Q4. The first quarter of 2009 is selected as the dividing point for the *before* and *after* crisis periods. Eleven time series are selected for the estimation, eight for the domestic economy, three for the foreign sector represented by the Euro area. Output (Y_t) is calculated as the real gross domestic product divided by the active labor market population. Inflation (π_t) is defined as a quarterly percentage change in the domestic CPI. The 3 month interbank rate is selected for the estimation as interest rate (i_t). The exchange rate (q_t) is defined as the real effective exchange rate between the domestic economy and the Euro area. Harmonized unemployment rate (U_t) is selected as a measure of the unemployment. The vacancies (V_t) are calculated as the number of vacancies divided by the active labor market population. The wages (w_t) are represented by the real unit labor costs. Finally, the hours worked (h_t) are defined as the average work time per worker. The three variables of the foreign sector are the output (Y_t^*), inflation (π_t^*) and interest

rates (i_t^*). The time series of the exchange rates are acquired from the EuroStat database. The hours worked are obtained from the official statistical databases of the examined countries. The rest of the data is selected from the OECD database. The data series are seasonally adjusted. Before the estimation the data series are stationarised using demeaning for the inflations and interest rates and Hodrick-Prescott filter with the standard smoothing parameter $\lambda = 1600$ for the other variables. The model is estimated using Bayesian techniques. The computations are carried out in the Dynare toolbox (version 4.4.3) for Matlab. Two chains of Metropolis-Hastings algorithm are generated for each estimation. Each chain contains 600.000 draws of which the initial 33% is dropped.

Several of the variables are calibrated according to the relevant literature and are the same for each four estimations. The discount factor is set to 0.99. The elasticity of labor in the production function is set to 0.667. The debt elasticity of risk premium is set to 0.001 as in Albertini et al. [1]. The scale parameter in the vacancy creation function is set to 0.05, as in Lubik [3]. Other parameters are derived from the data and are set to different values to provide more data related information for the estimations. The steady state value of unemployment is calculated as the sample mean of the unemployment rate. For the Czech economy, it is set to 0.07 *before* and 0.067 *after* crisis and for Hungary to 0.066 and 0.101 respectively. The import share of GDP is set to 0.333 and 0.4 for the Czech Republic, and to 0.4 and 0.467 for Hungary. The job separation parameter is calculated to get the sample mean of the vacancy rate and set to values between 0.3 and 0.4 for the four estimations.

4 Estimation results

Table 1 shows the parameter estimation results for both economies and both time periods. This table also contains the prior densities used for the estimations. These values are the same for each estimation to get the differences between the estimations based solely on the differences between the input data and not the model setting. The habit parameter ϑ is higher for the Czech Republic, which suggests a less smoothed consumption in Hungary. The estimates of this parameter for Hungary are similar to our previous paper [5] – 0.4621. Although the changes in the mean are noticeable, the standard deviations suggest statistical insignificance. The firms bargaining power ξ is estimated to be lower in the Czech economy. This parameter increased *after* the crisis in both countries, possibly due to the increase in the unemployment and decrease of vacancies. This shift on the labor market makes the vacancies more rare

		Czech Republic				Hungary			
		Posterior before		Posterior after		Posterior before		Posterior after	
Params	Prior density	Mean	*Stdev*	Mean	*Stdev*	Mean	*Stdev*	Mean	*Stdev*
Other parameters									
ϑ	$\beta(0.5, 0.15)$	0.5899	*0.0808*	0.5658	*0.1175*	0.4714	*0.0923*	0.4121	*0.1170*
ξ	$\beta(0.5, 0.2)$	0.2720	*0.1179*	0.3456	*0.1392*	0.3860	*0.1454*	0.4613	*0.1609*
ν	$\beta(0.5, 0.2)$	0.3868	*0.2030*	0.5342	*0.2165*	0.5193	*0.2147*	0.4805	*0.2178*
e	$\Gamma(1, 0.5)$	7.0881	*0.6502*	6.6384	*0.8388*	7.1303	*0.6189*	6.5596	*0.8702*
Price and wage setting									
γ_H	$\beta(0.75, 0.1)$	0.5916	*0.1317*	0.5507	*0.1200*	0.6410	*0.1168*	0.6739	*0.1092*
γ_F	$\beta(0.75, 0.1)$	0.7824	*0.0926*	0.7671	*0.0937*	0.7646	*0.0961*	0.7366	*0.1020*
γ_W	$\beta(0.75, 0.1)$	0.6103	*0.1222*	0.5512	*0.1275*	0.5406	*0.1262*	0.5920	*0.1265*
ψ_H	$\Gamma(50, 15)$	56.1169	*14.6623*	33.1562	*10.8325*	28.6902	*8.9787*	28.8904	*8.7966*
ψ_F	$\Gamma(50, 15)$	57.0627	*14.9510*	44.5975	*12.7647*	55.4398	*14.1978*	48.6207	*13.3753*
ψ_W	$\Gamma(50, 15)$	6.3773	*1.6856*	11.3269	*2.6489*	8.7053	*1.9456*	9.8835	*2.5985*
Monetary policy									
ρ_r	$\beta(0.5, 0.15)$	0.7337	*0.0563*	0.6916	*0.0597*	0.4337	*0.0881*	0.4539	*0.0951*
ρ_π	$\Gamma(1.5, 0.5)$	1.9257	*0.4433*	1.3039	*0.2879*	2.3758	*0.4979*	2.5195	*0.5536*
ρ_Y	$\mathcal{N}(0.25, 0.1)$	0.2929	*0.1003*	0.3225	*0.0927*	0.2902	*0.0984*	0.3308	*0.0983*
$\rho_{\Delta Y}$	$\mathcal{N}(0.25, 0.1)$	0.2830	*0.0988*	0.2778	*0.0928*	0.2927	*0.0986*	0.2678	*0.0987*
ρ_e	$\mathcal{N}(0.25, 0.1)$	0.2820	*0.0849*	0.0977	*0.0659*	0.3030	*0.0803*	0.2023	*0.0754*

Table 1 Estimation results (parameters)

so the firms have better conditions in the negotiating process. However, even the highest value of the bargaining power parameter is far below our previous estimate for Hungary (0.9252). This is given mainly because of the different setting of the prior values. The parameter of elasticity of the matching function with respect to the job seekers ν increased in the Czech Republic and reached the level of Hungary, where it remained relatively stable. In both countries, the vacancy creation elasticity e decreased *after* the crisis. This decrease could be due to the firms effort to lower costs during the recession. However, the values are still far above our prior and suggest an increasing cost of creating an additional vacant job position. The backward looking price and wage parameters γ are more stable than the price and wage adjustment costs ψ. The backward looking domestic price γ_H is higher in Hungary and the backward looking wage parameter γ_W is lower. The backward looking foreign price γ_F is around the same value for the four estimations. This is given by the same foreign sector for both countries and the similar development of the exchange rates. There are significant differences in the adjustment parameters ψ in the Czech Republic between the two periods. The domestic ψ_H and foreign ψ_F price adjustment costs were almost the same *before* the crisis, while the wage setting cost ψ_W was much smaller. *After* the crisis, the price parameters decreased – the domestic more than the foreign – and the wage parameter increased. On the other hand, these parameters are unaffected by the crisis in the Hungarian economy, while the domestic price adjustment cost is much smaller in Hungary and the wage parameter is larger *before* and smaller *after* the recession than in the Czech Republic. The Taylor-rule parameters ρ imply relatively low volatility of the interest rates in the Czech Republic (interest rate smoothing parameter $\rho_r = 0.7$) and considerably higher fluctuations in Hungary, where $\rho_r = 0.4$. However, the weight of the inflation ρ_π in the monetary authority's decision making is relatively higher in Hungary.

Next, we focus on the shock decomposition of outputs presented in Figure 1 for the Czech economy and Figure 2 for Hungary. These two figures show, that in the period directly *before* the crisis, there were huge productivity shocks that pushed the product upwards. These shocks were also accompanied by high preference shocks, and foreign and labor market shocks (mainly the shock of disutility form work). Figure 1 shows the main cause in the output drop in 2008-09 in the Czech Republic to be the high negative foreign and preference shocks. This proves that the recession came to the Czech Republic from outside (in our case from the Euro area). The decrease of the product in 2012 is also caused mainly by the foreign shocks. On the other hand, the decline of Hungarian output in 2008-09 was caused by a combination of three shocks – shock of disutility form work, preference shock and foreign shocks. A positive shock of disutility form work caused the population to be less willing to work. This increased the unemployment and thus, reduced the output. This figure also shows an attempt of the monetary authority in the last eight quarters of the observed period to boost the output.

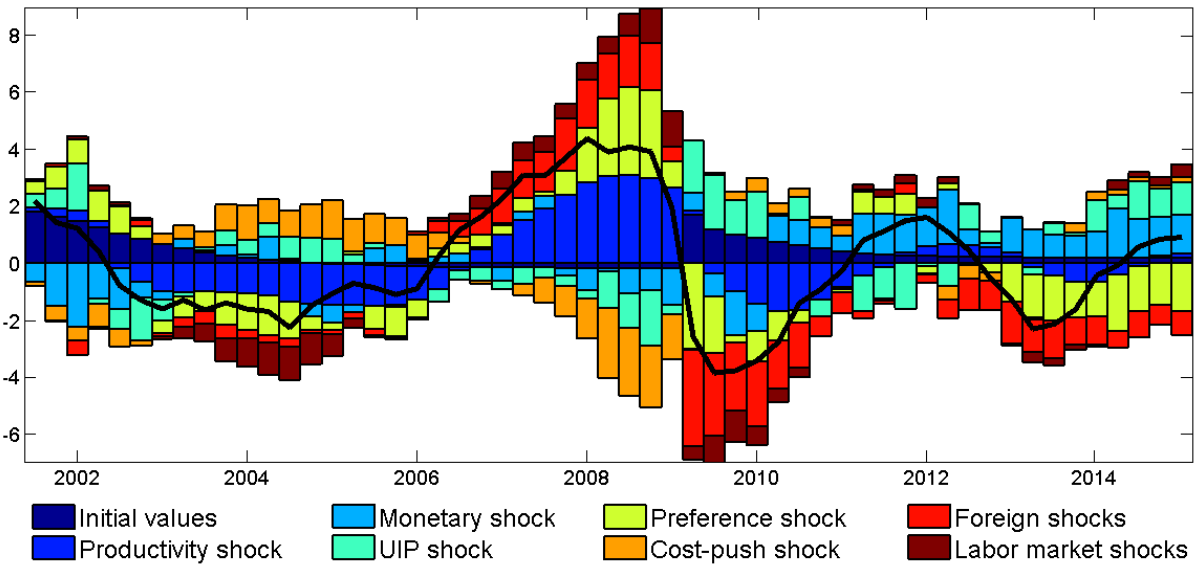


Figure 1 Shock decomposition of output for Czech Republic

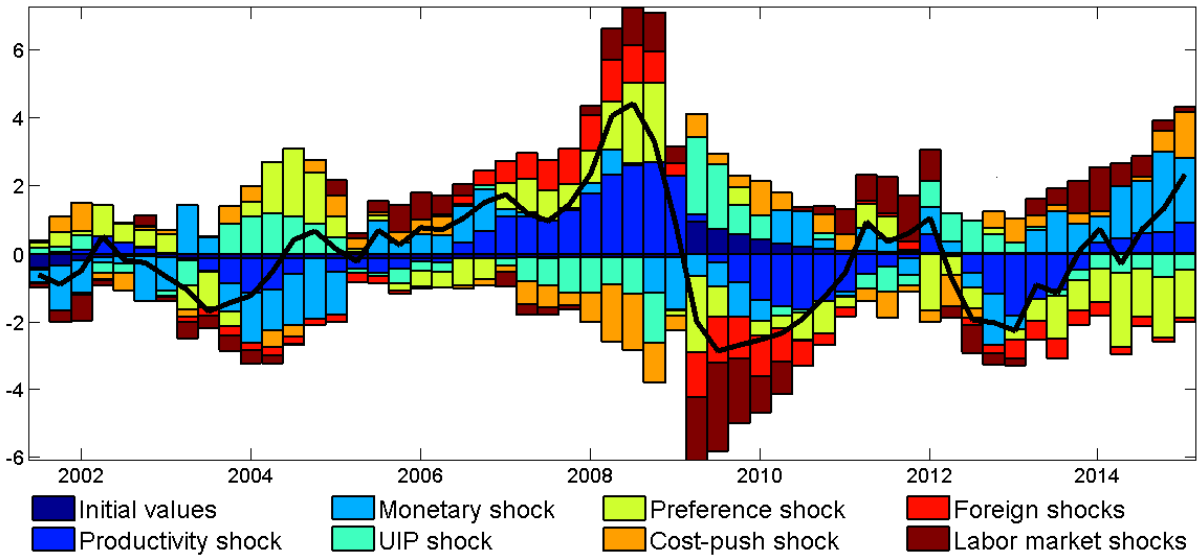


Figure 2 Shock decomposition of output for Hungary

5 Conclusion

In this paper we presented the estimation results of a small open economy DSGE model using the data of the Czech Republic and Hungary. The parameter estimates suggest that the structure of the Hungarian economy was affected less by the Great Recession than the Czech Republic. Also, we found significant rigidities present in both countries. The model provides a reasonable explanation to the causes of the output drop in both countries – in the Czech Republic it was caused mainly by the foreign and preference shocks, in the Hungarian economy by the disutility from work and preference shock.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1235/2014. This support is gratefully acknowledged.

References

[1] Albertini, J., Kamber, G. and Kirker, M.: Estimated small open economy model with frictional unemployment. *Pacific Economic Review* **17**(2), 2012, 326–353.

[2] Christoffel, K., Kuester, K. and Linzert, T.: The role of labor markets for euro area monetary policy. *European Economic Review*. **53**(8), 2009, 908–936.

[3] Lubik, T. A.: Estimating a search and matching model of the aggregate labor market. *Economic Quarterly* **95**(2), 2009, 101–120.

[4] Mortensen, D. T. and Pissarides, C. A.: Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, **61**(3), 1994, 397–415.

[5] Pápai, A. and Němec, D.: Labour market rigidities: A DSGE approach. *Proceedings of 32nd International Conference Mathematical Methods in Economics* (Talašová, J., Stoklasa, J., Talášek, T., eds.), Olomouc: Palacký University, 2014, 748–753.

[6] Rotemberg, J. J.: Cyclical wages in a search-and-bargaining model with large firms. In: *NBER International Seminar on Macroeconomics, 2006*, University of Chicago Press, 2008, 65–114.

[7] Thomas, C.: Search and matching frictions and optimal monetary policy. *Journal of Monetary Economics*. **55**(5), 2008, 936–956.

Problems Connected with Applying VaR for Determining Solvency Capital Requirement of Insurance Companies

Ondřej Pavlačka[1], Pavla Rotterová[2], Ondřej Nevídal[3]

Abstract. The aim of the paper is to point out some problems connected with the application of Value at Risk for determining Solvency Capital Requirement (SCR) of an insurance company within upcoming regulations Solvency II. We focus on the insurance risk. By several illustrative examples we show that SCR of the whole insurance portfolio is not generally bounded by the sum of SCRs of partial subportfolios, that SCR does not provide complete information about undertaken risk, that purchasing surplus reinsurance can increase SCR, and that the losses from dependent claims can be hidden.

Keywords: Solvency II, Value at Risk, solvency capital requirement, insurance.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

Determination of Solvency Capital Requirement (SCR) for an insurance company within upcoming regulations Solvency II (see [5] for more details) is based on the fact that the probability of a bankruptcy of an insurance company in a business year can be 0.5 % at most. From the mathematical point of view, SCR is given by a well-known risk measure called Value at Risk (VaR) with a level of confidence 0.995 and a time period one year (see e.g. [2]).

In the literature (see e.g. [1], [4]), it is presented that applying VaR for evaluating the risk capital brings some problems that should be taken into account: First, VaR is not a coherent risk measure (see [1] for more details) as it is not generally subadditive. This means that the VaR of a whole portfolio cannot be bounded by the sum of VaRs of particular subportfolios. Another problem is connected to the fact that VaR tells nothing about the possible big losses with probability of occurrence smaller than the chosen level of confidence (in our case 0.5 %). Such losses can significantly exceed VaR. Thus, VaR does not provide full information about undertaken risk. It could happen that two risk portfolios with the same VaR can have totally different behaviour with respect to the maximal possible losses.

The aim of the paper is to point out by several illustrative examples how the above mentioned general problems connected with VaR could affect an insurance company. The paper is organized as follows. In Section 2, the applied mathematical model for computing SCR is described. The illustrative examples mentioned above will be presented in Section 3. Finally, some concluding remark will be given in Conclusion.

2 Applied mathematical model

According to regulations Solvency II, the SCR will be computed as a sum of VaRs of market risk, credit risk, insurance risk and operational risk. In the paper, we will focus on the insurance risk, and for the sake of simplicity, we will consider only non-life insurance, namely property insurance.

Let us assume that we have an insurance portfolio X. By $L(X)$, we will denote the random variable expressing the total value of claims from policies belonging to X in a business year. The corresponding capital requirement $SCR(X)$ is then determined as follows:

$$SCR(X) = L(X)_{0.995} - \mu(X), \tag{1}$$

[1] Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, ondrej.pavlacka@upol.cz
[2] Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, pavla.rotterova01@upol.cz
[3] Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 12, 771 46 Olomouc, Czech Republic, ondrej.nevidal01@upol.cz

where $L(X)_{0.995}$ denotes the 0.995-quantile of $L(X)$, and $\mu(X)$ means the premium from X. For the sake of simplicity, as $\mu(X)$, we will further consider the average value of claims from X, i.e. $\mu(X) = E[L(X)]$. Thus, we can see that the capital equal to $\mu(X) + SCR(X)$ will cover the losses from insurance portfolio X with probability 99.5 %.

Further, we will construct also a so called *conditional value at risk* of an insurance portfolio X that expresses the expected capital necessary for covering the losses in case that $L(X)$ is greater than $L(X)_{0.995}$, i.e.

$$CVaR_{0.995}(X) = E[L(X)|L(X)>L(X)_{0.995}] - \mu(X).$$

In illustrative examples presented in Section 3, we will employ for evaluation of claims from a given insurance portfolio X in a business year the simulation model presented in [3] that is based on Monte Carlo method. The model generates the given number (in our case 500 000) of scenarios of claims from each policy belonging to X. The corresponding scenarios of $L(X)$ are then given as sums of claims from all policies. From these values, we can estimate $\mu(X)$ as the mean of the scenarios of $L(X)$, $L(X)_{0.995}$ as the corresponding empirical quantile of the scenarios of $L(X)$, and $CVaR_{0.995}(X)$ as the mean of the scenarios of $L(X)$ that are greater than $L(X)_{0.995}$.

In the model described in [3], the input data for each policy P_i belonging to an insurance portfolio X are the sum insured, the Probable Maximum Loss $PML(P_i)$ (for simplicity, we will assume here that the sum insured is equal to $PML(P_i)$ for each policy), the premium (in our case the expected claim) and the risk category given as a combination of a degree of frequency of claims and a degree of impact of claims. For the sake of simplicity, we will assume here that the claims from all policies are mutually independent (this assumption will be broken in Section 3.4). One scenario of claim from the policy P_i is in the model generated in the following way: To each degree of frequency, a corresponding probability distribution of a number of claims in one year is assigned. From this distribution, a number of claims is randomly generated. If it is equal to zero, then also the total claims from P_i are equal to zero. If the number of claims is positive, then for each claim a value from the unit interval [0,1] is randomly generated from a probability distribution corresponding to the degree of impact of P_i. This value expresses the ratio of a claim to $PML(P_i)$, so it is multiplied by $PML(P_i)$. Then, all individual claims are summed. This approach enables us to model the claims from an insurance portfolio with different policies.

In illustrative examples presented in Section 3, we will consider two degrees of frequency of claims – low frequency (F_1) and high frequency (F_2), and two degrees of impact - small impact (I_1) and large impact (I_2). The applied probability distributions of number of claims and of ratios of claims to PML are taken from [3]; they were fitted to data from industry insurance. In both degrees of frequency, the number of claims in one year is modelled by Poisson distribution with a parameter $\lambda = 0.07$ in case of F_1 and $\lambda = 0.17$ in case of F_2. For modelling the ratio of a claim to $PML(P_i)$, the following probability distributions on [0,1] are considered: First of all, it is set the maximal ratio of a standard claim to $PML(P_i)$ and the probability of its exceeding (the probability of a "catastrophe"). In case of small impact I_1, the maximal ratio of a standard claim to $PML(P_i)$ is 0.0001 and the probability of its exceeding is 2.7 %. In case of large impact I_2, the maximal ratio of a standard claim to $PML(P_i)$ is 0.01 and the probability of its exceeding is 5 %. Thus, first of all, it is decided whether a standard claim or a "catastrophe" occurs in a scenario. Then, for a standard claim as well as for a "catastrophe", the ratio of a claim to $PML(P_i)$ is modelled by a Beta distribution with modified boundaries – the boundaries for ratios of standard claims are 0 and the given maximal ratio of a claim to $PML(P_i)$, the boundaries for a "catastrophe" are the given maximal ratio of a standard claim to $PML(P_i)$ and 1. The parameters of Beta distribution are $\alpha = 2$ and $\beta = 4$ for a standard claim in I_1, $\alpha = 3$ and $\beta = 2$ for a standard claim in I_2, and $\alpha = 0.12$ and $\beta = 1.5$ for a "catastrophe". The densities are given in Fig. 1.

3 Some problems connected with applying VaR for determining SCR

In this section, we will illustrate by several examples some problems connected with the application of VaR for determining SCR of an insurance company. We will show that SCR of the whole insurance portfolio cannot be bounded by the sum of SCRs of partial subportfolios, that more risky insurance portfolio can have smaller SCR, that purchasing a surplus reinsurance can increase SCR, and that the loss from dependent claims can be hidden.

3.1 Lack of subadditivity

As it was mentioned in Introduction, VaR is not generally a subadditive risk measure. This means that the VaR of a whole portfolio cannot be generally bounded by the sum of VaRs of particular subportfolios, i.e. it does not hold that $VaR(X_1+X_2) \leq VaR(X_1) + VaR(X_2)$ for any risk portfolios X_1 and X_2.

Lack of this property means for an insurance company e.g. the following: Let an insurance company allocate the capital to several underwriters, and let each underwriter create such insurance portfolio that SCR corresponding to his/her portfolio is smaller than the capital allocated to him/her. Then it is not guaranteed that the SCR corresponding to the whole insurance portfolio created by putting all the underwriters' subportfolios altogether

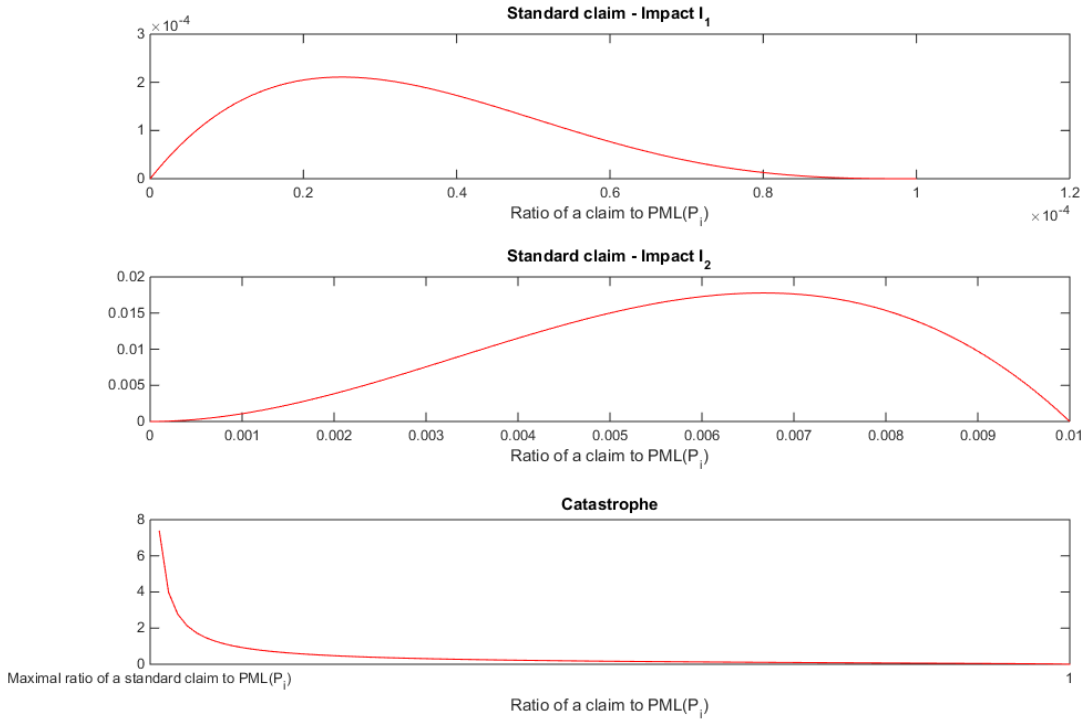


Figure 1 Probability densities of the ratios of claims to PML.

does not exceed the initial capital allocated to the underwriters. The problem is illustrated by the following example.

Example 1. Let us assume that we have a couple of exactly same insurance portfolios X_1 and X_2: both portfolios contain 10 000 policies, six of them have *PML* equal to 16 000 000 CZK, *PML*s of all other policies are equal to 100 000 CZK. All policies have low frequency of claims (degree F_1) and large impact of claims (degree I_2).

From the characteristics presented in Table 1, we can see that if we put X_1 and X_2 together, the corresponding SCR will be greater than the sum of SCRs of both subportfolios X_1 and X_2. It is caused by the fact that both subportfolios contain big losses hidden beyond the chosen level of confidence, but if we put them together, these losses come to the light.

	X_1	X_2	The sum of the characteristics	$X_1 + X_2$
$L(.)_{0,995}$	1 845 454	1 841 161	3 686 616	5 312 638
$\mu(.)$	764 507	764 507	1 529 014	1 529 013
$SCR(.)$	1 080 948	1 076 654	**2 157 602**	**3 783 625**
$CVaR_{0,995}(.)$	4 885 650	4 890 431	9 776 081	7 645 354

Table 1 Characteristics of insurance portfolios from Example 1

3.2 SCR does not provide complete information about the undertaken risk

Another problem connected with VaR is the fact that VaR does not tell anything about the possible big losses with probability of occurrence smaller than the chosen level of confidence. As determination of SCR is based on VaR, SCR does not provide full information about the undertaken risk by an insurance company. We will illustrate this problem on the following example where we will show that SCR of a potentially much more risky insurance portfolio will be smaller than SCR of an insurance portfolio that is safer.

Example 2. Let us assume the following couple of insurance portfolios:

- Insurance portfolio X_1 contains 10 000 policies that are completely homogenous. $PML(P_i)$ = 250 000 CZK, i = 1,…,10 000, and all policies have low frequency of claims (degree F_1) and large impact of claims (degree I_2).
- Insurance portfolio X_2 contains 10 000 policies. All but one policies have the same, low, *PML*: $PML(P_i)$ = 100 000 CZK, i = 1,…,9 999. *PML* of the last policy is non-adequately large: $PML(P_{10\,000})$ = 10 000 000 CZK. All policies have high frequency of claims (degree F_2) and large impact of claims (degree I_2).

The characteristics of both insurance portfolios are shown in Table 2. We can see that the premiums (the average claims in our case) from both portfolios are approximately the same. The $SCR(X_1)$ is significantly greater than $SCR(X_2)$. However, according to the standard deviations of claims as well as according to CVaRs, insurance portfolio X_2 is more risky than X_1. The problem is that the biggest claims from X_2 are hidden beyond the chosen level of confidence, so X_2 offers a chance for big revenue but there is a small probability of large loss. The histograms of the claims are shown in Fig. 2.

	X_1	X_2
$L(.)_{0,995}$	2 581 232	2 306 083
$\mu(.)$	1 744 812	1 738 727
SCR(.)	**836 420**	**567 356**
Standard deviation	275 692	328 454
$CVaR_{0,995}(.)$	963 535	2 423 363

Table 2 Characteristics of insurance portfolios from Example 2

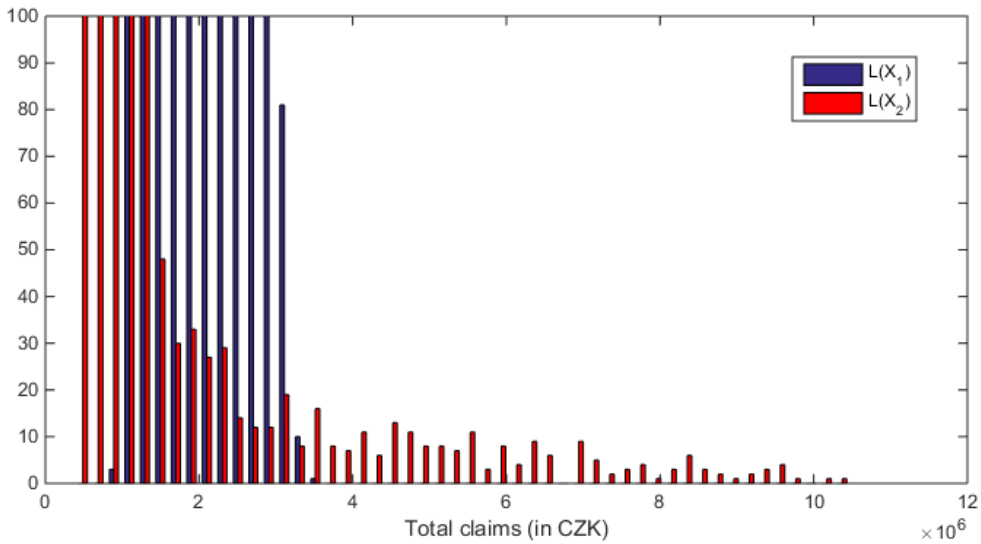


Figure 2 Detailed cut of histograms of $L(X_1)$ and $L(X_2)$

3.3 Ceding a risk can increase SCR

The anticipated property of any way of determination of the solvency capital requirement of an insurance company from insurance risk is the fact that purchasing reinsurance has a positive effect, i.e. if an insurance company cedes a risk to a reinsurer, its solvency capital requirement from insurance risk should be less (but it causes a growth of a credit risk). In this section, let us show an example that in case of SCR given by (1), purchasing reinsurance can in some cases increase SCR from insurance risk.

In the following example, we will consider *surplus reinsurance*. An insurance company may seek surplus reinsurance to limit the losses it might incur from a small number of large claims. Under a surplus reinsurance, the ceding insurance company decides on a "retention limit". The ceding insurance company retains the full amount

of each risk, with a maximum of the chosen retention limit per policy, and the rest of the risk is reinsured (for simplification, we will not consider the maximum automatic underwriting capacity of the reinsurer). For instance, if a retention limit is 100 000 EUR and the insurance company issues a policy for 100 000 EUR, it would keep all of the premiums and losses from that policy. If it issues a 200 000 EUR policy, it would give half of the premiums and losses to the reinsurer.

Example 3. Let us consider the following insurance portfolio X: 1 000 policies, 997 of them have *PML* equal to 100 000 EUR, the rest 3 policies have *PML* equal to 100 000 000 EUR. All policies have low frequency of claims (degree F_1) and small impact of claims (degree I_1). Let us assume that the insurance company that owes X has a possibility to purchase surplus reinsurance with a retention limit 100 000 EUR with no maximum underwriting capacity of the reinsurer.

In Table 3, we can see that purchasing surplus reinsurance would mean for the insurance company giving up most of the premiums. Since the big losses from the three policies with large *PML* are not much probable and are not fully displayed in $SCR(X)$, the resulting SCR in the case of purchasing the surplus reinsurance would be greater than SCR corresponding to X without purchasing the surplus reinsurance. As we can see from $CVaR_{0,995}(X)$, if the insurance company would not purchase the surplus reinsurance, it will face a much more risk.

	Case without surplus reinsurance	Surplus reinsurance purchased
$L(X)_{0,995}$	144 861	122 828
$\mu(X)$	58 949	14 748
$SCR(X)$	**85 912**	**108 080**
$CVaR_{0,995}(X)$	8 716 490	130 351

Table 3 Characteristics of the two cases from Example 3

3.4 Dependent claims could be hidden

In previous examples, we assumed that claims from the different policies are mutually independent. However, some catastrophic event, like flood, earthquake, etc., can occur in real life. This typically means that claims approaching *PML* are reported from almost all policies from the affected region. Fortunately, probability of occurrence of such disasters is very low. In the following example, we will show that setting the value of probability of such event is the key element for determination of SCR. If the probability is smaller than the level of confidence (0.5 %), the dependent claims remain hidden.

Example 4. Let us consider the following insurance portfolio X: 1 000 policies with *PML* equal to 100 000 EUR. All policies have low frequency of claims (degree F_1) and small impact of claims (degree I_1). Table 4 shows the characteristics of the three cases: First column shows the results if claims from the policies were mutually independent. Claims in the next two cases are dependent in the following way: With a certain probability p, a catastrophic event occurs such that from all policies from X the maximal losses (claims equal to *PML*) are realized. If the catastrophic event does not occur, the claims are generated standardly. Second column in Table 4 shows the case where p = 0.49 %, i.e. p is smaller than the level of confidence 0.5 %. As we can see, in such a case $SCR(X)$ is even negative as the catastrophic event is displayed in $L(X)_{0,995}$ only partly while it fully affects the premium $\mu(X)$. On the contrary, if p = 0.5 %, the catastrophic event is fully revealed in $SCR(X)$ (see the last column of Table 4).

	Independent claims	Dependent claims, p = 0.49 %	Dependent claims, p = 0.5 %
$L(X)_{0,995}$	122 148	199 938	50 119 266
$\mu(X)$	14 748	504 748	514 748
$SCR(X)$	**107 400**	**-304 810**	**49 604 518**

Table 4 Characteristics of the considered cases from Example 4

4 Conclusion

We have focused on the possible problems connected with the upcoming change of the way of determination of Solvency Capital Requirement of an insurance company. The determination will be based on the risk measure Value at Risk that is known for its problematic behavior in some cases. By several illustrative examples we show that SCR of the whole insurance portfolio cannot be generally bounded by the sum of SCRs of partial subportfolios, that SCR does not provide complete information about the undertaken risk, that purchasing surplus reinsurance can increase SCR, and that the losses from dependent claims can be hidden.

Further research will be aimed on the analysis of the possible indicators of the revealed problems

Acknowledgements

The research is supported by the *Insurance education support foundation* and by the grant IGA_PrF_2015_013 *Mathematical Models* of the *Internal Grant Agency of the Palacký University Olomouc*. The support is greatly acknowledged.

References

[1] Artzner, P., Delbaen, F., Eber, J.–M., Heath, D.: Coherent Measures of Risk. *Mathematical Finance* **9** (1999), 203–228.

[2] Hull, J. C.: *Risk Management and Financial Institutions*. Pearson Prentice Hall, Upper Saddle River (USA), 2007.

[3] Pavlačka, O.: *Aplikace teoretických postupů pro ocenění rizika při upisování pojistných smluv v oblasti velkých rizik*. Insurance education support foundation [online]. 2011 [cited 2015-05-15]. Available at: < http://www.nfvp.cz/aplikace-teoretickych-postupu.html>.

[4] Strnad, P.: Řízení tržních rizik pomocí Value at Risk – úskalí a problémy. *Economic Review* **1** (2009), 91-102.

[5] Solvency II. European Insurance And Occupational Pensions Authority. EIOPA [online]. 2015 [cited 2015-04-11]. Available at: < https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii>.

The profitability and technical efficiency of young farmers supported from Rural Development Programme

Marie Pechrova[1]

Abstract. The aim of the paper is to assess the profitability and technical efficiency of young farmers in order to conclude whether the financial support from Rural Development Programme (RDP) under measure *1.3.1 Setting up young farmers' businesses* had desirable effect. Financial indicators (return on assets – ROA, return on equity ROE and return on sales – ROS) and technical efficiency (using Data Envelopment Analysis) are calculated. The development is observed from 2008 to 2013 to match the major part of the functioning of the RDP. The sample consists of 11 young farmers for which overall technical efficiency (CCR model), pure technical efficiency (BCC model) and scale efficiency were calculated. It cannot be clearly concluded whether the farms improved after receiving the sup-port as sometimes ROA, ROE or ROS increased after obtaining the subsidy, but sometimes otherwise. The farmers were overall efficient from 48.5% and pure technical efficiency from 76.6%. Scale efficiency was lower (65.5%) as there were more farms with decreasing than those with increasing returns to scale. The effect of subsidies on technical efficiency is ambiguous. There was no statistically significant relation found between the number of years from subsidy receiving and technical efficiency.

Keywords: Data Envelopment Analysis, young farmer, Rural Development Programme, technical efficiency.

JEL classification: C61, H25
AMS classification: 90C05

1 Introduction

The age structure of the Czech workers in agriculture is not favourable see e. g. study of Simpach, Pechrova [11]. Majority of farmers belong to the category between 50 and 54 years. Therefore, the steps are taken to encourage the young people to enter the agribusiness. There are various ways through which the structural change is encouraged. Young farmers might have tax reliefs or they receive special financial support. Common Agricultural Policy of the European Union also provided special support from Rural Development Programme (RDP) for the period of 2007–2013 under measure *1.3.2 Setting up of young farmers' business.* For the next programming period 2014–2020 the support is expected to continue in the framework of operation *6.1.1 Establishing of young farmers' business.* A person which is younger than 40 years is considered to be a young farmer. "The granting of specific benefits to young farmers may facilitate both their initial establishment and the structural adjustment of their holdings after their initial setting up" (EC, 2005). There were six calls for submission of the application for a grant (1^{st} round in 2007, 3rd in 2008, 9^{th} in 2010, 12_{th} in 2011 and 16^{th} in 2012) announced by the MoA. In total, there were 3 606 projects in a value over 3.9 bld. CZK registered, from which 1 286 projects in a value of 1.3 bld. CZK were repaid by 30^{th} September 2014. In total, 1 105 young farmers were supported usually by 1.1 mil. CZK each. Figure 1 shows the development of the number of projects supported from RDP measure *1.3.2 Setting up of young farmers' businesses* in all rounds of calls for grants. The general assumption is that the public money should be efficiently spent. "There is therefore a need for methods and approaches which enable the evaluator to assess precisely the mechanisms by which beneficiaries are responding to the intervention" (Ratinger et al. [9]). According to Amores and Contrerase [2] the subsidies should not be firstly granted and then hoped that they will have desirable effect and will stimulate the performance

[1]Institute of Agricultural Economics and Information, Manesova 1453/75, Prague, pechrova.marie@uzei.cz

of the agricultural holding. They should rather be distributed as "annual reward" for achieved results. However, the support for young farmers is specific as they should only just set up their business and might not have previous results. Therefore, the efficiency can be assessed only later.

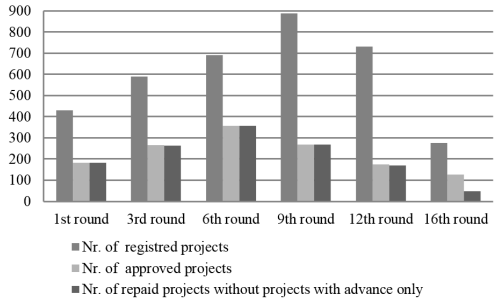


Figure 1 Nr. of subsidised projects of young farmers in 2007–2012; Source MoA (2014), own display.

The aim of the paper is to assess the profitability and technical efficiency of young farmers (≤ 40 years) supported from RDP 2007–2013 in the framework of measure *1.3.1 Setting up young farmers' businesses*. First, we describe used financial analysis indicators and method for calculation of the technical efficiency. Then the results are presented. Consequently they are discussed and conclusions are made.

2 Methodology

There were 1 105 farmers supported from RDP measure *1.3.1 Setting up young farmers' businesses* during the whole functioning period of the programme. Most of them were natural persons for which it is harder to obtain data. On the other hand, legal persons are obliged to report their financial statements to the business register and their data are publically accessible. Therefore, the sample consists of 11 young farmers legal persons to which the payments were granted (for maximally five years). The development is observed from 2008 to 2012 to cover the main period of functioning of the RDP. It is expected that the effect of provided finances might be delayed; therefore the year 2007 was not included. Also the last available data in business register were for year 2012. The accountancy data were obtained from Czech business register and about subsidies from State Agricultural Interventional Fund (SAIF). First, standard financial analysis indicators are utilized. The profitability and the impact of subsidies are assessed by financial indicators of profitability: return on assets (ROA), return on equity (ROE) and return on sales (ROS) (see e.g. Pechrova [7]). None of the indicators should achieve negative value. ROA is calculated as the division of earnings before interests and taxes and assets. It gives information how effectively are the total assets used. It shows the extent to which companies manage to generate profit from the available assets. Similarly ROE gives information about the ability of a company to generate profit, but for investors in this case. It is a key indicator for shareholders and partners. It is calculated as a share of $EBIT$ on equity. ROS tells how much profit generates 1 CZK of sales. Second, the technical efficiency of the beneficiaries is calculated and the impact of subsidies on it is assessed. Farrell [4] defined technical efficiency as the ability of a farm to produce the maximum feasible output from a given bundle of inputs or to use minimum feasible amounts of inputs to produce a given level of output. Depending on what type of definition is taken into account the "output-oriented" and "input-oriented" model is constructed. A compensation is accepted (higher outputs need more inputs maintaining the consumption efficiency) (Cooper et al. [3]). First approaches towards calculations of technical efficiency were non-parametric, based on linear programming, and deterministic. To this type of methods belongs Data Envelopment Analysis (DEA). Consequently parametric methods base on econometric modelling were elaborated. The most used is Stochastic Frontier Analysis. DEA enables to treat multiple inputs and outputs. The combinations of both for each decision making unit (DMU) create a field of points which DEA envelopes by a frontier so all units lie either on or under the frontier. DEA is more appropriate when there are less DMUs farms in our case as "it is not significantly affected by a small sample size, as long as the number of inputs is not too high in comparison to the sample size" (Pechrova [7]). The number of units should be higher than the number of inputs (n) multiplied by the number of outputs (m) or higher than three times the sum of inputs and outputs (1).

$$q \geq max\{n \times m, 3(n+m)\} \tag{1}$$

According to Speelman et al. [10], other advantage of DEA is that it permits the construction of a surface over the data, which allows the comparison of one production method with the others in terms of a performance index. This efficiency score vary between 0 (0%) and 1 (100%). DMU is (100%) efficient when there is no any other unit maintaining the same level of outputs with lower level of inputs. Units with the highest efficiency are located on the efficient frontier. Others DMUs should improve their performance. The disadvantage of DEA is its deterministic character. Because of that it is sensitive to extreme values, measurement errors and other noise in the data. Besides, as DEA assumes free disposability and convexity assumptions, it is further restricted by making an assumption on the shape of the convex hull or convex cone (Pechrova [8]). It means that an assumption about the returns to scale (RTS) of the production technology must be made. DEA models can be classified according to RTS as CCR and BCC. First one elaborated by Charnes, Cooper a Rhodes (CCR) calculates overall technical efficiency (TE) and assumes constant returns to scale (CRS). It implies that when the amount of input increase by one, also the amount of output increases by one. BCC model introduced by Banker, Charnes a Cooper (BCC) calculates pure technical efficiency (PTE) and assumes variable returns to scale (VRS). This suggest that when the amount of input increase by one, the amount of output can decrease or increase by any value. The increasing (IRS) or decreasing (DRS) RTS are distinguished. The paper calculates both models as scale efficiency (SE) can be derived from them as (2):

$$SE = TE(TE\ CRS)/PTE(TE\ VRS) \tag{2}$$

An input oriented two stage DEA was chosen, because the approach of minimization of inputs assumes that farms' inputs should be used as economically as possible (they represent costs to the company). Output oriented DEA, on the other hand, maximizes the outputs while the amount of inputs is given. CCR model maximizes the efficiency of the DMUs as the division of weighted outputs and weighted inputs. Constrains limit the efficiency of other DMUs to be ≤ 1. For each DMU_q ($q = 1...\ 11$), each input x is weighted by v_j, $j = 1, ...\ n$ and virtual input is gained, and each output y is weighted by u_i, $i = 1, ...\ m$ and virtual output is get. Using Charnes-Cooper transformation (see e.g. Jablonsky [6]) primal CCR model is formulated as (3):

$$\begin{aligned}
\max \quad & z = \sum_{i=1}^{m} u_i y_{iq}, \\
\text{s.t.} \quad & \sum_{i=1}^{m} u_i y_{ik} \leq \sum_{i=1}^{m} v_j x_{jk}, \qquad k = 1, 2, ..., 112, \\
& \sum_{j=1}^{n} v_j x_{jq} = 1, \\
& u_i \geq \varepsilon, \qquad i = 1, 2, ..., m, \\
& v_j \geq \varepsilon, \qquad j = 1, 2, ..., n,
\end{aligned} \tag{3}$$

where z is efficiency of DMU_q, x_{jk} is value of j^{-th} input for DMU_k, y_{ik} is value of i^{-th} output for unit DMU_k, ε is infinitesimal constant ensuring that all weights of inputs and outputs will be positive and included in a model. The value is usually 10^{-8}. Formulation of primal BCC model (4) is similar to CCR, but includes variable μ taking arbitrary value, which ensures the convexity. There are non-increasing RTS when $\mu \leq 0$ and non-decreasing RTS when $\mu \geq 0$.

$$\begin{aligned}
\max \quad & z = \sum_{i=1}^{m} u_i y_{iq} + \mu, \\
\text{s.t.} \quad & \sum_{i=1}^{m} u_i y_{ik} + \mu \leq \sum_{i=1}^{m} v_j x_{jk}, \qquad k = 1, 2, ..., 112, \\
& \sum_{j=1}^{n} v_j x_{jq} = 1, \\
& u_i \geq \varepsilon, \qquad i = 1, 2, ..., m, \\
& v_j \geq \varepsilon, \qquad j = 1, 2, ..., n, \\
& \mu = \text{any}
\end{aligned} \tag{4}$$

There are $n = 3$ inputs, and $m = 1$ output and 11 DMUs. As an output y was considered the performance (production) and as inputs x_1 production consumption, x_2 number of employees and x_3 land of the

agricultural holdings. As the number of observations was not equal for all farms, a pooled DEA model from average data values for years 2008–2012 was constructed. The calculations were done in Stata 11.2.

3 Results and Discussion

First, the financial indicators of young farmers supported from RDP are presented. Second, the technical efficiency is assessed, than the discussion follows. *EBIT* of the farms should be positive, but from 35 observations it achieved 6 times negative values. This influences the calculations of profitability indicators. As the amount of assets cannot take values lower than zero, *ROA* was also negative in 6 cases. On average, *ROA* achieved value 0.03 which tells that the farmers are not quite efficient in generating profit (*EBIT*) from their assets because the higher is the value, the better the firms uses their resources. As it can be seen from Figure 2, it was developing around zero during the examined period achieving negative value in year 2010. The company number 3 had negative value in both observed years, despite that it obtained subsidy in 2008. On the other hand, while the company number 1 had negative value of *ROA* in 2008, the fact that it received a subsidy in the same year might help it to gain positive *ROA* later. Firm number 6 also had negative *ROA* in 2010, but when it received subsidies in 2011, its value was positive in 2012. In majority of cases the farms had positive *ROE* as they equity was above zero. However, company number 2 achieved negative equity in 2008. It received subsidy for young farmer at that year and since that its equity (and hence *ROE*) was positive. On average, ROE took value of 0.25 suggesting that the share of *EBIT* on equity was only one quarter. This key indicator for owners of the farms telling them about profit generated on their own invested capital is not that positive. The investors have to invest 4 times more capital than they actually "receive back" in a form of profit. Third indicator was always on average positive. Some companies in some years even did not report any sales (nor of goods nor of products). They are excluded from the analysis. Only in 4 cases the individual values were negative. In 4 cases, the sales of farms were not reported at all, therefore, those were not included into analyses. *ROS* were the highest in 2009, when the *EBIT* and sales were almost equal. The development of all profitability indicators shows that the economic crisis in 2008 influenced them with two years delay.

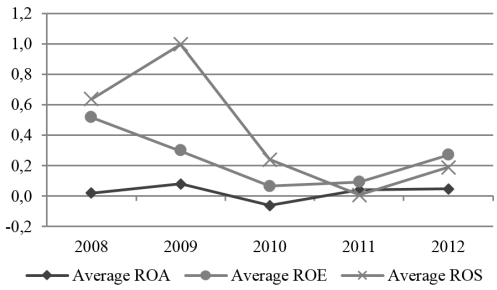


Figure 2 Development of farms' average *ROA*, *ROE* and *ROS* in 2008–2012; Source Own elaboration.

The amount of subsidies was 1.10 mil. CZK for all farms (except for farm 2 which received only 1.05 mil. CZK). Four farms received subsidies in 2008, three in 2011, two in 2007 and in 2010. Hence, each farm could utilize the RDP subsidies for different period. Basic year is 2013 (one year after 2012 which is the last year for which the accountancy data were available). However, no dependence can be seen or calculated. It seems that the development of particular indicator is not influenced by the fact whether the company obtained subsidies. It cannot be clearly concluded whether the effectiveness improved after receiving the support. Sometimes *ROA*, *ROE* or *ROS* increased after obtaining the subsidy, but sometimes otherwise. The subsidies were granted to the firms with negative *ROA* and *ROE*, but did not help in all cases to change them to positive values. Farms with hither *ROA* tend to have also higher *ROS* as the Spearman correlation coefficient achieved value 0.8743 and was statistically significant. On average, the farmers were overall efficient from 48.51% under CRS. It is due to the fact that there were only two farms 100% efficient. The lowest TE accounted only for 15.70%. Pure technical efficiency calculated under VRS was higher (76.55%) as more farms were 100% efficient. This is due to the fact that the envelope of the data is convex. Beside two farms efficient as under CRS there were five others. On the other hand, the same farm which had the lowest TE had also the lowest PTE (17.27%). In both cases there is still a space for improvement, because the production potential of farms is not fully utilized. Scale efficiency was lower than PTE (65.50%) as there were more (five) farms with decreasing than those

with increasing returns to scale (four). Based on the type of RTS an optimal size of the farm can be assessed. There were only two agricultural companies with equal TE and PTE and hence had optimal size. Those farms number 6 and 10 can serve as an example to others. It might be because firm number 6 benefited from subsidies of CZK 1.1 bld. for 5 years. However, the farm number 10 had subsidy only for 2 years. Two 100% efficient farms also differ in terms of production and production factors. Farm number 10 produces the goods in average value of 103 thous. CZK and spent 224 thous. CZK on the production, the other produce the amount of 1 430 thous. CZK (14 times more) with consumption of 6 854 thous. CZK (30 times more) on 407 hectares (54 times more). However, farm number 2 needs fewer employees (10 instead of 12 as it is in the smaller farm). Both are, nevertheless, the best.

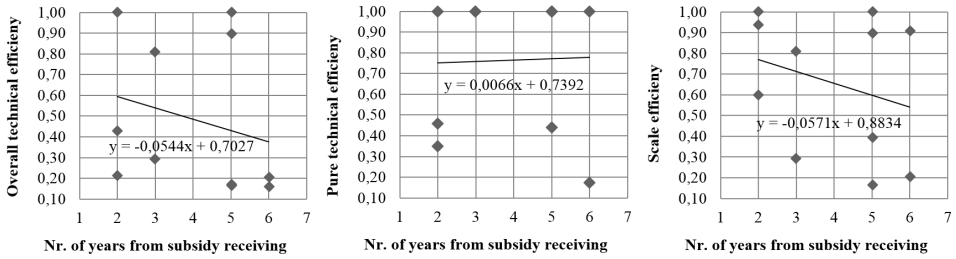


Figure 3 All types of technical efficiency

There is an opposite situation. Despite that the farmer number 5 received subsidy already in 2007 and could benefit from it also for the whole eligible five-year-period, the *TE* and *PTE* of this company was the lowest. Despite that, its *SE* was relatively high (90.89%). It produces almost 4 times more than farm number 10, but need also almost 4 times more land, twice as higher production consumption, but only 1 employee. Also other results show that the effect of subsidies on technical efficiency is ambiguous. Graph 3 below show the linear relation between the number of years from receiving of the subsidy by a young farmer and all types of technical efficiency. It is mild and not statistically significant. Also the Spearman correlation coefficients between the number of years since the subsidy was received by the farmer and the *TE* (-0.4897), *PTE* (0.0109) and SE (-0.3099) are not statistically significant. It seems that the relation is positive only in case of *PTE*. However, we have to conclude that there was no statistically significant relation found between the number of years from subsidy receiving and technical efficiency. The regression coefficients displayed in the graph were not statistically significant and hence could not serve as the indicator of dependence between the technical efficiency and year from receiving the subsidies. Contrary to that Pechrova [8] concluded that "to a certain level, the RDP subsidies have a positive and statistically significant impact on the technical efficiency" of Czech farm in years 2007–2013. However, she did not consider the division on young and old farmers. The correlation between economic indicators and technical efficiency was also not statistically significant. Therefore, other factors than the number of years of subsidy usage influence the technical efficiency. The reason why the young farmers are supported is their age. Therefore, the technical efficiency should be positively influence by it. For example Amudayi et al. [1] found that the effect of age was consistent with expectation that older farmers in Western Kenya could be technically less efficient than younger farmers in crop protection technology uptake. Also Speelman et al. [10] concluded similarly. He examined small-scale irrigation in South Africa and found that despite the fact that older farmers are more experienced in applying various practices; their technical efficiency is lower than of the younger farmers. Speelman et al. [10] suggest that it is due to the fact that older farmers are often less willing to adopt new ideas. Pius, Odjuvwuederhie [5] for Southeastern Nigeria found "farmers with more years of farming experience are relatively less efficient in yam production, and vice versa. The farmers' years of experience correlates with their ages. Furthermore, younger farmers are seemingly more progressive, as they demonstrate a greater willingness to adopt new practices that raise their overall level of efficiency". Even without subsidies the young farmers are technically more efficient than their older counterparts just because of their age. Therefore, non-economic reasons for support may be more significant for justification of the support for young farmers.

4 Conclusion

The aim of the paper was to assess the profitability and technical efficiency of young farmers in order to conclude whether the financial support from Rural Development Programme (RDP) under measure

1.3.1 Setting up young farmers' businesses had desirable effect. A profitability indicators and technical efficiency were calculated on a sample of 11 young farmers supported in the programme period 2007–2013. From financial analysis point of view, Returns on Assets (ROA), Returns on Equity (ROE) and Returns on Sales (ROS) should have improved after receiving the subsidies. However, it seems that the development of particular indicator is not influenced by the fact whether the company obtained subsidies. Sometimes ROA, ROE or ROS increased after obtaining the subsidy, but some-times otherwise. The subsidies were also granted to the firms with negative ROA and ROE, but did not help in all cases to change them to positive values. Similarly in the case of the technical efficiency, the effect is not clear. The technical efficiency (TE), pure technical efficiency (PTE) and scale efficiency (SE) of farms were calculated using input oriented non-parametric Data Envelopment Analysis (DEA). An average TE was calculated at the level of 48.51%, average PTE at 76.55% and average SE at 65.50%. There were two farms managed by young farmers 100% efficient in terms of TE and PTE (and hence working with optimal size i.e. SE equal to 100%). One farm profited of subsidies for 5 year another one for only 2 years. Contrary to that, there was one farm which obtained subsidy already in 2007 (which until 2013 is 6 years later), but its TE and PTE was the lowest. Also other relations are ambiguous. Besides, there was no statistically significant relation found between the number of years from subsidy receiving and technical efficiency. Therefore, non-economic reasons for support may play more important role in justification of the support for young farmers.

Acknowledgements

The research was financed from grant 6/2014 of the Institute of Agricultural Economics and Information.

References

[1] Amudavi et al.: Assessment of technical efciency of farmer teachers in the uptake and dissemination of pushpull technology in Western Kenya. *Crop Protection* **28**: 987-996, (2009).

[2] Amores, A., and Contreras, I.: New approach for the assignment of new European agricultural subsidies using scores from data envelopment analysis: Application to olive-growing farms in Andalusia (Spain). *European Journal of Operational Research* **193**: 718729, (2009).

[3] Cooper, W., et al.: *Data Envelopment Analysis: History, Models, and Interpretations.* Texas, (2007).

[4] Farrell, M. J.: The measurement of productive efficiency. *Journal of the Royal Statistical Society A* **120**: 253-281, (1957).

[5] Pius, I., Odjuvwuederhie, I.: Determinants of yam production and economic efficiency among smallholder farmers in Southeastern Nigeria. *Journal of Central European Agriculture* **7**: 337–342, (2006).

[6] Jablonsky, J.: MS Excel based software support tools for decision problems with multiple criteria. *Procedia Economics and Finance* **12**: 251–258, (2014).

[7] Pechrova, M.: Comparison of financial situation of the biodynamic farms in the CR. *The International Scientific Conference IMPROFORUM 2013*, 207–212, (2013).

[8] Pechrova, M.: The Influence of Location, Legal Form and Size on the Efficiency of the Functioning of Czech Local Action Groups. *Proceedings of the Agrarian Perspectives XXIII.*, 121–128, (2014).

[9] Ratinger, T., Medonos, T., Spicka, J., Hruska, M., Vilhelm, V.: The assessment of the effects of investment support measures of the Rural Development Programmes: the case of the Czech Republic. *Proceedings of the Joint Conference on 21st International Scientific Conference on Agrarian Perspectives / 131st EAAE Seminar*, 96–111, (2012).

[10] Speelman, S., D'Haese, M. and Buysse, J.: A measure for the efficiency of water use and its determinants, a case study of small-scale irrigation schemes in North-West Province, South Africa. *Agricultural Systems* **98**: 31-39, (2008).

[11] Simpach, O., Pechrova, M.: The Impact of Population Development on the Sustainability of the Rural Regions. *Proceedings of the Agrarian perspectives XXIII.*, 129–136 (2014).

Measuring solution robustness

Jan Pelikán[1], Maria Kobzareva[2]

Abstract. Nowadays a great research interest is devoted to uncertainty modeling, as changing data, uncertainty, and dynamic modifications of data can cause current solution achieved with optimization technique to become non-optimal and present poor results from the objective function point of view. One of the popular approaches which deals with problem with uncertain data is robust optimization, which is designed to find solutions that are robust, which means that the solution does not depend on changes of certain parameters. There are two types of robustness – quality and solution robustness. Quality robustness is a property of a solution whose quality, measured by the objective function value, does not deviate much from optimality when small changes in the problem data occur. Solution robustness can then be described as robustness in the solution space. This paper focuses on quality robustness – analysis of optimal solution and calculating its stability interval using proposed approaches based on interval programming. Stability of solution is calculated according to the possible changes in objective function coefficients only. Proposed method is compared with sensitivity analysis and parametric programming, main differences are outlined. The method is illustrates on a case study, results are provided.

Keywords: robustness, sensitivity analysis, interval programming, parametric programming

JEL Classification: C44, C6, D81

AMS Classification: 90C05, 90-08, 93B35

1 Introduction

Uncertainty is a topic which gets more and more attention in the last decades. Nowadays in the world that changes every day stability is a rare quality to find. Many methods and approaches work with optimization problems where data are considered to be deterministic, in other words we assume input data to be known precisely at the moment we solve the problem. Unfortunately, in reality it is rarely the case, as uncertain elements are the part of every problem, are we talking about vehicle routing problem, flow optimization problems, allocation problems and others. For example, if we are planning optimal route for a vehicle we have to take into account the fact that the route could be extended due to traffic jam, traffic light malfunction, traffic accident, road reconstruction or bad weather conditions. Similar situation could occur at the securities market, where predicted future profits of shares can be affected by a number of factors, which are impossible to predict, or when locating a material storage in allocation problem, where the demand at different centers can change unexpectedly. If we assume the input data are exact and will not change under any circumstances and in reality the changes in given inputs will occur, we can find ourselves in the situation when the generated optimal solution violates critical constraints, can become non-optimal and present non-efficient results from the objective function point of view. That's why in the changing world it is important to take into account all possible factors that can affect the generated solution when optimizing. That is the aspect this paper is going to work with – searching for an optimal solution under uncertainty and measuring solution robustness, which can calculate the possible change in objective function coefficients, for which the calculated solution will remain optimal.

This paper introduces new method for judging solution robustness based on interval theory. The idea is to calculate a certain interval for changes in objective function coefficients, which assures, that the solution, obtained with the new objective function coefficients is still optimal. The key benefit of proposed method is the fact, that we are not only looking into increases in objective function coefficients, but decrease of coefficients inside the interval is also possible. The other key benefit is the fact, that we can change all objective

[1] University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: pelikan@vse.cz

[2]University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq.4, Prague 3, Czech Republic, e-mail: mariakobzareva@gmail.com

function coefficients at the same time and the changes must be inside the interval but amount of changes must not be the same for all coefficients.

2 Preliminaries

For introduction we would like to present basic operations with intervals, which is the foundation for proposed method.

For the purposes of our study let us define an interval as a set (1)

$$\tilde{x} = [\underline{x}, \overline{x}] := \{x \in R, \underline{x} \leq x \leq \overline{x}\}, \quad (1)$$

where $\underline{x}, \overline{x}$ are elements from R and $\underline{x} \leq \overline{x}$. The interval is defined as thin when $\underline{x} = \overline{x}$ and wide when $\underline{x} < \overline{x}$. We can also understand interval as space, where unknown $\tilde{x}$ belongs.

2.1. Interval operations

Let us define summation (2) and subtraction (3) of intervals. For summation and subtraction of intervals lower and upper bound of an outcome interval can be defined as follows:

$$\tilde{x} + \tilde{y} = \left[\underline{x} + \underline{y}, \overline{x} + \overline{y}\right], \quad (2)$$

$$\tilde{x} - \tilde{y} = \left[\underline{x} - \underline{y}, \overline{x} - \overline{y}\right]. \quad (3)$$

Now let's define multiplication (4) and (5) of an interval with a number *k*.

$$for\ k \geq 0 \qquad k * \tilde{x} = \left[k * \underline{x}, k * \overline{x}\right], \quad (4)$$

$$for\ k < 0 \qquad k * \tilde{x} = \left[\, k * \overline{x}, k * \underline{x}\right] = (-k) * (-\tilde{x}). \quad (5)$$

To consider, whether one interval is greater than another we can use the following definition (6) and (7):

$$\tilde{x} \leq \tilde{y} \quad \text{is defined as} \quad \overline{x} \leq \underline{y}, \quad (6)$$

$$x \geq y \quad \text{is defined as} \quad \underline{x} \geq \overline{y}. \quad (7)$$

Negative interval (8) can be then described as follows:

$$-\tilde{x} = \left[-\overline{x}, -\underline{x}\right]. \quad (8)$$

3 Measuring solution robustness with interval programming

Let us now introduce the method for measuring optimal solution robustness based on operations with intervals and it's properties.

Let us assume that x_0 is a basic optimal solution for the following maximization problem:

$$\max\ z = c^T x,\ Ax \leq b,\ x \geq 0. \quad (9)$$

We will define cost robustness index ε of the optimal solution x_0 of the problem (9).

Definition 1. *For basic optimal solution x_0 of the problem (9) $\varepsilon \geq 0$ is **robustness index** if the solution x_0 is optimal for all*

$$c \in [{c_1}^0 - \varepsilon, {c_1}^0 + \varepsilon] \times [{c_2}^0 - \varepsilon, {c_2}^0 + \varepsilon] \times \ldots \times [{c_n}^0 - \varepsilon, {c_n}^0 + \varepsilon]. \quad (10)$$

Definition 2. *For basic optimal solution x_0 of the problem (9) $\varepsilon_p \geq 0$ is* ***percentage robustness index*** *if x_0 is optimal for all*

$$c \in {c_1}^0\ [1-\varepsilon_p, 1+\varepsilon_p] \times {c_2}^0[1-\varepsilon_p, 1+\varepsilon_p] \times \ldots \times {c_n}^0[1-\varepsilon_p, 1+\varepsilon_p]. \tag{11}$$

Cost robustness index ε assumes that a positive or negative change in any objective function coefficient (or at more coefficients at the same time) will not affect optimal solution of the problem, as long as the possible change is not greater than calculated robustness index ε.

We will use the interval operations for the estimation of ε robustness and ε_p percentage robustness index . Let the optimal solution x_0 correspond with the optimal simplex table in the form of the matrix (12)

$$\begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_2 & \ldots & \bar{\alpha}_n & \bar{b} \\ z_1 & z_2 & \ldots & z_n & z_0 \end{bmatrix}. \tag{12}$$

If the price vector is in interval (13)

$$c \in ([{c_1}^0 - \varepsilon, {c_1}^0 + \varepsilon] \times [{c_2}^0 - \varepsilon, {c_2}^0 + \varepsilon] \times \ldots \times [{c_n}^0 - \varepsilon, {c_n}^0 + \varepsilon]), \tag{13}$$

and the base price vector is

$$c_B \in \left(\left[{c_{B_1}}^0 - \varepsilon, {c_{B_1}}^0 + \varepsilon\right] \times \left[{c_{B_2}}^0 - \varepsilon, {c_{B_2}}^0 + \varepsilon\right] \times \ldots \times \left[{c_{B_m}}^0 - \varepsilon, {c_{B_m}}^0 + \varepsilon\right]\right), \tag{14}$$

then *z*-coefficients are

$$\tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n), \tag{15}$$

where

$$\tilde{z}_j = \tilde{c}_B \bar{a}_j - \tilde{c}_j. \tag{16}$$

The solution x_0 will still be optimal if $\tilde{z}_j \geq 0$ for all $j = 1,2,\ldots,n,$ where x_j is nonbasic. The coefficients $\tilde{z}_j \geq 0$ can be solved for ε, so that ε robustness index is the result of the condition $\tilde{z}_j \geq 0$, considering the fact that for basic x_j the condition is fulfilled as $\tilde{z}_j = 0$.

We can use the similar technique to obtain percentage robustness index ε_p.

The calculated cost robustness index presents the possible amount of changes (positive or negative) for any objective function coefficient (or more coefficients simultaneously), for which the optimal solution remains stable.

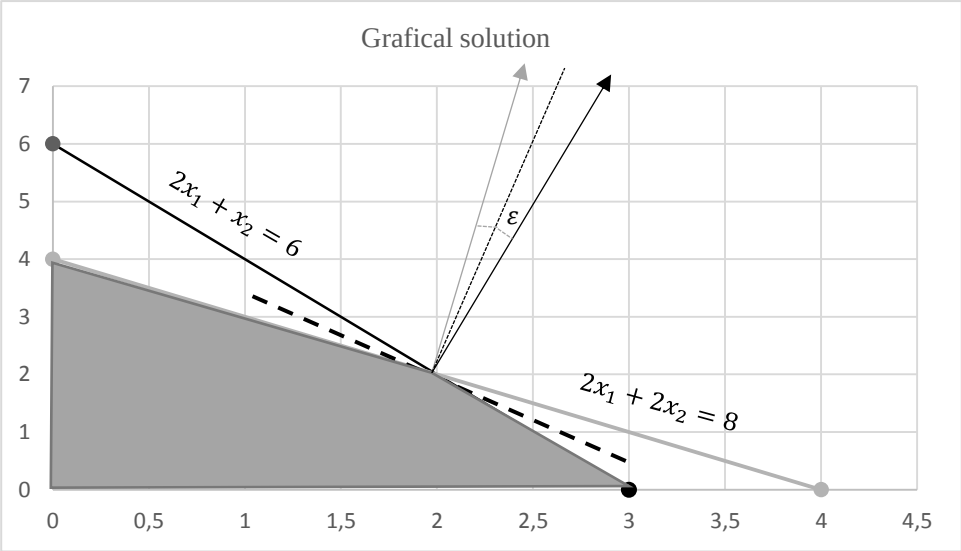
When comparing presented approach with existed approaches such as parametric programming and stability interval calculating we can remark the following: the strength of proposed method comparing to stability analyses of objective function coefficients is the fact, that we can change all objective function coefficients simultaneously, the changes must be inside the interval but amount of changes must not be the same for all coefficients. Comparing proposed method with parametric programming we can see that apart from parametric programming we are not bounded by a direction of change – the change is possible in any direction as long as we are moving inside the interval calculated for cost robustness index.

4 Case study

We will introduce the proposed method on a simple example which will help to illustrate the idea of proposed method. Let us assume the following linear optimization problem (17):

$$\begin{aligned} 2x_1 + x_2 &\leq 6, \\ 2x_1 + 2x_2 &\leq 8, \\ 3x_1 + 2x_2 &\ldots max. \end{aligned} \tag{17}$$

Graphical solution of the problem (17) is presented on the Graph 1.



Graph 1 Graphical solution

4.1. Calculation of ε robustness and ε_p percentage robustness index

The optimal solution for problem (17) is presented in Table 1.

	Variable	x_1	x_2	x_3	x_4	B
$[3-\varepsilon, 3+\varepsilon]$	x_1	1	0	1	-0.5	2
$[2-\varepsilon, 2+\varepsilon]$	x_2	0	1	-1	1	2
	$z_{(j)}$	0	0	1	0.5	10

Table 1 Optimal basic solution simplex table

From the final simplex table we can calculate $\varepsilon \geq 0$, according to (10) which will meet our constraints. We assume that for basic variables x_1, x_2 objective function coefficient $z_j$is always equal to zero, which means $z_1 = z_2 = 0$.

$$\tilde{z}_3 = 1 * [3-\varepsilon, 3+\varepsilon] - 1 * [2-\varepsilon, 2+\varepsilon] = [-3-\varepsilon, -3+\varepsilon] + [3-\varepsilon, 3+\varepsilon] = [1-2\varepsilon, 1+2\varepsilon],$$

$$[1-2\varepsilon, 1+2\varepsilon] \geq 0, \quad \text{is valid for} \quad 0 \leq \varepsilon \leq \frac{1}{2}.$$

$$\tilde{z}_4 = \left(-\frac{1}{2}\right) * [3-\varepsilon, 3+\varepsilon] + 1 * [2-\varepsilon, 2+\varepsilon] = \left[-\frac{3}{2} - \frac{\varepsilon}{2}, -\frac{3}{2} + \frac{\varepsilon}{2}\right] + [2-\varepsilon, 2+\varepsilon] = \left[\frac{1}{2} - \frac{3}{2}\varepsilon, \frac{1}{2} + \frac{3}{2}\varepsilon\right],$$

$$\left[\frac{1}{2} - \frac{3}{2}\varepsilon, \frac{1}{2} + \frac{3}{2}\varepsilon\right] \geq 0 \quad \text{is valid for} \quad 0 \leq \varepsilon \leq \frac{1}{3}.$$

From the calculations above it is obvious that the robustness index for the presented model is

$\varepsilon = min\left\{\frac{1}{2}, \frac{1}{3}\right\} = \frac{1}{3}$, which means that any changes of any objective function coefficients in the range $\left[-\frac{1}{3}, \frac{1}{3}\right]$ will not affect the optimal solution (see Graph 1).

From the final simplex table we can calculate $\varepsilon_p \geq 0$, according to (11) which will meet our constraints.

$$\tilde{z}_3 = 1 * 3 * [1-\varepsilon_p, 1+\varepsilon_p] - 1 * 2 * [1-\varepsilon_p, 1+\varepsilon_p] = [3-3\varepsilon_p, 3+3\varepsilon_p] + [-2-2\varepsilon_p, -2+2\varepsilon_p] = [1-5\varepsilon, 1+5\varepsilon],$$

$$[1-5\varepsilon_p, 1+5\varepsilon_p] \geq 0 \quad \text{is valid for} \quad 0 \leq \varepsilon_p \leq \frac{1}{5}.$$

$$\tilde{z}_4 = \left(-\tfrac{1}{2}\right) * 3\left[1-\varepsilon_p, 1+\varepsilon_p\right] + 1 * 2\left[1-\varepsilon_p, 1+\varepsilon_p\right] = \left[-\tfrac{3}{2}-\tfrac{3}{2}\varepsilon_p, -\tfrac{3}{2}+\tfrac{3}{2}\varepsilon_p\right] + \left[2-2\varepsilon_p, 2+2\varepsilon_p\right] = \left[\tfrac{1}{2}-\tfrac{7}{2}\varepsilon_p, \tfrac{1}{2}+\tfrac{7}{2}\varepsilon_p\right],$$

$$\left[\tfrac{1}{2}-\tfrac{7}{2}\varepsilon_p, \tfrac{1}{2}+\tfrac{7}{2}\varepsilon_p\right] \geq 0, \quad \text{is valid for} \quad 0 \leq \varepsilon_p \leq \tfrac{1}{7}.$$

The percentage robustness index is $\varepsilon_p = min\left\{\tfrac{1}{5}, \tfrac{1}{7}\right\} = \tfrac{1}{7} = 0{,}1428$, which means that any changes of any objective function coefficients in the range of 14, 28% will not affect the optimal solution.

5 Conclusion

This paper proposes a new method for measuring solution robustness, which helps to calculate solutions stability taking into account the fact that objective function coefficients could change due to unexpected circumstances. The method works with cost robustness index which declares the possible amount of changes (both positive and negative) for all coefficients in the objective function, for which the optimal solution remains stable. The paper introduces the method and describes it on a case study to illustrate how the cost robustness index is calculates. The paper also contains brief comparison of proposed method with parametric programming and stability interval calculation.

Acknowledgements

This research was supported by the Czech Science Foundation, grant No. P403/12/1947 and by the project F4/54/2015 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Bertsimas, D., Weismantel, R.: *Optimization over integers*, Dynamic Ideas, Belmont, Massachusetts, 2005.

[2] Kouvelis, P., Yu, G.: *Robust discrete optimization and its applications*, Kluwer Academic Publishers, 1997.

[3] Přívozník, L.: *Řešení soustav lineárních rovnic na bázi intervalové aritmetiky*, České vysoké učení technické v Praze, 2008.

[4] Sevaux, M., Sörensen, K.: *Robustness Analysis*: Optimisation, *EWG-MCDA Newsletter*, **Series 3**, No.10 (2004), p. 3-5.

Use of regression analysis to prediction of the attendance of cultural events (on example of the Moravian-Silesian Region)

Pavlína Pellešová[1], Alena Zedková[2]

Abstract. Cultural tourism is one the most widespread forms of tourism. It is apparent from the global number of trips motivated by "seeing the places" and culture, but also from researches carried out in the Czech Republic. Thus, also domestic tourists like travelling for cultural reasons.
The paper focuses on factors influencing attendance of selected cultural destinations in the Moravian-Silesian Region. The aim of the contribution is to construct a regression model for attendance of cultural events in the Moravian-Silesian Region and predict the development of number of their visitors. The partial aim is to highlight the importance of culture for tourism development and significance of analyses and researches in the field of culture. The paper describes the process of regression analysis in detail. Subsequently, the data on attendance of cultural events, the number of these events and the number of their organizers in the investigated region in years 2008 – 2013 are analyzed. The analysis results show that there is a strong correlation between the number of organized cultural events and the number of visitors. This means that with the increasing number of cultural events, the number of their visitors is rising, and conversely. Operators of monuments and cultural institutions can therefore be used the regression analysis for prediction of attendance, profits and other indicators and to identify the factors having the most influence of them.

Keywords: attendance, cultural destination, cultural tourism, regression analysis.

JEL Classification: C12, L83
AMS Classification: 62J05

Introduction

Cultural tourism represents cultural activities and experiences that attract and enhance visit of destinations. Cultural tourism is about enjoyment of the lifestyle of the local people, the local area and what gives it identity and character. It is such a form of tourism whose object is, among other aims, the discovery of monuments and sites [8].

Cultural tourism is often associated with a visit of a cultural destination, which is such a destination in whose offer the culture significantly or dominantly appears (according to Kesner [4]). Cultural destinations are both single establishments (museums, galleries, historical sites, open-air folk museums or specific exhibitions), and whole regions, towns and municipalities on whose territory visited historical monuments, museums, cultural landscape and other cultural attractions and events are located.

Cultural tourism belongs to the most popular forms of tourism. Its growth rate is faster than the growth of the world tourism [8]. According to OECD estimates [6], the share of cultural tourism amounts 40 % of international tourism, which is currently equivalent to almost 455 million trips; according to UNWTO [9], in 2014 the total number of tourist arrivals worldwide reached 1.138 billion.

Below we will deal with regression analysis and its subsequent use for predicting the attendance of cultural events in selected destination, namely the Moravian-Silesian Region. This region does not possess such a rich cultural heritage as other typical cultural destinations of the Czech Republic (e.g. Prague, South-Moravian Region or South-Bohemian Region), nevertheless, it has an interesting and unique cultural potential that is available in tourism [5]. The analysis of attendance can help to businessmen in tourism, operators of monuments and other cultural institutions in their next decisions. Further analysis of the selected region, for example, published in [2].

[1] Silesian university in Opava, School of business administration in Karvina, Department of tourism and leisure activities, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, pellesova@opf.slu.cz
[2] Silesian university in Opava, School of business administration in Karvina, Institute of doctoral studies, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, alena.zedkova@seznam.cz

1 Regression analysis

Regression analysis enables to measure the dependence between quantitative characters. One of them is referred as the dependent variable, which we denote *Y* and consider it as a random variable, and one or more characters are called independent variable denoted *X*, that are known in advance or given. In this context, we distinguish so-called simple regression measuring the dependence of only two quantitative characters, and multiple regression that determines the dependence of the dependent variable on several quantitative characters.

Regression equation has a following form:

$$y = f(x) + \varepsilon \tag{1}$$

- *y* and *x* represent the values of variables *Y* and *X*,
- *f* is the regression function,
- ε is residual (random error).

Assuming a linear regression function *f*, the equation of the regression line is:

$$f(x) = \beta_0 + \beta_1 x \tag{2}$$

- β_0 and β_1 are parameters (coefficients) of the regression function.

The regression function can also be nonlinear, e.g. parabola, hyperbola, logarithmic function, exponential function and others.

1.1 Simple linear regression

As already mentioned, the simple regression identifies the dependence of the dependent variable on the independent variable. The task of the linear regression analysis is to find linear regression function.

The equation of regression line:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, 2, \dots, n. \tag{3}$$

Estimates of regression parameters β_0 a β_1 are denoted b_0, b_1 and called regression coefficients. The regression coefficient b_0 is the point of intersection of the regression line with the vertical y-axis and regression coefficient b_1 indicates the slope of the line to the horizontal x-axis. Its size shows how the functional value *Y* will change, if the independent variable *x* changes by one unit. The regression coefficients are calculated using the least squares method.

The least squares method consist in finding the minimum value of residual sum of squares (S_R) of deviations of the values of the dependent variable y_i from the theoretical value $Y_i = f(x_i, b_0, b_1)$.

$$S_R = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i-1}^{n} (y_i - f(x_1, b_0, b_1))^2. \tag{4}$$

Function S_R attains its minimum for such values of b_0 and b_1 for which the partial derivatives are equal to zero.

$$\frac{\partial S_R}{\partial b_0} = 0 \quad ; \quad \frac{\partial S_R}{\partial b_1} = 0 \tag{5}$$

If we calculate the partial derivatives and put the linear regression function in relation SR, we obtain a system of equations with two unknowns (b_0 and b_1). The estimates b_0 and b_1 have following form:

$$b_1 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\overline{x^2} - \bar{x}^2} \tag{6}$$

$$b_0 = \bar{y} - b_1 \bar{x} \tag{7}$$

Then the resulting regression line has this form: $Y = b_0 + b_1 x$. (8)

With the use of the least squares method, we are able to express the total variability of the model (the dependent variable *Y*) and the intensity of the relationship between *Y* and *X* (namely through the coefficient of determination). First, we will focus on the total variability of the dependent variable *Y* which is characterized by a total sum of squares (S_y).

$$S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{9}$$

The theoretical sum of squares (S_T) characterizes a part of the total variability explained by the regression model.

$$S_T = \sum_{i=1}^{n} (Y_i - \bar{y})^2 \tag{10}$$

The residual sum of squares (S_R) characterizes an unexplained part of the total variability.

$$S_R = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (e_i)^2 \tag{11}$$

- e_i – residual.

Among the sums of squares, the following relationship applies: $S_y = S_T + S_R$. (12)

The **coefficient of determination** determines a tightness of the data to the regression model:

$$R^2 = 1 - \frac{S_R}{S_y}. \tag{13}$$

The coefficient of determination takes on values from the interval [0, 1]. The higher its value is, the better the regression model reflects reality; resp., by multiplying the coefficient of determination by the 100, we find out how many percent of the total variability is expressed by the regression model. The coefficient of determination is therefore an important indicator that reflects the suitability of the chosen regression model.

1.2 Multiple regression analysis

If the variable *Y* is influenced by several independent variables *X*, we talk about multiple regression. Multiple linear regression model is a generalization of the simple linear regression model and has this form:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_m x_m \tag{14}$$

As in the case of simple analysis, we assume that we are working with a sample and estimates of individual regression coefficients ($b_0, b_1, \ldots, b_m$). For their findings we apply the least squares method again. Process of the multiple regression analysis is thus similar like in a case of simple regression.

To confirm the correctness of the estimates of the regression coefficients, we can use hypotheses tests of statistical significance of regression coefficients, as in the case of simple, so in the case of multiple regression. If any of the regression coefficients proves to be statistically insignificant, it is suitable to exclude relevant variables from the model.

Test of statistical significance of regression coefficients determines whether the regression coefficients are indeed zero. We determine the null hypothesis H_0: $\beta_i = 0$ against to the alternative hypothesis H_1: $\beta_i \neq 0$.

Test criterion: $$T = \frac{b_i}{s(b_i)}, \tag{15}$$

- b_i – estimate of the regression coefficient,
- $s(b_i)$ – standard deviation of the regression coefficient b_i,

Critical value of the Student's distribution: $t_{n-k}(\alpha)$.

Critical region is defined as: $|T| > t_{n-k}(\alpha)$. Thus, if the test criterion is greater than the critical value, we reject the null hypothesis. In this case, the regression coefficient is statistically significant and can be included in the model.

We also **test the accuracy of choice of the whole regression model**, it means the appropriateness of choice of the independent variables. The null hypothesis assumes that the regression coefficients are zero ($\beta_1, \beta_2 \ldots \beta_n$) with the exception of the coefficient β_0. H_0: $\vec{\beta} = \vec{0}$. Then the alternative hypothesis is H_1: $\vec{\beta} \neq \vec{0}$.

Test criterion:

$$F = \frac{\frac{S_T}{k-1}}{\frac{S_R}{n-k}} \tag{16}$$

- n – number of observations,
- k – number of regression coefficients.

Critical value of Fisher's distribution at the significance level α: $C = F_{k-1,\, n-k}(\alpha)$. (17)

If the test criterion falls to the critical region (F > C), then we reject the null hypothesis and accept the alternative hypothesis. The conclusion is such that the independent variables x are selected appropriately and the whole model is created correctly [7]. Alternative methods of analysis have been published, for example [1].

2 Use of regression analysis to prediction of the attendance of cultural events

The following graphs show the attendance of cultural destinations (institutions and events) in the period from 2008 to 2013, i.e. in the period for which data are available continuously [1].

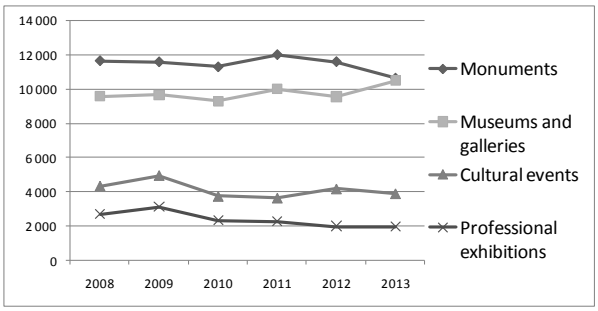


Figure 1 Attendance of cultural destinations in the Czech Republic in the years 2008 – 2013 (in thousands of persons)

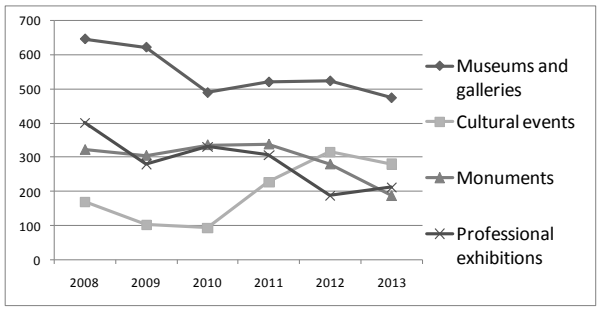


Figure 2 Attendance of cultural destinations in the Moravian-Silesian Region in the years 2008 – 2013 (in thousands of persons)

As the graphs show, the development of cultural destinations attendance in the Moravian-Silesian Region does not correspond to the nationwide development. Unlike the other regions, in the Moravian-Silesian Region the attendance of museums and galleries exceeds the attendance of monuments. It is due to the fact that the amount of monuments in the region is significantly lower than the number of museums and galleries. Some much visited castles are registered as affiliates of museums and galleries and thus also contribute to their higher attendance. In the whole republic, the attendance of cultural destinations has declining trend, which is reflected partly also in the Moravian-Silesian Region. The only exception is the attendance of cultural events which develops unevenly. Attendance of cultural events is very dependent on the number of these events realized in the region (correlation coefficient reached the value 0.98; as mentioned below, positive statistical correlation between the number of cultural events in the Moravian-Silesian Region and the number of their visitors was confirmed).

Year	Number of visitors (in thous. of persons) – Y	Number of cultural events – X_1	Number of organizers – X_2
2008	170	219	8
2009	102	205	8
2010	94	176	9
2011	227	300	10
2012	316	369	10
2013	281	357	10

Table 1 Data for regression analysis

To the regression analysis we use the data about attendance of cultural events in the Moravian-Silesian Region. The dependent variable (Y) is just the number of visitors of cultural events in the region. Independent vari-

ables (X) are the number of these events and the number of organizer of these events (these are only cultural events organized by historical monuments, such as concerts, exhibitions, folklore events, etc.). The data in Table 1 are taken from the Czech Statistical Office [1].

To calculate the regression analysis, we use MS Excel, supplement Data analysis, a tool Regression. Values of the regression coefficients are presented in the following table. Statistical significance of the regression coefficients was tested at a significance level $\alpha = 0.05$.

Regression coefficient	Value of a coefficient	T	$t_{n-2}(\alpha)$	Conclusion
b_0	-60.62	---	---	statistically significant
b_1	1.15	5.32	2.78	statistically significant
b_2	-5.87	-0.32	2.78	statistically insignificant

Table 2 Statistical significance of regression coefficients

Because the regression coefficient b_2 is not statistically significant, it should not be included in the model. We can say that the number of monuments in the Moravian-Silesian Region, which organize cultural events, does not affect the attendance of these events. From that reason we have to take into account only the number of cultural events and recalculate the model. Values of regression coefficients are shown in the Table 3.

Regression coefficient	Value of a coefficient	T	$t_{n-k}(\alpha)$	Conclusion
b_0	-98.94	---	---	statistically significant
b_1	1.097	9.96	2.78	statistically significant

Table 3 Statistical significance of regression coefficients – simple linear regression

The regression equation has the following form: $Y = 1.097X_1 - 98.94$ and it is graphically shown in the Figre 3.

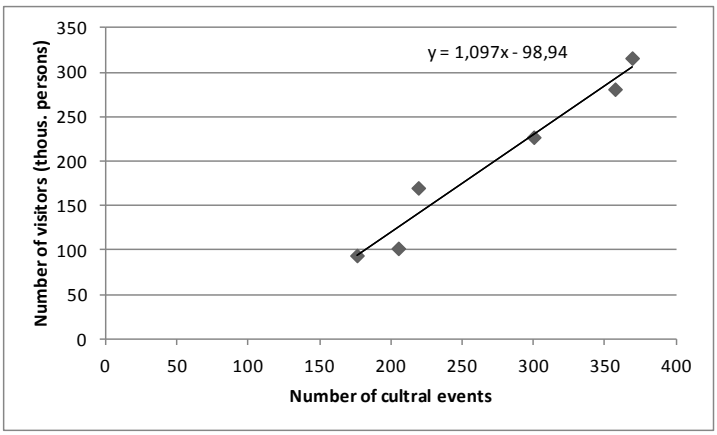


Figure 3 Regression line

The following table summarizes the basic characteristics of the model. Statistical significance of the whole model was tested at the significance level $\alpha = 0.05$.

R	R^2	F	$F_{k-1, n-k}(\alpha)$
0.98	0.96	99.17	7.71

Table 4 Evaluation of the regression model

The value of the correlation coefficient (R) is 0.98, which indicates a high degree of positive statistical correlation between the examined variables. The coefficient of determination (R^2) which explains tightness of the data to the regression line reached a value of 0.96. Thus, the regression model describes 96 % of the total variability of the data.

We can conclude that the model as a whole is statistically significant. The test of statistical significance has demonstrated that the test criterion falls to the critical region of Fisher's distribution (99.17 > 7.71). Therefore we can not accept the hypothesis of zero regression coefficients. Independent variable was chosen appropriately.

The regression model has the form: $Y = 1.097X_1 - 98.94$. If we assume that in the year 2014 the number of cultural events organized in the region will increase at 300, then the expected number of visitors will be 230.16 thousands: $y_{pred} = 1.097 \cdot x_{pred} - 98.94 = 1.097 \cdot 300 - 98.94 = 230.16$ thous.

From the above it is clear that regression analysis can be used to estimate the future attendance of cultural destination or other institutions. Due to the fact that the attendance of monuments and other cultural destinations in the Moravian-Silesian Region develops differently from the national development, we focused just on the issue of cultural events. Trend of their attendance is growing, to contrast to other cultural destination in the Moravian-Silesian Region. The use of regression analysis, resp. other statistical methods can contribute to more efficient management of cultural destinations.

Conclusion

Regression analysis can be used to analyze and predict of attendance of different cultural destinations. It can help operators of cultural monuments and other cultural institutions in their decision and find out what factors influence the attendance of destinations. However, the analysis need not confine itself to examining the factors influencing attendance, but also other facts that are important to the functioning of cultural destinations.

Attendance of cultural destinations is not only influenced by factors that can be statistically processed; often there are also psychological or psychographic factors. As reported by Kesner [4], so called barriers to participation in culture can affect the attendance of cultural destination. Available leisure time, information, cultural competence and own perception of destination image are examples of the mentioned barriers. For cultural and historical monuments, it is important to know their visitors. But a lot of valuable information they can also gain from so called non-visitors (people to whom the above-mentioned barriers impede visiting cultural destinations). Obtaining such information is essential for the monuments and it has a strategic importance, because from opinions of visitors and non-visitors it can be created strategies and goals and prepare specific cultural products etc. However, cultural destinations are often not managed comprehensively and they have a negative relation to planning. Thus, they neglect the importance of various analyses and researches. But these are the basis for strategic decision of cultural destinations and organizations.

Acknowledgements

The paper has originated with the support of the grant SGS/20/2014: Analysis of the business environment of the Karvina region.

References

[1] Botlík, J., Botlíková, M. Determination of precedence in the network model for the region´s analysis. In: *IBIMA Creating Global Competitive Economies: 2020 Vision Planning & Implementation* . Rome, Italy, 2013, s.777-786. ISBN 978-0-9860419-1-4.

[2] Botlík, J., Botlíková, M., Andrysek, L. Innovative Methods of Regional Availability Analysis as a Factor of Competitiveness, In*: 11th International Conference on Liberec Economic Forum*. Sychrov, 2013, s. 46-56

[3] ČSÚ. *Kulturní zařízení v ČR 2013* [online]. Praha: ČSÚ, 2015, updated on 15. 9. 2014 [cit. 2015-03-06]. Available at: http://www.czso.cz/csu/2014edicniplan.nsf/p/090004-14.

[4] Kesner, L. *Marketing a management muzeí a památek*. Praha: Grada, 2005. 304 s. ISBN 80-247-1104-4.

[5] Mynarzová, M. a Švajdová, L. The Effect of Economical Recesion on Economy Development of Moravian-Silesian Region. In: *Proceedings of International Scientific Conference on Hradec Economical Days 2010 - Economic Development and Management of the Region*. Hradec Králové: Univerzita Hradec Králové, Fakulta informatiky a managementu, 2010, s. 309-312. ISBN 978-80-7435-040-5.

[6] OECD. *The Impact of Culture on Tourism* [online]. Paris: OECD, 2009 [cit. 2015-03-06]. ISBN 978-92-64-05648-0. Available at: http://www.tava.gov.lv/sites/tava.gov.lv/files/dokumenti/petijumi/OECD_Tourism_Culture.pdf

[7] Tošenovský, F. *Statistical methods for economists*. Karviná: SU OPF, 2014. ISBN 978-80-7510-033-7.

[8] UNWTO. *Tourism 2020 Vision: Volume 7: Global Forecast and Profiles of Market Segments* [online]. Madrid: UNWTO, 2001 [cit. 2015-03-06]. 139 p. ISBN 92-844-0466-5. Available at: http://www.sete.gr/files/Media/Ebook/110303_Global%20Forecasts%20and%20Profiles%20of%20Market%20Segments.pdf

[9] UNWTO. *World Tourism Barometer: Volume 13 – January 2015* [online]. Madrid: UNWTO, 2015 [cit. 2015-03-06]. 68 p. ISSN 1728-9246. Available at: http://www.e-unwto.org/content/h7527k0518431038/fulltext?p=654b39b75dc64135bd94f2d424c7ff4d&pi=0#section=1190847&page=1

Multistage portfolio optimization with risk premium constraints

Barbora Petrová [1]

Abstract. A multistage stochastic optimization is a tool which enables to manage portfolio in constantly changing financial markets by periodically rebalancing its structure in order to achieve desired target. This paper presents a decision-making process where the objective function is to maximize investor's expected utility over a finite time horizon, namely we consider a class of non-separable multivariate utility functions. Features of utility functions already contain the information on investor's risk attitude thus basically no risk constraints are necessary. However, the solution cannot guarantee that the investor does not find herself in an undesirably risky position within the investment horizon. We therefore suggest a reformulation of the underlying problem by adding an extra constraint on an upper bound of risk premiums. The performance of the suggested model is demonstrated by optimizing the allocation of wealth at each time instance to two different assets one of which is assumed to be risk-free. The stochastic character of the problem is provided in the form of a scenario tree. Experiments are conducted to compare performance of the underlying formulation (considering no explicit constraints on the risk exposure) with the extended one (considering constraints on the risk exposure).

Keywords: Risk premium, utility function, portfolio optimization, multistage stochastic programming.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Multistage stochastic portfolio optimization is determined to find the optimal strategy of periodically rebalancing the structure of portfolio in order to reach desired target (the comprehensive overview on multistage optimization can be found in [6]). The objective of the problem is to maximize utility of investor's wealth received in the course of the investment horizon. Therefore we have to consider multidimensional utility functions (see, for instance, [1]), namely we focus on their non-separable class. Stochastic features are provided through scenario approach which stems from scenario tree generation. The generation is based on the moment matching method suggested in [7]. In the computational experiment we aim to compare performance of the optimal investment strategy stemming from two formulations. The first model did not take into account any explicit constraints on investor's risk exposure. This problem can provide a reasonable investment strategy since the information on the risk attitude is contained in the shape of utility function. However, the solution cannot guarantee that the investor does not find herself/himself in an undesirably risky position within the investment horizon. The second model considers the constraints in the form employing the risk premium concept. One dimensional risk premium was firstly introduced in [11] and later extended to multidimensional version (see, for instance, [1]).

2 Model formulation

Multistage portfolio optimization problems assume there are k risky assets and one riskless asset, often represented by a bond or a bank account paying a fixed interest rate, traded during a finite time horizon. In order to achieve a desired target the investor can rebalance her/his portfolio at fixed times of the considered horizon. In this paper we reduce the number of risky assets to a single one, which can be

[1] Charles University in Prague, Faculty of Mathematics and Physics, barbora.petrova@optipro.cz

interpreted as a stock index, and the investor's task reduces to allocation of her/his wealth into a risky and a riskless asset. The investor rebalances her/his portfolio so as she maximizes the expected utility over wealth achieved throughout the investment horizon. In this study we focus on a class of non-separable utility function. We note that any transaction costs associated with portfolio rebalancing are neglected.

Consider an investment horizon with T equidistant periods such that the rebalancing of the portfolio is held at their beginning. Let u be a T-dimensional non-separable utility function and let $w = (w_1, \ldots, w_T)$ be a T-dimensional wealth vector, where w_t, $t = 1, \ldots, T$ corresponds to wealth achieved at the end of time period t. We consider initial wealth w_0 to be given. The wealth fraction invested into asset k at the beginning of period t is denoted as x_{tk} (in our notation $k = 1$ corresponds to the riskless asset and $k = 2$ to the risky one). Similarly, we denote as r_{tk} the one-period return of the k-th asset corresponding to period t. Our aim is to find investment strategy represented by matrix $x = (x_{tk})$ which maximizes utility of achieved wealth. We conveniently denote as $x_{t\cdot}$ t-column of the investment strategy matrix x, i.e., $x_{t\cdot} = (x_{t1}, x_{t2})$. Analogous notations is adopted for $r_{t\cdot}$. The problem can be formulated as follows:

$$\begin{aligned}
&\underset{x,w}{\text{maximize}} && u\left(w_1, \ldots, w_T\right) && && \\
&\text{subject to} && w_t = w_{t-1} + r_{t\cdot}^T x_{t\cdot} && t = 1, \ldots, T, && (1)\\
& && w_{t-1} = 1^T x_{t\cdot} && t = 1, \ldots, T, && (2)\\
& && x_{t\cdot} \geq 0 && t = 1, \ldots, T. && (3)
\end{aligned}$$

Constraint (2) ensures that the whole wealth achieved at the end of time period $t-1$ is reallocated between risky and riskless assets at the beginning of time preiod t. Constraint (1) describes wealth development during the investment horizon corresponding to specific investment decisions. Finally constraint (3) guarantees that short positions are not allowed.

The above formulation relates to a deterministic problem where the return on assets is non-random. In further text we develop formulation where the return on riskless asset remains deterministic but the return on risky asset becomes a random process. In order to derive a stochastic formulation we adopt a scenario approach with a specific choice of a scenario tree.

2.1 Scenario tree generation and scenario approach formulation

Let $(R_1^*, \ldots, R_T^*)$ be the one-period return vector. We make a simplifying assumption that $R_1^*, \ldots, R_T^*$ are independent and identically distributed (their common distribution is a distribution of a random variable denoted as R^*). Moreover, we assume that R^* is discrete and acquires M different values $r_1^*, \ldots, r_M^*$ with probabilities $p_1^*, \ldots, p_M^*$. In order to generate a scenario tree one needs to estimate parameters r_m^*, p_m^*, $m = 1, \ldots, M$. These parameters are set as to fit the sample mean, sample upper and lower semi-variance and sample central moments of higher degrees. The scenarios three then branches out into M followers at each node and the ramification subjects to the same distribution. On the basis of this tree we define a scenario s to be a specific path from the initial node to a particular terminal node. In further text S denotes the set of all possible scenarios s.

With respect to multistage character of the problem, we have to consider the non-anticipativity conditions on the investment decisions and the consequal wealth in the following form:

$$x_{t\cdot}^s = x_{t\cdot}^q \qquad t = 1, \ldots, T,\ q = 1, \ldots, M^{t-1},\ s = (q-1)M^{T-t+1} + 1, \ldots, qM^{T-t+1}, \tag{4}$$

$$w_{t\cdot}^s = w_{t\cdot}^q \qquad t = 1, \ldots, T,\ q = 1, \ldots, M^{t},\ s = (q-1)M^{T-t} + 1, \ldots, qM^{T-t}. \tag{5}$$

The optimization problem (P1) using scenario approach can be formulated as:

$$\begin{aligned}
&\underset{x^s,w^s}{\text{maximize}} && \sum_{s\in S} p_s \cdot u\left(w_1^s, \ldots, w_T^s\right) && && \\
&\text{subject to} && w_t^s = w_{t-1}^s + (r_{t\cdot}^s)^T x_{t\cdot}^s && t = 1, \ldots, T,\ s \in S, && (6)\\
& && w_{t-1}^s = 1^T x_{t\cdot}^s && t = 1, \ldots, T,\ s \in S, && (7)\\
& && w_0^s = w_0 && s \in S, && (8)\\
& && x_{t\cdot}^s \geq 0 && t = 1, \ldots, T,\ s \in S, && (9)\\
& && (4), (5) && &&
\end{aligned}$$

where p_s denotes a probability of scenario s. We refer to this formulation later in the computational part in order to compare its solution with a solution of the model incorporating constraints on the risk exposure.

2.2 Concept of risk premiums

Investor's risk exposure at a specific investment position can be evaluated using various tool. When dealing with investor's preferences captured by means of utility functions there is another convenient method to evaluate risk exposure which utilizes risk premium framework. A risk premium associated with a random variable is defined as the maximum amount of money that the investor would be willing to pay in order to avoid the random investment and receive its expected value instead. Considering univariate utility function u, the risk premium Π at wealth level w associated with random variable X can be formally defined as:

$$\mathbb{E}_x\ u(w+X) = u\ (w + \mathbb{E}_x X - \Pi).$$

There are several ways how to extend the above stated definition for multidimensional utility functions. One possible way is to define a risk premium at time instance t to be the maximum amount of money the investor would be willing to pay in order to avoid a one-period random investment occurring at this time. In our specific model we aim to measure risk exposure at each node of the scenario tree. To this purpose, we firstly define risk premium associated with time instance t and scenario s for all $t = 1, \ldots, T$ and $s \in S$ and secondly we propose a formula to calculate the premium at each node.

Consider particular scenario s, wealth vector associated with scenario s, w^s, and investment strategy associated with scenario s, x_t^s. We define risk premium Π_t^s at time instance t as the solution of the following equation:

$$u\left(w_1^s, \ldots, w_{t-1}^s, w_{t-1}^s + \mathbb{E}_R R^T x_{t\cdot}^s - \Pi_t^s, w_{t+1}^s, \ldots, w_T^s\right) = \mathbb{E}_R\ u\left(w_1^s, \ldots, w_{t-1}^s, w_{t-1}^s + R^T x_{t\cdot}^s, w_{t+1}^s, \ldots, w_T^s\right). \quad (10)$$

We emphasize that the risk premium only measures the investor's risk exposure at specific time, however in reality the investor cannot pay this amount in order to substitute the random revenue by the certain one.

For each node we define the risk premium Π_t^q, $t = 1, \ldots, T$, $q = 1, \ldots, M^{t-1}$ as:

$$\Pi_t^q = \sum_{s=(q-1)M^{T-t+1}+1}^{qM^{T-t+1}} p_s^n \Pi_t^s, \qquad p_s^n = \Big(\sum_{s=(q-1)M^{T-t+1}+1}^{qM^{T-t+1}} p_s\Big)^{-1} p_s, \quad (11)$$

where Π_t^s solves equation (10) and p_s^n is a conditional probability. In other words, the premium corresponding to a particular node is a weighted sum of all risk premiums Π_t^s where the summation is taken over all scenarios which have the node lying on their path.

2.3 Multistage model with risk premium constraints

In this section we finally arrive at the multistage portfolio optimization model considering risk exposure constraints. The model stems from formulation (P1) adding constraints on risk premium at each node derived in the previous section. Thus the final model (P2) has the form:

$$\begin{array}{lll} \underset{x^s, w^s, \Pi_t^q}{\text{maximize}} & \displaystyle\sum_{s\in S} p_s \cdot u\left(w_1^s, \ldots, w_T^s\right) & \\ \text{subject to} & (6), (7), (8), (9), (4), (5) & \\ & \Pi_t^q \le C & t = 1, \ldots, T,\ q = 1, \ldots, M^{t-1}, \\ & (11), (10). & \end{array} \quad (12)$$

C is a deterministic constant restricting investor's risk exposure at each node.

3 Computational experience

The portfolio optimization is carried on with the exponential utility function of the following form:

$$u(w_1, \ldots, w_T) = -\exp\left(-\alpha \sum_{t=1}^{T} v^t w_t\right), \tag{13}$$

where parameter α specifies the strength of investor's risk aversion and parameter v is an appropriate discount factor which captures time value of investor's wealth. In particular, the risk aversion parameter is set as $\alpha = 1.5 \cdot 10^{-4}$ and the factor as $v = 0.99$. With the special choice of the utility function given as above, the optimization problem (P2) simplifies in the following sense:

Proposition 1. *Consider the utility function to be given by equation (13). Constraint (12) in problem (P2) is equivalent to:*

$$x_{t2}^s \sum_{m=1}^{M} p_m^* r_m^* + \frac{1}{\alpha v^t} \log\left(\sum_{m=1}^{M} p_m^* \exp\left(-\alpha v^t r_m^* x_{t2}^s\right)\right) \leq C, \qquad t = 1, \ldots, T, s \in S. \tag{14}$$

Proof. The proof stems directly form substituting the utility function into equation (10) and the fact that the resulting premium is a function of single variable $x_{t\cdot}^s$ and thus subjects to the same non-anticipativity conditions as $x_{t\cdot}^s$. □

We recall that we consider two types of assets one of which is risky and the other one riskless. The riskless investment is represented by a regular bank account earning zero interest rate ($r_f = 0$) and the risky asset is taken to be the Prague stock exchange index (PX). In order to design a scenario tree we consider a serie of monthly development of PX index. On the basis of this data set we construct the distribution of the one-period rate of return with three possible outcomes ($M = 3$). To proceed with the experiment it remains to define the investment horizon and the investor's initial wealth. Thus, in this paper we present solutions to problems (P1) and (P2) formulated as five-stage optimization problems in which the investor disposes with 1000 at the beginning of the horizon.

Table 1 summarizes percentage of the initial wealth invested in the risky asset for different choices of risk constraint C and for the model excluding the constraint (we refer to this model by symbol '/' in all tables). Distinctive changes in the ratio can be observed for risk constraint lower than 0.2, higher limitations on risk exposure cause little changes in diversification which consequently yields to the same expected wealth achieved at the end of the first investment period as indicated in Table 2. Note that the proportion of wealth invested in risky asset at a particular time instance can oscillate with C raising as shown in Table 1. We recall that this behaviour is acceptable since the objective of the optimization problem is not the maximization of utility of the expected wealth at particular moment but the maximization of utility of discounted wealth achieved throughout the whole investment horizon.

C	0.1	0.2	0.3	0.4	/
x_{12}	0.5553	0.6853	0.6834	0.6840	0.6827

Table 1: First stage decision (proportion of the initial wealth invested in the risky asset) for different choices of the risk constraint.

Figure 1 displays expected wealth development achieved at different stages together with its 0.95-confidence interval for problems (P1) and (P2) with risk constraint equal to 0.1. When limiting the risk exposure substantially the investor achieves slightly lower level of expected wealth at each stage however its variance decreases.

Tables 2, 3 summarizes expected wealth and its variance at each stage. We observe only slight differences between the expected wealth which might be caused by the nature of the scenario tree. As one examines decision made for higher stages, differences between expected wealth become more significant. Focusing on the other table, we conclude that variance increases as the risk constraint is less limiting which is in line with the fact that the investor is allowed to enhance her/his diversification. One can observe the reverse trend only for the first investment period, however drops in variance are related to slight decrease in percentage of the initial wealth invested in the risky asset (see Table1).

C	stage 1	stage 2	stage 3	stage 4	stage 5
0.1	1001.22	1002.44	1003.68	1004.87	1005.57
0.2	1001.50	1003.24	1004.98	1006.67	1007.66
0.3	1001.50	1003.38	1005.48	1007.37	1008.59
0.4	1001.50	1003.38	1005.58	1007.57	1008.84
/	1001.50	1003.43	1005.64	1007.63	1008.91

Table 2: Comparison of the wealth mean for different choices of the risk constraint.

C	stage 1	stage 2	stage 3	stage 4	stage 5
0.1	1345.74	2705.07	4078.14	5410.31	6221.61
0.2	2049.37	4767.15	7510.74	10179.41	11801.13
0.3	2038.19	5231.40	9235.91	12793.17	15218.48
0.4	2041.62	5240.56	9643.98	13559.74	16229.65
/	2033.78	5453.71	9895.44	13829.64	16541.79

Table 3: Comparison of the wealth variance for different choices of the risk constraint.

The sum of discounted wealth achieved throughout the investment horizon has to raise as we relax limitation on the risk premium since it is the argument of strictly increasing utility function considered as the object of maximization. The graphical development of the sum of discounted wealth is illustrated in Figure 2. We conclude that the investor achieves slightly lower discounted wealth throughout the horizon when limiting her/his risk exposure, but its variance decreases considerably. Table 4 displays the expected sum of discounted wealth together with its 0.95% confidence interval.

Figure 3 depicts dependence of the expected terminal wealth on the value of the risk constraint. We can conclude as the constraint exceeds threshold 0.35 the terminal wealth stabilizes around 1008.5 in the expectation. In our specific case the function is strictly increasing however we note again that in general this statement does not have to be satisfied.

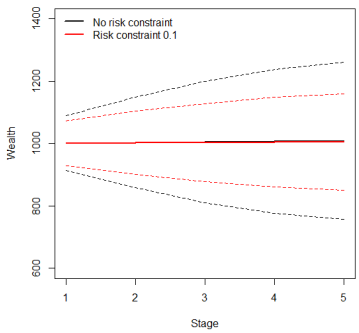


Figure 1: Expected wealth development with 0.95-confidence interval.

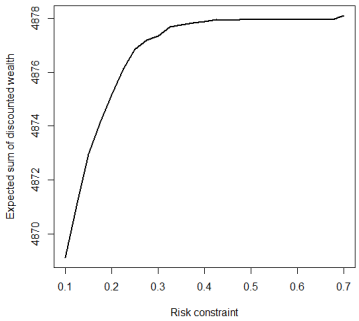


Figure 2: Expected sum of discounted of wealth as a function of the risk constraint.

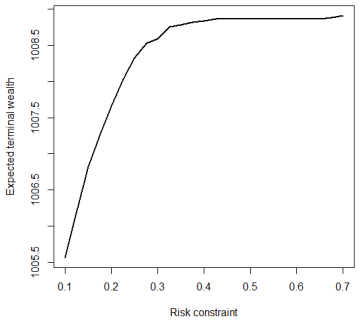


Figure 3: Expected terminal wealth as a function of the risk constraint.

4 Conclusion

In this paper we presented a multistage portfolio optimization problem having the maximization of investor's utility of wealth achieved throughout the horizon as its objective. We considered the exponential class of utility functions. We compared performance of two different formulations. The first model did not take into account any explicit constraints on investor's risk exposure, the other one considered these constraints in the form of limitation on risk premiums. The obtained results are in line with the expectation. When limiting the risk exposure, the investor achieves slightly lower level of expected wealth throughout the horizon however its variance considerably decreases. Markable change of the first stage decision can be observed for considerable strict limitation on the risk exposure, which is a consequence of special structure of the scenario tree.

C	mean	lower CI	upper CI	length of CI
0.1	4869.12	4354.33	5383.91	1029.58
0.2	4875.17	4188.28	5562.06	1373.77
0.3	4877.34	4133.26	5621.43	1488.18
0.4	4877.87	4120.69	5635.04	1514.34
/	4878.10	4111.88	5644.33	1532.45

Table 4: Expected sum of discounted wealth together with its 0.95% confidence interval for different choices of the risk constraint.

In the future work, we intend to raise the number of risky asset the investor could include in her/his portfolio. It will be necessary then to generate multidimensional scenario tree, which simulates development of return on all risky assets together with their interaction captured via correlation. One possible way is to adopt procedure suggested in [7]. Another field of interest is to consider different choice of the family of utility functions. As we saw in the computational part, the optimization problem simplifies when the exponential class is used, however no such simplification is possible with different choice of the family. Even more interesting is to consider unknown utility function and to modify the problem in the sense of stochastic dominance (see [4]), [8], [9], [10]). The problem can be examined in terms of robustness by considering a class of distribution for the scenario tree or its contamination (see, for instance, [3] or [5]). Simple modification of the problem can be made by including transaction costs since in general rebalancing of the portfolio comes in hand with extra fees. Finally, we remark that the risk premium at each node can be also given as the maximum of risk premiums corresponding to scenarios having the node on their path. This formulation may lead to bilevel optimization, for more details we refer to [2].

Acknowledgements

I would like to express my gratitude to my advisor doc. RNDr. Ing. Miloš Kopa, PhD., who offered invaluable assistance and guidance. The paper is supported by the grant No. P402/12/G097 of the Czech Science Foundation and SVV 2015 grant no. 2602255 of Charles University in Prague.

References

[1] R. Ambarish and J. G. Kallberg. Multivariate risk premiums. *Theory and Decision*, 22:77–96, 1987.

[2] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153:235–256, 2007.

[3] J. Dupačová and M. Kopa. Robustness in stochastic programs with risk constraints. *Annals of Operations Research*, 200:55–74, 2012.

[4] J. Dupačová and M. Kopa. Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research*, 234(2):434–441, 2014.

[5] J. Dupačová and V. Kozmík. Structure of risk-averse multistage stochastic programs. *OR Spectrum*, 37(3):559 – 582, 2015.

[6] J. Dupačová, J. Hurt, and J. Štěpán. *Stochastic modeling in economics and finance*. Kluwer, 2002.

[7] K. Høyland, M. Kaut, and S. W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2-3):169–185, 2003.

[8] M. Kopa and T. Post. A general test for ssd portfolio efficiency. *OR Spectrum*, 37(3):703 – 734, 2015.

[9] T. Post and M. Kopa. General linear formulations of stochastic dominance criteria. *European Journal of Operational Research*, 230(2):321 – 332, 2013.

[10] T. Post, Y. Fang, and M. Kopa. Linear tests for decreasing absolute risk aversion stochastic dominance. *Management Science*, 61(7):1615 – 1629, 2015.

[11] J. W. Pratt. Risk aversion in the small and the large. *Econometrica*, 32:122–136, 1964.

On return rate estimated by intuitionistic fuzzy probabilistic set

Krzysztof Piasecki[1]

Abstract. In general case return rate is defined as a function of present value (PV) and future value (FV). The FV is burdened with risk of quantified uncertainty. A formal model of this uncertainty is to present the FV as a random variable. On the other side, it is fully justified by economic reasons that imprecisely assessed PV may be described as intuitionistic fuzzy subset. Then the return rate is determined as intuitionistic fuzzy probability set. Then return risk consists of quantified uncertainty risk, imprecision risk and knightian risk. The quantified uncertainty risk is characterized in usual way as variance of distribution of precisely determined return rate. Subsequently the expected return rate is estimated as intuitionistic fuzzy set. Its membership function is the image of imprecision risk. Imprecision is composed of ambiguity and indistinctness. The ambiguity risk is evaluated by energy measure. The indistinctness risk is characterized by entropy measure. The hesitation function of expected return rate can be interpreted as an image of knightian uncertainty. As a result, knightian uncertainty risk may be rated by indeterminacy measure. This way, four-dimensional image of risk is introduced. This perception of risk endangering a security will lead to the formulation of new approaches to risk management.

Keywords: return rate, intuitionistic fuzzy probabilistic set, knightian uncertainty, quantified uncertainty, imprecision, ambiguity, indistinctness, insolubility

JEL Classification: C02, G11
AMS Classification: 03E72, 91G99

1. Research problem

The current equivalent value of payments available in a fixed point in time is called a present value (PV for short) of this payment. In financial arithmetic any PV is used as tool for dynamic assessment of the money value. The starting point for financial arithmetic development was the interest theory. The considerations set out in this article had their genesis in acceptance of view that - without the interest theory - PV of future cash flow may be imprecise. The natural consequence of this approach is the assessment of PV by means of fuzzy numbers. In [17] it is shown that in such case, if future value (FV for short) is determined as a random variable then return rate is given as fuzzy probabilistic set [6].In [17] and [19] it is shown that the fuzzy probabilistic return rate may be applied for financial decision making.

On the other hand Atanassov [1] has defined intuitionistic fuzzy sets as an extension of the fuzzy set concept. In this way we obtain a new tool for description of interaction between imprecision and knightian uncertainty. Therefore, some kinds of PV may be assessed as intuitionistic fuzzy subsets. An example of such approach is justified by economic reasoning [18].

The main aim of this article is to determine return rate for the case when FV is given as random variable and PV is assessed by intuitionistic fuzzy subset in the real line. Moreover, some model of risk burdening this return rate will be studied.

2. Intuitionistic fuzzy sets

2.1. Basic concepts

Let us take any space $\mathbb{X}$. Each element of space $\mathbb{X}$ is interpreted as objects of any cognitive-application activities. The basic tool for imprecise classification of these objects is the concept of fuzzy set A which may be described as the set of ordered pairs

[1]Poznan University of Economics,Department of Operations Research, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: k.piasecki@ue.poznan.pl

$$\tilde{A} = \{(x, \mu_A(x)) : x \in \mathbb{X}\}. \quad (1)$$

where $\mu_A : \mathbb{X} \longrightarrow [0,1]$ is its membership function. In multi-valued logic terms, the value $\mu_A(x)$ is interpreted as the truth value of the sentence $x \in A$.The family of all fuzzy subsets in the space $\mathbb{X}$ will be denoted by symbol $\mathcal{F}(\mathbb{X})$.

Atanassov [1] defines intuitionistic fuzzy set (for short IFS) Aas the set of ordered triples

$$A^* = \{(x, \mu_A(x), \nu_A(x)) : x \in \mathbb{X}\}, \quad (2)$$

where nonmembership function $\nu_A : \mathbb{X} \longrightarrow [0,1]$fulfills the condition

$$\nu_A(x) \leq 1 - \mu_A(x) \quad (3)$$

for each $x \in \mathbb{X}$. In multi-valued logic terms, the value $\nu_A(x)$ is interpreted as the truth value of the sentence $x \notin A$.The family of all IFS in the space $\mathbb{X}$ we denote by symbol $\mathcal{I}(\mathbb{X})$.

We define hesitation function $\pi_A : \mathbb{X} \longrightarrow [0,1]$ by the identity

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)\ . \quad (4)$$

Value $\pi_A(x)$ indicates the degree of our hesitation in assessment of the relationship between the element $x \in \mathbb{X}$ and IFSA. For this reason, the hesitation function π_A may be interpreted as a image of knightian uncertainty [10].

Let us consider $B \in \mathcal{F}(\mathbb{X})$ described by its membership function $\mu_B \in [0;1]^{\mathbb{X}}$. This fuzzy subset can beidentified with IFS represented by the set of ordered triples

$$B^* = \{(x, \mu_B(x), 1 - \mu_B(x)); x \in \mathbb{X}\}, \quad (5)$$

The hesitation function of the above IFS identically fulfils the condition

$$\pi_B(x) = 0\ . \quad (6)$$

It implies that the application of fuzzy sets to create real object model is connected with implicit acceptance of a strong assumption proclaiming that we are always able to decide on the fulfilment of each elementary state's requirements postulated to it. However, as we know from everyday observations usually it is not, and our settlements are burdened with a noticeable hesitation margin. This means that the extension of the fuzzy sets class to IFS class extends the capabilities of a reliable imprecision description. The way to obtain this extension is quite easy, because at the moment the IFS had been extensively and comprehensively theoretically studied (e.g.[3]).

For any$A, B \in \mathcal{I}(\mathbb{X})$set theory operations and relations are defined in the following way

$$A \subset B \Leftrightarrow \mu_A \leq \mu_B \wedge \nu_A \geq \nu_B\ , \quad (7)$$

$$A^C = \{(x, \nu_A(x), \mu_A(x)) : x \in \mathbb{X}\}, \quad (8)$$

$$A \cup B = \{(x, \mu_A(x) \vee \mu_B(x), \nu_A(x) \wedge \nu_B(x)) : x \in \mathbb{X}\}, \quad (9)$$

$$A \cap B = \{(x, \mu_A(x) \wedge \mu_B(x), \nu_A(x) \vee \nu_B(x)) : x \in \mathbb{X}\},. \quad (10)$$

According to Atanassov [2], for any IFS$A \epsilon \mathcal{I}(\mathbb{X})$ we determine the biggest fuzzy subset$\Box A \subset A$ and the smallest fuzzy subset $\Diamond A \supset A$. We have here

$$\Box A = \{(x, \mu_A(x), 1 - \mu_A(x)) : x \in \mathbb{X}\}, \quad (11)$$

$$\Diamond A = \{(x, 1 - \nu_A(x), \nu_A(x)) : x \in \mathbb{X}\}. \quad (12)$$

2.2. Imprecision and insolubility

IFS's are applied for description of imprecise information under knightian uncertainty. Many researchers of this subject (e.g. [9]) distinguish two components of imprecision. They state that in the general case imprecision is composed of ambiguity and indistinctness. The information ambiguity is interpreted as a lack of clear recommendation one alternative between the various given alternatives. The information indistinctness we interpret as the lack of explicit distinction between the given information and its negation. Intensification of the information imprecision decreases this information usefulness. This leads to the problem of imprecision evaluation. The hesitation function describes information insolubility which is interpreted as the lack of

possibility to decide on the fulfilment of each elementary state's requirements postulated to it. This insolubility causes knightian uncertainty.

In this paper, we restrict ourselves to the case of the information described by the IFS in the real number space $\mathbb{R}$. In this case, the starting point for measurement is the fuzzy set measure proposed by Khalili [8].This measure can be generalized to the case of any IFS. Thus, the measure for any $A \epsilon \mathcal{I}(\mathbb{R})$ is given as a value

$$w(A) = \int_{-\infty}^{+\infty} \mu_A(x)dx \,. \tag{13}$$

The Khalili's measure does not exist for each IFS. Therefore, we use this measure to define a normalized IFS measure which does always exist. Normalized measure for any $A \epsilon \mathcal{I}(\mathbb{R})$ is determined by correspondence

$$m(A) = \lim_{y \to +\infty} \frac{\int_{-y}^{y} \mu_A(x)dx}{1+\int_{-y}^{y} \mu_A(x)dx}. \tag{14}$$

In the first step, we will focus our attention on the ambiguity measurement. Here the right tool is the arbitrary energy measure $d: \mathcal{I}(\mathbb{R}) \to [0;1]$ fulfilling axioms given by de Luca and Termini [13]. Energy measure of IFS increases with increasing ambiguity of information represented by this set. Thus this measure increases with the worsening of this information quality. For our considerations it is sufficient this measure exemplification is given by the identity

$$d(A) = m(A) \tag{15}$$

The indistinctnees may be measured by any entropy measure defined for the fuzzy case as function fulfilling axioms proposed by de Luca and Termini [12].This definition was generalized to IFS case by Szmidt and Kacprzyk [20].Entropy measure of IFS increases with increasing indistinctnees of information represented by this set. Thus this measure increases with the worsening of this information quality. In our considerations we will use introduced by Kosko [11] the most popular entropy measure $e: \mathcal{I}(\mathbb{X}) \to [0;1]$ given by the identity

$$e(A) = \frac{d(A \cap A^C)}{d(A \cup A^C)}. \tag{16}$$

It is worth noting that the used above sets $A \cap A^C$ and $A \cup A^C$ are, respectively, the W-empty set and W-universe described in [15].This observation may be a suggestion that further studies on IFS measurement.

The insolubility may be evaluated by any entropy measure fulfilling axioms given by Burillo and Bustince [4].The Burillo's and Bustince's definition of entropy measure is not equivalent to the de Luca's and Termini's definition of entropy measure. The Burillo's and Bustince's entropy measure is used to measurement of phenomena different from the phenomena measured by de Luca's and Termini's entropy measure. To emphasize these facts, we propose for arbitrary Burillo's and Bustince's entropy measure different name, that is, an ignorance measure. In our considerations we will use the ignorance measure $k: \mathcal{I}(\mathbb{X}) \to [0;1]$ given by the identity

$$k(A) = d(\diamond A) - d(\Box A). \tag{17}$$

3. Return rate

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the security considered here is determined by two values:

- anticipated FV $V_t \in \mathbb{R}^+$,
- assessed PV $V_0 \in \mathbb{R}^+$.

The basic characteristic of benefits from owning this instrument is a return rate r_t given by the identity

$$r_t = r(V_0, V_t). \tag{18}$$

In the general case, the function: $r: \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is a decreasing function of PV and an increasing function of FV. It implies that for any FV V_t we can determine inverse function $r_0^{-1}(\cdot, V_t): \mathbb{R} \to \mathbb{R}^+$. In the special case we have here:

- simple return rate

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1 \tag{19}$$

- logarithmic return rate

$$r_t = ln\frac{V_t}{V_0}. \tag{20}$$

The FV is at risk of uncertainty. A formal model of this uncertainty is the presentation of FV V_t as a random variable $\tilde{V}_t: \Omega = \{\omega\} \longrightarrow \mathbb{R}^+$. The set Ω is a set of elementary states of the financial market.

In the classical approach to the problem of the return value estimation of the security PV is identified with the observed market price $\check{C}$, what we denote as

$$V_0 = \check{C}. \tag{21}$$

The return rate is then a random variable which is at uncertainty risk. This random variable is determined by the identity

$$\tilde{r}_t(\omega) = r\left(\check{C}, \tilde{V}_t(\omega)\right). \tag{22}$$

In practice of financial markets analysis, the uncertainty risk is usually described by probability distribution of return rates. At the moment, we have an extensive knowledge on this subject. Let us assume that this probability distribution is given by cumulative distribution function $F_r: \mathbb{R} \longrightarrow [0; 1]$. We assume here that the variance σ^2 of this distribution exists. Then, the probability distribution of FV is described by conditional cumulative distribution function $F(\cdot|\check{C}): \mathbb{R}^+ \longrightarrow [0; 1]$ given as follows

$$F(x|\check{C}) = F_r\left(r(\check{C}, x)\right). \tag{23}$$

In [18] it is well justified that PV may be assessed by IFS

$$V = \{(x, \mu_V(x), \nu_V(x)): x \in \mathbb{R}^+\}. \tag{24}$$

According to the Zadeh extension principle, for each fixed elementary state $\omega\epsilon\Omega$ of financial market, membership function $\rho(\cdot|\omega): \mathbb{R} \to [0; 1]$ and nonmembership function $\varphi(\cdot|\omega): \mathbb{R} \to [0; 1]$of return rate are determined by the identities

$$\rho(r|\omega) = \max\left\{\mu_V(y): y\epsilon\mathbb{R}^+, r = r\left(y, \tilde{V}_t(\omega)\right)\right\} = \mu_V\left(r_0^{-1}\left(r, \tilde{V}_t(\omega)\right)\right), \tag{25}$$

$$\varphi(r|\omega) = \min\left\{\nu_V(y): y\epsilon\mathbb{R}^+, r = r\left(y, \tilde{V}_t(\omega)\right)\right\} = \gamma_V\left(r_0^{-1}\left(r, \tilde{V}_t(\omega)\right)\right). \tag{26}$$

It means that the return rate considered here is represented by intuitionistic fuzzy probabilistic set [14]. For this reason, this return rate is called intuitionistic fuzzy return. In the special case we have:

– for the simple return rate

$$\rho(r|\omega) = \mu_V\left((1 + r)^{-1} \cdot \tilde{V}_t(\omega)\right), \tag{27}$$

$$\varphi(r|\omega) = \gamma_V\left((1 + r)^{-1} \cdot \tilde{V}_t(\omega)\right). \tag{28}$$

– for logarithmic return rate

$$\rho(r|\omega) = \mu_V\left(e^{-r} \cdot \tilde{V}_t(\omega)\right), \tag{29}$$

$$\varphi(r|\omega) = \gamma_V\left(e^{-r} \cdot \tilde{V}_t(\omega)\right). \tag{30}$$

For any intuitionistic fuzzy return, expected return is IFS $R \in \mathcal{I}(\mathbb{R})$, given as follows

$$R = \left\{\left(x, \mu_R(x|\check{C}), \nu_R(x|\check{C})\right): x \in \mathbb{R}\right\}, \tag{31}$$

where the membership function $\rho(\cdot|\check{C}): \mathbb{R} \to [0; 1]$and nonmembership function $\varphi(\cdot|\check{C}): \mathbb{R} \to [0; 1]$ are determined by the identities

$$\rho(r|\check{C}) = \int_{-\infty}^{+\infty} \mu_V(r_0^{-1}(r, y))dF_V(y|\check{C}), \tag{32}$$

$$\varphi(r|\check{C}) = \int_{-\infty}^{+\infty} \nu_V(r_0^{-1}(r, y))dF_V(y|\check{C}). \tag{33}$$

Let us note that in the above model, the expected rate return rate is dependent on observable market price $\check{C}$. This corresponds to the practice of investment decision-making. Thus, we can treat this property as the added value of the proposed model. The image of the expected return rate formulated in such way can play mentioned role in

financial market analysis, which is identical to the role played there by the expected return rate described by a real number. On the other hand, thanks to the application of IFS as a model of expected value we obtain much richer image of the risk burdening the return rate.

4. Multidimensional risk image

In classical financial market theories return rate is at uncertainty risk. The uncertainty risk follows from the lack of investor's knowledge about future states of financial market. This lack of knowledge results in lack of investor's confidence regarding the future profits or losses. In our model, we evaluate the uncertainty risk using the variance σ^2 of return rate distribution given by cumulative distribution function $F_r: \mathbb{R} \longrightarrow [0;1]$. Similarly as in the case of precisely defined return rate, there are such probability distributions of return for which the return rate variance does not exist. We then replace this distribution with a distribution truncated on both sides, for which the variance always exists. This procedure finds its justification in the theory of perspective [7]. Among other things, this theory describes the behavioural phenomenon of the rejection of extremes.

In the previous chapter we propose such model of the expected return rate which includes the impact of the observable current market price. In this way we have increased the cognitive value of the financial instrument description. However, this increase in the description usefulness has its price. This price is existence of new kinds of risk. The ambiguity risk and the indistinctness risk follow from the imprecision of information about expected return rate. The information insolubility causes insolubility risk. The question arises here whether, from the investor's point of view, these risks are significant.

When we use imprecise images of security then we cannot precisely indicate the recommended investment alternative. Then each of investment alternatives is recommended to some extent. On the other hand, the investor can choose only one alternative. The investors are personally responsible for that choice. Investors shift a part of this responsibility onto advisers or onto the applied forecasting tool. For this reason, the investors restrict their choice of investment decisions to alternatives recommended in the greatest degree. In this way the investor minimizes his individual responsibility for financial decision making. It shows that imprecision risk assessment is relevant for the analysis of investment processes.

An increase of ambiguity risk means that the number of recommended investment alternatives increases too. This increases the chance of selecting the recommended alternative, which involves the lost opportunity cost. In this way, the investor's responsibility increases with increase in ambiguity risk. For any expected return rate $R \in \mathcal{I}(\mathbb{R})$ its ambiguity risk will be evaluated by energy measure $d(R)$ determined by the identity (15).

An increase in indistinctness risk means that distinctions between recommended and unrecommended alternatives are more blurred. It causes a higher probability of choosing unrecommended alternatives. In this way the indistinctness risk increases impacts on the increase in investor's responsibility. For any expected return rate $R \in \mathcal{I}(\mathbb{R})$ its indistinctness risk will be evaluated by entropy measure $e(R)$ determined by the identity (16).

An increase in insolubility risk lowers the responsibility which may be shifted onto advisers or onto the applied forecasting tool. This will automatically increase the investor's personal responsibility for the decision. For any expected return rate $R \in \mathcal{I}(\mathbb{R})$ its insolubility risk will be evaluated by ignorance measure $k(R)$ determined by the identity (17).

These observations show that an increase in imprecision risk or in insolubility risk significantly worsens the quality of information supported the investor's decision making process. Using the four-dimensional risk image $\left(\sigma^2, d(R), e(R), k(R)\right)$ facilitates risk management. Here it is desirable to minimize each of the four risk assessments.

Using the four-dimensional risk image enables investigation of relationships between different types of risk. Here we can observe empirical interaction between some risk kinds. Moreover, there is a formal correlation between the uncertainty risk and ambiguity risk. The number of recommended alternatives increases with ambiguity risk. In this way, there is more certainty that the recommended alternatives include the best investment decision. This means that the uncertainty risk decreases. In summary, the uncertainty risk and the ambiguity risk are negatively correlated. Moreover, it is very easy to check, that for some kinds of IFS the insolubility risk may be a decreasing function of indistinctnees risk.

In comparison with the classical Markowitz theory, imprecision is a new aspect of risk assessment. It should be here to answer the question whether such extension of the risk assessment is purposeful. The usefulness of taking into account imprecision in risk study is well justified by the following three arguments.

Firstly, it is mentioned above that some kinds of risk may be negatively correlated. This means that it is possible to reduce the uncertainty risk of the forecast by the appropriate reduction of the information quality.

Only imprecision risk control and insolubility risk control will allow to avoid the undervaluing of assessments of uncertainty risk.

Secondly, if we take into account the imprecision risk then we can reject investment alternatives which are attractive from the viewpoint of the classical Markowitz theory, but, unfortunately, the information gathered about them are poor quality.

Thirdly, in [16] it is shown, that using imprecise PV assessment can explain the phenomenon of market equilibrium on the efficient financial market remaining in the state of financial imbalance.

5. Final remarks

Discussion carried out in this article is a continuation of research on problems of imprecise returns which have been initiated in [17]. From the view point of the finance theory, a new element in this discussion is the insolubility risk. Despite this modernization all the rich empirical knowledge gathered about the uncertainty risks of can be used in the proposed model. This is a highly advantageous feature of the proposed model, because it brings us closer to the possibility of real applications.

The results obtained here provide a basis for generalization of the investment decision making normative theory of described in [17] and [19]. On the other side, these results can be directly used in the decision making models described in [5].

Acknowledgements

The project was financed by funds of National Science Center - Poland granted on the basis the decision number DEC-2012/05/B/HS4/03543.

References

[1] Atanassov K., Stoeva S.: Intuitionistic fuzzy sets, In: *Proceedings of Polish Symposium on Interval and Fuzzy Mathematics*, (Albrycht J., Wiśniewski H., Eds.) Poznań,(1985), 23-26.
[2] Atanassov K.:New variant of modal operators in intuitionistic fuzzy modal logic, *BUSEFAL*54 (1993), 79-83.
[3] Atanassov K.: *Intuitionistic Fuzzy Sets*, Springer-Verlag, Heidelberg(1999).
[4] Burillo P., Bustince H.: Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets, *Fuzzy Sets and Systems*78 (1996), 305-316.
[5] Deng-Feng Li.: *Decision and Game theory in Management with Intuitionistic Fuzzy Sets.* Springer, Berlin - Heidelberg,(2014).
[6] Hiroto K.: Concepts of probabilistic sets, *Fuzzy Sets and Systems* 5(1981), 31-46.
[7]Kahneman D., Tversky A., Prospect theory: an analysis of decision under risk, *Econometrica* 47 (1979), 263-292. DOI: 10.2307/1914185
[8] Khalili S.: Fuzzy measures and mappings, *J. Math. Anal. Appl*. 68(1979), 92-99.
[9] Klir G.J.: Developments in uncertainty-based information, *Advances in Computers* 36, (1993), 255-332.
[10] Knight, F. H.: *Risk, Uncertainty, and Profit*, Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA, (1921)
[11] Kosko B.: Fuzzy entropy and conditioning, *Information Sciences* 40, (1986), 165-174. DOI:10.1016/0020-0255(86)90006-X
[12] de Luca A., Termini S.: A definition of a non-probabilistic entropy in the settings of fuzzy set theory, *Information and Control* 20, (1972), 301-312. DOI:10.1016/S0019-9958(72)90199-4
[13] de Luca A., Termini S.: Entropy and energy measures of fuzzy sets, *Advances in fuzzy set theory and applications* 20, (1979), 321-338.
[14] Qiansheng Zhang, BaoguoJia, Shengyi Jiang: Interval-valued intuitionistic fuzzy probabilistic set and some of its important properties, *Proceedings of the 1st International Conference on Information Science and Engineering ICISE2009*, Guangzhou, (2009).4038-4041, DOI:10.1109/ICISE.2009.692
[15] Piasecki K.: New concept of separated fuzzy subsets, In: *Proceedings of Polish Symposium on Interval and Fuzzy Mathematics*, (Albrycht J., Wiśniewski H., eds.) Poznań, (1985), 193-196, DOI:10.13140/RG.2.1.3188.9440
[16] Piasecki K., Behavioural Present Value, *SSRN Electronic Journal* (2011), DOI:10.2139/ssrn.1729351
[17] Piasecki K.: Effectiveness of securities with fuzzy probabilistic return, *Operations Research and Decisions* 2, (2011), 65-78.
[18] Piasecki K.: Intuitionistic assessment of behavioural present value, *Folia Oeconomica Stetinensia* 13(21) (2013), 49-62, DOI: 10.2478/foli-2013-0021
[19] Piasecki K.: On imprecise investment recommendations, *Studies in Logic, Grammar and Rhetoric* 37(50)(2014), 179-194, DOI: 10.2478/slrg-2014-0024
[20] Szmidt E., J. Kacprzyk J.: Entropy for intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 118(2001), 467-477.

Integrated measure of risk in a cumulated-surplus-based financial plan optimization model for a 2-person household

Radoslaw Pietrzyk [1], Pawel Rokita [2]

Abstract. Long term financial planning for a household is aimed at preservation of desired life standard in the whole life cycle, including retirement, under the constraint that realization of other goals is also provided for. The possibility of achieving this depends, however, on a number of stochastic factors. They may be of different nature, and, therefore, a household financial plan is exposed to many different types of risk (e.g., life-length risk, market risk). Measuring the risk should be treated as an important part of financial planning, allowing to better address preferences of a household. Encompassing in a measure at least the most significant types of risk is, however, a challenge. A measure of risk suited to household financial planning will not be value oriented, nor cash-flow oriented. It should rather be in a way related to the question about threats to life standard and to ability of the household to accomplish its goals. In this article, there is made an attempt to construct an original risk measure that fulfills the condition.

Keywords: personal finance, household financial planning, integrated measure of risk.

JEL Classification: C61, G11, D14
AMS Classification: 91B16, 91G10

1 Introduction

When constructing a financial plan for a household or individual, many kinds or risk need to be taken into consideration (comp., e.g., Mitchell et al. [9]). But there are no household-specific integrated measures of risk that would allow to easily compare financial plans in respect of their total risk.

Financial plan optimization procedures may use direct or implicit information on risk, and give as a result such plans that offer the best tradeoff between risk and, for instance, some aggregate reflecting expected standard of living. There is a rich literature about optimization of life-long financial plans, with different risk factors worked into the model. The list includes the classical model by Yaari [14], as well as its various modifications and augmentations. Besides stochastic nature of survival, to the factors of risk that are included in financial planning belong, for instance, returns on assets ([2], [12], [8]), stochastic labor income ([5]), health condition ([13]). There are also generalizations to the case of a two-person household (married couple) – comp. [6], [7], [3].

Optimization of a financial plan using a formalized procedure is, however, often just half of the job. This is because the very constraints of the optimization problem sometimes also need to be revised and adjusted. Some decisions cannot be automatized. For instance, optimization of a bunch of goals would be really hard, if not impossible. This is due to non-separable and, thus, not perfectly transitive preferences ([1]). Only household members may change goals, in terms of the desired time and magnitude (or, as to the extent to which the goals must be realized). The more so, only the household members may forego some goals or replace them with some other ones. This puts the decision maker in a situation of choice between a number of plans, each being optimal for its particular set of constraints. The main difference between these plans will be the degree to which the originally planned goals may be realized and the risk that the plan will fail in the future in some scenarios. The second may be measured. This, however, requires risk measures that indeed address the needs of decision making process by households. The aim of constructing these measures is to provide households with a tool (or tools) that inform them how likely they are to fail accomplishing the life objectives assumed in a given financial plan if they adopt this particular plan and keep sticking to it throughout the whole life cycle.

[1] Wroclaw University of Economics, Department of Financial Investments and Risk Management, ul. Komandorska 118/120, 53-345 Wroclaw, Poland, radoslaw.pietrzyk@ue.wroc.pl
[2] Wroclaw University of Economics, Department of Financial Investments and Risk Management, ul. Komandorska 118/120, 53-345 Wroclaw, Poland, pawel.rokita@ue.wroc.pl

2 Outline of the household financial plan optimization task

As it is explained in more details by Pietrzyk and Rokita in [10] and [11], the goal function of household financial plan optimization task, called there *value function of the household*, is based on expected discounted utility of consumption and bequest. A general concept of the value function presents equation (1).

$$V\left(c_0, v, \boldsymbol{\kappa}; \mathbf{Z}\right) =$$

$$\sum_{z_j^*=LB(z_j)}^{UB(z_j)} \sum_{z_k^*=LB(z_k)}^{UB(z_k)} p_{jk}^* \cdots \sum_{z_l^*=LB(z_l)}^{UB(z_l)} p_l^* \cdots \sum_{z_m^*=LB(z_m)}^{UB(z_m)} \sum_{z_k^*=LB(z_n)}^{UB(z_n)} p_{mn}^* \cdots \left(\begin{array}{c} \alpha \left(\sum_{t=0}^{T_B^*} \frac{1}{\left(1+r_C\right)^t} u\left(C\left(t;\mathbf{Z}^*\right)\right)\left(\gamma(t)+\delta(t)\right)\right) + \\ \beta \frac{1}{\left(1+r_B\right)^{T_B^*}} u\left(B\left(T_B^*;\mathbf{Z}^*\right)\right) \end{array} \right) \quad (1)$$

where:

$c_0, v, \boldsymbol{\kappa}$ – decision variables:

c_0 – consumption rate at start of the plan,

v – proportion of household retirement investment contribution assigned to private pension plan of Person 1 (pension plan contribution proportion of Person 2 is just $1-v$),

$\boldsymbol{\kappa}$ – vector of own-means-proportions for financing of all financial goals planned by the household;

$\mathbf{Z}$ – random vector of risk factors,

$z_i \equiv \mathbf{Z}(i)$ – i-th element of vector $\mathbf{Z}$

(amongst elements z_i of the vector $\mathbf{Z}$ there are also times of death of the two main household members, i.e., D_1 – time of death of Person 1, D_2 – time of death of Person 2),

$\mathbf{Z}^* \equiv \left[z_1^* \ldots z_n^* \right]$ – a particular realization of the random vector $\mathbf{Z}$,

z_i^* – a particular value of the i-th element of vector $\mathbf{Z}$

(amongst particular realizations z_i^* there are also realizations of times of death, i.e., D_1^* and D_2^*),

$LB\left(z_i\right)$ – lower bound of the range of z_i scenarios,

$UB\left(z_i\right)$ – upper bound of the range of z_i scenarios,

$p_{ij}^* \equiv P\left(z_i = z_i^*, z_j = z_j^*\right)$ – joint probability of particular realizations of risk factors z_i and z_j if they are treated as dependent in the model,

$p_i^* \equiv P\left(z_i = z_i^*\right)$ – probability of a particular realization of risk factors z_i if it is treated in the model as independent from other risk factors

(e.g., in the equation (1), risk factors z_j and z_k are dependent of each other but independent of the rest, z_l is independent of any other, z_m and z_n are dependent on each other and independent of any of the other ones),

T_B^* – time of household end ($T_B^* \equiv \max\{D_1^*, D_2^*\}$) for a given scenario $\mathbf{Z}^*$,

α – propensity-to-consume parameter,

β – bequest preference parameter,

$u(.)$ – utility of consumption,

$C\left(t;\mathbf{Z}^*\right)$ – consumption at a moment t and under a scenario $\mathbf{Z}^*$,

$B\left(T_B^*;\mathbf{Z}^*\right)$ – bequest at the end of the scenario $\mathbf{Z}^*$ (at the time T_B^*, under the scenario $\mathbf{Z}^*$),

$\gamma(t)$, $\delta(t)$ – functions reflecting subjective severity of premature death and longevity, respectively, based on life-length risk aversion parameters (γ^*, δ^*) declared by the household,

r_C – rate used for discounting utilities of consumption,

r_B – rate used for discounting utility of bequest.

Optimization of the household financial plan consists in maximizing the function $V(.)$ given by the formula (1) under several constraints. The budget constraint is the most obvious one. But in the model by Feldman, Pietrzyk and Rokita [4] and Pietrzyk and Rokita [10], [11] there is also a constraint that all financial goals are met, as well as a constraint of minimum acceptable consumption.

Whereas the value function of the household is based on consumption and bequest, verification of optimization results and measuring risk of the plan is based on net cash flows. To be more precise, the considered financial category is then cumulated net cash flow (i.e., cumulated surplus or cumulated shortfall).

3 Dealing with time value of money if the impact of cash-flow timing is not obvious

When analyzing potential threats to financial liquidity of the household, particularly if caused by deep shortfall, recovery from which may require time and efforts, applying just classical discounting of cashflow may be not the most comprehensive way of reflecting time value of money. In the model, comparing cash flows from different periods is facilitated in two ways. The first is that all cash flows are expressed in real terms assuming some long-run inflation rate forecast. This would, however, be insufficient because time value of money is dependent also on other components of interest rates than just inflation. On the other hand, discounting all cash flows at some interest rate would obviously attach higher weight to the cash flows of the nearest future and low weight to the once that are more distant in time from the moment zero. This is fully appropriate in overwhelming majority of situations in finance but not particularly in weighting shortfalls in the life-cycle of a household. This is because severity of shortfalls, especially – deep shortfalls, is not monotonous with age. A negative cumulated net cash flow is most problematic for very young persons and for elderly ones.

To take into account the non-monotonous nature of the relationship between age and "burdensomness" of finding financing for shortfalls, all negative cumulated cash flow values are in the model charged additionally at a *contingency financing rate*, composed of the following elements:

$$\eta_{c_t} = \eta_t + \lambda_t + \zeta_t + \xi_t \ , \tag{2}$$

where: η_t – market benchmark interest rate (in real terms) at t; λ_t – credit spread (risk premium plus bank margin) for regular financing at t; ζ_t – additional margin for contingency financing at t; ξ_t – burdensomeness charge (dependent on household age) at t.

An example of the shape of the relationship between household age and contingency financing rate is shown in the **Figure 1**.

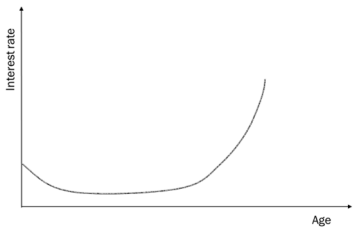


Figure 1 Correction of credit costs with household age taken into account.

Having added the additional charge for contingency financing to the negative cumulated net cash flows, no limitations on the size of cumulated shortfall is imposed here. The only exception from this rule is here the third of the proposed measures of risk, that is – household default probability.

Summing up, cumulated net cash flow is calculated in the following way:

- real values of subsequent net cash flows are added,
- in periods when the cumulated net cash flow is negative, additional charge, equal to interest for annual contingent loan, is added.

The procedure may be illustrated with the following formula (3):

$$CSp_0 = Sp_0, \quad CSp_t = \begin{cases} CSp_{t-1} + Sp_t & \text{if} \quad CSp_{t-1} >= 0 \\ CSp_{t-1} \cdot \left(1 + \eta_{c_t}\right) + Sp_t & \text{if} \quad CSp_{t-1} < 0 \end{cases} \tag{3}$$

where: Sp_t – net cash flow (in real terms) at t, called here surplus; CSp_t – cumulated surplus at t.

Does it mean that no traditional discounting is used in the model? Not exactly – compare formula (1) in which utility of consumption and bequest is discounted).

4 Household financial plan measures of risk

As it has already been mentioned the underlying model of household financial plan is based on cashflows and one of its major outcome is term structure of cumulated net cashflows of the household. Therefore, it is postulated that the risk measures to be proposed are all suited to this way of description of household financial situation. The nature of the problem implies negative concept of risk rather than neutral, which, in turn, elicits the choice of downside risk measures. The following three concepts of household financial plan risk measures are discussed: Residual Wealth at Risk, Lifetime Cumulated Net Cash Flow at Risk and Household Default Probability.

As the models proposed by Pietrzyk and Rokita ([10], [11]) are discrete and based on a finite number of survival scenarios, estimation of the measure is also performed for a grid of discrete multivariate scenarios, belonging to a hypercube formed by confidence intervals for all of the considered risk factors.

Residual Wealth at Risk (*RWaR*)

This measure is very similar to Cash Flow at Risk (CFaR) with the difference that cumulated net cash flow is the risk variable here and that realization of the considered random variable may refers to different points in time (household end may be different in different scenarios). The course of actions to obtain the measure is as follows:

1) in each scenario (of the dimension of 4 or 6 – dependent on the number of risk factors used) there is calculated a deviation from the applied benchmark of cumulated net cash flow at household end,
2) a left-tail quantile of deviations, corresponding to a pre-set tolerance level (small significance level), is determined on the basis of the scenarios,

This measure of risk can be describe by the following formula:

$$P(CSp_{T_B} \le XCSp_{T_B} - RWaR) = q\,, \tag{4}$$

where: $XCSp_{T_B}$ – benchmark of cumulated net cash flow at household end; CSp_{T_B} – cumulated net cash flow at household end; *RWaR* – Residual Wealth at Risk; q – *RWaR* tolerance level (a small significance level).

Lifetime Cumulated Net Cash Flow at Risk (*LCNCFaR, LCaR*)

In search of a measure that encompasses to a higher extent all threats throughout the whole life time of the household, a new proposition is formulated. Its concept assumes that both the moment when a shortfall is encountered and the length of period during which the household remains in the situation of a negative cumulated cash flow are important and should be taken into account. If in any of the considered scenarios there is a cumulated shortfall in whichever period, it is treated as a realization of risk.

The proposition is to do the following:

1) in each of the considered scenarios, all cumulated shortfalls are summed up (cumulating of cumulated shortfalls),
2) From all scenarios there is taken a quantile of the sum of cumulated shortfalls, corresponding to a predefined small tolerance level.

This measure of risk can be describe by the following formulas:

$$P(CCSh \le -LCaR) = q\,, \tag{5}$$

where q is *LCaR* tolerance level (a small significance level).

The Lifetime Cumulated Net Cash Flow at Risk is a quantile of sum of cumulated net cash flows throughout the whole life time, multiplied by (–1), so that it gives a positive value if the quantile is negative (a shortfall).

Sum of cumulated shortfalls (*CCSh*) is defined for each scenario **Z**. Cumulated shortfalls are defined in the formula **Błąd! Nie można odnaleźć źródła odwołania.**. And sum of cumulated surplus is calculated for all periods from the starting point of the plan until the end (T_B^*) of a given scenario ($\mathbf{Z}^*$)

$$CSh_t = \begin{cases} CSp_t & \text{if} \quad CSp_t < 0 \\ 0 & \text{if} \quad CSp_t \geq 0 \end{cases} \qquad (6) \qquad CCSh^{(\mathbf{z}^*)} = \begin{cases} \sum_{t=1}^{T_B^*} CSh_t^{(\mathbf{z}^*)} \end{cases} \qquad (7)$$

where: $CSh_t^{(\mathbf{z}^*)}$ – cumulated shortfall at a moment t, under a scenario $\mathbf{Z}^*$, $CCSh^{(\mathbf{z}^*)}$ – sum of cumulated shortfalls from the start of the plan until the end of the particular scenario $\mathbf{Z}^*$,

The measure takes into account the information about time and magnitude of shortfalls, but it does not include explicitly the information if the household is able to make up for the shortfall in subsequent periods or not.

Household Default Probability (*HDP*)

The Lifetime Cumulated Net Cash Flow at Risk, presented above, supplies a decision maker with no bankruptcy threat signals. Household Default Probability, in turn, takes into account only such negative values of cumulated net cash flow, that exceed some threshold above which the household is not able to increase its indebtness. The threshold is determined for each period in time and in each of the considered scenarios. The course of actions to construct the measure is as follows:

1) A default threshold is determined at the moment t_0 (e.g., credit worthiness of the household at t_0); then, the default threshold changes in time with financial situation of the household,
2) Scenarios in which cumulated shortfall exceeds the default threshold are identified as default scenarios,
3) Probabilities of default scenarios are summed up,
4) Probability that any of the default scenario realizes is the measure of risk.

$$HDP = P\left(CSh < DTh\right), \qquad (8)$$

where *DTh* is default threshold.

When constructing the measure one needs to define a rule of default threshold determination. The proposition is that threshold is calculated on the basis of household potential free cash flow (*PFC*), costs of contingency loan servicing and the household age. The default threshold changes over time. An important element that is taken into consideration when determining default threshold are loans for contingency financing of cumulated shortfall, and a minimum possible payment for a given age (assuming that expected further life length at a given moment is treated as the maximum crediting period). The minimum possible payment for contingency loans during a period cannot exceed potential free cash flow. It must be emphasized that all payments resulting from planned loans are already deduced by definition from the potential free cash flow (comp. eq. (9)):

$$PFC_t = Ic_t - C_{min_t} + NIt_t + ND_t + CSp_t, \qquad (9)$$

where: Ic_t – income at a moment t, C_{min_t} – minimum consumption at t, NIt_t – net cash flow from investment at t, ND_t – net cash flow from planned debts at t.

The default threshold at a given moment is set as such value of shortfall that cost of debt servicing connected with this shortfall during the next period exceeds potential free cash flow.

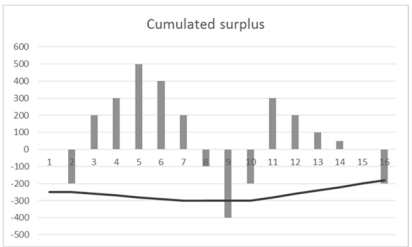


Figure 2 A trajectory of cumulated surplus (real values, additionally modified by contingency financing charge) with the default threshold marked with the line.

In the **Figure 2** there is an example of a scenario for which cumulated surplus trajectory takes on negative values is several periods and the default threshold is exceeded in periods number 9 and 16. Of course, only one exceedance of the threshold is sufficient to recognize the whole scenario as a default one.

This measure does not contain the information by what amount of a shortfall the threshold is exceeded nor how many times would it be exceeded given a default scenario.

5 Summary

All the proposed measures of risk are based on integrated information about potential performance of a household financial plan. The underlying financial category (risk variable) is cumulated net cash flow of the household, under the constraint that all financial goals are met and at least minimum acceptable consumption is realized. Therefore, any negative values of this variable indicate threats to life objectives of household members. This makes the approach based on cumulated net cash flow well suited to household needs.

The measures may find application particularly at the stage of plan revision. The plan being a result of the optimization process performs differently under different scenarios. Even if it is the best plan possible for a given bunch of household goals, it may be inacceptable due to a poor performance in some scenarios. The household needs a tool supporting its judgement whether the scenarios are not too likely.

There are three measures presented, because it would be hard to grasp all important risk information in one figure. The first (*RWaR*) reflects threat to the final effect. Only implicitly, there is also contained information from the whole path of the process, but the stress is placed on the end of the planning period. The second (*LCaR*) aggregates information about all potential shortfalls that may be encountered during the life of the household. Lengths of periods during which the household remains in the state of shortfall is also taken account of. The last (*HDP*) just informs how likely the household is to lose its financial liquidity whenever along the line.

Acknowledgement: The research project was financed by The National Science Centre (NCN) grant, on the basis of the decision no. DEC-2012/05/B/HS4/04081.

References

[1] Arrow K.J., Intriligator M.D. (eds.). Handbook of Mathematical Economics, vol. II. North-Holland Publishing Company (1982).
[2] Bodie, Z., Detemple, J. B., Otruba, S., Walter, S. Optimal consumption – portfolio choices and retirement planning. Journal of Economic Dynamics and Control, 28(6), 1115–1148 (2004).
[3] Brown, J.R., Poterba, J.M. Joint Life Annuities and Annuity Demand by Married Couples. The Journal of Risk and Insurance. 67(4), 527–554 (2000).
[4] Feldman, L., Pietrzyk, R., Rokita, P.: A practical method of determining longevity and premature-death risk aversion in households and some proposals of its application. In: Data Analysis, Machine Learning and Knowledge Discovery, Studies in Classification, Data Analysis, and Knowledge Organization (Spiliopoulou, M., Schmidt-Thieme, L., Janning, R., eds.). Springer International Publishing, 2014, 255–264.
[5] Geyer, A., Hanke, M., Weissensteiner, A. Life-cycle asset allocation and consumption using stochastic linear programming. The Journal of Computational Finance, 12(4), pp. 29-50 (2009).
[6] Hurd, M. Mortality Risk and Consumption by Couples. NBER working paper series, 7048. Cambridge, MA: National Bureau of Economic Research (1999).
Available at: <http://www.nber.org/papers/w7048.pdf> [Accessed: 10 Oct. 2012].
[7] Kotlikoff, L.J., Spivak, A. The Family as an Incomplete Annuities market. Journal of Political Economy, 89(2), 372–391 (1981).
[8] Merton, R.C. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory, 3, pp. 373–413 (1971).
[9] Mitchell, O. S., Bodie, Z., Hammond, B. P., Zeldes, S. Innovations in Retirement Financing. Philadelphia: University of Pennsylvania Press (2002).
[10] Pietrzyk, R., Rokita, P. Stochastic Goals in Household Financial Planning. Statistics in Transition. New Series /In Press/ (2015).
[11] Pietrzyk, R., Rokita, P. On a concept of household financial plan optimization model. Res. Pap. Wroc. Univ. Econ. Press (2015).
[12] Richard, S.F. Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model. Journal of Financial Economics, 2, 187–203 (1975).
[13] Scholz, K., Seshadri, A. Health and Wealth In a Lifecycle Model. [online] University of Wisconsin-Madison and NBER and University of Wisconsin-Madison: Working Paper (2012). Available at <http://www.ssc.wisc.edu/~/Research/Health and Wealth v16.pdf> [Accessed: 23 Nov. 2013].
[14] Yaari, M. E.. Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. The Review of Economic Studies, *32*(2), 137–150 (1965)

Determinants of Manufacturing Bank Deposits: Evidence from Slovakia

Renáta Pitoňáková[1]

Abstract. Industry is a core of Slovak economy, its share in GDP is about 23 %. Within industry, manufacturing assumes key position in employment and GDP. The paper searches for determinants of manufacturing bank deposits within ARDL framework using monthly data 2009 M1 – 2013 M12. In the long-run bank deposits decline with rising long-term interest rate, wages and prices and, increase with higher income. Lagged values of income, wages and prices are important determinants of deposits in the short-run. The process of adjustment is quite fast; more than 50 % of disequilibrium is corrected within one month.

Keywords: ARDL framework, industry, manufacturing, bank deposits, income, interest rate.

JEL Classification: G3
AMS Classification: 91B84

1 Introduction

The level of financial infrastructure is of great importance for enterprises. Financial institutions provide companies with necessary finances and offer products for a secure investment. For cashless payment and settlement, bank accounts are inevitable products. According to [2], bank deposits have peculiar features which combine liquidity, profitability and security and, to the investor, they are the most secured and liquid financial assets available. Deposits are low cost finances for credit institutions obviously involving checking accounts, savings accounts, time deposits and the money market accounts[2]. Secure assets are less profitable. Interest rates on bank deposit are considered a minimum acceptable rate of return. Therefore management of a corporate is to reasonably decide the amount of finances held as bank deposits. Banks offer deposits of short- and long-term maturity. A sound financial market stimulates enterprises to use bank deposits as short-term rather than long-term opportunities. Bank deposits are in such case preferably utilized for payment and settlement purposes and serve as a reserve for unexpected needs and occasions. According to [9], there are a number of reasons why companies prefer liquid assets. Psychological and business incentives to liquidity are analysed by [5]. In compliance with [5], Transaction motive, Precautionary and Speculative motive adding the Compensation balances[3] are focus points in [9]. According to [5], within the Transaction motive entities hold liquid assets to bridge the interval between the receipt of income and its disbursement. The aggregate of liquidity depends on the amount of income and the normal length of the interval between its receipt and its disbursement. Within the Precautionary motive liquidity is held to provide for contingencies requiring sudden expenditure and for unforeseen opportunities of advantageous purchase, and also to hold asset of which the value is fixed in terms of money to meet a subsequent liability fixed in terms of money. The Speculative motive expresses desire to hold resources in liquid form to take advantage of future changes in the rate of interest or bond prices.

The aim of this article is to identify determinants of bank deposits focusing on manufacturing companies due to their importance in economy of Slovakia. The structure of the article is the following. After the introductory part there is an overview of literature and corresponding working papers. Part three describes variables and model. The fourth one presents achieved results, the last one concludes.

[1] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 821 05 Bratislava,
renata.pitonakova@fses.uniba.sk.

[2] Money market account may require a note for withdrawals.

[3] Compensation balances: banks provide a range of financial services, many of which are “free” as long as the company keeps a positive bank balance.

2 Overview of literature and corresponding papers

It is of importance for each corporate management to decide the minimum amount of liquidity it has to hold. Theory indicates some types of liquid assets according to: interest rates on overnight and short-term deposits; transaction costs on conversion deposits into cash and vice versa and liquidity of short-term deposits (demand and supply on the market). Beside those, there are additional factors determining the level of liquid assets in a corporate industry such as: income, GDP, exporting or importing, inflation, economic expectations, employment, labour productivity, and level of wages in given industry (proxy for costs). Economic growth drives business, increases income, employment, and labour productivity. Rising interest rates on deposits stimulate companies to hold deposits with banks. Inflation changes prices of labour factors and companies' outcomes (changes of relative prices of products and services) and hampers companies react adequately. Financial assets are due to rising inflation no longer preferable. Working papers dealing with bank deposits and, focusing on manufacturing in Slovakia, are missing. Existing information are published by the National Bank of Slovakia (NBS) in the majority as part of analysis included in financial reports of the NBS without any detailed empirical investigation. Papers dealing with bank deposits are in general elaborated mostly by foreign researchers, e. g. demand for commercial bank deposits in Lebanon were inspected by [3]. Domestic factors such as economic activity, prices, and the interest differential between the Lebanese pound and the U.S. dollar were significant in explaining deposit demand. The determinants of private-sector deposits to domestic commercial banks in Greece were examined in [6]. The study documents a strong positive relation between bank deposits and bank credit to the domestic banking sector in the periods before and after the eruption of the sovereign debt crisis. A strong positive influence exists between bank deposits and the level of gross national product. Paper elaborated by [4], focused upon determinants of non-resident deposits in commercial banks in Lebanon. A relationship between macroeconomic determinants and bank deposits in Nigeria were modelled in [2]. The results show that bank investment, bank branches, interest rate and the general price level are important determinants of bank deposits.

3 Description of variables and model

Data were taken from databases of National Bank of Slovakia (NBS), and Statistical Office of the Slovak Republic. Figures 1 – 4 present manufacturing bank deposits in nominal values in EUR. Figure 1 depicts total deposits (DT), overnight accounts (DON) and deposits with agreed maturity (DM). Figure 2 presents deposits with agreed maturity (DM), and short-term deposits up to 1 year (DMS), and Figure 3 long-term deposits over 1 year up to 5 years (DMM), Figure 4 long-term deposits over 5 years (DML).

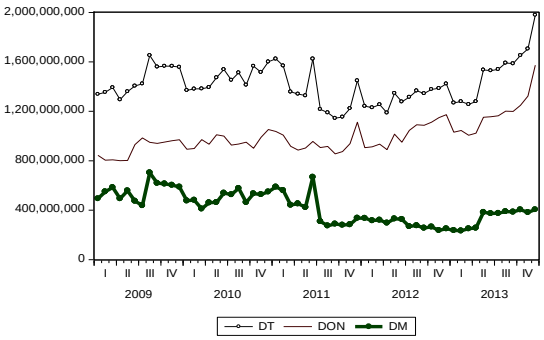


Figure1 Total deposits (DT), overnight accounts (DON) and deposits with agreed maturity (DM).

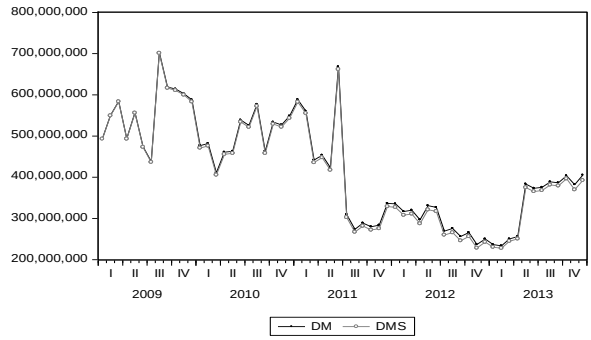


Figure 2 Deposits with agreed maturity (DM), and short-term up to 1 year (DMS).

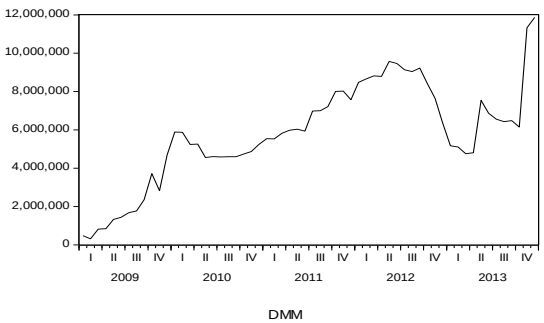


Figure 3 Long-term deposits over 1 year up to 5 years (DMM).

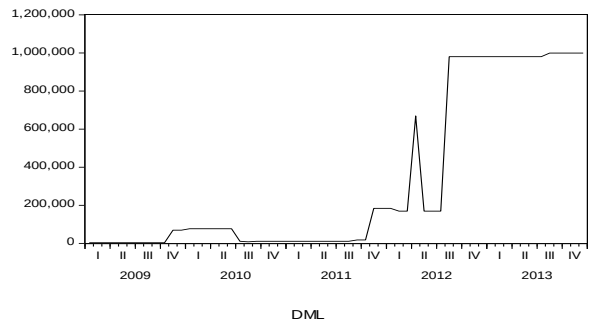


Figure 4 Long-term deposits over 1 year up to 5 years (DML).

In modelling, following variables for manufacturing companies were used: employment, short-term interest rate (EONIA, EURIBOR), income, labour productivity, producer price index, wages, long-term interest rate and export. From these, income, wages, long-term interest rate and producer price index significantly influence manufacturing bank deposits. The share of overnight deposits on entire manufacturing bank deposits is about 80 %. In spite of the low interest rate on these deposits (from 0, 01 % p.a.to 0, 10 % p.a.) they are due to credit possibility a preferable bank product. Figure 1 proves, that manufacturing companies hold finances (liquidity)[4] with banks mainly for payment and settlement purposes (transaction motive) as the majority of deposits falls on accounts with immediate access (DON). From deposits with agreed maturity (DM) (time or term deposits) those with maturity up to 1year (DMS) are by companies most preferable. These enable flexible deposit manipulation when new or better financial possibilities occur on the financial or money market (speculative motive). Medium-term deposits 1-5 years (DMM) and long-term deposits over 5 years (DML) reflect companies precautionary behaviour rather than their investment appetite.

For our investigation we apply ARDL technique[5]presented by [7], [8]. ARDL approach enables to identify cointegration relation in small samples and can be applied to regressors of I(1) and I(0). In ARDL procedure variables can have different number of lags. The ARDL approach requires the existence of long-term relationship among variables (based upon F-test), the estimation of coefficients of the long-term relation and the estimation of short-term elasticity of variables with ECT (Error Correction Term) of the ARDL model [1]. According to [7], if the computed Wald or F-statistic falls outside the critical value bounds, a conclusive inference can be drawn without needing to know the integration/cointegration status of the underlying regressors. However, if the Wald or F-statistic falls inside these bounds, inference is inconclusive and knowledge of the order of the integration of the underlying variables is required before conclusive inferences can be made.

A basic form of the ARDL model is as follows:

$$Y_t = \beta_0 + \sum_{k=1}^{r} \chi_k L^k Y_t + \sum_{m=0}^{s} \delta_m L^m X_t + u_t \quad (1)$$

where: β_0 - intercept; χ_k , δ_m - coefficients; L – lag operator (e. g. $LX_t = X_{t-1}$); u_t – error term.

As it is well known the ARDL model helps us to study

a) the short-run reaction and
b) the long-run reaction that are important in many economic problems.

ARDL can be reproduced to ECM[6] ARDL version by transforming variables (1) into differences and lags. Equation (2) displays unrestricted ARDL ECM structure that we apply for identifying short and long-term relations among variables.

$$\Delta D_t = \alpha_0 + \sum_{j=1}^{m} \beta_j \Delta D_{t-j} + \sum_{j=0}^{m} \gamma_j \Delta I_{t-j} + \sum_{j=0}^{m} \delta_j \Delta IR_{t-j} + \sum_{j=0}^{m} \theta_j \Delta W_{t-j} + \sum_{j=0}^{m} \mu_j \Delta P_{t-j} + \lambda_1 D_{t-1} + \lambda_2 I_{t-1} + \lambda_3 IR_{t-1} + \lambda_4 W_{t-1} + \lambda_5 P_{t-1} + u_t \quad (2)$$

where: D - Deposits of manufacturing companies[7] (seasonally not adjusted, transformed into log, deflated by Producer Price Index); I - Income of manufacturing[8] (seasonally not adjusted, transformed into log, deflated by Producer Price Index); IR - Long-term interest rate (yield to maturity of the 10 years government bond); W - Average wages of employees of industry[9] (indices the sppy = 100); P - Industrial producers price index[10].

The restricted ARDL ECM model involving error correction term is the following:

$$\Delta D_t = \alpha_0 + \sum_{j=1}^{m} \beta_j \Delta D_{t-j} + \sum_{j=0}^{m} \gamma_j \Delta I_{t-j} + \sum_{j=0}^{m} \delta_j \Delta IR_{t-j} + \sum_{j=0}^{m} \theta_j \Delta W_{t-j} + \sum_{j=0}^{m} \mu_j \Delta P_{t-j} + \omega ECT_{t-1} + u_t$$

[4] Bank deposits of different maturities represent liquidity as this can be withdrawn from bank without or with charges relatively easily.
[5] Estimations in E-Views.
[6] ECM includes variables in first differences and an error-correction term [10], [11].
[7] Total credit institutions: Deposits for manufacturing companies according to Classification of Economic Activities: C. Deposits involve: Overnight deposits, Deposits with agreed maturity (short-term up to 1 year; long-term over 1 year up to 5 years, and over 5 years).
[8] Income (Receipts of own output and goods in industry in EUR (www. statistics.sk)).
[9] By sections and special aggregates of the industrial branches of SK NACE Rev. 2 classification.
[10] By economic activities (NACE Rev. 2) without excise duties - total (base year 2005=100).

(3)

where: Δ - first difference operator; ω – speed of adjustment; ECT - Error correction term: lag residuals from the long- run relationship.

To prove the cointegration relation among variables the ECT coefficient has to be negative and significant. ARDL procedure needs to set appropriate lags of variables. Following AIC criterion two lags should be appropriate (Table 1).

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-179.9611	NA	0.000574	6.725859	6.908344	6.796428
1	-3.510762	314.4025	2.34e-06	1.218573	2.313482*	1.641983*
2	27.96252	50.35725*	1.89e-06*	0.983181*	2.990514	1.759433
3	47.39193	27.55444	2.46e-06	1.185748	4.105505	2.314842
4	66.93216	24.15883	3.37e-06	1.384285	5.216467	2.866221
5	91.27550	25.67116	4.23e-06	1.408164	6.152769	3.242941

Table 1 Lag Selection Criteria

* indicates lag order selected by the criterion;
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 2 presents the results from ADF and PP unit root tests: D is I(1), I is I (0), IR is I(1), W is I(0), and P is I(1).

ADF						**PP**					
	D	I	IR	**W**	P		D	I	IR	W	P
c	0.3139	0.0152	0.6577	**0.0000**	0.6646	c	0.3139	**0.0211**	0.7995	**0.0000**	0.6440
ct	0.7262	**0.0001**	0.7247	**0.0000**	0.9381	ct	0.8408	**0.0076**	0.7339	**0.0000**	0.9880
c-diff	**0.0000**	**0.0010**	**0.0001**	**0.0000**	**0.0009**	**c-diff**	**0.0000**	**0.0000**	**0.0000**	**0.0001**	**0.0003**
ct-diff	**0.0000**	**0.0000**	**0.0004**	**0.0000**	**0.0024**	**ct-diff**	**0.0000**	**0.0000**	**0.0001**	**0.0001**	**0.0017**

Table 2 Unit root tests (p-values)

To test for cointegration, the F test was applied to test the null hypotheses of no cointegration against the alternative of cointegration.

H_0: $\lambda 1 = \lambda 2 = \lambda 3 = \lambda 4 = \lambda_5 = 0$

$H_{1:}$ $\lambda 1 \neq \lambda 2 \neq \lambda 3 \neq \lambda 4 \neq \lambda_5 \neq 0$

Table 3 involves Critical Value Bounds for the F-Statistic used for testing the Existence of a long-run relationship.

	Bound Testing							
	90%		95%		97,5%		99%	
K	I(0)	I(1)	I(0)	I(1)	I(0)	I(1)	I(0)	I(1)
4	2.45	3.52	2.86	4.1	3.25	4.49	3.74	5.6

Table 3 Pesaran Critical Value Bounds for the F-Statistic

The computed F-statistic (10.12) is higher than the upper bound values tabulated by [8], for unrestricted intercept and no trend for four variables (without lagged dependent variable). This confirms long-run equilibrium among variables and justifies the application of an ARDL procedure.

4 Interpretation of Results

Table 4 presents the estimations of the ECM ARDL modelling (equation 2 and 3) after applying General to Specific Method. The coefficient of the lagged dependent variable (Deposits) is negative and significant, variables are cointegrated.

	ECM ARDL unrestricted	**ECM ARDL restricted**
Variable	Δ D	Δ D
c	9.812*** [3.19]	
D(-1)	-0.836*** [-6.60]	
IR(-1)	-0.070*** [-3.53]	
I(-1)	0.708*** [4.17]	
P (-1)	-0.045*** [-5.51]	
W(-1)	-0.038*** [-4.36]	
Δ I(-1)	-0.430*** [-3.27]	
Δ I(-2)	-0.233* [-1.94]	
Δ W	-0.010** [-2.36]	
Δ W(-1)	0.020*** [3.06]	0.007* [1.94]
Δ W(-2)	0.009** [2.07]	
Δ P(-1)	0.034* [1.94]	
ECT (-1)		-0.586*** [-4.79]
R^2	0.57	0.30
Adj R^2	0.46	0.29
Note: ***,**,* imply significance at 1%, 5%, 10% levels respectively t-statistic in parentheses		

Table 4 Results of ECM ARDL modelling

Calculated elasticity of variables from the ECM ARDL unrestricted model (2) identifies impact of explanatory variables upon dependent variable. This is done by dividing the coefficients of explanatory variables IR (-1); I (-1); P (-1) and W (-1)) by negative value of the coefficient of dependent variable (D (-1)). The results are presented in table 5.

Variable	Elasticity
IR (-1)	-0.0844
I (-1)	0.8468
P (-1)	-0.0540
W (-1)	-0.0465

Table 5 Long run Effect of Explanatory Variables

Results indicate motives that stimulate companies to hold liquid assets. It is obvious that companies do not hold bank accounts for investment reasons, therefore only a small amount of liquid assets are held on long-term accounts. If there is a better investment opportunity, companies mobilize finances on short-term accounts for their transformation into more profitable investment tools, e. g. bonds if the long-term rate of return rises. The semielasticity is -0. 08. Improved economic conditions result in higher receipts of manufacturing. Companies increase bank deposits that are, later on used for making necessary payments in appropriate time e. g. pay-

ments for wages, material. Inflation corrodes deposits, increases prices of inputs and outputs, declines savings, hence companies are willing to increase consumption rather than pool liquid assets with banks. Rising wages indicate favourable economic conditions, but as such they reduce financial sources of manufacturing. Interestingly, in the short-term, higher income from previous months decline manufacturing bank deposits. One of possible reasons might be the fact, that, due to inflation, companies expect to pay higher prices for inputs and in addition feel constrained to pool liquid assets on bank accounts under declining value. Lagged wages increase bank deposits in the short-run as companies accumulate liquidity necessary for paying higher wages later on. The coefficient of adjustment (ECT) is negative and significant. It indicates how quickly variables restore equilibrium. The speed of adjustment is high; more than 58 % of disequilibrium in the previous month will be corrected in the current month. Table 6 presents the results from residual tests.

LM-test (Breusch-Godfrey)	0.487	(0.783)
Heteroscedasticity Test: Breush-Pagan-Godfrey	9.099	(0.612)

Table 6 Results from residual tests

Note: P-values in parentheses.

The p-values of the autocorrelation (BG-test) and heteroscedasticity (BP-test) show that the null hypothesis of non-autocorrelation and homoscedasticity is not rejected.

5 Conclusion

Bank deposits are a major source of low cost funds in banking industry. It is a safe form of assets economic entities can use for cashless payments or to hold for a short-run. The management of bank deposits is of great importance for effective allocation of liquid assets. Central banks influence the actions of commercial banks among others in their credit and deposit interest rates policy. Cash flow of a company indicates the necessary amount which corresponds to the needs resulting from the transaction motive. Moreover, company has to set forth the minimum of finances referring beside transaction motive to precautionary and speculative motives regarding unexpected fluctuations in netto cash flow, attitude of management toward risk taking and, possibilities for utilize unexpected situations on the market. Knowledge on what drives industry companies to invest on bank accounts is of a great importance for the financial institutions. It can help them set appropriate marketing tools for getting finances from private sector. It is relevant also for industry companies when deciding about investment possibilities. It is obvious, that companies, including those, operating in manufacturing branch in Slovakia, prefer checking accounts (overnight accounts) for payment and settlement purposes, short-term deposits for speculative reasons and long-term deposits for precautionary motives.

Article prepared within APVV-14-0020 research project.

References

[1] Adeniji, S.: Investigating the Relationship between Currency substitution, Exchange Rate and Inflation in Nigeria: An Autoregressive Distributed Lag (ARDL) Approach. *MPRA*, 2013.
[2] Eriemo, N.: Macroeconomic Determinants of Bank Deposits in Nigeria. *Journal of Economics and Sustainable Development*, vol. 5 (10), 2014.
[3] Finger, H., and Hesse, H.: Lebanon-Determinants of Commercial Bank Deposits in a Regional Financial Center. *IMF WP*/09/195, 2009.
[4] Kanj, O., and Khoury, El, Rim.: Determinants of Non-Resident Deposits in Commercial Banks: Empirical Evidence from Lebanon. *International Journal of Economics and Finance*, vol. 5(12), 2013.
[5] Keynes, J. M. *The General Theory of Employment, Interest and Money*. Available on: http://cas.umkc.edu.
[6] Monokrousos, P., and Thomakos, D.: A Technical study on the determinants and Outlook of private sector deposits in the Greek banking system. *ECONOMY and MARKETS*, vol. V (8), 2011.
[7] Pesaran, M. H., and Shin, Y.: An autoregressive distributed lag modeling approach to cointegration analysis. *Centennial Volume of Ragnar Frisch*, eds. S. Strom, A. Holly and P. Diamond, Cambridge University Press, Cambridge, 1999.
[8] Pesaran, H. M., Shin, Y., and Smith. J. R.: Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics*, 289 – 326, 2001.
[9] Pike, R., and Neale, B.: *Corporate Finance and Investment Discussion & Strategies*, fifth Edition, FT Prentice Hall, 2006.
[10] Vogelvang, B.: *Econometrics*. FT Prentice Hall, 2005.
[11] Želinský, T., Gazda, V., and Výrost, T.: *EKONOMETRIA*. Košice: TU, 2010.

Productivity and real exchange rates in the euro area

Klára Plecitá[1], Ladislava Issever Grochová[2]

Abstract. According to the Balassa-Samuelson hypothesis, which is regularly used as an explanation for real exchange rate deviations from the Purchasing Power Parity, growth of productivity in the tradable sector leads (ceteris paribus) to real exchange rate appreciation. However the empirical evidence for the positive impact of growth in productivity of tradables on real exchange rate is at least mixed and depends on productivity proxies used to test this hypothesis.
In our paper, we use a panel estimation framework to investigate the impact of labour productivity shocks on real effective exchange rates among individual euro area countries over the period 1999 – 2013. Besides the commonly used indicator of the overall tradable-sector productivity we also apply a set of sector-specific indicators of labour productivity for agriculture, construction, manufacturing and services, which enables us to assess the impact of productivity shocks on real exchange rates stemming both from the tradable and the non-tradable sector of the economy.
We find evidence that the intra-euro-area real effective exchange rate is affected by productivity shocks stemming both from the tradable and non-tradable sector of the economy, which contradicts the Balassa-Samuelson hypothesis.

Keywords: sectoral productivity, panel data, real exchange rate, euro area.

JEL Classification: C23
AMS Classification: 62P20

1 Introduction

The Purchasing Power Parity (PPP) is one of the oldest and mostly used concepts for the long-run real exchange rate determination. According to the absolute version of the PPP the price of a common basket of goods will be identical when expressed in the same currency. In other words, the real exchange rate will be constant. However in reality the real exchange rate can deviate from the PPP in the short run due to various monetary shocks and in the long run due to persistent real shocks [10]. When monetary shocks to the real exchange rate predominate, the PPP can be used successfully as an equilibrium concept for the real exchange rate in the long run. However if shocks to the real exchange rate are caused by real factors, like changes in the productivity differential, changes in the world's demand for country's specific products etc., then the PPP is a poor instrument for real exchange rate determination.

More than 50 years ago Balassa [1] and Samuelson [17] argued, that the PPP does not hold in the long run due to different productivity growth rates in the tradable sector among various countries. According to the Balassa-Samuelson hypothesis, the main source of innovation and productivity growth in an economy lies in the tradable sector – the non-tradable goods are assumed to be more labor intensive with relatively low potential for technological progress. Since both Balassa and Samuelson assumed that the law of one price holds due to arbitrage in the highly competitive tradable sector, relative growth of productivity in the tradable sector will not affect prices of tradable goods. Growth of productivity in the tradable sector thus leads (ceteris paribus) to wage increases, which in turn lead to higher prices of non-tradables (since there is no equivalent increase in non-tradables productivity) and real exchange rate appreciation. The outcome of the Balassa-Samuelson hypothesis is that relatively less developed countries with relatively low levels of GDP per capita will catch up with the relatively more developed countries and their real exchange rate will appreciate relatively more than the PPP theory predicts.

Even if the positive effect of the relative productivity growth on the real exchange rate (appreciation) is broadly accepted, the empirical evidence for the Balassa-Samuelson effect is at least mixed. On the basis of large literature survey on the Balassa-Samuelson effect, Peltonen and Sager [15] and Lee and Tang [12] conclude, that the validity of the Balassa-Samuelson hypothesis was confirmed on the basis of cross-sectional analysis of post-war data, whereas the results of time-series analysis were inconclusive. According to Peltonen and Sager [15] the empirical results are sensitive to the choice of the productivity measure, numéraire currency (euro is more supportive than US dollar) and countries analyzed (the wider sample, the stronger effect).

[1] Mendel University in Brno, Department of Finance, Zemědělská 1, Brno, klara.plecita@mendelu.cz.
[2] Mendel University in Brno, Department of Economics, Zemědělská 1, Brno, ladislava.grochova@mendelu.cz.

Among the other widely accepted long-run real determinants of real exchange rates belong terms of trade and net foreign assets [16]. Terms of trade play a specific role in real exchange rate determination, since they mirror both the demand-side and supply-side shocks [4] and thus there are more possible channels through which they can affect real exchange rates. Typically terms of trade are expected to have a positive effect on real exchange rates. Terms of trade can affect real exchange rates for example through a commodity price shock or through a shift in foreign demand to higher-value exports. A favorable shift in consumer preferences concerning domestic tradables leads to increases in relative prices of domestic exports, which trigger a substitution effect - domestic producers shift their focus to the traded sector, which initiates first wage increases in the tradable sector and later (due to the wage equalization process) in the non-tradable sector. This increases the overall domestic price level and appreciates the real exchange rate. Nevertheless, the substitution effect can be overridden by a wealth effect, which causes real exchange rate depreciation and the impact of a commodity-price shock depends upon commodity dependency of the domestic economy. Thus the empirical evidence on the impact of terms of trade on the real exchange rate can be ambiguous (see Frait *et al.* [5] for an overview).

Traditionally it is expected that improvement in net foreign assets increases household wealth, which enables higher absorption and current account deficits. The effect of improving (raising) net foreign assets on real exchange rate is positive, either by increasing the overall domestic price level or by an appreciation of the nominal exchange rate [2].

Even if the introduction of the common currency euro, together with the free movement of goods, services and capital inside the single EU market, has been considered as a necessary condition for the goods arbitrage, the empirical evidence provides limited support for PPP in the euro area (see Lopez and Papell [14] and Koedijk *et al.* [9]). Thus in our paper we use a panel estimation framework to estimate the relationship between real effective exchange rates for individual euro-area countries vis-à-vis the rest of the euro area (hence intra-euro-area real effective exchange rates, *reer*) and their long-term real fundamentals, namely productivity differentials. Although countries in the euro area use the common currency euro, it is still meaningful to analyze the behavior of their intra-euro-area real effective exchange rates, as they mirror not only the price competitiveness of individual member states, but also their internal and external balance.

2 Data and methodology

Our empirical analysis is conducted within a panel framework, using yearly data over 1999 – 2013 and covering 12 euro area countries (EA12), which adopted the common currency euro (in cash) in 2002. The countries under consideration are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain. Since not all data are individually available for Belgium and Luxembourg, in our analysis we use joint data for the Belgium-Luxembourg Economic Union (BLEU).

The real effective exchange rate (*reer*) vis-à-vis the rest of the EA12 is calculated using the methodology described by Schmitz *et al.* [18] and since *reer* is quoted in indirect form, an increase (decrease) in *reer* represents an appreciation (depreciation). The data on nominal effective exchange rates and harmonized indices of consumer prices used for *reer* calculation are obtained from the European Commission and Eurostat, respectively. Moreover the same weights used by the European Commission to calculate the nominal effective exchange rates are also used to calculate the relative terms of trade (*tot*). Terms of trade are defined as a ratio of domestic export prices to domestic import prices and the source of the data is the Ameco database. Relative terms of trade (*tot*) are then calculated as a ratio of domestic terms of trade to the geometrically averaged terms of trade for the rest of the EA12.

We calculate labor productivity as real value added per employment in the following sectors: agriculture, forestry and fishery products (hereafter agriculture), manufacturing industry (hereafter manufacturing), building and construction (hereafter construction) and services. The data are taken from the Ameco database. Agriculture and manufacturing are both considered as tradable sectors, whereas construction and services both represent non-tradable sectors. Hence tradable-sector productivity is calculated as real value added per employment in agriculture and manufacturing and non-tradable-sector productivity consists of real value added per employment in services and construction. In our estimation process we apply three different approaches to capture the effect of labor productivity on the intra-euro-area real effective exchange rate: (1) we estimate the Balassa-Samuelson effect traditionally via tradable-sector productivity differential (*t*); (2) we estimate the effect of both the tradable-sector productivity differential (*t*) and the non-tradable-sector productivity differential (*nt*) on *reer;* and (3) we estimate the elasticity between *reer* and productivity differentials in individual sectors – agriculture (*agri*), manufacturing (*ind*), construction (*con*) and services (*serv*). All sector-productivity differentials are calculated as domestic productivity divided by the average productivity in the rest of the EA12.

Preliminary to the estimation process, we test all variables for unit roots using a test developed by Levin, Lin and Chu [13]. The optimal number of lags is chosen with Bayesian information criterion. All variables are found to be stationary or trend stationary (see Table 1). In a panel regression we estimate following 3 equations:

$$\ln reer_{it} = \mu + \beta_0 \ln tot_{it} + \beta_1 nfa_{it} + \beta_2 \ln t_{it} + \varepsilon_{it} \tag{1}$$

$$\ln reer_{it} = \mu + \beta_0 \ln tot_{it} + \beta_1 nfa_{it} + \beta_2 \ln t_{it} + \beta_3 \ln nt_{it} + \varepsilon_{it} \tag{2}$$

$$\ln reer_{it} = \mu + \beta_0 \ln tot_{it} + \beta_1 nfa_{it} + \beta_2 \ln agri_{it} + \beta_3 \ln ind_{it} + \beta_4 \ln const_{it} + \beta_5 \ln serv_{it} + \varepsilon_{it} \tag{3}$$

In order to choose the correct model we run panel diagnostic tests, concretely the F-test, Breusch and Pagan's [3] Lagrange multiplier (LM) test and Hausman test. The first compares a fixed effect model and OLS to see how much the fixed effect model can improve the goodness-of-fit, whereas the second contrasts a random effect model with OLS. Finally, the similarity between random and static fixed effect estimators is tested by a Hausman test used to decide whether random effects (under the null hypothesis) or fixed effects (under the alternative hypothesis) are more efficient. The structure of error terms is verified with Modified Wald test [6] and Wooldridge test [19].

variable	adjusted t-statistic	p-value
ln*reer*	-5.3174	0.0000
ln*tot*	-2.6457	0.0041
nfa	-2.6017	0.0046
ln*tnt*	-2.7692	0.0028
ln*t*	-2.2884	0.0111
ln*agri*	-3.0829	0.0010
ln*ind*	-2.1095	0.0175
ln*const*	-1.5226	0.0639
ln*const* (detrended)	-3.4231	0.0003
ln*serv*	-0.1420	0.4435
ln*serv* (detrended)	-2.9664	0.0015

Table 1 Levin, Lin and Chu panel unit root test (H_0: Panels contain unit roots)

3 Empirical results and discussion

Before we investigate how real factors, especially labour productivity differentials, affect *reer* (the real effective exchange rates of individual euro-area countries vis-à-vis the rest if the euro area), it is worth noting, that on basis of the Levin, Lin and Chu test (see Table 1), we have rejected the null hypothesis about nonstationarity in *reer*. This can be seen as evidence in favour of multi-country version of PPP among the euro-area countries, which implies that shocks to the intra-euro-area real effective exchange rate may be only temporary. Even if this finding is in contrast to our previous results - in Issever Grochová and Plecitá [8] using quarterly data and shorter time span, we have failed to rejected nonstationarity in *reer* - it seems to be in line with empirical literature on PPP, which provides limited support for PPP in the euro area (see Lopez and Papell [14]and Koedijk et al. [9]).

Results of our panel estimations of equations (1-3) are reported in Table 2. In all three models relative terms of trade *tot* and net foreign assets *nfa* are statistically significant and of similar magnitude. The elasticity of relative terms of trade is positive, which means that an improvement (increase) in *tot* leads to appreciation of *reer*, thus indicating that in the case of the euro area, the substitution effect dominates the wealth effect. Economic theory expects a positive relationship between net foreign assets and real exchange rates and this expectation is often affirmed by empirical analysis. However, similarly to Issever Grochová and Plecitá [8] and Borgy et al. [2] our results indicate, that in the case of the euro-area countries, negative relationship between *reer* and *nfa* may exist. In our opinion, as we have argued in Issever Grochová and Plecitá [8], this negative relationship between *reer* and *nfa* may mirror the accumulation of economic imbalances inside the euro area. The logic of this argument is as follows [8]: *"around the creation of the euro area the catching-up (peripheral) countries were expected to have higher growth potential than the core countries, therefore the catching-up countries were more appealing to foreign capital. Moreover, the decrease of interest rates in the catching-up countries and the increase of external credit from the core countries with higher savings helped to increase their aggregate demand*

(mainly consumption) and fuel speculative bubbles, which raised the domestic price level and wages in both the non-traded and traded sector. This led in the catching-up countries to the appreciation of their real effective exchange rates vis-à-vis the rest of the EA12, to the loss of their price competitiveness in the tradable sector and to persistent current account deficits." Since our data cover only 16 years, it is possible that the effect of accumulation and de-accumulation of macroeconomic imbalances inside the euro area has overcast the theoretically expected positive effect of *nfa* on *reer*. Thus it is reasonable to expect, that over time (with longer time series available) the estimated sign of *nfa* will be positive.

	Model 1	**Model 2**	**Model 3**
ln*tot*	0.0684 (0.0176)***	0.0952 (0.0160)***	0.1000 (0.0165)***
nfa	-0.0001 (0.0000)***	-0.0001 (0.0000)***	-0.0001 (0.0000)***
ln*t*	-0.0127 (0.0049)**	0.0029 (0.0045)	
ln*nt*		-0.0748 (0.0085)***	
ln*agri*			0.0093 (0.0018)***
ln*ind*			-0.0105 (0.0053)*
ln*const*			-0.0152 (0.0024)***
ln*serv*			-0.0224 (0.0101)*
constant	4.6130 (0.0019)***	4.6160 (0.0014)***	4.6140 (0.0017)***
N	165	165	165
R-sq	0.4053	0.3307	0.5254
F-test	24.1000	24.1000	15.7600
p-value	0.0000	0.0000	0.0000
Modified Wald test for groupwise heteroskedasticity	772.4400	4600.7500	2072.0100
p-value	0.0000	0.0000	0.0000
Wooldridge test for autocorrelation	459.3460	459.3460	558.3850
p-value	0.0000	0.0000	0.0000
Breusch and Pagan Lagrangian multiplier test for random effects	264.9900	259.1000	185.2600
p-value	0.0000	0.0000	0.0000
Hausman test for fixed effects	23.1500	12.5000	9.1200
p-value	0.0000	0.0140	0.1668

Note: Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001

Table 2 Estimation results for *reer*

The main difference between Models 1 – 3 lies in the proxy used to capture the effect of labour productivity on real effective exchange rates among the euro-area members. The original Balassa-Samuelson hypothesis assumes that the only source of productivity growth lies in the tradable sector. Thus in Model 1 we use only one indicator of sectoral productivity - the tradable-sector productivity differential *t*. Even if statistically significant, the negative elasticity between the tradable-sector productivity differential and *reer* contradicts the Balassa-Samuelson effect.

Within Model 2 we allow not only the tradable but also the non-tradable sector to be the source of technological innovation and productivity growth. Thus we expand the Model 1 by adding another productivity proxy– the productivity differential in the non-tradable sector *nt*. Similarly to Model 1, our results of Model 2 also contradict the prediction of the Balassa-Samuelson effect. The estimated elasticity for the tradable sector productivity differential in Model 2 is positive as predicts the Balassa-Samuelson hypothesis, albeit statistically insignificant. And the elasticity for the non-tradable sector productivity differential is negative and statistically significant. Thus our estimated results suggest, that in contrast to the Balassa-Samuelson hypothesis, changes in the relative productivity stemming from the tradable sector do not affect intra-euro-area real effective exchange rates, whereas relative increase in inter-country non-tradable productivity decreases domestic prices and leads to real effective exchange rate depreciation.

In Model 3 we use industrial productivity proxies, i.e. individual labour productivity differentials for agriculture, industry, construction and services. The estimated elasticities for agriculture and industry, which in our model together represent tradable sector, are both statistically significant; however with opposing signs. The estimated positive elasticity of agriculture, which is fully in line with predictions of the Balassa-Samuelson effect, and the negative elasticity for industry, which contradicts the Balassa-Samuelson hypothesis, help to explain, why in Model 2 the overall tradable productivity *t* is statistically insignificant. Moreover the statistically significant negative elasticities of construction and services, which in our model represent non-tradable sector, contradict the Balassa-Samuelson effect as well.

Reflecting all estimated sectoral labour productivities from Table 2, we can conclude, that in contrast to the assumption of the Balassa-Samuelson effect, relative productivity changes stemming from construction and services (i.e. non-tradable sector) have an impact, albeit negative, on intra-euro-area real effective exchange rates. The positive effect of inter-country changes in tradable productivity, which is predicted by the Balassa-Samuelson hypothesis, is identified only in the case of agriculture in Model 3, whereas the other tradable-sector productivity proxies used are either statistically insignificant or have a negative sign.

Likewise results are presented for example in Peltonen and Sager [15], Lee and Tang [12] and Gubler and Sax [7]. As mention Peltonen and Sager [15] who have reviewed substantial part of literature focusing on the Balassa-Samuelson effect, the support in favour of the Balassa-Samuelson hypothesis is found mainly in studies, which are using very large samples of countries at various levels of economic development. However if the number of countries under consideration is limited and the countries can be considered similarly developed, as is in our case, the estimated results often contradict the Balassa-Samuelson hypothesis.

The lack of evidence in favour of the Balassa-Samuelson effect is in literature often explained on the basis of New Open Macroeconomics, which takes into account international pricing decisions. Typical explanation why the Balassa-Samuelson hypothesis does not hold is pricing to the market (Lee and Tang [12], Peltonen and Sager [15]). Since the tradable-goods prices are affected also by the cost of distribution service (non-tradable), the prices of tradables cannot be equalised internationally and firms thus apply pricing to the market. Therefor not only tradables productivity but also non-tradables productivity will affect real exchange rate (Lee and Tang [12]). Lee and Tang [12] and Lee [11] explain negative relationship between inter-country productivity differential and real effective exchange rate on the basis of high elasticity of substitution between domestic and foreigner tradables. In case of a positive domestic productivity shock in tradables, domestic producers have to reduce their prices in order to persuade consumers to absorb their higher output. Thus an increase in productivity in the domestic tradable sector can lead to real depreciation. Another explanation, why improvement in productivity leads to real depreciation comes from Gubler and Sax [7] and is based on the idea of skill-biased technological change. They show that introduction of technological innovations requires high-skilled labour, which leads to lower wages for low-skilled labour and ultimately to a lower domestic price level and real exchange rate depreciation. In our opinion all of these explanations seem to be reasonable for the euro area countries and thus justify our empirical results.

4 Conclusion

In this paper we used a panel estimation framework to investigate the impact of labour productivity, net foreign assets and terms of trade on real effective exchange rates among individual euro area countries over the period 1999 – 2013. Our special focus was on labour productivity shocks, thus we used different labour productivity proxies in our models. Besides the commonly accepted indicator of the overall tradable-sector productivity we also applied a set of sector-specific indicators of labour productivity for agriculture, construction, manufacturing and services, which enabled us to assess the impact of productivity shocks on real exchange rates stemming both from the tradable and the non-tradable sector of the economy. In our results we found no evidence in favour of the Balassa-Samuelson effect. Moreover, in contrast to assumptions of the Balassa-Samuelson hypothesis we identified non-tradable sector as an important source of productivity shocks to real effective exchange rates

among the euro-area countries. All labour productivity proxies in our models, with the exception of the agriculture productivity differential, exhibit negative effect on the intra-euro-area real effective exchange rates.

References

[1] Balassa, B.: The Purchasing Power Parity Doctrine: A Reappraisal. *Journal of Political Economy* **72** (1964), 584–596.

[2] Borgy, V., Lahreche-Revil, A., and López-Villavicencio, A.: Exchange rate dynamics in monetary union: an empirical assessment. In: *Ten years of a common currency: achievements and policy challenges in the euro area*, 2008.

[3] Breusch, T. S., and Pagan, A. R.: The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* **47** (1980), 239-253.

[4] DeLoach, S.: More Evidence in Favor of the Balassa Samuelson Hypothesis. *Review of International Economics* **9** (2001), 336–342.

[5] Frait, J., Komárek, L., and Melecký, M.:. The real exchange rate misalignment in the five central European countries. In: *Warwick economic research papers*, No 739, 2004.

[6] Greene, W.: *Econometric Analysis*. Upper Saddle River, NJ: Prentice–Hall, 2000.

[7] Gubler, M., and Sax, C.: The Balassa-Samuelson effect reversed: New evidence from OECD countries. In: *WWZ Discussion paper*, No 9, 2011.

[8] Issever Grochová, L., and Plecitá, K.: Behavioral Equilibrium Exchange Rates: Re-estimation for the Euro Area Countries. forthcoming.

[9] Koedijk, K. G., Tims, B., and Dijk, M. A.: Purchasing Power Parity and the Euro Area. *Journal of International Money and Finance* **23** (2004), 1081–1107.

[10] Lafrance, R., and Schembri, L.: Purchasing-Power Parity: Definition, Measurement, and Interpretation. In: *Bank of Canada Review*, Ottawa, 2002, 27–32.

[11] Lee, J.: Transfer Effect in National Price Levels. *Review of World Economics* **143** (2007), 534–556.

[12] Lee, J., and Tang, M.: Does Productivity Growth Appreciate the Real Exchange Rate? *Review of International Economics* **15** (2007), 164–187.

[13] Levin, A., C.F. Lin, and Chu, C.S. J.: Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* **108** (2002), 1-24.

[14] Lopez, C., and Papell, D. H.: Convergence to Purchasing Power Parity at the Commencement of the Euro. *Review of International Economics* **15** (2007), 1–16.

[15] Peltonen, T. A., and Sager, M.: Productivity shocks and real exchange rates: a reappraisal. In: *European Central Bank Working Paper Series*, No 1046, 2009.

[16] Salto, M., and Turrini, A.: Comparing alternative methodologies for real exchange assessment. In: *Economic papers*, No 427, 2010.

[17] Samuelson, P.: Theoretical Notes on Trade Problems. *Review of Economics and Statistics* **46** (1964), 145–154.

[18] Schmitz, M., De Clercq, M., Fidora, M., Lauro, B., and Pinheiro, C.: Revisiting the effective exchange rates of the euro. In: *European Central Bank Occasional Paper Series*, No 134. 2012.

[19] Wooldridge, J. M.: *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press, 2002.

Identification of a time-varying model using the wavelet approach and the AR process

Jitka Pomenkova[1], Eva Klejmova[2]

Abstract. The aim of the paper is to give recommendation for working with time frequency modeling of macroeconomic time series on the basis of a comparative study. We investigated the wavelet analysis and the time-varying autoregressive process. We focused on two main areas - sample size for an available data set and its shortening and optimization of parameters of mentioned methods. In the case of the time-varying autoregressive process, we investigated optimization of parameters such as lag length, windowing and overlap. In the wavelet analysis approach, we investigated the type of wave and scale. Performance of the methods was presented on the gross domestic product data of USA, United Kingdom and Korea. These representatives were chosen from the perspective of available sample size and for the reason that the countries represent economy in different geographic areas. An advantage of the wavelet analysis is better time resolution. An autoregressive process provides better frequency resolution, but it is quite sensitive to sample size.

Keywords: Wavelet transform, autoregressive process, optimization

JEL classification: C18, C63, E32
AMS classification: 62P20, 62M10, 90C90, 91B84

1 Introduction

Cyclical movements and cyclical comovements have been an issue of great interest in a lot of macroeconomic literature for several decades. Traditionally, movements and co-movement measurements were performed in the time domain. A subsequent methodological approach has been oriented towards the frequency domain, which provides an additional look into the time series structure. Here, the standard analytical steps are based on spectral and cross spectral analysis. In the last decade the time-frequency approaches have predominated. It is generally argued that the combination of time and frequency tools provides a more efficient means of statistical analysis. This reflects the fact that frequency analysis, as well as the time-frequency analysis of a time series and processes, are instruments that has been used in interdisciplinary analyses for a long time. It has been applied in engineering, natural sciences or social and economic sciences.

For estimating the time-frequency representation of a time series there are several approaches, mainly 4 are used which are wavelet analysis [11], the multiple window method using Slepian sequences (MWM) [20], time-frequency varying autoregressive (AR) process spectrum estimation and time-frequency Fourier transform estimation [16]. Comparison of these main methods can be found in the work of Blumenstein et al. [4] which applies them on an example of industry production in EU27 countries. Moreover, Klejmová [12] propose a comparison of the same methods on a simulated signal and technical application focused on method optimization. Among works using the time-frequency approach, we can find Woźniak and Paczyński [19] calculating the spectrum using a time-varying autoregressive process and consequently study the patterns of comovement via time-frequency coherences between economic cycles in the euro area and some of the European Union member states as well. Hughes Hallett and Richter [8] uses time-varying spectral methods for testing links and leadership/dependency relationships between the People's Republic of China, the United States, and other large Asian countries. In Hughes Hallett and Richter [9]

[1]Faculty of Electrical Engineering and Communication, Brno University of Technology, Dept.of Radio Electronics, Technicka 12, 616 00 Brno, pomenkaj@feec.vutbr.cz
[2]Faculty of Electrical Engineering and Communication, Brno University of Technology, Dept.of Radio Electronics, Technicka 12, 616 00 Brno, klejmova@phd.feec.vutbr.cz

they explain why volatility transfer from high frequency to low frequency cycles can and did occur during the period commonly referred to as the "great moderation". For this reason they identify frequency cycles using time-frequency empirical techniques.

In most economic applications, wavelet analyses predominate. As Jiang and Mahadevan [11] wrote, the advantage of the wavelet analysis is that it can capture the features of a non-stationary time series due to simultaneous time-frequency decomposition of inputs. Ftiti et al. [7] examine the comovement dynamics between OECD countries with the US and Europe. They use the co-spectral analysis applied on the wavelet transform of a time series as well. They put in evidence the existence of both long run and short-run co-movement. Aguiar-Conraria and Soares [1] use the wavelet analysis to study business cycle synchronization across the EU-15 and Euro-12 countries. They analyze the wavelet power spectrum, wavelet coherency and also phase-differences. Additionally, Aguiar-Conraria and Soares [2] generalize the concept of simple coherency to partial wavelet coherency and multiple wavelet coherency analogous to partial and multiple correlations. Furthermore, Berdiev and Chang [3] illustrate the importance of examining the strength of business cycle synchronization using a time-frequency framework. They investigated the synchronization of growth cycles between China, Japan, the United States and other Asia-Pacific countries using wavelet analyses. Rua [17] focuses on wavelet-based measure of synchronization to capture both time and frequency varying features for a set of 23 countries. He finds that worldwide synchronization has increased over the last decades and it has attained an unprecedented level during the Great Recession. Moreover, Fidrmuc et al. [6] apply the wavelet spectrum analysis to study globalization and business cycles in China and G7 countries. They find a significant relationship between the time-varying wavelet measure of synchronization and trade only for business-cycle frequencies. The co-movements at longer frequencies are negatively related to trade, so that the overall co-movements and trade tend not to be significantly related. Maršálek et al. [13] proposes an original method for filtering-out global shocks from the time series. This approach is based on the continuous wavelet transform, its inverse and the comovement measurement in the time-frequency domain. Its application also enables to uncover detailed development of the business cycle synchronization in time.

The aim of the paper is to give recommendation for working with time-frequency modeling of macroeconomic time series on the basis of a comparative study. We investigate the wavelet analysis and the time-varying autoregressive process. We follow two main areas of focus - sample size for an available data set and optimization of parameters of mentioned methods. In the case of the time-varying autoregressive process, we investigate optimization of lag length, windowing and overlap. In the wavelet analysis approach we investigate the type of wave and scale. The paper is structured as follows. In section two we provide a methodological background. Section three is focused on data description and presentation of results. The last section summaries the results and discusses conclusions.

2 Methodical Background

2.1 Wavelet transform

In order to describe parameters of signal, not only the time but also the frequency domain wavelet transform and its modifications can be used. One of these modifications commonly used to assess cyclical movements in different types of macroeconomics time series is the continuous wavelet transform (CWT). It can be described as an integral of the analyzed signal with a base function (mother wavelet)[10]:

$$S_{CWT}(a,\tau) = \int_{-\infty}^{\infty} s(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t}{a} - \tau\right) dt, \quad a > 0, \tau \in R, \tag{1}$$

where $s(t)$ is the time series, $\psi\left(\frac{t}{a} - \tau\right)$ is a scaled version of the mother wavelet, τ denotes the time shift, and a denotes the scale (or frequency). The numerator of the fraction $\sqrt{a}$ ensures conservation of energy [10]. Several families of mother wavelets can be used. For purposes of this paper we selected the Daubechies, complex Morlet and Morlet wavelet (for details see [18]).

2.2 Autoregressive (AR) process

Another approach for time-frequency modeling is the time varying Autoregressive (AR) process. This parametric method creates a model generating an input signal using white noise. The estimation of the

time-frequency model is then done according to formula [16]:

$$\widehat{S}_s(f) = \frac{\widehat{\sigma}_w^2}{\left|1 + \sum_{k=1}^{p} \widehat{a_p}(k) e^{-j2\pi fk}\right|^2}, \tag{2}$$

where $\widehat{a_p}(k)$ are estimates of the AR parameters, p is the lag order and σ_w^2 is white noise variance. To obtain AR model parameters, the Burg method, Yule-Walker method, unconstrained least-squares method or sequential estimation methods are commonly used [16]. For AR modeling, selecting lag order p is an important aspect. Selecting a low level order leads to excessive smoothing of the spectrum. Furthermore, if the level of p is selected too high, a non-significant spectral coefficient can be seen as a high peak. Several information criteria can be used to ensure optimal selection (for details see [16]).

2.3 Cross-spectrum

To evaluate the similarity between two signals, the cross-spectrum representation is widely used. It can be obtained by the following formula: [11]

$$\widehat{S}_{xy}(a,b) = \widehat{S}_x(a,b)\widehat{S}_y(a,b) \tag{3}$$

where, $\widehat{S}_x(a,b)$and $\widehat{S}_y(a,b)$ are frequency representations of signal x resp. y. If these representations are complex numbers (depends on used method) the resulting cross-spectrum is also complex. Division of the amplitude and phase part is then possible [11].

$$A_{xy}(a,b) = \left|\widehat{S}_{xy}(a,b)\right|, \qquad \Phi_{xy}(a,b) = tan^{-1}\left[\Im(\widehat{S}_{xy}(a,b))/\Re(\widehat{S}_{xy}(a,b))\right] \tag{4}$$

where $A_{xy}(a,b)$ is the amplitude of the complex signal $\widehat{S}_{xy}(a,b)$, in some applications denoted as cospectrum, and $\Phi_{xy}(a,b)$ represents its phase (quadrature spectrum). Symbols $\Im$ and $\Re$ denote the imaginary part and real part of the cross-spectrum $\widehat{S}_{xy}(a,b)$. In the case of two identical signals, the cross-spectrum is equal to one. If the input signals differs greatly, their cross-spectrum approaches zero.

3 Application

3.1 Data and settings

We use seasonally adjusted quarterly data of gross domestic product (GDP), and volume index in OECD reference year 2005 (OECD 2015) of the United States (US) and United Kingdom (UK) in 1956/01-2014/04, Korea in 1970/02-2014/02. All variables are in first differences of logarithms.

After transformation (first differences of logarithms) the data was analyzed using the continuous wavelet transform. We investigated the impact of scale value and type of wave. In the first step, the scale setting was established for 64, 128 and 256 values which are commonly used. With respect to sample size, data structure and expected events, we established 128 as the optimal scale. From the group of waves, we used three mother wavelets, namely Morlet, complex Morlet and Daubechies. The Morlet wavelet was chosen as it is commonly used for this type of application. The complex Morlet wavelet is based on the standard Morlet with the advantage of providing complex results making it possible to obtain the phase part (quadrature) of the spectrum. Evaluation of this part of spectrum will be subjected to further research. The Daubechies wavelet of order 5 was chosen as it is the second most commonly used and also for its ability to assess lower frequencies and business cycles. The results of time-frequency wavelet representations of cospectra for USA, UK and Korea are given in figure 1.

In the case of the cospectra estimate via the time-varying AR process, we used the Burg approach for coefficient estimates on 20 samples with 50% overlay and Hann window. The optimal value of lag order was based on AIC criteria [12]. The results of the time-frequency cospectra using the AR process for USA, UK and Korea is given in figure 2. The cospectrum (fig. 1 and 2) is denoted on the z-axis for specific periods (x-axis) and periodicities (y-axis). The figure shows a two-dimensional projection of three-dimensional charts.

3.2 Results

Comparison of results can be done in the following way. The first is a comparison among chosen mother wavelets, the second is among data. Focusing on cospectra for complex Morlet waves, we can identify the comoving areas across a range of periods (i.e. 1/frequencies) from very long (up to 8 years), across business cycles (1.5 to 8 years) to short cycles (1.5 years and less). In some cases (USA-UK, USA-Korea) the most significant comovement can be found around a 5 year long period while in another case (UK-Korea) the most significant comovement can be found for very long periods. In the case of the Morlet mother wave, the results of USA-UK show the most significant comovement for very long periods (more than 20 years). We can also find some comovement among 2.5 and 9 years which is not so significant in this case. The comovement for UK-Korea is significant among 2.5-20 years and in the case of USA-Korea there is a range of significant periods starting at 2.5 years and including very long periods (more than 20 years). Applying the Daubechies mother wave produces similar results to the Morlet mother wave, but there is not such a significant comovement for shorter periods, i.e. for periods between 2.5 and 20 years. The complex Morlet wave offers the possibility to study phase shift which is not possible by using the Morlet or Daubechies wave. Applying Morlet and Daubeschie on our data revealed rather longer cycles. Therefore, the authors suggested to remove such components via an advanced filtering method by using the Baxter-King or Christiano-Fitzgerald band-pass filter (Poměnková, J. [15]) to get better visibility of other comovement periodicities. The frequency range that we want to remove can be easily established by specifying the range before using the proper band-pass filter.

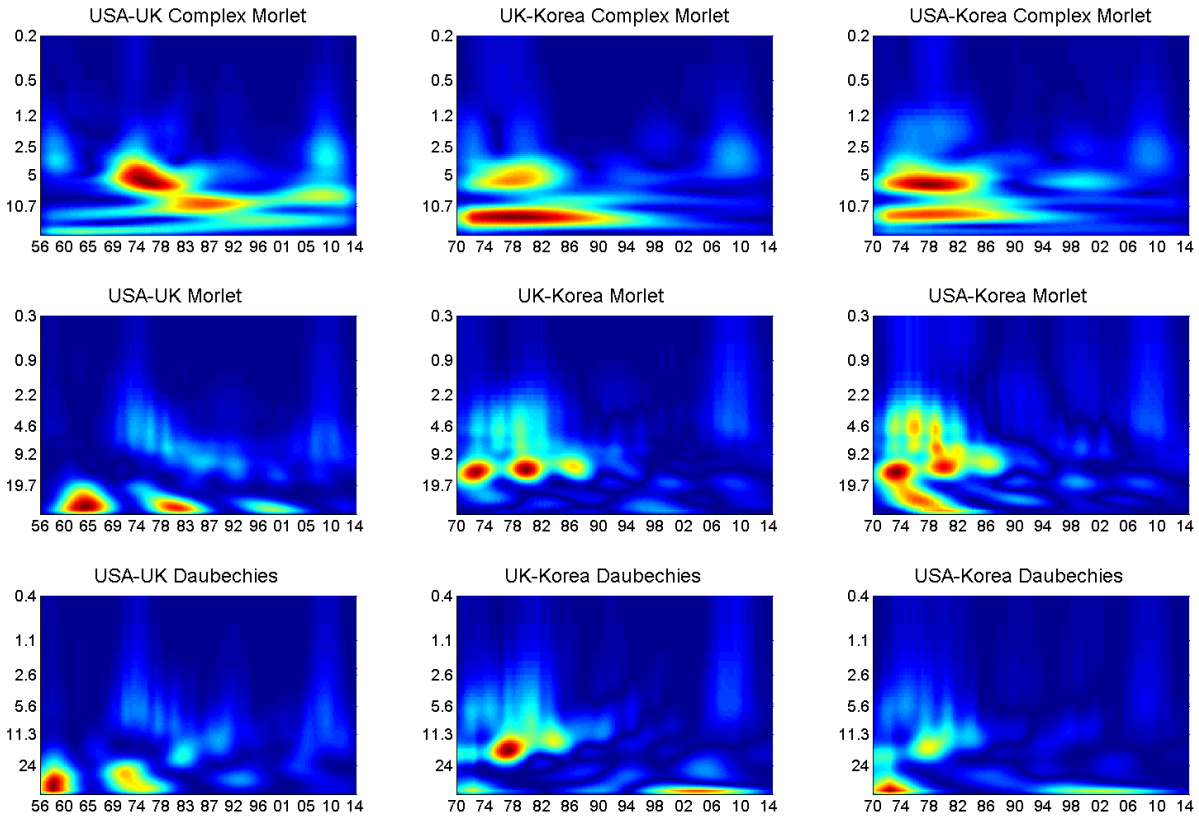


Figure 1 CWT cospectrum (x-axis represent time in years and y-axis represent time cycles in years)

With respect to the methodical results described above, we can capture the following economic events in comovement. Using the complex Morlet wave reveals the existence of significant comovement between the USA and UK several times, namely 1969-1980, in business cycle frequencies (8-2.5 years) which can be caused by oil crises in the US (1973-1975, 1979). This event is also visible in the case of the Morlet and Doubechies wave application, but as mentioned previously, its visibility is suppressed by permanent components. In such cases, we recommend filtering out these components via a band-pass filter and analyze comovement by focusing on a specific range of frequencies. In 1958 we can find a recession which occurs in both the US and UK for applying the complex Morlet wavelet with lower significance in time evolution, but for the Morlet and Daubechies wave the results showed better visibility. We can say that both the complex Morlet and Morlet wave's application provide complementary points of

view confirmed by the results for the Daubechies wave. Additional comovement was found for periods around 11 years between 1980-1992 (Complex Morlet) which is probably due to the early recession in 1980 and consequently in 1990. Applying the Morlet and Daubechies wave does not indicate such strong comovement which can be caused by a permanent component. The recession in 2008 is less visible among all events. Cospectra for USA-Korea and UK-Korea show similar results. All mother wave results reveal comovement in 5 year period between 1974-1982. The cause can be found in the USA oil crisis which has a global impact on different economic sectors and aspects. In Korea, we can admit that this crisis was deepened by the Gwangju uprising in 1980. Additionally, the complex Morlet wave application shows comovement of a 19 year period between the USA, UK and Korea in 1970-1990. This result is also visible for the Daubechies wave, but it is not so significant. Similarly to the USA and UK case, an advantage of the complex Morlet is the possibility to analyze phase shift, which can reveal additional information.

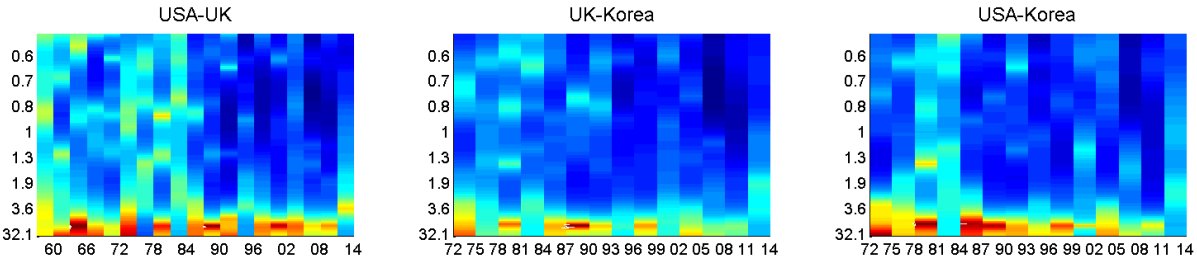


Figure 2 AR cospectrum (x-axis represent time in years and y-axis represent time cycles in years)

Figure 2 shows results for measuring comovement via cospectrum for the AR process. Usage of this approach confirms the results given by the wavelet transform and also confirms expectations of better frequency resolution as Klejmová [12] wrote. For all countries, we can identify strong significant comovement in a wide range of frequencies, predominately in long periods. Opposite to the wavelet analysis, we can also see significant comovement in rapid changing periods (shorter than 2.5 years). Unfortunately, with respect to the sample size, better time resolution is impossible. This confirms the fact given in Blumenstein [4] which recommends, for the AR approach, rather monthly data or higher frequency data (weekly, daily etc.). We can admit that in such a case the AR method will confirm results from a wavelet analysis much better than from quarterly data.

4 Conclusion

The wavelet approach for measuring comovement shows good ability to capture rapid changes in time. However, its ability to identify length of business cycles strongly depends on mother wavelet selection. Our results suggest that the complex Morlet wavelet was able to better capture short business cycles and for this we recommend its usage for data were long cycles are not present. The second possibility is to remove permanent components via a band-pass filter with respect to the aim of the consequent analysis. Another property of this wavelet is its ability to provide complex output, making the evaluation of phase spectrum possible. The Daubechies wavelet provided good results for long business cycles. However, short business cycles can be shown as less significant. In the case of comovement measured via the AR process, our results confirm better frequency resolution, but in the case of a small sample size (which our quarterly data is) the time resolution is not as good as for higher frequency data. Therefore, we suggest using these methods for time series where monthly data are available and in cases when frequency resolution is more important than time resolution.

Acknowledgements

Research described in this paper was financed by the Czech Ministry of Education in the frame of National Sustainability Program under grant LO1401, the BUT Internal Grant Agency under project no. FEKT-S-14-2281. For research, infrastructure of the SIX Center was used.

References

[1] Aguiar-Conraria, L. and Soares, M. J.: Business cycle synchronization and the Euro: A wavelet analysis, *Journal of Macroeconomics* **33** (2011), 477-489.

[2] Aguiar-Conraria, L. and Soares, M. J.: The continuous wavelet transform: Moving beyond uni-and bivariate analysis, *Journal of Economic Surveys* **28** (2014), 344-375.

[3] Berdiev, A. N. and Chang, Ch-P.: Business cycle synchronisation in Asia-Pacific: New evidence from wavelet analysis, *Journal of Asia Economics* **37** (2015), 20-33.

[4] Blumenstein, J., Poměnková, J. and Maršálek, R.: Comparative Study Of Time- Frequency Analysis Approaches With Application To Economic Indicators. In: *Proceedings of the 26th European Conference on Modelling and Simulation ECMS 2012* (Troitzsch, K., Mhring, M., Lotzmann, U., eds.). University Koblenz Landau, Koblenz (2012), 291-297.

[5] Crowley, P. and Hughes Hallett, A.: Volatility transfers between cycles: A theory of why the "great moderation" was more mirage than moderation, *Research Discussion Papers* **23** (2014), Bank of Finland.

[6] Fidrmuc, J., Korhonen, I. and Poměnková, J.: Wavelet spectrum analysis of business cycles of China and G7 countries, *Applied Economic Letters* **21** (2014), 1309-1313.

[7] Ftiti, Z., Tiwari, A. and Belanés, A.: Tests of Financial Market Contagion: Evolutionary Cospectral Analysis V. S. Wavelet Analysis, *Computational Economics* **1** (2014); DOI: 10.1007/s10614-014-9461-8.

[8] Hallett, A. H., and Richter, Ch.: Trans-Pacific Economic Relations and US-China Business Cycles: Convergence within Asia versus US Economic Leadership, *ADBI WP* **292** (2011), Asian Development Bank Institute.

[9] Hallett, A. H, and Richter, Ch.: Volatility transfer between cycles: A theory of why the "great moderation" was more mirage than moderation, *BOFIT Discussion Papers* **23** (2014), Bank of Finland Research.

[10] Ján, J.: *Číslicová filtrace, analýza a restaurace signálu*. VUTIUM Brno, 2002.

[11] Jiang,W. and Mahadevan, S.: Wavelet spectrum analysis approach to model validation of dynamic systems, *Mechanical Systems and Signal Processing* **25** (2011), 575-590.

[12] Klejmová, E.: Comparison of parametric and non parametric methods for signal time-frequency description. In: *Proceedings of the 25th International Conference Radioelektronika 2015* (Pidanic, J., eds.). Univerzita Pardubice, 2015, 272-275.

[13] Maršálek, R., Poměnková, J. and Kapounek, S.: A Wavelet-Based Approach to Filter Out Symmetric Macroeconomic Shocks, *Computational Economics* **44** (2014), 477-488.

[14] Organisation for Economic Co-operation and Development: National Accounts [online database]. (2015) [cit. 2015-02-3]. Available at: http:\\stats.oecd.org\Index.aspx?DatasetCode=SNA_TABLE1.

[15] Poměnková, J.: Business cycle identification, *Journal of Economics* **60** (2012), 899-917.

[16] Proakis, J. G., Rader, Ch. M, Ling, F. L., Nikias, Ch. L., Moonen, M. and Proudler, J. K.: *Algorithms for Statistical Signal Processing*. Prentice Hall, 2002.

[17] Rua, A.: Worldwide synchronization since the nineteenth century: a wavelet-based view, *Applied Economics Letters* **20** (2013), 773-776.

[18] Vetterli, M. and Herley, C.: Wavelets and filter banks: theory and design, *Signal Processing, IEEE Transactions* **40** (1992) , 2207-2232.

[19] Woźniak, P. and Paczyński, W.: Business Cycle Coherence between Euro Area and the EU New Member State: a Time-Frequency Analysis. Center for Economic Research and Graduate Education and the Economics Institute (CERGE-EI), Prague.

[20] Xu, A., Haykin, S. and Racine, R. J.: Multiple Window Time-Frequency Distribution and Coherence of EEG Using Slepian Sequences and Hermite Function.*IEEE Transactions of Biomedical Engineering* **46** (1999), 861-866.

Benford's Law and Opendata portal of Ministry of Finance of Czech Republic

Alena Pozdilkova[1], Jaroslav Marek[2]

Abstract. The intuitive mind is that, in large sets of numbers, each integer from 1 through 9 has the same probalities of occuring on the first digit of numbers. But Benfords Law reports that the digit 1 leads approximately 30 % of the time and each successive digit is less common, with 9 occurring less than 5 % of the time. The main goal of our article is to aply Benford's Law for analysing Finance Operation datasets from Opendata portal of Ministry of Finance of Czech Republic. A chi-square goodness of fit test for testing null hypothesis is realised and verificated, that data follow the Benford distribution. Further the finance operation is compared in each of the years 2010-2014 from the perspective of Benford's law. There will be also compared different approaches to the analysis of the tested data and compared the results of different testing methods, for example Pearson's chi-squared test of goodness of fit and Z statistics. Benfords Law by the current theory is considered to be a powerful tool allowing quick screening of data for anomalies. The secondary aim of the article is to draw attention to the fact, that some anomalities can be false positive.

Keywords: Benford's law, Pearson's chi-squares test statistics of goodness of fit, Opendata portal of Ministry of Finance of Czech Republic.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Simon Newcomb (1835-1909) has probably been the first to discover the phenomenon that would later be called Benfords law, or at least the first, who published something about it. [12] He was a famous American astronomer, his most famous work regarding planetary theories and astronomical constant derivation. He found the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster ones wear out the last ones. However, this sentence is a lot more general, as the distributions of all the digits, and even that of the whole mantissa, can be derived. His article has a kind of genius sense, but as often in that kind of discovery it is not well explained at all because of its apparent lack of mathematical background. Like Newcomb before him, Frank Benford observed the difference of dirtiness on the logarithm pages. He then tried to reproduce in a now famous experiment, so he collected 20229 observations of natural numbers and computed the frequency of the first digit for each different dataset and also the average of all the datasets together. Benford observed that the logarithmic law was better fitted by the more random numbers in his experiment. Benford tried to explain the phenomenon by investigating the set of natural integers, in an attempt to prove that it comes naturally from our number system. The problem that he encountered is that this set has no asymptotic natural frequency.

1.1 Benford's law

Theorem 1 (Benford's law [5]). *The logarithmics density function for the first digit D is defined by:*

[1]University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Physics, Studentska 95, 532 10 Pardubice, Czech Republic, alena.pozdilkova@upce.cz
[2]University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Physics, Studentska 95, 532 10 Pardubice, Czech Republic, jaroslav.marek@upce.cz

$$P(D = d) = \log\left(1 + \frac{1}{d}\right), \ d \in \{1, \ldots, 9\}. \tag{1}$$

The logarithmics joint distribution of the first significant digits $D_1, D_2, \ldots D_k$ *is defined by:*

$$P(D_1 = d_1, D_2 = d_2, \ldots, D_k = d_k) \quad = \quad \log\left(1 + \left(\sum_{i=1}^{k} 10^{k-i} d_i\right)^{-1}\right), \tag{2}$$

where $d_1 \in \{1, \ldots 9\}$ *and all other* $d_j \in \{0, \ldots, 9\}$*, for every* $k \in N$*.*

1.2 Theoretical probabilities

	2^{nd} digit										
1^{st} digit	$D_2 = 0$	$D_2 = 1$	$D_2 = 2$	$D_2 = 3$	$D_2 = 4$	$D_2 = 5$	$D_2 = 6$	$D_2 = 7$	$D_2 = 8$	$D_2 = 9$	$\sum$
$D_1 = 1$	0.0414	0.0378	0.0348	0.0322	0.0300	0.0280	0.0263	0.0248	0.0235	0.0223	0.30103
$D_1 = 2$	0.0212	0.0202	0.0193	0.0185	0.0177	0.0170	0.0164	0.0158	0.0152	0.0147	0.17609
$D_1 = 3$	0.0142	0.0138	0.0134	0.0130	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.12494
$D_1 = 4$	0.0107	0.0105	0.0102	0.0100	0.0098	0.0095	0.0093	0.0091	0.0090	0.0088	0.09691
$D_1 = 5$	0.0086	0.0084	0.0083	0.0081	0.0080	0.0078	0.0077	0.0076	0.0074	0.0073	0.07918
$D_1 = 6$	0.0072	0.0071	0.0069	0.0068	0.0067	0.0066	0.0065	0.0064	0.0063	0.0062	0.06695
$D_1 = 7$	0.0062	0.0061	0.0060	0.0059	0.0058	0.0058	0.0057	0.0056	0.0055	0.0055	0.05799
$D_1 = 8$	0.0054	0.0053	0.0053	0.0052	0.0051	0.0051	0.0050	0.0050	0.0049	0.0049	0.05115
$D_1 = 9$	0.0048	0.0047	0.0047	0.0046	0.0046	0.0045	0.0045	0.0045	0.0044	0.0044	0.04576

Table 1 Teoretical probabilities of first two digits (see formula (2) for $k = 2$)

2 Testing od statistical hypothesis

This section includes methodology of statistical tests.

Commonly used tests for conformance with Benfords Law are the Pearson's χ-squared test of goodness of fit, the Kolmogorov-Smirnov test, Z statistics, the mean absolute deviation test, the test based on the logarithmic basis of Benford's law, the Kuiper test. Such tests are not too powerfull and tests are too conservative. See [2], [11].

Some of the less known tests, cf. [11], are the Mantissa Arc test, the max test, the distance test.

2.1 Pearson's χ-squared test of goodness of fit

It is advisable to test the hypothesis that a random sample correspond to Benford's law. The goodness of fit has test statistic

$$\chi^2 \quad = \quad \sum_{i=0}^{t} \frac{(X_i - np_i)^2}{np_i} \sim \chi^2_{t-1}, \tag{3}$$

where p is the number of unknown parameters, $t - 1$ is the number of degrees of freedom, X_i are an observed frequency (i.e. count) for bin i, np_i are an expected (theoretical) frequency for bin i asserted by the null hypothesis.

If χ^2 exceeds the value $\chi^2_{t-1}(\alpha)$ quantile of χ^2 distribution on chosen level of significance α, we reject the null hypothesis.

We can apply a test to the first digits, but we can meet even with testing on the first two digits. The first case leads to $t = 9$ classes, the second case points to $t = 90$ classes.

2.2 Z statistics

The Z statistics takes into account the absolute magnitude of the difference, the size of the data sets, and the expected proportion. Test statistics is

$$Z_k = \frac{|AP_k - EP_k| - \frac{1}{2N}}{\sqrt{\frac{EP_k(1-EP_k)}{N}}}, \tag{4}$$

where AP are the actual proportion, EP are the expected proportion, and N is the number of records.

The P-values should not be compared to the level of significance, $\alpha = 0.05$, since there are nine comparisons. The probability of at least one Type I error in a series of 9 test is $1 - 0.95^9 = 0.37$. See [2].

The process of reducing the level of significance used in each of the nine tests is based on Bonferronis inequality, cf. [3], pp. 14, 481. Each P-value is compared to $\frac{\alpha}{9} = \frac{0.05}{9} = 0.0056$, giving an approximate overall probability of rejection of 0.05. Since $P(|Z| > 2.77) = 0.0056$, any Z-score greater in absolute value than 2.77 implies rejection of the null hypothesis. If all P-values are greater than 0.0056 (and all Z-scores are less than 2.77 in absolute value), we do not reject the null hypotheses that each proportion is the corresponding Benford proportion.

3 Application of test on Financial operation of Ministery of Finance of Czech republic

This section includes a numerical study of transfers. The data are partially presented in table 2 containing theoretical and empirical probabilities. Additional tables containing the numerical results of the statistical tests.

3.1 Empirical probabilities

The Table 2 contains obtained frequencies of first digits for 9 classes of first digits at Financial transfers of Ministry of Finance of Czech republic in the years 2010–2014. Benford's law gives the expected frequencies at Tab. 2 written in round brackets.

digit/year	2010	2011	2012	2013	2014
n	10888	9406	8739	8859	7949
1	3036 (3277.6)	2525 (2831.5)	2365 (2630.7)	2201 (2666.8)	1894 (2392.9)
2	1761 (1917.3)	1627 (1656.3)	1463 (1538.9)	1657 (1560.0)	1557 (1399.7)
3	1273 (1360.3)	1180 (1175.2)	1001 (1091.9)	1303 (1106.8)	1343 (993.1)
4	1402 (1055.2)	1169 (911.5)	1081 (846.9)	884 (858.5)	1040 (770.3)
5	850 (862.1)	737 (744.8)	802 (692.0)	758 (701.5)	539 (629.4)
6	797 (729.0)	662 (629.7)	585 (585.1)	622 (593.1)	421 (532.2)
7	657 (631.4)	567 (545.5)	522 (506.8)	590 (513.7)	422 (461.0)
8	566 (556.9)	448 (481.1)	446 (447.0)	420 (453.1)	379 (406.6)
9	546 (498.2)	491 (430.4)	474 (399.9)	424 (405.4)	354 (363.7)

Table 2 Frequencies of first digits: values for years 2010–2014

Graphical comparing of theoretical and empirical probabilities

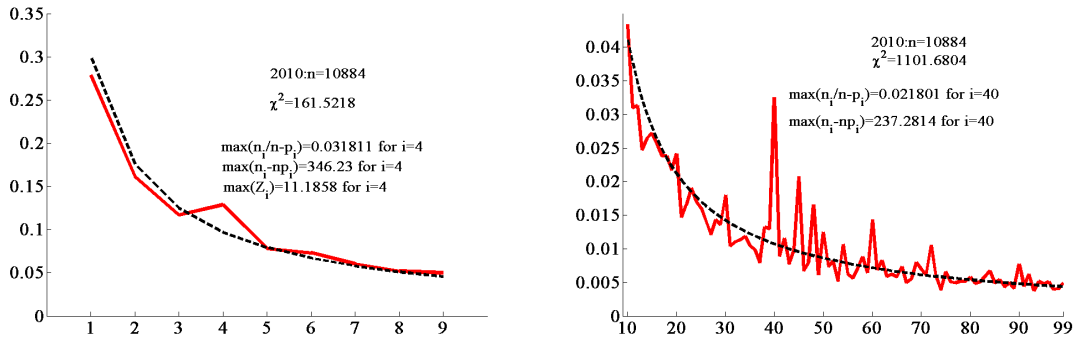


Figure 1 Year 2010: comparing of empirical and theoretical probabilities

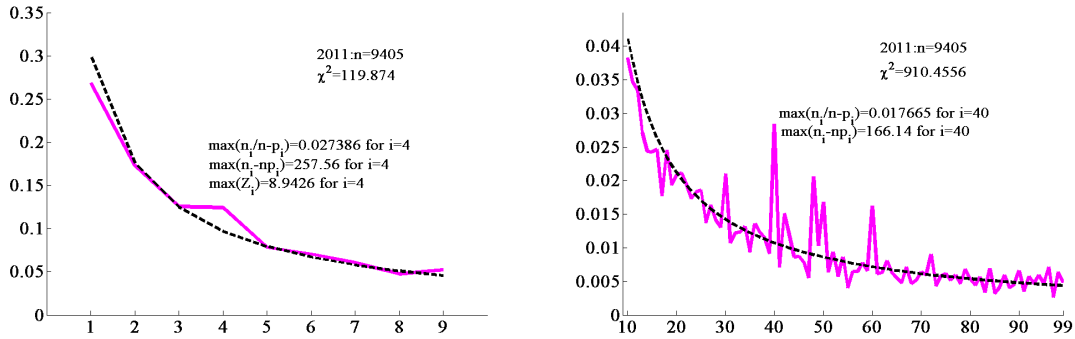


Figure 2 Year 2011: comparing of empirical and theoretical probabilities

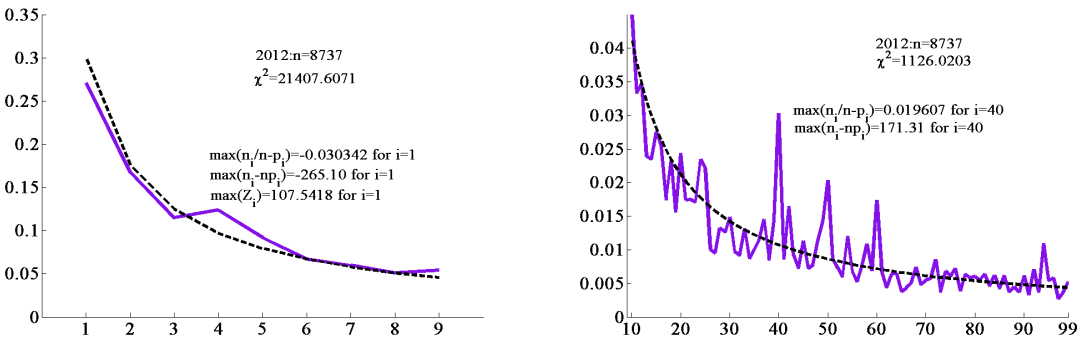


Figure 3 Year 2012: comparing of empirical and theoretical probabilities

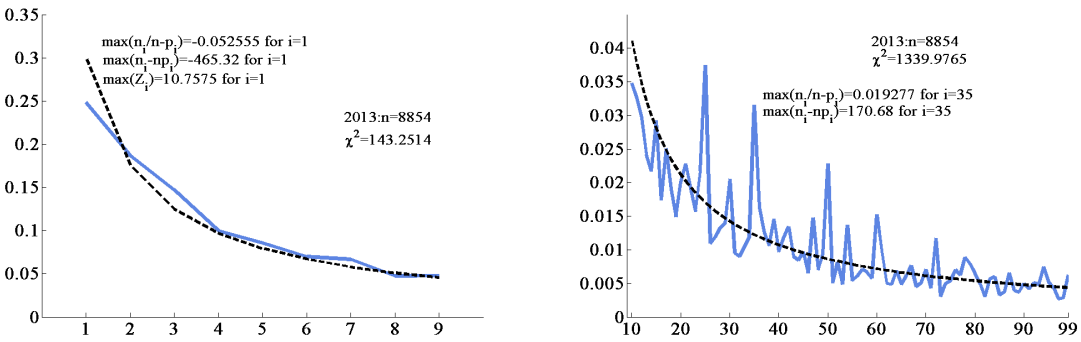


Figure 4 Year 2013: comparing of empirical and theoretical probabilities

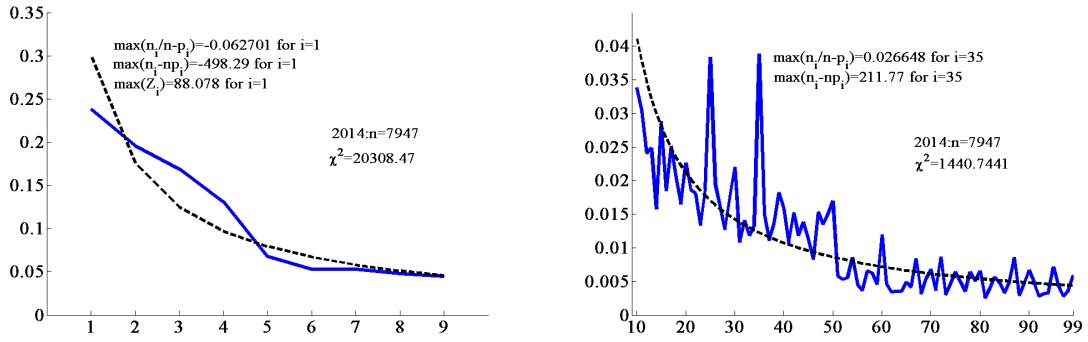


Figure 5 Year 2014: comparing of empirical and theoretical probabilities

3.2 Results of testing null hypothesis

Application of χ^2 test and Z test Critical value for 9 classes is $\chi^2_{9-1}(0.05) = 15.507$, critical value for 90 classes is $\chi^2_{90-1}(0.05) = 112.2$. The value of Pearson's χ^2 test statistics of goodness of fit and the value of Z test for all years cannot pass null hypothesis. Compare critical values of χ^2 test and Z test with Tables 3 and 4. We reject conformance to Benford at all years.

We can find several possible explanations for rejecting the null hypothesis:

1. transaction data really do follow Benford's Law, but due to random chance this population of observations does not.
2. Underlying statistical assumptions go to a greater chance of making a Type I error.
3. Repetition of some transfers might change a probability of occuring of some particular digits.
4. Alternative hypothesis is true. There is an indication of fraudulent observations.

year	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
2010	5.0028	3.8899	2.4882	11.1858	0.4192	2.5830	1.0177	0.3592	2.0631
2011	6.8605	0.7617	0.1232	8.9426	0.2940	1.2941	0.9082	1.5019	2.9436
2012	107.5418	63.5022	40.6663	48.4214	33.8368	21.9804	19.4418	15.6168	18.8530
2013	10.7575	2.6754	6.2609	0.8966	2.2013	1.1163	3.4347	1.5390	0.9085
2014	88.0780	73.8110	65.2652	49.7813	19.6657	13.4060	14.8364	13.0301	12.3265

Table 3 Values of Z_i statistics: values for years 2010–2014

year	9 classes	90 classes
2010	162.46	1101.7
2011	119.83	910.5
2012	134.54	1126.0
2013	143.47	1340.0
2014	380.91	1440.7

Table 4 Values of χ^2 statistics: values for years 2010–2014

Deviations from expected curves are caused by high specificity of selected data. It can be tracked, what the greatest deviations (peaks) are caused after a detailed analysis of data of all evaluated periods. In all periods globally provided FKSP contribution reached 4.000 CZK to most of employees in years 2010–2012 (deflection curves for the two initial digits of 40). In years 2013–2014 this contribution was reduced to 3500 CZK (which caused a shift in the largest displacement to values 35). Other variations are caused, for example, subsidies for election (very significant deviation of 30 in 2011) or by regular monthly payments of the same amount for lease or operation of certain products.

4 Conclusion

Benford's Law is an empirical generalization that says that in many sets of measured data, the greater the initial digit, the less likely it appears at the beginning of numbers. Cf. [10]. It is reported that for the adequate force of law just to have a random selection of at least a range of thousands.

An analysed dataset contain the financial transaction of Ministry of Finance of Czech Republic, as of january of 2010. All the tests reject null hypothesis. Repetition of some transfers might change a probability of occuring of some particular digits. In all the years we find the class, which contributed to the rejection. For example, in 2010–2012 it is the number 40. The increase in the frequency in this class are caused by high specificity of selected data, because globally provided FKSP contribution reached 4.000 CZK to most of employees in years 2010–2012. Other variations are caused, for example, subsidies for election (very significant deviation of 30 in 2011) or by other regular mothly payments. All these variations always occur in only one combination of the first two digits and thus affect the final shape of the curve.

Acknowledgements

Supported by institutional support of the University of Pardubice.

References

[1] Cho, W. K. T. and Gaines, B. J.: Breaking the (Benford) law: Statistical fraud detection in campaign finance, The American statistician. **61** (2007), 218–223.

[2] Cleary, R., Thibodeau, J. C.: Applying Digital Analysis using Benfords Law to Detect Fraud: A Practice Note about the Dangers of Type I and Type II Errors. *Auditing: A Journal of Practice & Theory*, **24**, No. 1, (2005), pp. 77–81

[3] Hogg, R. V., McKean, J. W., Craig, A. T.: *Introduction to Mathematical Statistics.* Pearson Prentice Hall, Upper Saddle River, New Jersey, 2005.

[4] Leemis, L. M., Schmeiser, B. W., Evans, D. L.: Survival Distributions Satisfying Benfords Law. *The American Statistician.* **54** (2000), 236–241.

[5] Nigrini, M.J.: *Benford's law. Applications for Forensic Accounting, Auditing, and Fraud detection.* Wiley and Sons, New Jersey, 2012.

[6] Nigrini, M.J.: A Taxpayer Compliance Application of Benford's law. *Journal of the American Taxation Association.* **18** (1) (1996), 74.

[7] Placek, M.: The Application of Benfords Law on Government Macroekonomics data: Case of Trade Balance. *Trends Economics and Management* **VIII** (2014), 45–51.

[8] Placek, M.: Applying Benford's Law by Testing the Government macroeconomics data. *Acta academica karviniensia* **3** (2014), 148–160.

[9] Portal Ministerstva financi, [online], cit. 15.5.2015, URL: http://data.mfcr.cz/cs/dataset/prehled-faktur-ministerstva-financi-cr

[10] Seibert, J., Zahrádka, J.: Zákon první číslice a jeho aplikace [First digit law and its application]. *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration.* **1** (2014), 75–83.

[11] Morrow, J.: Benfords Law, Families of Distributions and a Test Basis [online], cit. 15.5.2015, URL: http://www.johnmorrow.info/projects/benford/benfordMain.pdf

[12] Jamain, A.: Benfords Law [online], cit. 15.5.2015, URL: http://wwwf.imperial.ac.uk/ ~ nadams/classificationgroup/Benfords-Law.pdf

Nonlinear Endogenous Model of Business Cycle

Pavel Pražák [1]

Abstract. This paper deals with a nonlinear dynamic model of a closed economic unit that exhibits a cyclical development. The assumptions of the model are based on the continuous multiplier-accelerator model of A. W. Phillips. Then a nonlinear investment function is introduced. It is assumed that this function depends on GDP output level and the rate of output change. The model is formulated as a nonlinear ordinary differential equation of the second order. Both the local and the global analysis of the equation is then introduced. Finally it is shown that there exists Poincaré-Andronov-Hopf bifurcation. It means that there can be observed a cyclical development of the given economic unit.

Keywords: bifurcation, business cycle, investment, nonlinear model, output.

JEL classification: C61, E32
AMS classification: 34C07, 37G15

1 Introduction

Movements in output of an economic system around its trend is called business cycle. It can be measured by fluctuations in real gross domestic product, industrial production, employment, sales, personal incomes and other macroeconomic variables, [2]. All this fluctuations are considered subject to their trends as can be seen on Figure 1. Business cycle is usually characterized by four phases: recession, recovery, growth, and decline, for more details see [10]. According to [4] and [3] business cycle theories can be divided into two main groups. The first group concentrates on endogenous factors in economic units. These models are characterized by the assumption that the output of a given economy system is not in its natural level. Traditionally, it was argued that recessions are periods in which the economy system is below its potential output. In these models it was assumed that the fluctuations are based on changes in investment. In particular, the lags and nonlinearities in investment cause the economic system to generate an oscillation in in its output, [9]. The second group concentrates on exogenous factors in economic unit. Advocates of this approach has developed models of real business cycle. Among the basic assumptions of these models is that the output is always in its natural level and that the fluctuations are caused by exogenous shocks in level of technological progress, [6]. In this paper we concentrate on a nonlinear variant of endogenous models of business cycle that is based on ideas of [3].

1.1 Model of Endogenous Cycle

The essential endogenous model of business cycle is the model that applied multiplier analysis and accelerator principle. It was developed by Samuelson in [11] and it was presented as a discrete time model with difference equations. The hypothesis of multiplier suggested that consumers spend a fixed fraction of their incomes, so that any initial income change leads to a convergent geometric series of subsequent spending, which multiplies up the initial change by a factor reciprocal to the fraction saved, [9]. According to the principle of acceleration, capital is assumed to be needed in a fixed proportion to the output to be produced, so investments, by definition the change in capital stock, are directly proportional to the change of output, [9]. Later the model was reformulated as a continuous time model. This modification was developed by Phillips in [7] and a linear ordinary differential equation was used. In this paper we shall deal with this continuous model with a general nonlinear investment function.

[1]University of Hradec Králové, Faculty of Informatics and management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, Czech Republic, pavel.prazak@uhk.cz

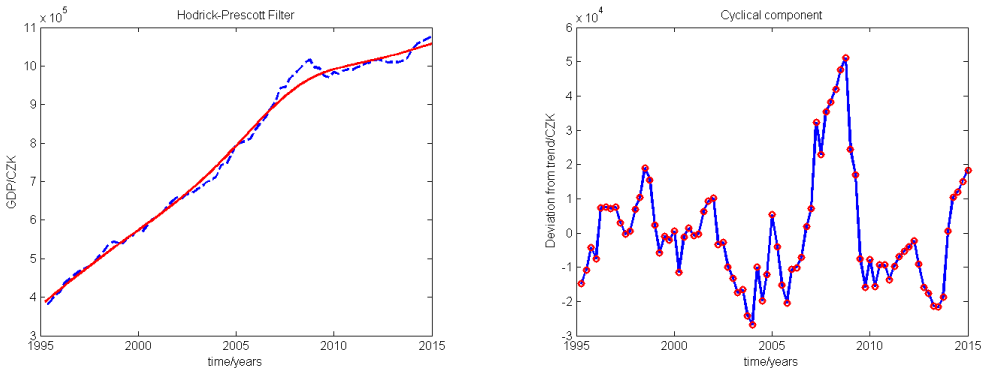


Figure 1 The left panel: The Czech Republic RGDP and a model of its trend determined by Hodrick-Prescott filter. The right panel: A model of the cyclical component of the Czech Republic RGDP. Source: Data of Czech Stastistical Office (Quaterly National Accounts, Seasonally Adjusted) and author's arrangement using MATLAB.

1.2 Phillips Linear Model of a Closed Economic Unit

At first we formulate the assumptions of the linear model, [7], [1], [8]:

- The aggregated demand $Z = Z(t)$ equals to the sum of the aggregated consumption $C = C(t)$ and investment $I = I(t)$, it means that for $t \in [t_0, \infty)$ is

$$Z(t) = C(t) + I(t). \tag{1}$$

- The consumption C is proportional to the aggregated output – income $Y = Y(t)$, it means that for $t \in [t_0, \infty)$ is

$$C(t) = cY(t), \tag{2}$$

  where the coefficient of proportionality – multiplier $c \in (0, 1)$ is called the marginal propensity to consume. It is convenient to introduce the marginal propensity to save $s = 1 - c \in (0, 1)$ and instead of (2) it is possible to write $C(t) = (1 - s)Y(t)$.

- The investment decision $B = B(t)$ at time t is proportional to the speed of production $\dot{Y}(t)$ at the time t, it means that for all $t \in [t_0, \infty)$ hold

$$B(t) = v\dot{Y}(t), \tag{3}$$

  where $v > 0$ is the proportionality factor. This factor is called the accelerator coefficient and its dimension is the time unit.

- The state of investment expenditures $I = I(t)$ is lagged behind the investment decision B. It is considered that the lag equals to one unit of time, it means that the actual speed of the output can influence investment till next period and we can write $1 \cdot \dot{I} = B - I$. If (3) is used we gain

$$\dot{I}(t) = v\dot{Y}(t) - I(t) \tag{4}$$

  for $t \in [t_0, \infty)$. Since the equation (4) has to be dimensionally consistent we notice that it is implicitly divided by the unit of time and there should be written $v/1$ and $I/1$ instead of the short notation v and I. This fact is necessary to hold in mind because of dimension of v and I.

- It is considered that the state of output Y reach the state of demand Z in the time period τ, $\tau \in (0, 1)$. It means that the supply of output Y adjusts with a lag τ to excess demand $Z - Y$ which can be written as

$$\tau\dot{Y}(t) = Z(t) - Y(t), \tag{5}$$

  where $t \in [t_0, \infty)$. For more details see [8].

Now we can replace (1) and (2) to (5) and we gain

$$\tau\dot{Y} = cY + I - Y = I - sY.$$

From this equation we have either

$$I = \tau\dot{Y} + sY \tag{6}$$

or

$$\tau\ddot{Y} = \dot{I} - s\dot{Y} \tag{7}$$

if it is in addition considered that the functions $Y, \dot{Y}$ and I are differentiable. Substituting (4) into (7) we find $\tau\ddot{Y} = v\dot{Y} - I - s\dot{Y}$. Instead of I we now replace (6) and we gain $\tau\ddot{Y} = v\dot{Y} - \tau\dot{Y} - sY - s\dot{Y}$, which is the linear differential equation of the second order and it is possible to write it as

$$\tau\ddot{Y} + (\tau + s - v)\dot{Y} + sY = 0. \tag{8}$$

The solution to equation (8) represents the time-path of output Y. The details can be found in [1] or [8]. It can be shown that the harmonic periodical solution to output Y can be observed only if $v = \tau + s$, which is quite strict, edge and unlikely condition, see [1]. It follows that the linear model cannot satisfactory explain all phenomena of business cycles. An improvement and modification has to be suggested.

2 Poincaré-Andronov-Hopf bifurcation

Most systems of ordinary differential equations usually contain parameters. For different values of these parameters, the behaviour of the solutions to the system can be qualitatively different. The changes of the nature of critical points and branching of solutions when parameters passes certain values are called bifurcation, [5]. Consider the system of ordinary differential equations

$$\dot{x} = f(\mu, x), \tag{9}$$

where $x \in \mathbb{R}^n$ is a vector of n state variable, $n \in \mathbb{N}$, $\mu \in \mathbb{R}^r$, is a vector of r parameters, $r \in \mathbb{N}$, and $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is assumed to be sufficiently smooth. The stationary point usually depends on parameters and is therefore denoted by (x°, μ°), i.e.

$$f(\mu^\circ, x^\circ) = 0 \in \mathbb{R}^n. \tag{10}$$

Bifurcation theory is interested in how the orbit structure near x° changes when μ is varied. For instance as the parameter μ, $\mu \in \mathbb{R}$, varies through zero the number of stationary points can vary, a periodical orbits can emerge or it can disappear etc. In such a case the value $\mu = 0$ is called the bifurcation point. The Poincaré – Andronov – Hopf bifurcation is a sample of bifurcation when the equilibrium changes into a limit cycle. More precisely, let

$$\dot{x} = A(\mu) \cdot x + F(\mu, x) \tag{11}$$

be a $C^k(\mathbb{R} \times \mathbb{R}^2)$, $k \geq 3$, planar vector field depending on a parameter μ, $\mu \in \mathbb{R}$, such that $F(\mu, 0) = 0$ and $D_2F(\mu, 0) = 0$ for all sufficiently small $|\mu|$. Assume that the linear part $A(\mu)$ has the pair of complex conjugate eigenvalues $\alpha(\mu) \pm i\beta(\mu)$ with $\alpha(0) = 0$ and $\beta(0) \neq 0$. Suppose further that

$$\alpha'(0) \neq 0. \tag{12}$$

Then, in any neighborhood $\mathcal{U}$ of $0 \in \mathbb{R}^2$ and any given $\mu_0 > 0$ there is a μ^*, $|\mu^*| < \mu_0$ such that the differential equation $\dot{x} = A(\mu^*) \cdot x + F(\mu^*, x)$ has a nontrivial periodic orbit in $\mathcal{U}$. Although the question of the the stability of the limit cycle is not possible to solve by the linearized system we can make the following observation about the stability of the stationary point $0 \in \mathbb{R}^2$ of the system (11).

Remark 1. Suppose that $\alpha'(0) > 0$. This assumption can be rewritten as

$$\lim_{\mu \to 0} \frac{\alpha(\mu) - \alpha(0)}{\mu} = \lim_{\mu \to 0} \frac{\alpha(\mu)}{\mu} > 0, \tag{13}$$

which means that there is a deleted neighborhood $\mathcal{P}$ of the point $\mu = 0$ such that

$$\frac{\alpha(\mu)}{\mu} > 0 \tag{14}$$

for all $\mu \in \mathcal{P}$. Now it follows that $\alpha(\mu) < 0$ for $\mu < 0$ and $\alpha(\mu) > 0$ for $\mu > 0$, which means that the origin is asymptotically stable for $\mu < 0$ and unstable for $\mu > 0$. A similar observation can be accomplished if $\alpha'(0) < 0$.

3 Model

The modification of the original Phillips model with multiplier and accelerator will be introduced now. The model is based on ideas given in [3]. In addition to basic linear model we assume that the planned investment B are given by a nonlinear function of the output Y and its speed $\dot{Y}$. Shortly we will write that $B = B(Y, \dot{Y})$. It will be assumed that $B \in C^1(\mathbb{R} \times \mathbb{R})$. Now equation (4) which relates to the lagged real investment can be rewritten as

$$\tau_i \dot{I} = B(Y, \dot{Y}) - I, \tag{15}$$

where τ_i, $\tau_i > 0$, is the length of the average investment lag. If we apply (15) and repeat the same steps as in the subsection 1.2 we gain the equation

$$\tau\tau_i \ddot{Y} + (\tau + s\tau_i)\dot{Y} + sY - B(Y, \dot{Y}) = 0, \tag{16}$$

which is a nonlinear representation of (8).

At first we characterize a stationary point of the latter equation. Let us suppose that there is just one constant solution to (16) that will be denoted $Y(t) = Y^\circ$, $t \in [t_0, \infty)$. Because $\dot{Y}(t) = 0$, $t \in [t_0, \infty)$, this constant solution can be characterized by the equation

$$sY^\circ - B(Y^\circ, 0) = 0. \tag{17}$$

For further analysis we introduce the difference from stationary solution as $y = Y - Y^\circ$. Now (16) can be rewritten as

$$\ddot{y} + \alpha(\tau + s\tau_i)\dot{y} + \alpha s y - \alpha b(y, \dot{y}) = 0, \tag{18}$$

where

$$\alpha = 1/\tau\tau_i > 0 \text{ and } b(y, \dot{y}) = B(y + Y^\circ, \dot{y}) - sY^\circ.$$

If we consider (17) we can observe that $b(0, 0) = 0$. Now the Taylor expansion of the function $b(y, \dot{y})$ at the point $(0, 0)$ can be considered

$$b(y, \dot{y}) = b_1 y + b_2 \dot{y} + e(y, \dot{y}), \tag{19}$$

where the term $e(y, \dot{y})$ contains higher powers of y, $\dot{y}$ and the following relations are valid: $e(0, 0) = 0$ and

$$\lim_{(x_1, x_2) \to (0,0)} \frac{e(x_1, x_2)}{|x_1| + |x_2|} = 0.$$

The terms b_1 and b_2, respectively, can be written as $b_1 = D_1 B(Y^\circ, 0)$, and $b_2 = D_2 B(Y^\circ, 0)$, respectively. In (19) the linear and nonlinear part was separated. Using (19) the equation (18) can be rewritten as

$$\ddot{y} + \alpha(\tau + s\tau_i - b_2)\dot{y} + \alpha(s - b_1)y - \alpha e(y, \dot{y}) = 0. \tag{20}$$

Setting $x_1 = y$ and $x_2 = \dot{y}$, the equation (20) can be rewritten as an autonomous system in $\mathbb{R}^2$

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\alpha(s - b_1) & \alpha\mu \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ e(x_1, x_2) \end{pmatrix}, \tag{21}$$

where $\mu = b_2 - \tau - s\tau_i$.

3.1 Local Analysis

The stationary point (μ°, x°) of the system (21) is $x^\circ = (x_1^\circ, x_2^\circ) = (0, 0)$ and $\mu^\circ = 0$. First we will deal with the linear part of the equation (21) in this stationary point. The characteristic equation of its matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -\alpha(s - b_1) & \alpha\mu \end{pmatrix},$$

is

$$\lambda^2 - \alpha\mu\lambda + \alpha(s - b_1) = 0. \tag{22}$$

Using Viéte's relation we find that

$$\lambda_1\lambda_2 = \alpha(s - b_1),\ \lambda_1 + \lambda_2 = \alpha\mu.$$

According to the standard Keynesian assumption the saving function is steeper than the investment function, cf. [3], [1], which means that in a neighborhood of the stationary point x° it is possible to consider that $s > b_1$. Now we can observe that $\lambda_1\lambda_2 > 0$, which means that the stationary point x° is not a saddle point. If $\mu > 0$, the real part of the roots λ_1, λ_2 of characteristic equation (22) are also positive. It means that the stationary point x° is not stable. If $\mu < 0$, real parts of both roots λ_1, λ_2 of characteristic equation (22) are also negative. It means that the stationary point x° is asymptotically stable. The characteristic behaviour of the stationary point (μ°, x°) can change, which means that the value $\mu^\circ = 0$ of the parameter μ is a bifurcation point of the equation (21).

3.2 Limit Cycle

If the discriminant of the characteristic equation (22) of matrix **A** is applied its roots can be written as

$$\lambda(\mu) = \frac{1}{2}(\alpha\mu \pm \sqrt{\alpha^2\mu^2 - 4\alpha(s - b_1)}).$$

Let us remind that $\alpha > 0$ and $s - b_1 > 0$. For $\mu = 0$ we get pure imaginary roots

$$\lambda_1(0) = -i\sqrt{\alpha(s - b_1)} \text{ and } \lambda_2(0) = i\sqrt{\alpha(s - b_1)}. \tag{23}$$

By the direct computation we can verify that

$$\frac{d}{d\mu}[Re\ \lambda(\mu)]_{|\mu=0} = \frac{\alpha}{2} > 0. \tag{24}$$

From (23) and (24) it is clear that the planar system (21) satisfies the assumptions of Poincaré – Andronov – Hopf theorem. It means that this system has a periodic orbit near its stationary point (x°, μ°), where $x^\circ = (0, 0)$ and $\mu^\circ = 0$, for some small values of μ. In more detail: for each neighborhood $\mathcal{U}$ of the stationary point x° and for all neighborhood $\mathcal{V}$ of parameter value $\mu = 0$ there is $\mu \in \mathcal{V}$ such that system (21) has a nonconstant periodical solution, whose trajectory is a closed curve in $\mathcal{U}$. Its amplitude converge to 0 as μ converge to 0. To find the stability of this periodical solution it is necessary to consider further properties of the nonlinear investment function.

3.3 Discussion

Now it is clear that the first coordinate $x_1 = x_1(t)$ of the solution of the system (21) is a periodical function. Because we put $y(t) = x_1(t)$ and $y(t) = Y(t) - Y^\circ$, there is a periodical solution for output $Y(t) = Y^\circ + y(t)$, which oscillates around the equilibrium Y°. The amplitude of this solution depends on the values of parameters given in equation (16).

Because (24) is valid we can use properties given in Remark 1. Recall that $\mu = b_2 - \tau - s\tau_i$, where τ is a length of lag of the output Y behind demand Z, further τ_i is a length of the average investment lag and $b_2 = D_2B(Y^\circ, 0)$. If $\mu < 0$, which means that $b_2 < \tau + s\tau_i$, then the equilibrium Y° of output is asymptotically stable. On the other side if $\mu > 0$, which means that $b_2 > \tau + s\tau_i$, then the equilibrium Y° of output is unstable.

4 Conclusion

In this paper, a possible generalization of the multiplier-acceleration model of a closed economic unit was described. The original linear model represented a simple endogenous business cycle model that was based on Keynesian accelerator - multiplier and delayed investments. This model cannot completely explain all business cycle phenomena because it requires quite strict condition to produce a periodical development of the output of the given closed economic unit. The model described in this paper uses almost all assumptions as the original model but in addition it also reproduces business cycles that arise from nonlinear relationships between economic aggregates. In particular, it was considered that

there is a nonlinear relationship between output and its rate of change. This nonlinear relation was used in accelerator principle and it is the cause of the emergence of periodical behaviour of output of the considered economic unit. The lags and nonlinearities can be a source of some kind of market imperfection. The market imperfection can be a cause of the endogenous business cycle. The amplitude of the periodical function relates on initial conditions. Because we do not know the specific rule for nonlinear function it is not possible to find all properties of the periodic solution to the suggested model.

Although we concentrated on endogenous model of business cycle as in [4], [3], [7], we think that exogenous real shocks play an important role in business cycles as in [6]. However, we do not think that the external shocks are the only reason for business fluctuations. In this way, the interaction of demand and supply in many markets would not be important perfectly, which would be a contradiction with the essential parts of the economic theory, [2]. What we would like to emphasis is that it is not possible to simply reject the role to endogenous sources of business fluctuations that are caused by nonlinear feedbacks within the economic system itself. In the future we would like to deal with a model that would combine both approaches to business cycle, endogenous models and real business cycle models, respectively.

Acknowledgements

Support of the Specific research project Models and algorithms for optimization problems in terms of uncertainty II of the Faculty of Informatics and Management of University of Hradec Králové is kindly acknowledged.

References

[1] Allen, R. G. D.: *MacroEconomic Theory, A Mathematical Treatment*. Macmillan & Co Ltd, London, 1970.

[2] Blanchard, O.: *Macroeconomics*. Prentice Hall, New Jersey, 2000.

[3] Chiarella, C.: *The elements of a Nonlinear Theory of Economic Dynamics*. Springer-Verlag, Berlin, 1990.

[4] Gandolfo, G.: *Economic Dynamic*. SpringerVerlag, Berlin, 2010.

[5] Hale, J. and Koçak, H.: *Dynamics and Bifurcation*. SpringerVerlag, Berlin, 1991.

[6] Kydland, F., E. and Prescott, E. C.: Time to Build and Aggregate Fluctuations. *Econometrica,* **50**(6), (1982), 1345-1370.

[7] Phillips, A. W.: Mechanical Models in Economic Dynamics. *Economica*, **17**(67), (1950), 283-305.

[8] Pražák, P.: Phillips' Model of a Closed Economic Unit, *WDS'03 - Proceedings of Contributed Papers*(J. Šafránková, ed.), Matfyzpress, Prague, 2003, 66–71.

[9] Puu, T.: Short History of the Multiplier-Accelerator Model. In: *Business Cycle Dynamics Models and Tools* (T. Puu and I. Sushko, eds.), Springer-Verlag, Berlin, 2006, 79–112.

[10] Romer, D.: *Advanced Macroeconomics*. Mc Grow-Hill Companies, Inc., New-York, 2012.

[11] Samuelson, P. A.: Interactions between the Multiplier Analysis and the Principle of Acceleration, *Review of Economic Statistics*, **21**(2) (1939), 7578.

Intuitionistic Fuzzy Cognitive Maps for Corporate Performance Modeling

Ondřej Procházka[1], Petr Hájek[2]

Abstract.
Fuzzy cognitive maps (FCMs) extend cognitive maps to design both the concepts and the relations between the concepts more accurately. The original causality may be represented by a fuzzy relation to express the uncertainty in causal connections. Intuitionistic FCMs (iFCMs) generalize the FCMs with a strong capacity to express uncertainty and to handle hesitancy in decision making. Despite the increasing body of business decision-making literature using intuitionistic fuzzy sets, no attention has been paid to business applications of iFCMs. This paper develops an iFCM model for corporate performance modeling. Here we combined text analysis with knowledge acquisition from experts to assess the connections in the iFCMs. In the analysis of the designed system, we focus on the target concepts represented by sales, costs and profitability. We tested two models of iFCMs, one with hesitancy in connections and the other one with hesitancy in both connections and concept values. We show that the latter approach better models increasing uncertainty in the concept values of business performance.

Keywords: intuitionistic fuzzy sets, fuzzy cognitive maps, corporate performance.

JEL Classification: C63, L21
AMS Classification: 03E72

1 Introduction

In the 1970s, Axelrod [4] introduced cognitive maps. He used cognitive maps for representing social scientific knowledge. Cognitive maps are defined as signed digraphs, in which nodes represent variable concepts and edges represent oriented causal connections. In traditional (crisp) cognitive maps, positive edges have value +1 and the negative ones -1 (zero denotes no edge). However, to express causality using only three values {-1, 0, +1} can be inaccurate in most application domains. Therefore, fuzzy cognitive maps (FCMs) were introduced by Kosko [14]. He combined cognitive maps with the theory of fuzzy sets so that Axelrod's cognitive maps were extended to design both the concepts and the relations between the concepts more accurately. Thus, the original causality may be represented by a fuzzy relation on causal concepts to express the uncertainty in the strengths of the causal connections. This quality has been demonstrated useful in many applications, including financial and business modeling [13, 15, 16, 19, 23]. The applications of FCMs in economy and business have been reviewed, for instance, by [20, 22]. FCMs have also been used for analyzing collaborative planning [5], proactive balance scorecard development [8] or as a decision-support tool for financial planning [26].

Several extensions of FCM have been proposed such as rule-based FCMs [6], dynamic random FCMs [1, 7] or intuitionistic FCMs (iFCM) [11]. The latter approach extends the FCMs with intuitionistic fuzzy sets [2], a generalization of fuzzy sets. Intuitionistic fuzzy sets and systems have attracted considerable interest in business and finance due to their strong capacity to express uncertainty in decision making [9, 10]. In this domain, the precise determination of the concept and relation values may be problematic due to uncertainties associated with dynamic environments, linguistic uncertainties or disagreement among experts. The intuitionistic approach was included into FCMs to handle the hesitancy in decision making. More specifically, iFCM is a decision-making model that can express additional level of uncertainty represented by the degree of expert's hesitancy. Thus, the relations between concepts can describe not only the strength of causality but also the degree of certainty assigned by experts. Despite the increasing body of business decision-making literature dedicated to intuitionistic fuzzy sets, no attention has been paid to business applications of iFCMs. This paper aims to fill this gap and develop an iFCM model for corporate performance modeling.

The remainder of this paper is divided into four sections. The next section lays out the foundation of iFCM and describes the intuitionistic operators on fuzzy sets. The design of iFCM for corporate performance modelling

[1] Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentska 84, 532 10 Pardubice, Czech Republic, st47576@student.upce.cz.
[2] Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentska 84, 532 10 Pardubice, Czech Republic, petr.hajek@upce.cz.

is mentioned in section 3. The results of experiments and what-if analyses are discussed in section 4. The last section concludes this paper and discusses future works.

2 Intuitionistic FCM

The theory of intuitionistic fuzzy sets was introduced by [2]. Fuzzy sets can be represented by membership functions $\mu_A(x)$ that define the degree of membership for each element $x \in E$, $\mu_A(x)$: $E \rightarrow [0,1]$. The intuitionistic approach adds a non-membership function, which describes certainty that an element does not belong to the set. Specifically, an intuitionistic fuzzy set A can be defined as follows. Let x be an element of a set E. Then the intuitionistic fuzzy set A can be defined as:

$$A = \left\{ \left\langle x, \mu_A(x), \gamma_A(x) \right\rangle \middle| x \in E \right\}, \tag{1}$$

where $\mu_A(x)$ is a membership function and $\gamma_A(x)$ is a non-membership function. The summation of membership and non-membership functions gives the amount of certainty. The supplement to one then represents the uncertainty about the element x, this is whether it belongs to the set A or not. This uncertainty or indeterminacy is called hesitancy π_A. Therefore, the hesitancy can be calculated according to the following equation:

$$\pi_A = 1 - \mu_A(x) - \gamma_A(x). \tag{2}$$

The hesitancy represents the indeterminacy degree of the membership of element x. With the intuitionistic approach, the description of uncertainty can be more intuitive [21]. The operators for intuitionistic fuzzy sets were defined in [3]. Here, summation and multiplication operators are mentioned:

$$A \oplus B = \left\{ \left\langle x, \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x), \gamma_A(x) \cdot \gamma_B(x) \right\rangle \middle| x \in E \right\}, \tag{3}$$

$$A \otimes B = \left\{ \left\langle x, \mu_A(x) \cdot \mu_B(x), \gamma_A(x) + \gamma_B(x) - \gamma_A(x) \cdot \gamma_B(x) \right\rangle \middle| x \in E \right\}. \tag{4}$$

The negation operator is defined as follows:

$$\neg A = \left\{ \left\langle x, \gamma_A(x), \mu_A(x) \right\rangle \middle| x \in E \right\}. \tag{5}$$

As mentioned in the previous section, introducing intuitionistic fuzzy sets in FCMs enables stronger expression of uncertainty in both concepts and relations. In addition to the value of membership function, the concepts and relations are also described by the values of non-membership functions. In FCMs, the state of nodes is represented by a state vector $\mathbf{s}^k$ with values s_i^k for each concept, where k denotes the number of iteration (step in time). The concept value for the next iteration k+1 can be calculated as follows:

$$s_i^{k+1} = f\left(s_i^k + \sum_{\substack{j=1 \\ j \neq i}}^{N} s_j^k \times w_{ji} \right), \tag{6}$$

where indexes i and j refer to concepts, and the whole FCM contains N concepts. The concepts in FCMs are connected by edges corresponding to the relations between the concepts. Weight w_{ji} is used to assess the edge from concept j to concept i, this is the weight determines the strengths of the causalities. To calculate the new value of concept, all edges connected to the concept must be involved. Activation function f is used to transform the values of concepts into [0,1] range. Several types of activation functions can be used such as linear, sigmoid or hyperbolic tangent.

In [21], two types of iFCMs are defined, iFCM-I and iFCM-II. iFCM-I enables modeling hesitancy only in the causal relations between concepts. The reasoning process in iFCM-I is based on the following equation [11]:

$$s_i^{k+1} = f\left(\left(2s_i^k - 1\right) + \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(2s_j^k - 1\right) \cdot \xi_{ji} \cdot w_{ji}^{\mu} \cdot \left(1 - w_{ji}^{\pi}\right) \right), \tag{7}$$

where w^{μ}_{ji} denotes the weight of influence, w^{π}_{ji} is the weight of hesitancy, and ξ_{ji} indicates the positive ($\xi_{ji} = 1$) or negative ($\xi_{ji} = -1$) sign of the weight.

iFCM-II extends the complexity of the iFCM-I model by adding hesitancy to concept values. Therefore, iFCM-II can be used to express the hesitancy of both the concepts and the relations between the concepts. In other words, both concepts and relations are regarded as intuitionistic fuzzy sets in iFCM-II and their values are calculated as follows:

$$\left\{\left\langle v^{\mu}, v^{\gamma}\right\rangle\right\}_i^{k+1} = F\left(\left\{\left\langle v^{\mu}, v^{\gamma}\right\rangle\right\}_i^{k} \oplus \left(\bigoplus_{\substack{j=1 \\ j\neq i}}^{N}\left(\left\{\left\langle v^{\mu}, v^{\gamma}\right\rangle\right\}_j^{k} \otimes \left\{\left\langle w^{\mu}, w^{\gamma}\right\rangle\right\}_{ji}\right)\right)\right). \tag{8}$$

The value of concepts in eq. (8) is defined by two values, v^{μ} and v^{γ}, which represent the values of membership and non-membership functions. Values w^{μ} and w^{γ} are used to assign the values of membership and non-membership functions to relations. The value of concept's membership function can be defined as [21]:

$$\left(v_i^{\mu}\right)^{k+1} = f_{\mu}\left(\left(v_i^{\mu} + \left(1 - v_i^{\mu}\right)\cdot m_{iN}\right)^{k}\right), \tag{9}$$

where m_{iN} is the summation operator of all concepts with the influence on the target concept, and function f_{μ} represents the activation function for the membership value of concept i. The summation m_{iN} can be calculated recursively for all $j = 1,2, \ldots, N$ as follows:

$$m_{ij} = \begin{cases} v_j^{\mu} w_{ji}^{\mu}, & j = 1 \\ m_{i(j-1)} + v_j^{\mu} w_{ji}^{\mu} - m_{i(j-1)} v_j^{\mu} w_{ji}^{\mu}, & j > i \end{cases} \tag{10}$$

This relation uses eq. (3) for summation, and it describes the summation of influence of all concepts j connected to concept i. To calculate the non-membership value of concept, the multiplication operator (eq. (4)) can be used as follows:

$$\left(v_i^{\gamma}\right)^{k+1} = f_{\gamma}\left(\left(v_i^{\gamma} \cdot \prod_{\substack{j=1 \\ j\neq i}}^{N}\left(v_j^{\gamma} + w_{ji}^{\gamma} - v_j^{\gamma} w_{ji}^{\gamma}\right)\right)^{k}\right), \tag{9}$$

where function f_{γ} represents the activation function for the non-membership values of concept i.

3 Design of iFCM for Corporate Performance Modeling

Three broad ways to construct FCMs are mostly used in the literature [18]: (1) knowledge acquisition from experts via questionnaires or interviews, (2) coding FCM from text analysis, or (3) by drawing FCM from data. Causal relations in corporate performance modeling evolve dynamically and strongly depend on context. Therefore, automatic knowledge acquisition is preferable. However, there are several difficulties associated with the two latter approaches. Text analysis requires understanding of the natural language semantics. Moreover, causal relations may be stated implicitly in the text. Drawing FCMs from data, on the other hand, requires data that reliably approximate the investigated concepts [24]. This is increasingly difficult in business modeling because corporate performance is largely associated with intangible assets. To address these issues, we combined text analysis with knowledge acquisition from experts to assess the weights of relations in the iFCM. First, the concepts were extracted from text documents using latent semantic analysis (LSA) [12]. Specifically, we used a set of annual reports of U.S. chemical companies as a source of text data. The set of annual reports from the year 2013 for 140 companies was drawn from the U.S. Securities and Exchange Commission EDGAR System. For the analysis, we used only the Management Discussion and Analysis sections of the reports because management discusses past and present corporate performance in these sections. To perform LSA, linguistic pre-processing was performed that included tokenization, lemmatization and discarding the stop-words and the least frequent words in the corpus of documents. The detection of synonyms represented another issue to be addressed. We used WordNet ontology [17] both to detect the synonyms and to find the correct sense of the words for the domain (the highest score for the Economy domain were chosen). Using the *tf.idf* weighting scheme, we performed singular value decomposition as a standard method to extract the concepts in LSA. Finally, the resulting con-

cepts were interpreted by means of the keywords with the highest weights. The extracted set of concepts c_1 to c_{18} is partially specific for chemical industry (see Figure 1). In the next step, three domain experts were asked to assign the directions and weights to the causal relations. Seven values of linguistic variable (very low, low, medium low, moderate, high, medium high and very high) were used to express both the strength of influence and the hesitancy of the experts. The opinions of the experts were pooled using the center of gravity defuzzification [25] method to obtain the final pairs $<w^{\mu}, w^{\pi}>$, $w^{\gamma} = 1 - (w^{\mu} + w^{\pi})$.

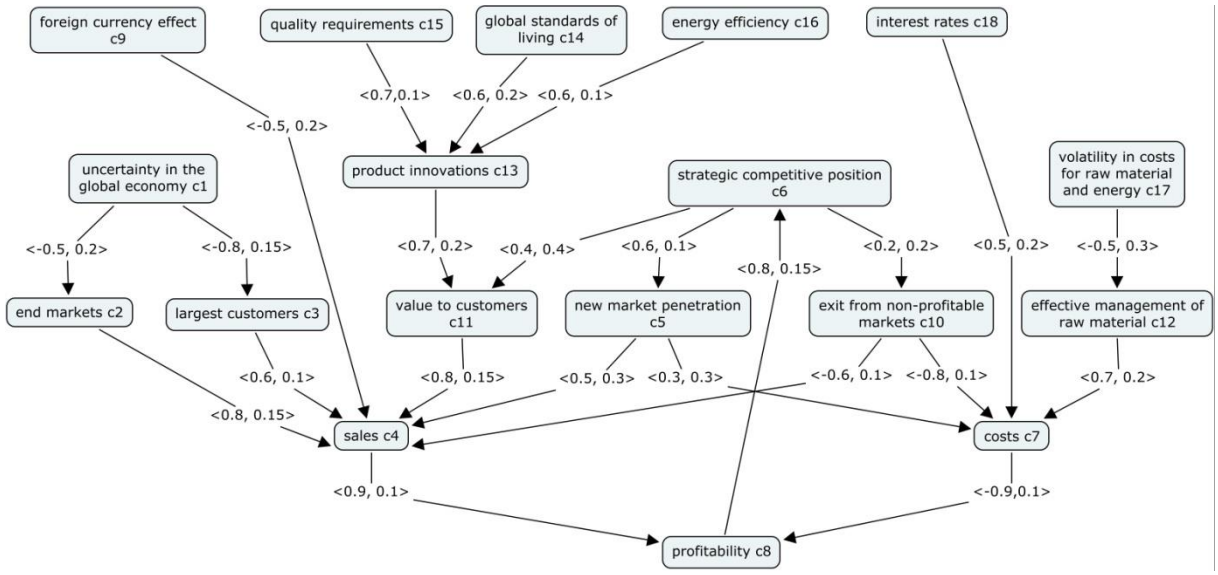


Figure 1 iFCM of corporate performance modeling. The weights of influence are represented by $<\xi_{ji}w^{\mu}, w^{\pi}>$, where $\xi_{ji} = 1$ ($\xi_{ji} = -1$) indicates a positive (negative) influence.

4 Experimental Results

First, we performed experiments with the iFCM-I model. The values of concepts were randomly generated 100 times. To maintain clarity, we report only the average values hereinafter. We focused on the target concepts (c4 sales, c7 costs and c8 profitability) because they represent critical indicators of financial performance. For activation functions, we used sigmoid (Figure 2 left) and hyperbolic tangent functions (Figure 2 right). Although hyperbolic tangent functions were used by [21], they appear to be unsuitable for our case, because concept values exceed the range [0,1] (Figure 2 right). This occurs because hyperbolic tangent transforms negative input value into negative output values. The negative values of business indicators c4 sales and c7 costs are not allowed in this case. Therefore, sigmoid functions provide more realistic results than hyperbolic tangent functions. Using sigmoid functions, the values of the target concepts led to the stable values of 0.5. In terms of convergence, sigmoid functions led to convergence to a fixed equilibrium point 0.5 for all the target concepts, while hyperbolic tangent functions led to c4 = -0.8, c7 = -1.0 and c8 = -0.8.

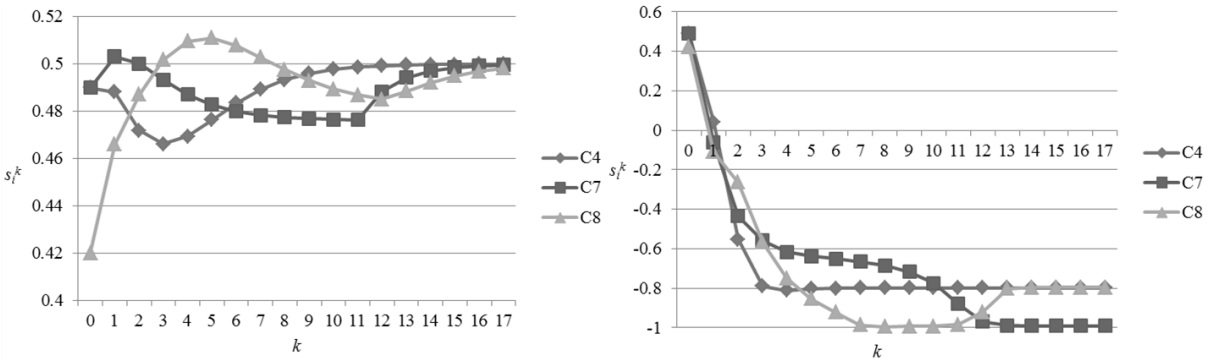


Figure 2 Concept values in iFCM-I, sigmoid (left) and hyperbolic tangent (right) activation function.

We performed further experiments with iFCM-II using hyperbolic tangent and sigmoid functions. This model uses the values of non-membership functions for concepts as well. Therefore, iFCM-II model extends the iFCM-I by the degree of non-membership defined for the concepts. As can be seen from Figure 3, the hesitancy term has different development for each type of activation function. For sigmoid function, hesitancy decreases below zero. This is due to high values of non-membership values. In fact, in this case it holds that $\mu_A(x) + \gamma_A(x) >$

1. In contrast, hesitancy increased from zero up to 0.7 for hyperbolic tangent activation function because $\gamma_A(x)$ = 0 after several iterations. This is much closer to business reality since hesitancy is supposed to increase with increasing projection horizon.

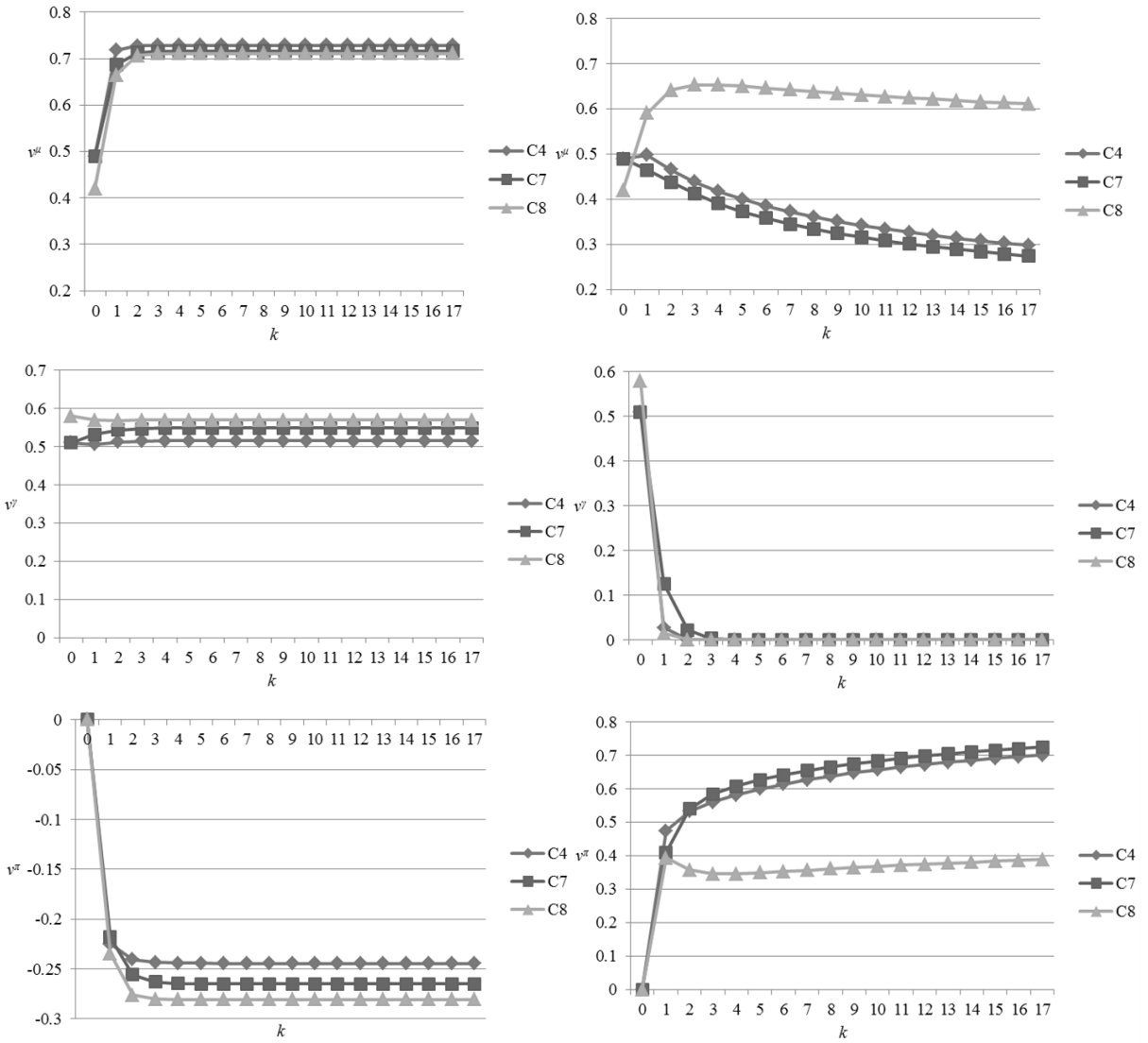


Figure 3 Values of membership, non-membership and hesitancy of target concepts in iFCM-II for sigmoid (left) and hyperbolic (right) activation functions.

5 Conclusion

Taken together, this study presents a design and simulation of iFCMs in the area of economical performance modeling. However, several limitations to this study need to be acknowledged. Most importantly, the iFCMs fail to provide both temporal concept and dynamic inference mechanism. The simulations conducted in this study showed that the input concepts are modified according to their previous values and input signal from connected nodes. This becomes problematic for concepts without additional inputs (for example c1 uncertainty in the global economy or c9 foreign currency effect), where the update provided by experts is necessary. Although the convergence of the values of membership and non-membership functions strongly depend on the structure of iFCM, we observed that hyperbolic tangent activation function better corresponded to reality in case of more complex iFCM-II and it also converged faster than the sigmoid function. In this case, the hesitancy was increasing through time and the value of non-membership function was decreasing, which corroborates the findings obtained by [21] in medical decision making. Further experimental investigations are needed to estimate the role of intuitionistic fuzzy operators and we also recommend a future use of alternative operators. To address the issue of temporal concepts and dynamic inference mechanism, we are currently in the process of investigating dynamic iFCMs. Finally, we encourage future research in incorporating contextual knowledge in the form of setting the dynamic character of concepts.

Acknowledgements

This work was supported by the grant No. SGSFES_2015001 of the Student Grant Competition and by the scientific research project of the Czech Sciences Foundation Grant No: 13-10331S.

References

[1] Aguilar, J.: A dynamic fuzzy-cognitive-map approach based on random neural networks. *International Journal of Computational Cognition* **1(4)** (2003), 91–107.
[2] Atanassov, K. T.: Intuitionistic fuzzy sets. *Fuzzy Sets and Systems* **20(1)** (1986), 87–96.
[3] Atanassov, K. T.: New operations defined over the intuitionistic fuzzy sets. *Fuzzy Sets and Systems* **61(2)** (1994), 137–142.
[4] Axelrod, R. M.: *Structure of decision: the cognitive maps of political elites*. Princeton University Press, Princeton, 1976.
[5] Baykasoglu, A., Durmusoglu, Z. D., and Kaplanoglu, V.: Training fuzzy cognitive maps via extended great deluge algorithm with applications. *Computers in Industry* **62(2)** (2011), 187–195.
[6] Carvalho, J. P., and Tomè, J. A.: Rule based fuzzy cognitive maps-expressing time in qualitative system dynamics. In: *The 10th IEEE International Conference on Fuzzy Systems*. IEEE, 2001, 280–283
[7] Carvalho, J. P.: On the semantics and the use of fuzzy cognitive maps and dynamic cognitive maps in social sciences. *Fuzzy Sets and Systems* **214** (2013), 6–19.
[8] Chytas, P., Glykas, M., and Valiris, G.: A proactive balanced scorecard. *International Journal of Information Management* **31(5)** (2011), 460–468.
[9] Hájek, P., and Olej, V.: Defuzzification methods in intuitionistic fuzzy inference systems of Takagi-Sugeno type: The case of corporate bankruptcy prediction. In: *The 11th Int. Conf. on Fuzzy Systems and Knowledge Discovery* (FSKD'14), Xiamen, 2014, 240–244.
[10] Hájek, P., and Olej, V.: Adaptive intuitionistic fuzzy inference systems of Takagi-Sugeno type for regression problems. In: *Artificial Intelligence Applications and Innovations.* Springer, Berlin, 2012, 206–216.
[11] Iakovidis, D. K., and Papageorgiou, E.: Intuitionistic fuzzy cognitive maps for medical decision making. *IEEE Transactions on Information Technology in Biomedicine* **15(1)** (2011), 100–107.
[12] Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman, R. A.: Indexing by latent semantic analysis. *JAsIs* **41(6)** (1990), 391–407.
[13] Kardaras, D., and Mentzas, G.: Using fuzzy cognitive maps to model and analyze business performance assessment. In: *Proc. of Int. Conf. on Advances in Industrial Engineering–Applications and Practice II*. 1997, 63–68.
[14] Kosko, B.: Fuzzy cognitive maps. *International Journal of Man-Machine Studies* **24(1)** (1986), 65–75.
[15] Koulouriotis, D. E., Diakoulakis, I. E., and Emiris, D. M.: Learning fuzzy cognitive maps using evolution strategies: a novel schema for modeling and simulating high-level behavior. In: *Proceedings of the 2001 Congress on Evolutionary Computation*, IEEE, 2001, 364–371.
[16] Kwahk, K. Y., and Kim, Y. G.: Supporting business process redesign using cognitive maps. *Decision Support Systems* **25(2)** (1999), 155–178.
[17] Miller, G. A.: WordNet: a lexical database for English. *Communications of the ACM* **38(11)** (1995), 39–41.
[18] Özesmi, U., and Özesmi, S. L.: Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological Modelling* **176(1)** (2004), 43–64.
[19] Papageorgiou, E. I.: A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing* **11(1)** (2011), 500–513.
[20] Papageorgiou, E. I.: Review study on fuzzy cognitive maps and their applications during the last decade. In: *Business Process Management*. Springer, Berlin, 2013, 281–298.
[21] Papageorgiou, E. I., and Iakovidis, D. K.: Intuitionistic fuzzy cognitive maps. *IEEE Transactions on Fuzzy Systems*, **21(2)** (2013), 342–354.
[22] Papageorgiou, E. I., and Salmeron, J. L.: A review of fuzzy cognitive maps research during the last decade. *IEEE Transactions on Fuzzy Systems* **21(1)** (2013), 66–79.
[23] Procházka, O., and Hájek, P.: Modelling knowledge management processes using fuzzy cognitive maps. In: *10th International Conference on Knowledge Management in Organizations*, 2015, in press.
[24] Schneider, M., Shnaider, E., Kandel, A., and Chew, G.: Automatic construction of FCMs. *Fuzzy sets and systems* **93(2)** (1998), 161–172.
[25] Van Broekhoven, E., and De Baets, B.: Fast and accurate center of gravity defuzzification of fuzzy system outputs defined on trapezoidal fuzzy partitions. *Fuzzy Sets and Systems* **157(7)** (2006), 904–918.
[26] Xirogiannis, G., Glykas, M., and Staikouras, C.: Fuzzy cognitive maps in banking business process performance measurement. In: *Fuzzy Cognitive Maps*. Springer, Berlin, 2010, 161–200.

A polynomial case of generalized linear-fractional programming without nonnegativity condition

Miroslav Rada[1], Michal Černý[2]

Abstract. The usability of the linear-fractional programming (LFP) is often hindered by the fact that the assumption of positive denominators is too restrictive. This holds even for generalized LFP (GLFP), where the assumption on denominators is weakened to nonnegativity.

It is known that GLFP without nonnegativity condition can be solved via decomposition to some subproblems in GLFP form. However, in general, the number of these subproblems is exponential in the number of fractions in the problem.

In this paper, we focus on some polynomial cases of the above decomposition. In particular, it can be shown that the decomposition is related to enumeration of fulldimensional cells of hyperplane arrangements. We also discuss properties of the arrangements that are sufficient for the decomposition to be polynomial.

Keywords: generalized linear-fractional programming, no nonnegativity assumption, polynomial cases, arrangements of hyperplanes.

1 Motivation and structure of the paper

1.1 Linear-fractional programming – the easy problem

Consider classical linear-fractional problem (LFP) in form

$$\min_{x\in\mathbb{R}^n}\left\{\frac{d^\mathsf{T}x+c}{g^\mathsf{T}x+f}:\ Ax\le b,\ g^\mathsf{T}x+f>0\right\}, \tag{1}$$

where $c,f\in\mathbb{R}$, $g,d\in\mathbb{R}^n$, $b\in\mathbb{R}^k$ and $A\in\mathbb{R}^{k\times n}$. In fact, it consist of optimization of the ratio of two given linear functions over a polyhedron, while the denominator of the objective function must be positive. LFPs are of huge popularity due to two reasons: at first, it often occurs in practice, in particular, it is one of the main techniques in data envelope analysis (DEA), and moreover, it can be easily transformed to linear programming problem (LP) with Charnes-Cooper transformation: we set $y=\frac{x}{g^\mathsf{T}x+f}$ and $\alpha=\frac{1}{g^\mathsf{T}x+f}$ and rewrite (1) to

$$\min_{y\in\mathbb{R}^n,\alpha\in\mathbb{R}}\left\{c^\mathsf{T}y+\alpha d:\ Ay\le\alpha b,\ g^\mathsf{T}y+\alpha h=1,\ \alpha\ge 0\right\}. \tag{2}$$

The optimal solution of (1) can be obtained from the optimal solution (y,α) of (2) as $x=y/\alpha$ as long as $\alpha>0$; it is also obvious that the objective values of (1) and (2) are equal for every $\alpha>0$. The case of $\alpha=0$ means that minimum of (1) does not exist, however, the optimal value of (2) is infimum of (1), actually.

1.2 Generalized linear-fractional programming – the harder problem

Sometimes, we need the more general problem

$$\inf_{x\in\mathbb{R}^n}\sup_{i\in[m]}\left\{\frac{c_ix+d_i}{g_ix+f_i}:Ax\le b,Gx+f\ge 0\right\}. \tag{3}$$

[1]University of Economics, Prague, Faculty of Finance and Accounting, nám. W. Churchilla 4, 130 67 Praha 3, miroslav.rada@vse.cz

[2]University of Economics, Prague, Faculty of Informatics and Statistics, nám. W. Churchilla 4, 130 67 Praha 3, cernym@vse.cz

The goal is to find minimal upper bound of all the m given ratios of affine functions simultaneously. Herein, $f, d \in \mathbb{R}^m$ are given constant vectors and c_i and g_i are rows of matrices $C, G \in \mathbb{R}^{m\times n}$. The constraints $Ax \leq b$ have the same meaning as in (1). To avoid confusion at this point, the symbol $[m]$ stands for the set $\{1, \ldots, m\}$.

The problem (3) can be rewritten in a more convenient and shorter form

$$\inf_{x\in\mathbb{R}^n, t\in\mathbb{R}} \{t : \ Cx + d \leq t(Gx + f), \ Ax \leq b, \ Gx + f \geq 0\}. \tag{4}$$

The problems (3) and (4) are generally referred to as generalized linear-fractional problems (hereinafter GLFP). There are several economic applications leading to GLFPs, the von Neumann growth model [7] serves as example. A more comprehensive study on applications can be found in [9].

Although the problem (3) cannot be linearized, many algorithms for this problem were developed, for example interior point methods by Nesterov and Nemirovski [6] or by Freund and Jarre [4]. However, these algorithms were originally designed for searching for maxima and minima, instead of suprema and infima, that we assume in (3) and (4). Infima and suprema mean theoretically no problem, however in practice, some numerical instabilities may arise. In fact, the problem (3) can be also solved with arbitrary precision using binary search algorithm.

1.3 Generalized linear-fractional programming without nonnegativity assumption on denominators – the hardest problem

In some applications, the condition of $Gx + f \geq 0$ in the formulations of the problems (3) and (4) is a bit restrictive. For example, the problem (4) without this nonnegativity condition emerged recently in multicriteria parametric optimization [5]. Recall how the problems without nonnegativity conditions read: for (3) it is

$$\inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{c_i x + d_i}{g_i x + f_i} : \ Ax \leq b \right\} \tag{5}$$

and, for the shorter form (4), we have

$$\inf_{x\in\mathbb{R}^n, t\in\mathbb{R}} \{t : \ Cx + d \leq t(Gx + f), \ Ax \leq b\}. \tag{6}$$

The used symbols are the same as in previous section.

The bad news here are that the formulations (5) and (6) are not equivalent. The main difference is that we are minimizing the upper bound of ratios in (5), unlike in (6), where the constraints can bound t both from above and below in general.

For those problems in their full generality, no polynomial algorithms are known and it may be difficult to solve to optimum. However, under certain assumptions, the problem (5) is solvable efficiently – and here comes out our paper. Hence, from on now, we will denote (5) by the acronym GGLFP (as generalized GLFP).

The problem (6) seems to be much difficult to tract. In fact, as far as we know, there are only heuristics algorithms without any convergence results, see [5].

1.4 Goal and structure of the paper

The aim of our paper is twofold. First, we show the relation between the problem (6) and the enumeration of cells of hyperplane arrangements. We show that our GGLFP can be decomposed to some (in general exponentially many) GLFPs. Second, we want to show that under certain assumption on G and f, the number of GLFPs to be solved is polynomially bounded and thus GGLFP becomes polynomially solvable.

The rest of the paper is structured as follows: Section 2 lays some basics for the sections following. In Section 3, we describe the decomposition of GGLFP to solvable ones: in Subsection 3.1, we show how we decompose and why it works, and in Subsection 3.2, we reformulate the subproblems to GLFP. Section 4 introduces the notion of hyperplane arrangements and connects it to the GGLFP. The polynomial cases of GGLFP are discussed in Section 5.

2 Preliminaries

In the whole paper, the following symbols (for input data) will be used:

- a matrix $A \in \mathbb{R}^{k\times n}$ and a vector $b \in \mathbb{R}^k$,
- matrices $C, G \in \mathbb{R}^{m\times n}$ and vectors $d, f \in \mathbb{R}^m$, and
- the row vectors g_i for $i \in [m]$, these are rows or the matrix G.

Convention. To avoid confusion, we adopt the convention that infima are sought over extended reals, i.e. $\inf \mathbb{R} = -\infty$ and $\inf \emptyset = \infty$.

Some notation. For given vector $a \in \mathbb{R}^n$, the symbol $\mathrm{diag}(a)$ stands for the diagonal $n \times n$ matrix built by the elements of a. Unless said otherwise, vectors are understood columnwise. The symbol $S := \{-1, 1\}^m$ denotes the space of all m-sized ± 1 vectors. An element $s \in S$ is called *sign vector*. The symbol $[m]$ means the set of the first m naturals, $\{1, \ldots, m\}$.

For a given $s \in S$, let $\mathcal{C}(s)$ denote the set

$$\{x \in \mathbb{R}^n : \mathrm{diag}(s)(Gx + f) \geq 0\}. \tag{7}$$

Recall the definitions of GLFP and GGLFP and the complexity of GLFP:

Definition 1 (GLFP)**.** A *generalized linear-fractional problem* (GLFP) is the problem in the form (3), i.e.

$$\inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{c_i x + d_i}{g_i x + f_i} : Ax \leq b, Gx + f \geq 0 \right\}.$$

Lemma 1 ([4, 6])**.** *There exists a (weakly) polynomial algorithm for* GLFP.

Definition 2 (GGLFP)**.** A *generalized* GLFP (GGLFP) is the problem in the form (6), i.e.

$$\inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{c_i x + d_i}{g_i x + f_i} : \ Ax \leq b \right\}.$$

3 Decomposition procedure of GGLFP

In this section, we describe how GGLFP can be decomposed to at most 2^m subproblems, that can be solved efficiently.

3.1 Formulation of subproblems

Let $s \in S$ be given. Define t_s as

$$t_s := \inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{c_i x + d_i}{g_i x + f_i} : \ Ax \leq b, \ x \in \mathcal{C}(s) \right\}. \tag{8}$$

About the problem of finding infimal t_s from the all 2^m possible sign vectors s, which reads

$$\inf_{s\in S} \{t_s\}, \tag{9}$$

one can observe the following:

Proposition 2. *The optima of* (5) *and* (9) *are equal.*

Proof. The infimum (9) is sought among infima over every possible $\mathcal{C}(s)$. Obviously, the $\bigcup_{s\in S} \mathcal{C}(s) = \mathbb{R}^n$, hence, the least of the infima t_s is also the infimum over the whole $\mathbb{R}^n$. □

Corollary 3. *For solving* GGLFP (5)*, it is sufficient to solve* 2^m *problems of form* (8)*.*

3.2 Solving the subproblems

We are to solve (8) for a given $s \in S$. The main trick is easy: we multiply every ratio in (9) by 1 in form s_i/s_i. We get

$$t_s = \inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{s_i(c_i x + d_i)}{s_i(g_i x + f_i)} : \ Ax \le b, \ x \in \mathcal{C}(s) \right\}. \tag{10}$$

Proposition 4. *The problem* (10) *is* GLFP.

Proof. Set $C' := \operatorname{diag}(s)C$, $G' := \operatorname{diag}(s)G$, $d' := \operatorname{diag}(s)d$ and $f' := \operatorname{diag}(s)f$. Then (10) can be rewritten as

$$t_s = \inf_{x\in\mathbb{R}^n} \sup_{i\in[m]} \left\{ \frac{c_i' x + d_i'}{g_i' x + f_i'} : \ Ax \le b, \ G'x + f \ge 0 \right\},$$

which has the form (3). □

4 Hyperplane arrangements

Note that not every $\mathcal{C}(s)$ in the decomposition from Section 3 must be nonempty. The subproblems (8) with empty $\mathcal{C}(s)$ do not have to be examined at all, since their infima are obviously equal to ∞. In fact, the number of examined $s \in S$ can be reduced significantly in some cases. For the explanation, we introduce the notion of hyperplane arrangements.

Definition 3. The family of hyperplanes $\mathcal{A}(G, f) := \{\{x \in \mathbb{R}^n : g_i x = -f_i\} : i \in \{1, \dots, m\}\}$ forms an *arrangement of hyperplanes*, which is a subdivision of $\mathbb{R}^n$ into regions of the form $\mathcal{C}(s)$ (recall (7)). If the set $\mathcal{C}(s)$ is fulldimensional for a given $s \in S$, we say that s *generates a cell* of the arrangement $\mathcal{A}(G, f)$. The set $\mathcal{C}(s)$ itself is called *a cell*. The set of all sign vectors that generate a cell of $\mathcal{A}(G, f)$ is denoted by $\mathfrak{C}$.

The arrangement $\mathcal{A}(G, 0)$ is called *central*.

Lemma 5 ([1, 2])**.** *The number of cells of an arrangement of m hyperplanes in $\mathbb{R}^n$ is at most*

$$\sum_{i=0}^{n} \binom{m}{i}. \tag{11}$$

Furthermore, there are arrangements, for which the bound is attained.

Lemma 6 ([3])**.** *A central arrangement of m hyperplanes in $\mathbb{R}^n$ has at most $\sum_{i=0}^{n-1} 2\binom{m-1}{i}$ cells.*

Corollary 7. *The number of cells of an arrangement $\mathcal{A}$ of m hyperplanes in $\mathbb{R}^n$ is $\mathcal{O}(m^n)$. If $\mathcal{A}$ is central, the number of cells is $\mathcal{O}(m^{n-1})$.*

The relation between arrangements and the decomposition from Section 3 states Proposition 8:

Proposition 8. *The following holds:*

$$\min_{s\in S}\{t_s\} = \min_{s\in\mathfrak{C}}\{t_s\}.$$

Proof. Since every cell $\mathcal{C}(s)$ contains its boundary, it is obvious that $\bigcup_{s\in C(G)} \mathcal{C}(s) = \mathbb{R}^n$. The rest of the proof is analogous to proof of Proposition 2. □

Corollary 9. *For solving* GGLFP*, it is advantageous to be able to enumerate cells of hyperplane arrangement efficiently, since the number of subproblems* (8) *to be solved is reduced from 2^m to $\mathcal{O}(m^n)$.*

4.1 Algorithms for enumeration of cells of hyperplane arrangements

Cell enumeration for arrangements is a standard problem with numerous applications in data analysis, uncertainty or optimization.[1] Many algorithms for this problem were developed. Due to the exponential size of the output, one cannot expect the effectivity of the algorithms in the traditional sense, where the time complexity is measured by input size only. Instead, so-called *output-sensitivity* is often considered desirable. Output-sensitive algorithm is polynomial in size of input or output, or, more strictly, polynomial in size of output and *linear* in size of input. Also, *compactness* may be advantageous; which means that the space complexity is polynomial in input size only, so the output does not have to be remembered. These properties are closely related to the concept of P-enumerability, introduced by Valiant [10].

From the algorithms known, we mention the algorithm [2], that enumerates not only cells of arrangements, but also the faces of arrangement that have lower dimensions. Other notable algorithms are compact output-sensitive algorithms [8] and [3]. These papers constructively prove Lemma 10:

Lemma 10 ([8, 3])**.** *Let an arrangement $\mathcal{A}(G,f)$ of hyperplanes be given by G and f. There is a compact output-sensitive algorithm that enumerates all cells of $\mathcal{A}(G,f)$ in time $\mathcal{O}(|\mathfrak{C}|\ m\ lp(G,f))$ using space $\mathcal{O}(lp(G,f))$, where $lp(G,f)$ is time complexity or space complexity of solving a linear program with data G,f and $|\mathfrak{C}|$ is the number of cells of $\mathcal{A}(G,f)$.*

5 Polynomial cases of the decomposition

Putting together Propositions 2 and 8, Corollary 7 and Lemmas 1 and 10, we immediately conclude some polynomial cases for GGLFP:

Lemma 11.

(a) If m is $\mathcal{O}(1)$, then GGLFP *(5) can be solved in (weakly) polynomial time.*
(b) If n is $\mathcal{O}(1)$, then GGLFP *(5) can be solved in (weakly) polynomial time.*

Proof. If m is constant, it is sufficient to use the polynomial GLFP algorithm on the 2^m subproblems (8).

If n is constant, it is sufficient solve $\mathcal{O}(m^n)$ subproblems (8) according to Proposition 8 and Corollary 7.

The weakness of polynomiality follows from the usage of weakly polynomial algorithms for GLFP. □

The more interesting result is contained in Lemma 12.

Lemma 12. *Let d be rank of the $m\times(n+1)$ matrix $(G\ f)$ built by concatenation of G and f. If d is $\mathcal{O}(1)$, then* GGLFP *(5) can be solved in (weakly) polynomial time.*

Proof. First, convert the problem (3) to the form

$$\inf_{x\in\mathbb{R}^{n+1}} \sup_{i\in[m]} \left\{ \frac{c_i' x}{g_i' x} :\ A'x \le b' \right\}, \qquad C' = (C\ d),\ G' = (G\ f),\ A' = \begin{pmatrix} A & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix},\ b' = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Then, it is necessary to enumerate cells of $\mathcal{A}(G',0)$ for the decomposition. Since G' is a $m\times(n+1)$ matrix of rank d, $\mathcal{A}(G',0)$ can be realized in $\mathbb{R}^{d-1}$: a regular matrix $P\in\mathbb{R}^{(n+1)\times(n+1)}$ must exist such that $G'P = (G''\ 0)$, where $G''\in\mathbb{R}^{m\times d}$. Obviously, the arrangement $\mathcal{A}(G'',0)$ has the same cells as $\mathcal{A}(G',0)$. Since $\mathcal{A}(G'',0)$ is central arrangement of m hyperplanes in $\mathbb{R}^d$, the number its cells can be bounded by $\mathcal{O}(m^{d-1})$ due to Lemma 6.

The rest of proof is analogous to the proof of Lemma 11. □

[1] Although, some of the applications lead on enumeration of vertices of a zonotope, what is actually the same problem from dual point of view.

6 Conclusion

We faced the problem of minimizing the supremum of a finite number of fractions, where both the numerator and denominator were affine functions. We showed a method, how such problem can be decomposed into (in general exponentially many) solvable generalized linear-fractional problems. We showed that the decomposition is closely related to enumeration of fulldimensional cells of hyperplane arrangement. We described several cases, in which the studied problem can be solved in polynomial time. The case from Lemma 12 is especially interesting.

Our further research in this area shall comprise the generalization of our decomposition technique to the more general problem (6). This is is closely related to the solving of the problem

$$\inf_{t_1,t_2,x} \{t_1 - t_2 : \ (Cx+d)_i \le t_1(Gx+f)_i, \ i \in I, \ (Cx+d)_j \le t_2(Gx+f)_j, \ j \notin I, \ Ax \le b, \ x \in \mathcal{C}(s)\}$$

for a given $s \in S$ and $I := \{i : s_i = 1\}$. Unfortunately, it is unclear whether it can be solved efficiently.

Acknowledgements

Many thanks belong to our colleague and friend Milan Hladík. The idea of decomposition in Section 3 arose (in a bit different form) in our joint work on range sets in parametric multicriteria optimization.

The work of the first author was supported by Institutional Support of University of Economics, Prague, VŠE IP 100040. The work of the second author was supported by Internal Grant Agency of University of Economics, VŠE IGA F4/54/2015.

References

[1] Buck, R. C. Partition of space. *Amer. Math. Monthly.* 1943, 50, pp. 541–544.

[2] Edelsbrunner, H. – O'Rouke, J. – Seidel, R. Constructing arrangements of lines and hyperplanes with applications. *SIAM J. Comput.* May 1986, 15, 2, pp. 341–363. ISSN 0097-5397. Online: http://dx.doi.org/10.1137/0215024.

[3] Ferrez, J. A. – Fukuda, K. – Liebling, T. M. Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm. *European Journal of Operational Research.* 2005, 166, 1, pp. 35 – 50. ISSN 0377-2217. Online: http://www.sciencedirect.com/science/article/pii/S0377221704003352. Metaheuristics and Worst-Case Guarantee Algorithms: Relations, Provable Properties and Applications.

[4] Freund, R. W. – Jarre, F. An Interior Point Method for Multifractional Programs with Convex Constraints. *J. Optim. Theory Appl.* April 1995, 85, 1, pp. 125–161. ISSN 0022-3239. Online: http://dx.doi.org/10.1007/BF02192302.

[5] Hladík, M. – Rada, M. – Sitarz, S. Range sets in multiobjective linear programming and parametric polytopes intersection problem. *preprint submitted to EJOR.* 2015.

[6] Nesterov, Y. – Nemirovskii, A. An interior-point method for generalized linear-fractional programming. *Mathematical Programming.* 1995, 69, 1-3, pp. 177–204. ISSN 0025-5610. Online: http://dx.doi.org/10.1007/BF01585557.

[7] von Neumann, J. A model of general economic equilibrium. *The Review of Economic Studies.* 1945, 13, 1, pp. 1–9.

[8] Rada, M. – Černý, M. An incremental compact output-sensitive algorithm for enumerating cells of hyperplane arrangement: a rival for Reverse Search. *preprint submitted to SIAM Journal on Discrete Mathematics.* 2015.

[9] Schaible, S. Fractional programming: applications and algorithms. *European Journal of Operational Research.* 1981, 7, 2, pp. 111–120.

[10] Valiant, L. G. The complexity of enumeration and reliability problems. *SIAM Journal on Computing.* 1979, 8, 3, pp. 410–421.

Application of Benford's Law for the analysis of the reliability of production quality data

Angelina Rajda-Tasior[1]

Abstract. Each production process is exposed to a large amount of problems and these are often associated with breakdowns and quality defects. In order to isolate possible conflicts within data sets, techniques such as Benford's Law can be used to confirm the probity of quality data. Within the accompanying literature we are informed that quality is not produced, but is monitored. If we assume that the records of defects can be considered a strategic tool then analysis of this data can be used to develop continuous improvement strategies. One of the methods of analysis of the reliability of whole sets, using statistics and probability theory is Benford's law.

In the available literature we can find many examples of the application of Benford's law to assess the reliability of different sets of data, mainly in searching for fraud and signs of manipulation. The analysis of compliance data using Benford's distribution could also be applied in the study of the reliability of the defects detected during quality control. The aim of this paper is the verification of the effectiveness of Benford's law in different sets of data. The hypothesis is that the collected data conforms to Benford's law. In this paper empirical and simulation data is used. This study concerns the analysis of first significant digits based on the daily reject record sets of defective products.

Keywords: Benford's law, analysis of data sets, numerical methods, simulation methods, defects analysis

JEL Classification: C44

1. Introduction – Benford's Law

In [1] Benford's formula states that the probability of the leading digit being of a certain value can be described by the following function

$$p(d) = \log_{10}(1+\frac{1}{d}), \quad gdzie \;\; d \in \{1,2,...,9\} \tag{1}$$

In accordance with formula (1) the distribution of the first significant multi-digit is showed in table (1). As can be seen, 1 occurs with a 30% frequency, and subsequent digits have a progressively decreasing frequency (i.e. 9 occurs only with 4,6% frequency).

First digit	1	2	3	4	5	6	7	8	9
Probability of a digit	30,1%	17,6%	12,5%	9,7%	7,9%	6,7%	5,8%	5,1%	4,6%

Table 1. Benford's distribution.

[1] University of Economics in Katowice, Department of Statistics, 1 Maja 50, 40-287 Katowice, angelina.rajda@gmail.com

Benford's law, also called the First-Digit Law [8], refers to the frequency distribution of digits. In this distribution, 1 occurs as the leading digit about 30% of the time, while larger digits occur in that position less frequently: 9 as the first digit less than 5% of the time. This law of anomalous numbers as Benford called it, states that the logarithm's first digits of certain data is uniformly distributed. These predicted probabilities for the first significant digit have been shown to hold, not only for some theoretical distributions [4], but also for real data sets [3]. Benford' law began to be used in the verification of a number of cases of fraudulent statistical material. It helped to reveal data manipulation by testing whether or not the data's first digits confirm to Benford's distribution. Benford's law has been invoked as evidence of fraud in the 2009 Iranian elections and also used to analyze other election results.

2. Practical application of Benford's law

Farbaniec et al. [2] shows that Benford's law can be used in different areas. In economic studies for:

- ✓ Detection of fraudulent data or unintentional errors in accountancy
- ✓ Detection of tax fraud [7]
- ✓ Analysis of stock exchange data [5]
- ✓ Analysis of data about product prices which have been auctioned on the Internet
- ✓ Assessment which pertains to the correctness of compensation estimates in insurance companies
- ✓ Analysis of the reliability of the amount of the financial penalties are imposed in court proceedings
- ✓ Assessment of credibility rates pertaining to fines and penalties ruled in legal proceedings

In quantitative studies for:

- ✓Verifying the correctness of statistical data
- ✓Testing the correctness of econometric models i.e. forecasting procedure
- ✓Optimalization of calculations in solving transport and logistics problems

In technical studies for:

- ✓Designing mass memory for computers
- ✓Being able to distinguish real photography from computer generated graphics
- ✓Analysis of the capacity of files transferred via Internet, including the time of transfer
- ✓Analysis of the performance of numerical algorithms

In medical studies for:

- ✓ Analysis of the clinical effectiveness of medicines
- ✓ Analysis of the reliability of the emission of toxic pollutants

For all other non-classified areas:

- ✓ Analysis of subsidies for political parties
- ✓ Number of references in publications
- ✓ Number of injured and killed in traffic accidents
- ✓ Number of return customers to the stores
- ✓ Association rules regarding the number of purchased products
- ✓ Sports statistics

3. Testing for compliance with Benford's law

In the available literature it is possible to find examples of Benford's law to assess the reliability of data sets. Basically these analyses are comprehensive and they come from a process that produces Benford's law. This paper demonstrates the analysis of two sets of independent natural data as an example (length of rivers and areas of countries). The aim of the article is to analyse defect data that was obtained from clients in the form of complaints. The research contains one set of empirical quality data and one set generated from a simulation implemented in the R program for statistical computing. Based on analysis of the empirical and simulation models the study is designed to test the following hypothesis (2):

$$H_0 : F(x) = F_0(x)$$
$$H_1 : F(x) \neq F_0(x) \qquad (2)$$

The null hypothesis is that the first digits in a set of data follow the Benford's law. The alternative hypothesis is that the first digits in a set of data do not follow Benford's law. Data is tested for goodness of fit using the Pearsonian χ2-test and the critical value for testing is pre-computed at an $\alpha = 0{,}05$ level of significance.

3.1. Case of area of all the countries in the world

The first study concerns the geographical areas of countries of the world. The set contains 194 multi-digit numbers (Table 2) and these are analyzed for the prevalence of the first digit on the significant position (Figure 1). In order to analyse the compatibility of distributions χ^2 test was applied. P-value = 0.89 at the assumed significance level $\alpha = 0{,}05$ means no grounds to reject H_0.

No.	Contry	Area [km^2]
1.	Rosja	17 075 400
2.	Kanada	9 976 140
3.	Chińska Republika Ludowa	9 598 077
4.	Stany Zjednoczone	9 373 967
	⋮	
192.	Nauru	21
193.	Monako	1,95
194.	Watykan	0,4

Table 2 Areas of countries.

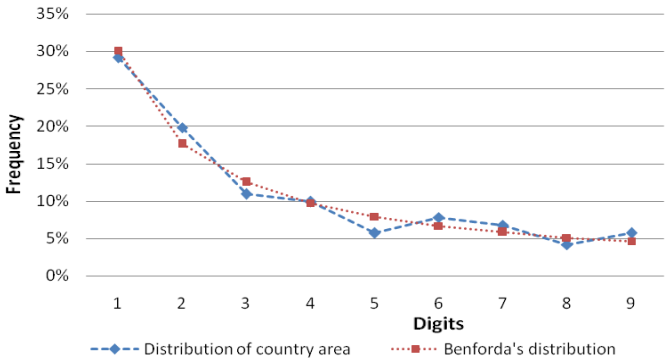


Fig. 1. Distribution of countries's area.

3.2. Case of length of the rivers in Europe

Another study concerns the length of the rivers in Europe. For a data set of 151 (Table 3), the analysis shows the results (Figure 2). Value $p = 0.43$ means there is no reason to reject H_0 that the distribution of the set conforms to Benford's distribution.

No.	River	Lenght [km]
1.	Adour	309,2
2.	Adyga	409
3.	Arno	241
4.	Bieła	1430
	⋮	
149.	Ural	2428
150.	Ver	20
151.	Wardar	388

Table 3. Length of rivers in Europe.

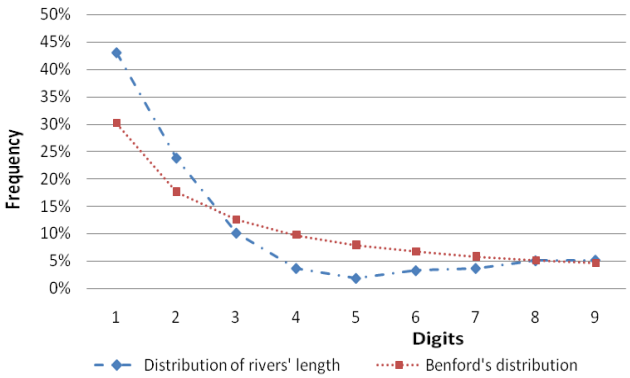


Fig. 2. Distribution of length of rivers in Europe.

4. Analysis of defectiveness.

Throughout this paper, let $E_i = \{e_1, e_2, ..., e_N\}$ denote a set of N articles that are checked in a process quality control. Individual elements within a production batch are measured and the results are assigned a value from the formula (3). Each article may be accepted or rejected according to its deviation from a specification, a standard, or an expectation.

$$E_i = \begin{cases} 0, & \text{when } e_1 \text{ fulfills the requirements} \\ 1, & \text{when } e_1 \text{ is defective} \end{cases} \tag{3}$$

Let $Z = \{z_1, z_2, ..., z_n\}$ be a set of defects in e_i. If the E_i consist of N articles, the result of a quality control will be Z which is the size of all defects observed in the articles in one month (formula 4).

$$Z = \sum_{i=1}^{n} z_i \tag{4}$$

For example one set of articles is E_1 = {70 sofas, 20 armchairs, …, 30 sofa corners} and number of defects Z = {10 broken legs, 20 crack frames, 38 blocked of hinges, …, 45 blemish on the fabric}.

5. Analysis of case study. Empirical and simulated data

The empirical analysis relates to data about defects. The data set consist of 300 records relating to the daily amount of complaints received within a manufacturing company (Fig. 3). The first digits in this set of data do not follow Benford's law. Results show that despite the fact that 1 is the most common digit and 2 and 3 occurring less so, digits 8 and 9 are more frequent than 5 and 7. Value p = 0,00 on assumed level of significance α = 0.05 leads to a rejection H_0 in favor of H_1. However, this result may be due to a relatively low data set.

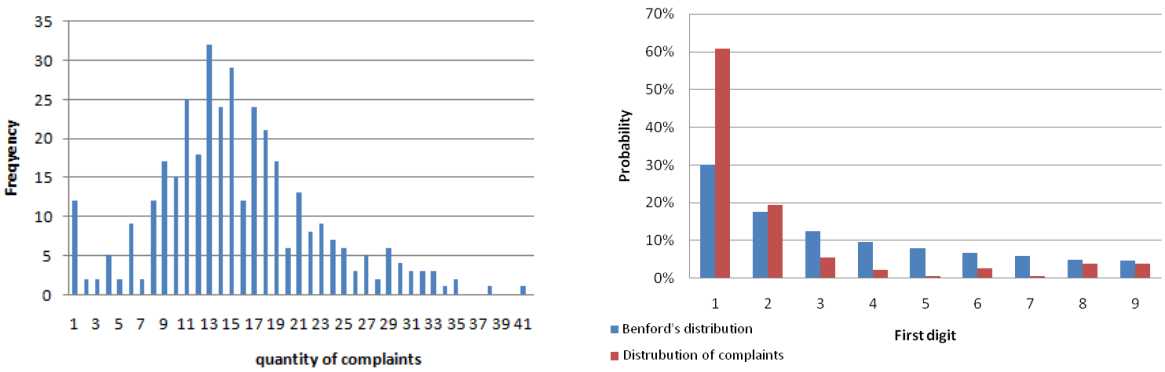


Fig. 3. Distribution of complaints (left side) and comparison to the Benford's law.

Because Benford's law does not work well on small sample sizes a larger sample set is needed. To ensure that data about defects follows Benford's law an increase in simulated data is required. The purpose of the analysis of simulations was to randomly generate from the binomial distribution to obtain an increased data set (E_i) with different numbers of products (N). An important method was to obtain the size of defects (Z) from Poisson distribution. To the model is added one parameter m which controls the intensity of defects. All simulation parameters are chosen to match well to the actual observed process control. In this way 10,000 sets of data was simulated with random defects. The next step was to find the distribution of significant figures for simulated defects.

Results

The principle idea in Benford's law is that the frequency distribution of the first digits from numerical data of certain processes should confirm to a discrete distribution known as

the Benford's distribution [6]. The main objective of this study was to verify the hypothesis that the distribution of quality control data (about defects) matches with Benford's distribution. If the null hypothesis is not rejected in this study, this would indicate that in every research of quality data if the null-hypothesis is rejected, it can be assumed that data manipulation has occurred and a more stringent audit should be conducted to quality controllers. Goodness-of-fit tests have been used to assess if the data's first digits confirm to described distribution. Simulation parameters are chosen to be well matched to the actual observed parameters of the control process. From research both empirical and simulated data has skewed to the right. However, the collection of simulated data after research does seem to obey the significant-digit law (Figures 4-5) but $\chi 2$-test for Benford's distribution is 268.1 with p-value less than 0,05 that means that the null hypothesis indicates a reject and the alternative hypothesis indicates to accept. Any simulation of the size of defects in products is very difficult, but very interesting and will be continued in future research.

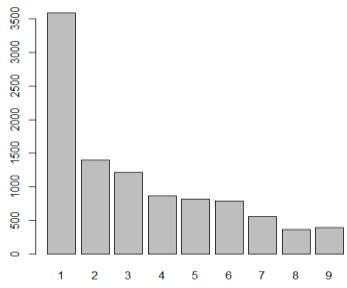


Figure 4 Frequency chart of simulation data.

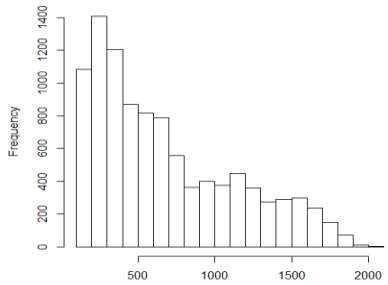


Figure 5 Histogram of simulation data distribution.

References

[1] Benford F.: The *Law of Anomalous Numbers, Proceedings of the American Philosophical Society*, Vol. 78, No. 4/1938

[2] Farbaniec M., Grabiński T., Zabłocki B., Zając W.: *The use of Benford's law in the analysis of the accuracy of financial data on the example of information on trade in goods. The role of science in economic and social.* Conference materials, 2011

[3] Hill T. P.: *The first digital phenomenon*, American Scientist, 86/1998

[4] Leemis L.M., Schmeiser B.W., Evans D. L.: *Survival distributions satisfying Benford's law*. The American Statistician 54(3)/2000

[5] Ley E.: *On a pecular distribution of the U.S. stock indicies digits*, American Statistician, 1996

[6] Newcomb S.: *Note on the Frequency of Use of the Different Digits in Natural Numbers*, American Journal of Mathematics, Vol. 4, No. 1,1881

[7] Nigrini, M.J. & Mittermaier, L.J.: *The use of Benford's law as an aid in analytical procedures*, Auditing: A Journal of Practice and Theory, 1997

[8] www.benford.pl

Linear programming problem with IF coefficients

Jaroslav Ramík[1]

Abstract. The paper is concerned with linear programming problems whose input data may be intuitionistic fuzzy (IF) while the values of variables are always real numbers. We propose rather general approach to this type of problems based on level sets, and present recent results for problems in which the notions of feasibility and optimality are based on the fuzzy relations of possibility and necessity. An illustrative example is supplied.

Keywords: linear programming, IF linear programming, fuzzy optimization.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Given the practical relevance of linear programming, it is not surprising that attempts to extend linear programming theory to problems involving uncertain data have been appearing since the early days of fuzzy sets, and, later of IF sets [1]-[2], [5], [6].

The intuitionistic fuzzy set (IFS), sometimes called Atanassov IFS, is an extension of fuzzy set [1], where the degree of non-membership denoting the non-belongingness to a set is explicitly specified along with the degree of membership of belongingness to the set. Unlike the fuzzy set, where the non-membership degree is taken as one minus the membership degree, in IFS, the membership and non-membership degrees are more or less independent and related only by that the sum of two degrees must not exceed one.

To obtain a meaningful extension of linear programming to problems involving IF data, one has to specify a suitable class of permitted IF numbers, introduce fundamental arithmetic operations with such numbers, define inequalities between IF numbers, and clarify the meaning of feasibility and optimality in IF linear programming problems (IFLP problems). Up till now, IF linear programming is being developed mostly in one direction based on special indexes that transform each IF number into a couple of real numbers. However, this strong simplification looses a detailed structure of the IF number, see [3], [2].

Throughout the paper we assume that some or all of the data defining the problem may be IF while the values of variables are always real numbers. For problems with IF decision variables, see e.g. [3].

2 Preliminaries

We shall use the following canonical form of a linear programming problem.

Given real numbers $b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n, a_{11}, a_{12}, \ldots, a_{mn}$,

$$\begin{aligned} &\text{maximize} && c_1x_1 + c_2x_2 + \cdots + c_nx_n && && (1)\\ &\text{subject to} && a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \;\leq\; b_i, && i = 1, 2, \ldots, m, && (2)\\ & && x_j \;\geq\; 0, && j = 1, 2, \ldots, n. && (3) \end{aligned}$$

The set of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers that simultaneously satisfy inequalities (2) and (3) is called the *feasible region* of problem (1)-(3), and the elements of feasible region are called *feasible*

[1]Silesian University in Opava, School of Business Administration in Karviná, University Sq. 1934/3, 733 40 Karvina, Czech Republic, ramik@opf.slu.cz

solutions. A feasible solution $\hat{x}$ such that no feasible solution x satisfies

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n > c_1\hat{x}_1 + c_2\hat{x}_2 + \cdots + c_n\hat{x}_n$$

is called an *optimal solution* of (1)-(3), and the set of all optimal solutions is called the *optimal region*.

Here, we follow the opinion that fuzzy sets are special nested families of subsets of a set, see [5].

Definition 1. A *fuzzy subset* of a nonempty set X (or a *fuzzy set* on X) is a family $\{A_\alpha\}_{\alpha\in[0,1]}$ of subsets of X such that $A_0 = X, A_\beta \subset A_\alpha$ whenever $0 \le \alpha \le \beta \le 1$, and $A_\beta = \cap_{0\le\alpha<\beta}A_\alpha$ whenever $0 < \beta \le 1$. The *membership function of* A is the function μ_A from X into the unit interval $[0, 1]$ defined by $\mu_A(x) = \sup\{\alpha \mid x \in A_\alpha\}$.

Similarly, IF sets are special couples of nested families of subsets of a set as follows.

Definition 2. An *IF subset* C of a nonempty set X (or an *IF set* on X) is a couple of families $C = (A, B)$, $A = \{A_\alpha\}_{\alpha\in[0,1]}$ and $B = \{B_\alpha\}_{\alpha\in[0,1]}$, where A_α, B_α are subsets of X such that

$$\begin{aligned}
A_0 = X, A_\beta \subset A_\alpha \quad & \textit{whenever} \quad 0 \le \alpha \le \beta \le 1,\\
A_\beta = \cap_{0\le\alpha<\beta}A_\alpha \quad & \textit{whenever} \quad 0 < \beta \le 1,\\
B_0 = X, B_\beta \subset B_\alpha \quad & \textit{whenever} \quad 0 \le \alpha \le \beta \le 1,\\
B_\beta = \cap_{0\le\alpha<\beta}B_\alpha \quad & \textit{whenever} \quad 0 < \beta \le 1,\\
A_\alpha \subset B_\alpha \quad & \textit{whenever} \quad 0 \le \alpha \le 1.
\end{aligned}$$

The *membership function of* C is the function μ_C from X into the unit interval $[0, 1]$ defined by $\mu_C(x) = \mu_A(x) = \sup\{\alpha \mid x \in A_\alpha\}$, and the *non-membership function of* C is the function ν_C from X into the unit interval $[0, 1]$ defined by $\nu_C(x) = 1 - \mu_B(x)$ where $\mu_B(x) = \sup\{\alpha \mid x \in B_\alpha\}$.

Remark 1. Let A be a subset of a set X and let $\{A_\alpha\}_{\alpha\in[0,1]}$ be the family of subsets of X defined by $A_0 = X$ and $A_\alpha = A$ for each positive α from $[0, 1]$. It can easily be seen that this family is a fuzzy set on X and that its membership function is equal to the characteristic function of A; we call it the *crisp fuzzy set* on X, see [4] for details.

Remark 2. An IF set (A, B), where $A = \{A_\alpha\}, B = \{B_\alpha\}$, is given by two fuzzy sets. The first one, A, represents the membership, the other one, B, represents the non-membership of the IF set. It is worth noting that an IF set (A, A), where $A = \{A_\alpha\}$ is equivalent to the fuzzy set A.

Remark 3. Notice that by the last inclusion in Definition 2, i.e. $A_\alpha \subset B_\alpha$ whenever $0 \le \alpha \le 1$, we obtain the standard condition for IF sets, see [1],

$$\mu_A(x) + (1 - \mu_B(x)) \le 1 \text{ for all } x \in X.$$

We denote the collection of all fuzzy sets on X by $\mathcal{F}(X)$, similarly, the collection of all IF sets on X by $\mathcal{F}^I(X)$. For each $\alpha \in [0, 1]$, the set $[A]_\alpha = \{x \in X \mid \mu_A(x) \ge \alpha\}$ is called the *α-cut of fuzzy set A*. Similarly, for each $\alpha, \beta \in [0, 1]$, the set $\{x \in X \mid \mu_C(x) \ge \alpha, \nu_C(x) \le \beta\}$ is called the *(α, β)-cut of* IF set $C = (A, B)$ and it is denoted by $[C]_{\alpha,\beta}$. Notice that $[C]_{\alpha,\beta} = \{x \in X \mid \mu_A(x) \ge \alpha, \mu_B(x) \ge 1 - \beta\}$. If $\alpha = \beta$ we simply say that $[C]_{\beta,\beta}$ is the β-cut of IF set $C = (A, B)$ instead of (β, β)-cut of IF set and we simply write $[C]_\beta$ instead of $[C]_{\beta,\beta}$. Notice that

$$[C]_\beta = \{x \in X \mid \mu_A(x) \ge \beta, \mu_B(x) \ge 1 - \beta\}. \tag{4}$$

If X is a nonempty subset of a real finite dimensional topological space, then a fuzzy set A in X is called *closed, bounded, compact* or *convex* if the α-cut $[A]_\alpha$ is a closed, bounded, compact or convex subset of X for every $\alpha \in]0, 1]$, respectively. Similarly, an IF set $C = (A, B)$ in X is called *closed, bounded, compact* or *convex* if the (α, β)-cut $[C]_{\alpha,\beta}$ is a closed, bounded, compact or convex subset of X for every $\alpha, \beta \in]0, 1[$, respectively.

Following the terminology of [4], we say that a fuzzy subset A of $\mathbb{R}$ is a *fuzzy quantity* whenever A is normal, compact and its membership function μ_A satisfies the following condition: there exist $a, b, c, d \in \mathbb{R}$, $-\infty < a \le b \le c \le d < +\infty$, such that

$$\begin{aligned}
\mu_A(t) = 0 \quad & \text{if } t < a \text{ or } t > d,\\
\mu_A \quad & \text{is strictly increasing and continuous on the interval } [a, b],\\
\mu_A(t) = 1 \quad & \text{if } b \le t \le c,\\
\mu_A \quad & \text{is strictly decreasing and continuous on the interval } [c, d].
\end{aligned} \tag{5}$$

Moreover, we define a *zero-cut of* A as $[A]_0 = [a, d]$, i.e. $[A]_0$ is a compact interval. The set of all fuzzy quantities is denoted by $\mathcal{F}_0(\mathbb{R})$.

Notice that $\mathcal{F}_0(\mathbb{R})$ contains well known classes of fuzzy numbers: crisp (real) numbers, crisp intervals, triangular fuzzy numbers, trapezoidal and bell-shaped fuzzy numbers etc. However, $\mathcal{F}_0(\mathbb{R})$ does not contain fuzzy sets with "stair-like" membership functions. In a similar way, we shall say, that an IF set $C = (A, B)$ of $\mathbb{R}$ is an *IF quantity* whenever A and B are fuzzy quantities. The set of all IF quantities is denoted by $\mathcal{F}_0^I(\mathbb{R})$.

Recall that the binary relations on X are subsets of the Cartesian product $X \times X$ and that the fuzzy sets on $X \times X$ are called the *fuzzy binary relation* on X, or simply *fuzzy relation* on X. Because the set of the conventional binary relations on X can be embedded into the set of IF relations on X, we obtain also extensions of conventional binary relations on X to IF relations on X.

3 Intuitionistic Fuzzy Linear Programming

We distinguish the fuzzy quantities from real numbers by writing the tilde above the corresponding symbol. The same tilde is used for denoting fuzzy quantities, each IF quantity is, however, given by a couple of fuzzy quantities. Thus we write $\tilde{c}_j = (\tilde{c}_j^{(1)}, \tilde{c}_j^{(2)}), \tilde{a}_{ij} = (\tilde{a}_{ij}^{(1)}, \tilde{a}_{ij}^{(2)})$ and $\tilde{b}_i = (\tilde{b}_i^{(1)}, \tilde{b}_i^{(2)})$ and consequently $\mu_{\tilde{c}_j^{(k)}} : \mathbb{R} \to [0,1]$, $\mu_{\tilde{a}_{ij}^{(k)}} : \mathbb{R} \to [0,1]$ and $\mu_{\tilde{b}_i^{(k)}} : \mathbb{R} \to [0,1]$, respectively, for $k = 1, 2$ and $i \in \mathcal{M} = \{1, 2, \ldots, m\}$ and $j \in \mathcal{N} = \{1, 2, \ldots, n\}$.

Notice that if $\tilde{c}_j$ and $\tilde{a}_{ij}$ are IF quantities, then, for every $(x_1, x_2, \ldots, x_n)$ from $\mathbb{R}^n$, the IF subsets $\tilde{c}_1 x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_n x_n$ and/or $\tilde{a}_{i1} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in} x_n$ of $\mathbb{R}$ defined by the Extension principle are again IF quantities.

As the canonical IF counterpart of the canonical linear programming problem (1)-(3) we consider the problem

$$\begin{array}{lllll} \text{"maximize"} & \tilde{c}_1 x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_n x_n & & & \\ \text{"subject to"} & (\tilde{a}_{i1} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in} x_n) & \tilde{P}_i & \tilde{b}_i, & i \in \mathcal{M}, \\ & x_j & \geq & 0, & j \in \mathcal{N}, \end{array} \tag{6}$$

where $\tilde{c}_j = (\tilde{c}_j^{(1)}, \tilde{c}_j^{(2)}), \tilde{a}_{ij} = (\tilde{a}_{ij}^{(1)}, \tilde{a}_{ij}^{(2)})$, $\tilde{b}_i = (\tilde{b}_i^{(1)}, \tilde{b}_i^{(2)})$ and $\tilde{P}_i = (\tilde{P}_i^{(1)}, \tilde{P}_i^{(2)})$.

For each $i \in \mathcal{M}$, the IF quantities $\tilde{a}_{i1} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in} x_n$ and $\tilde{b}_i$ from $\mathcal{F}_0^I(\mathbb{R})$ are compared by an IF relation $\tilde{P}_i$ on $\mathcal{F}_0^I(\mathbb{R})$, and, where the meanings of "subject to" and "maximize"; that is, the meanings of feasibility and optimality, remain to be specified.

Here, we shall study the case in which all $\tilde{P}_i$ appearing in the constraints of problem (6) are the same. Namely, let $\tilde{P} = (\tilde{P}^{(1)}, \tilde{P}^{(2)})$ be an IF relation on $\mathcal{F}_0^I(\mathbb{R})$ and let us assume that $\tilde{P}_i = \tilde{P}$ for all $i \in \mathcal{M}$.

Feasibility.

Let $\alpha, \beta \in]0, 1[$. By an (α, β)-*feasible region* of problem (6) we understand the (α, β)-cut $[\tilde{X}]_{\alpha,\beta}$ of the IF subset $\tilde{X} = (\tilde{X}^{(1)}, \tilde{X}^{(2)})$ of $\mathbb{R}^n$ whose $\mu_{\tilde{X}}$ is given by $\mu_{\tilde{X}} = \mu_{\tilde{X}^{(1)}}$, where

$$\mu_{\tilde{X}^{(1)}}(x) = \begin{cases} \min\limits_{1 \leq i \leq m} \mu_{\tilde{P}^{(1)}}(\tilde{a}_{i1}^{(1)} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in}^{(1)} x_n, \tilde{b}_i^{(1)}) & \text{if } x_j \geq 0 \text{ for all } j \in \mathcal{N}, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

and whose non-membership function $\nu_{\tilde{X}}$ is given by $\nu_{\tilde{X}} = 1 - \mu_{\tilde{X}^{(2)}}$, where

$$\mu_{\tilde{X}^{(2)}}(x) = \begin{cases} \min\limits_{1 \leq i \leq m} \mu_{\tilde{P}^{(2)}}(\tilde{a}_{i1}^{(2)} x_1 \tilde{+} \cdots \tilde{+} \tilde{a^{(2)}}_{in} x_n, \tilde{b}_i^{(2)}) & \text{if } x_j \geq 0\ ,\ j \in \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The elements of (α, β)-feasible region are called (α, β)-*feasible solutions* of problem (6), and $\tilde{X}$ defined by (7), (8) is called the *feasible region* of problem (6). When the data in (6) are crisp, then $\tilde{X}$ becomes the feasible region of the canonical linear programming problem (1)-(3). In this paper we consider the case $\alpha = \beta$. Then (β, β)-feasible region is simply called β-feasible region and (β, β)-feasible solutions of problem (6) are simply called β-feasible solutions; they have the membership grade at least β and non-membership grades at most $1 - \beta$.

Optimality.

The set of IF values of the objective function is not linearly ordered, and that the relation for making comparison of elements of this set may be independent of that used in the notion of feasibility. We propose to use the well known notion of α-efficient (α-nondominated) solution of the FLP problem, see [5]. First we observe that a feasible solution $\hat{x}$ of non fuzzy LP problem (1)-(3) is optimal exactly when there is no feasible solution x such that $cx > c\bar{x}$. This suggests to introduce a suitable intuitionistic fuzzy extensions of " $>$ ". Let $\tilde{Q} = (\tilde{Q}^{(1)}, \tilde{Q}^{(2)})$ be an IF relation on $\mathcal{F}_0^I(\mathbb{R})$ and let $\alpha \in]0,1[$. If $\tilde{a} = (\tilde{a}^{(1)}, \tilde{a}^{(2)})$ and $\tilde{b} = (\tilde{b}^{(1)}, \tilde{b}^{(2)})$ are IF quantities, then we write

$$\tilde{a}\ \tilde{Q}_\alpha\ \tilde{b}, \text{ if } (\mu_{\tilde{Q}^{(1)}}(\tilde{a}^{(1)}, \tilde{b}^{(1)}) \geq \alpha \text{ and } \mu_{\tilde{Q}^{(2)}}(\tilde{a}^{(2)}, \tilde{b}^{(2)}) \geq \alpha) \tag{9}$$

and call $\tilde{Q}_\alpha = (\tilde{Q}_\alpha^{(1)}, \tilde{Q}_\alpha^{(2)})$ the *α-relation on $\mathcal{F}_0^I(\mathbb{R})$ associated to $\tilde{Q}$*. We also write

$$\tilde{a}\ \tilde{Q}_\alpha^*\ \tilde{b}, \text{ if } (\tilde{a}\ \tilde{Q}_\alpha\ \tilde{b} \text{ and } \mu_{\tilde{Q}^{(1)}}(\tilde{b}^{(1)}, \tilde{a}^{(1)}) < \alpha \text{ and } \mu_{\tilde{Q}^{(2)}}(\tilde{b}^{(2)}, \tilde{a}^{(2)}) < \alpha), \tag{10}$$

and call $\tilde{Q}_\alpha^*$ the *strict α-relation on $\mathcal{F}_0^I(\mathbb{R})$ associated to $\tilde{Q}$*.

Now, let $\alpha, \beta \in]0,1[$. We say that a β-feasible solution $\hat{x}$ of (6) is *(α,β)-maximal solution* of (6) if there is no β-feasible solution x of (6) different from $\hat{x}$ such that

$$\sum_{j\in\mathcal{N}} \tilde{c}_j^{(1)}\hat{x}_j\ \tilde{Q}_\alpha^{*(1)}\ \sum_{j\in\mathcal{N}} \tilde{c}_j^{(1)}x_j, \text{ and } \sum_{j\in\mathcal{N}} \tilde{c}_j^{(2)}\hat{x}_j\ \tilde{Q}_\alpha^{*(2)}\ \sum_{j\in\mathcal{N}} \tilde{c}_j^{(2)}x_j. \tag{11}$$

In what follows, we use the natural extensions of binary relations " $\leq$ " and " $\geq$ " on $\mathbb{R}$ to fuzzy relations on $\mathcal{F}(\mathbb{R})$ that are based on the possibility and necessity relations Pos and Nec defined on the set of all fuzzy sets $\mathcal{F}(\mathbb{R})$ by

$$\mu_{Pos}(\tilde{u}, \tilde{v}) = \sup\{\min(\mu_{\tilde{u}}(x), \mu_{\tilde{v}}(y)) | x, y \in \mathbb{R}, x \leq y\}, \tag{12}$$

$$\mu_{Nec}(\tilde{u}, \tilde{v}) = \inf\{\max(1 - \mu_{\tilde{u}}(x), 1 - \mu_{\tilde{v}}(y)) | x, y \in \mathbb{R}, x \leq y\}, \tag{13}$$

where $\tilde{u}, \tilde{v} \in \mathcal{F}(\mathbb{R})$. Equivalently, we write $\tilde{u} \preceq^{Pos} \tilde{v}$ and $\tilde{u} \preceq^{Nec} \tilde{v}$, instead of $\mu_{Pos}(\tilde{u}, \tilde{v})$ and $\mu_{Nec}(\tilde{u}, \tilde{v})$, respectively, and by $\tilde{u} \succeq^{Pos} \tilde{v}$ we mean $\tilde{v} \preceq^{Pos} \tilde{u}$.

Proposition 1. *Consider IFLP problem (6), let $\beta \in]0,1[$.*
Let $P = (P^{(1)}, P^{(2)}), P^{(1)} = P^{(2)}) = \preceq^{Pos}$ ($\preceq^{Nec}$). A vector $x = (x_1, \ldots, x_n)$ is a β-feasible solution of the FLP problem (6) if and only if it is a nonnegative solution of the system of inequalities

$$\sum_{j\in\mathcal{N}} \tilde{a}_{ij}^{(1)\mathrm{L}}(\beta)x_j \leq \tilde{b}_i^{(1)\mathrm{R}}(\beta), i \in \mathcal{M}, (\sum_{j\in\mathcal{N}} \tilde{a}_{ij}^{(2)\mathrm{R}}(1-\beta)x_j \leq \tilde{b}_i^{(2)\mathrm{L}}(1-\beta), i \in \mathcal{M}, \text{ respectively}).$$

Remark 4. By analogy, we can easily derive more two results with either $P = (P^{(1)}, P^{(2)})$, $P^{(1)} = \preceq^{Pos}$, $P^{(2)} = \preceq^{Nec}$, or $P = (P^{(1)}, P^{(2)})$, $P^{(1)} = \preceq^{Nec}$, $P^{(2)} = \preceq^{Pos}$, see [4]. The same holds for Proposition 2.

Proposition 2. *Consider IFLP problem (6), let $\alpha, \beta \in]0,1[$.*
Let $P = (P^{(1)}, P^{(2)}), P^{(1)} = P^{(2)}) = \preceq^{Pos}$, ($\preceq^{Nec}$), $Q = (Q^{(1)}, Q^{(2)})$, $Q^{(1)} = Q^{(2)}) = \preceq^{Pos}$, ($\preceq^{Nec}$). Let $c_j \in \mathbb{R}$ be such that $\tilde{c}_j^{(1)\mathrm{L}}(\alpha) \leq c_j \leq \tilde{c}_j^{(1)\mathrm{R}}(\alpha)$ for all $j \in \mathcal{N}$.
If $x^* = (x_1^*, \ldots, x_n^*)$ is an optimal solution of the LP problem

$$\begin{array}{ll} \textit{maximize} & \sum_{j\in\mathcal{N}} c_j x_j \\ \textit{subject to} & \sum_{j\in\mathcal{N}} \tilde{a}_{ij}^{(1)\mathrm{L}}(\beta)x_j \leq \tilde{b}_i^{(1)\mathrm{R}}(\beta), i \in \mathcal{M}, (\sum_{j\in\mathcal{N}} \tilde{a}_{ij}^{(2)\mathrm{R}}(1-\beta)x_j \leq \tilde{b}_i^{(2)\mathrm{L}}(1-\beta), i \in \mathcal{M}, \text{ resp.}), \\ & x_j \geq 0, \quad j \in \mathcal{N}, \end{array} \tag{14}$$

then x^* is an (α,β)-maximal solution of IFLP problem (6).

4 Illustrating example

In this section we discuss a simple illustrative example to clarify the introduced concepts and results, to provide some interpretation and features of possible applications, see also [4].

Consider n products $G_1, G_2, \cdots, G_n$ that should be manufactured. The manufacturing process is composed of m sub-processes S_i, $i = 1, \cdots, m$. The estimated processing resources (e.g. processing time, materials, etc.) for manufacturing a batch of product G_j for sub-process S_i are $\tilde{a}_{ij}$ units. The total working resource for sub-process S_i is restricted by $\tilde{b}_i$ units. The profit rate (1000 EUR/batch) of product G_j is estimated as $\tilde{c}_j$. All mentioned parameters $\tilde{a}_{ij}$, $\tilde{b}_i$ and $\tilde{c}_j$ are IF quantities. We shall find how many products G_j should be manufactured in order to "maximize" the total profit. We formulate the following IFLP problem.

$$\begin{array}{lrcll} \text{"}\tilde{Q}\text{-maximize"} & \tilde{c}_1 x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_n x_n & & & \\ \text{"subject to"} & (\tilde{a}_{i1} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in} x_n) & \tilde{P} & \tilde{b}_i, & i = 1, \cdots, m, \\ & x_j & \geq & 0, & j = 1, \cdots, n, \end{array} \tag{15}$$

where $\tilde{c}_j = (\tilde{c}_j^{(1)}, \tilde{c}_j^{(2)})$, $\tilde{c}_j^{(k)} = (c_j^{(k)L}, c_j^{(k)}, c_j^{(k)R})$, $k = 1, 2$, $\tilde{a}_{ij} = (\tilde{a}_{ij}^{(1)}, \tilde{a}_{ij}^{(2)})$, $\tilde{a}_{ij}^{(k)} = (a_{ij}^{(k)L}, a_{ij}^{(k)}, a_{ij}^{(k)R})$ $k = 1, 2$, and $\tilde{b}_i = (\tilde{b}_i^{(1)}, \tilde{b}_i^{(2)})$, $\tilde{b}_i^{(k)} = (b_i^{(k)L}, b_i^{(k)}, b_i^{(k)R})$, $i = 1, \cdots, m$, $j = 1, \cdots, n$, are triangular fuzzy quantities with triangular piecewise linear membership functions.
From now on, we shall consider the case $n = m = 2$ with the following triangular fuzzy quantities

$$\begin{array}{llll} \tilde{c}_1^{(1)} = (3, 4, 5), & \tilde{c}_1^{(2)} = (2, 4, 6), & \tilde{c}_2^{(1)} = (2, 4, 6), & \tilde{c}_2^{(2)} = (1, 4, 8), \\ \tilde{a}_{11}^{(1)} = (1, 3, 5), & \tilde{a}_{11}^{(2)} = (0, 3, 6), & \tilde{a}_{12}^{(1)} = (1, 1, 1), & \tilde{a}_{12}^{(2)} = (1, 1, 1), \\ \tilde{a}_{21}^{(1)} = (1, 3, 5), & \tilde{a}_{21}^{(2)} = (0, 3, 7), & \tilde{a}_{22}^{(1)} = (3, 3, 3), & \tilde{a}_{22}^{(2)} = (3, 3, 3), \\ \tilde{b}_1^{(1)} = (8, 11, 14), & \tilde{b}_1^{(2)} = (6, 11, 16), & \tilde{b}_2^{(1)} = (11, 12, 15), & \tilde{b}_2^{(2)} = (10, 12, 17). \end{array} \tag{16}$$

Notice that $\tilde{a}_{12}$ and $\tilde{a}_{22}$ are crisp IF numbers.

(I) Assume that $\tilde{P}$ and $\tilde{Q}$ are IF relations, $\tilde{P} = (\preceq^{Pos}, \preceq^{Pos})$ and $\tilde{Q} = (\preceq^{Pos}, \preceq^{Pos})$.
Given $\alpha, \beta \in]0, 1[$. In order to obtain (α, β)-maximal solution of IFLP problem (15), by Proposition 2, we have to solve the following LP problem, where $\tilde{c}_j^{(1)\mathrm{L}}(\alpha) \leq c_j \leq \tilde{c}_j^{(1)\mathrm{R}}(\alpha)$ for $j = 1, 2$:

$$\begin{array}{ll} \text{maximize} & z = c_1 x_1 + c x_2, \\ \text{subject to} & \tilde{a}_{11}^{(1)\mathrm{L}}(\beta) x_1 \tilde{+} \tilde{a}_{12}^{(1)\mathrm{L}}(\beta) x_2 \leq \tilde{b}_1^{(1)\mathrm{R}}(\beta), \\ & \tilde{a}_{21}^{(1)\mathrm{L}}(\beta) x_1 \tilde{+} \tilde{a}_{22}^{(1)\mathrm{L}}(\beta) x_2 \leq \tilde{b}_2^{(1)R}(\beta), \\ & x_1, x_2 \geq 0, \end{array} \tag{17}$$

As $\tilde{c}_j^{(1)\mathrm{L}}(\alpha)$ is a "pessimistic" value of j-th unit profit, we call the optimal solution $x^P = (x_1^P, x_2^P)$ of problem (17) with $c_j = \tilde{c}_j^{(1)\mathrm{L}}(\alpha)$, the *pessimistic* (α, β)-maximal solution of (15) (with $\preceq^{Pos}$). Analogically, as $\tilde{c}_j^{(1)\mathrm{R}}(\alpha)$ is an "optimistic" value of j-th unit profit, we call the optimal solution $x^O = (x_1^O, x_2^O)$ of problem (17) with $c_j = \tilde{c}_j^{(1)\mathrm{R}}(\alpha)$, the *optimistic* (α, β)-maximal solution of (15) (with $\preceq^{Pos}$).
Particularly, to obtain the pessimistic (α, β)-maximal solution of (15) we solve the following LP problem.

$$\begin{array}{ll} \text{maximize} & z = (3 + \alpha) x_1 + (2 + 2\alpha) x_2, \\ \text{subject to} & (1 + 2\beta) x_1 + x_2 \leq 14 - 3\beta, \\ & (1 + \beta) x_1 + 3 x_2 \leq 15 - 3\beta, \\ & x_1, x_2 \geq 0. \end{array} \tag{18}$$

On the other hand, to obtain the optimistic (α, β)-maximal solution of (15) we solve the corresponding LP problem. The numerical results for some values of α and β, $\alpha = \beta$, can be seen in Table 1.

(II) Now, assume that $\tilde{P} = (\preceq^{Nec}, \preceq^{Nec})$ and $\tilde{Q} = (\preceq^{Nec}, \preceq^{Nec})$, see (13).
Given $\alpha, \beta \in]0, 1[$. In order to obtain (α, β)-maximal solution of IFLP problem (15), by Proposition 2, we have to solve the following LP problem, where $\tilde{c}_j^{(2)\mathrm{L}}(1 - \alpha) \leq c_j \leq \tilde{c}_j^{(2)\mathrm{R}}(1 - \alpha)$ for $j = 1, 2$:

$$\begin{array}{ll} \text{maximize} & z = c_1 x_1 + c x_2, \\ \text{subject to} & \tilde{a}_{11}^{(2)\mathrm{R}}(1 - \beta) x_1 \tilde{+} \tilde{a}_{12}^{(2)\mathrm{R}}(1 - \beta) x_2 \leq \tilde{b}_1^{(2)\mathrm{L}}(1 - \beta), \\ & \tilde{a}_{21}^{(2)\mathrm{R}}(1 - \beta) x_1 \tilde{+} \tilde{a}_{22}^{(2)\mathrm{R}}(1 - \beta) x_2 \leq \tilde{b}_2^{(2)L}(1 - \beta), \\ & x_1, x_2 \geq 0, \end{array} \tag{19}$$

Let $x^P = (x_1^P, x_2^P)$ be the optimal solution of problem (17) with $c_j = \tilde{c}_j^{(2)\mathrm{L}}(1-\alpha)$, i.e. the pessimistic (α,β)-maximal solution of (15) (with $\preceq^{Nec}$). Analogically, let $x^O = (x_1^O, x_2^O)$ be the optimal solution of problem (17) with $c_j = \tilde{c}_j^{(1)\mathrm{R}}(1-\alpha)$, i.e. the optimistic (α,β)-maximal solution of (15) (with $\preceq^{Nec}$). Particularly, to obtain the pessimistic (α,β)-maximal solution of (15) we solve the following LP problem.

$$\begin{array}{ll} \text{maximize} & z = (4-2\alpha)x_1 + (4-3\alpha)x_2, \\ \text{subject to} & (3-2\beta)x_1 + x_2 \le 11 + 3\beta, \\ & (2-\beta)x_1 + 3x_2 \le 12 + 3\beta, \\ & x_1, x_2 \ge 0. \end{array} \tag{20}$$

On the other hand, to obtain the optimistic (α,β)-maximal solution of (15) we solve the corresponding LP problem. The numerical results for some values of α and β, $\alpha = \beta$, can be seen in Table 2.

Table 1.

$\alpha=\beta$	x_1-Pos-O	x_2-Pos-O	z-Pos-O	x_1-Pos-P	x_2-Pos-P	z-Pos-P
0.0	13.500	0.500	70.500	14.000	0.000	42.000
0.2	8.600	1.360	48.896	8.600	1.360	30.784
0.4	6.150	1.730	37.286	6.150	1.730	25.754
0.6	4.680	1.904	29.731	4.680	1.904	22.941
0.8	3.700	1.980	24.252	3.700	1.980	21.188
1.0	3.000	2.000	20.000	3.000	2.000	20.000

Table 2.

$\alpha=\beta$	x_1-Nec-O	x_2-Nec-O	z-Nec-O	x_1-Nec-P	x_2-Nec-P	z-Nec-P
0.0	3.000	2.000	20.000	3.000	2.000	20.000
0.2	3.700	1.980	25.784	3.700	1.980	20.052
0.4	4.680	1.904	33.126	4.680	1.904	20.307
0.6	6.150	1.730	43.052	6.150	1.730	21.026
0.8	8.600	1.360	57.952	9.571	0.000	22.971
1.0	3.000	0.500	85.000	14.000	0.000	28.000

References

[1] K.T. Atanassov: Intutionistic fuzzy sets. *Fuzzy Sets and Systems* **20** 87–96 (1986)

[2] D. Dubey, S. Chandra, and A. Mehra: Fuzzy linear programming under interval uncertainty based on IFS representation. *Fuzzy Sets and Systems* **188** 1 68–87 (2012)

[3] A. L. Nachammai, P. Thangaraj: Solving intuitionistic fuzzy linear programming problem by using similarity measures, *European Journal of Operational Research* **72** 2 204–210 (2012)

[4] J. Ramík: Duality in fuzzy linear programming with possibility and necessity relations. *Fuzzy Sets and Systems* **157** 1283–1302 (2006)

[5] J. Ramík, M. Vlach: *Generalized concavity in optimization and decision making* (Kluwer Academic Publishers, Boston-Dordrecht-London 2001) (2001)

[6] R. R. Yager: Some aspects of intuitionistic fuzzy sets. *Fuzzy Optimization and Decision Making* **8** 67–90 (2009)

Acknowledgement: This work was supported by the GACR project No. 14-02424S.

Asymmetric behaviour of collateral constraint: applications on the Czech economy

Vlastimil Reichel[1], Miroslav Hloušek [2]

Abstract.

This paper analyses influence of asymmetric behaviour of collateral constraint through the lens of the DSGE model with occasionally binding constraint on housing wealth. The Czech economy data show positive but asymmetric behaviour between housing prices and consumption. It can be explained by following mechanism. When the prices of houses fall, the housing wealth decreases, which decreases consumption substantially as constraint on borrowing is binding. On the other hand, rise in house prices increases housing wealth, the constraint becomes slack and consumption increases only slightly. Such mechanism is incorporated into DSGE model that is estimated on data of the Czech economy using Bayesian techniques. The results from model simulation show that the occasionally binding constraint is able to replicate asymmetric behaviour of house prices and consumption found in the data.

Keywords: occasionally binding constraint, asymmetric behaviour, DSGE model, housing sector.

JEL classification: E370
AMS classification: 91B64

1 Introduction

Asymmetric behaviour is still central topic of macroeconomics and macroeconomic modelling. Specific attention has been devoted to behaviour of the variables connected with housing market, concretely housing prices and consumption. This issue was studied by Guerrieri and Iacoviello [2] who found asymmetric link between house prices and economic activity in USA data during housing boom and subsequent recession and explained it by occasionally binding collateral constraints on housing wealth. Their work was inspiration for our paper but we are using different model and apply it on Czech data.

First, we identified asymmetry on Czech data of housing prices and consumption. Then we estimated DSGE model with two collateral constraints using Bayesian methods.

This benchmark model assumes symmetric behaviour of housing prices and consumption for positive and negative house price shock which contradicted the data. After introduction of occasionally binding constraint that is binding for drop in housing prices, but slack for rise in housing prices, the model was able to replicate behaviour of Czech data.

2 Asymmetry of housing prices and consumption on Czech data

First, we look at behaviour of housing prices and consumption during the last fifteen years. Figure 5 shows scatter plot of the real house prices and consumption,[1] both time series expressed as year-on-year growth rates. The fitted curve is represented by following polynomial:

$$\Delta c = \underset{(0.003)}{0.028} + \underset{(0.035)}{0.232}\Delta q + \underset{(0.211)}{0.227}\Delta q^2 - \underset{(1.000)}{2.244}\Delta q^3 + \epsilon, \tag{1}$$

[1]Masaryk University, Lipová 41a, Brno, Czech Republic, reichel.v@mail.muni.cz
[2]Masaryk University, Lipová 41a, Brno, Czech Republic, hlousek@econ.muni.cz
[1]See section 5 for details about the data.

where Δc is growth rate of consumption and Δq is growth rate of the housing prices. Standard deviations are quoted in parenthesis. All the parameters are significant at 5 % except of parameter of Δq^2 that is not significant at conventional levels. The figure show asymmetry between the housing prices and consumption: consumption increases less then proportionally for the rise in the housing prices than for the drop in the housing prices.

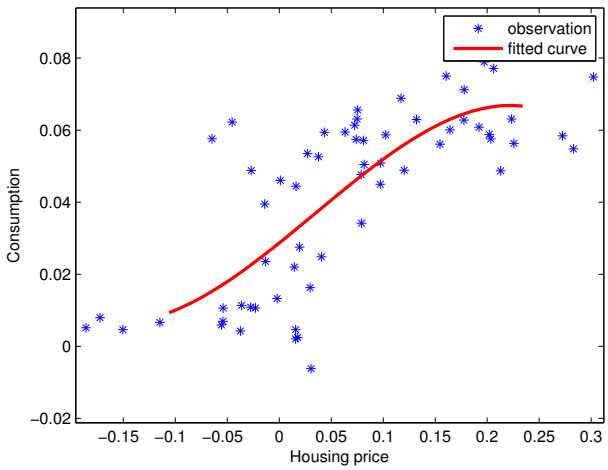


Figure 1 Asymmetry in data

3 Mechanism of asymmetry

The mechanism that is able to explain the asymmetrical behaviour is occasionally binding constraint on housing wealth. Let assume consumer who is credit constrained and the amount of the loan depends on the value of her house that serves as collateral. If there is a fall in housing prices, she gets less money through the loan and decreases consumption by the same amount. The credit constraint is still binding. This situation is described in left panel of Figure 2.

On the other hand, if the housing prices increases, the consumer is better off, she gets more money and increases her consumption. The rise in consumption is by smaller amount because now she can save something, the constraint becomes slack. This is depicted in the right panel of Figure 2. Based on this mechanism we can observe asymmetrical behaviour of consumption in reaction to changes of housing prices.

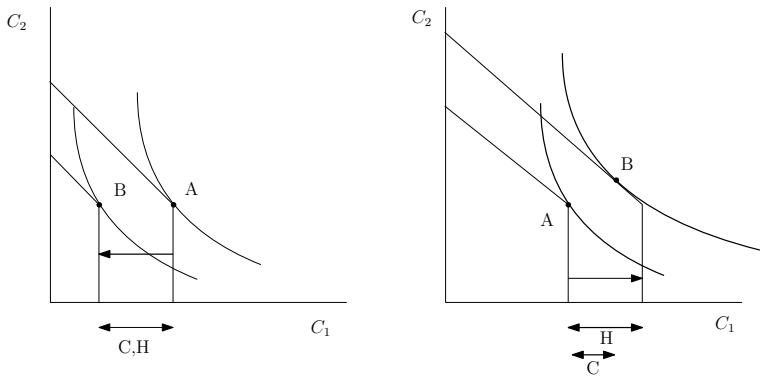


Figure 2 Illustration of occasionally binding constraint

4 Model

The model that we use is borrowed from Iacoviello [1] and ranks among small-scale models. We describe the model only verbally by help of Figure 3, for more details see the original paper. The model is composed of four sectors: households, entrepreneurs, retailers and the central bank. Households are divided into two types: patient and impatient. Patient households work, consume, accumulate housing and also own capital and supply funds to firms and to impatient households. Impatient households also work, consume and accumulate housing, but they are credit constrained. Entrepreneurs use labour, capital and real estates for production of intermediate goods. Entrepreneurs are credit constrained as well and the amount of their loans depends on the value of houses as in the case of impatient households. Retailers buy homogeneous intermediate goods, differentiate it by labelling and then sell it at monopolistically competitive market. They are price makers and there is nominal rigidity in the price setting which results in the Phillips curve. The central bank sets interest rate according Taylor rule with interest rate smoothing and weight only to inflation.

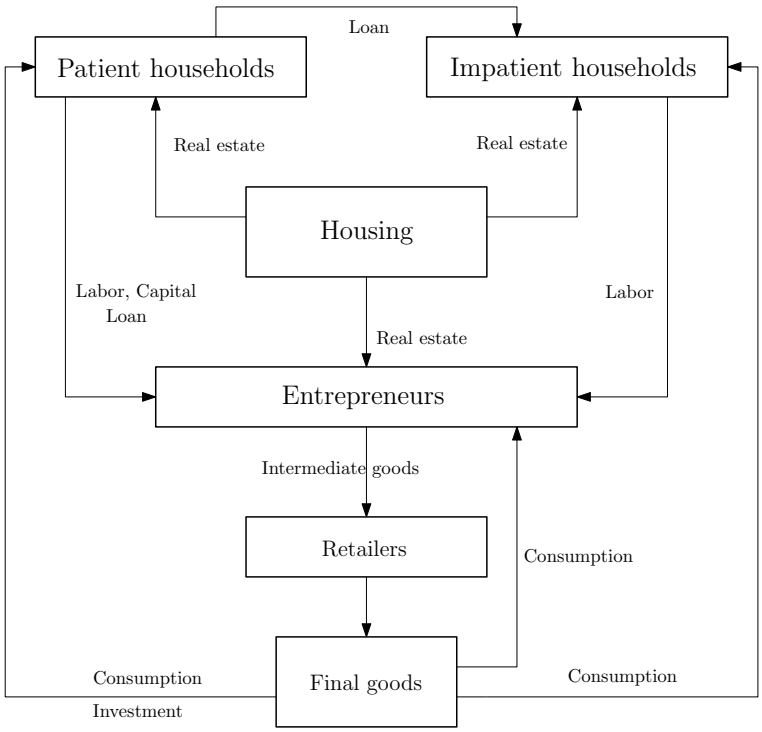


Figure 3 Model structure

5 Data and estimation

The model is estimated using data for following model variables: output Y_t, consumption C_t, investment I_t, real house prices q_t, inflation p_t and nominal interest rate R_t. The data were obtained from the Czech Statistical Office and the Czech National Bank databases and they cover period 1999:Q1 to 2014:Q1. Specifically, output is the gross domestic product (GDP), investment is gross fixed capital formation, consumption is measured by expenditures of households, interest rate is represented by 3M Pribor, inflation rate is q-on-q change of consumer price index (CPI) and the real house prices are represented by index of realized prices of flats deflated with CPI. Time series are all in logarithms, except of nominal interest rate. We use Hodrick Prescott filter (with $\lambda = 1600$, see Hodrick and Prescott [5]) to obtain the cyclical component of the variables.

Some of the model parameters are calibrated as in Hloušek [4]. The rest of the model parameters are estimated by Bayesian techniques. The prior means, standard deviations and posterior means together with 95 % confidence intervals of estimated parameters are quoted in Table 1. The setting of priors is based on Iacoviello [1] and Hloušek [4].

	Prior distribution			Posterior distribution		
Parameter	Density	Mean	S.D.	Mean	95% HPDI	
α	B	0.60	0.10	0.55	0.40	0.70
ψ	N	2.00	0.50	0.20	0.14	0.24
m_1	B	0.80	0.10	0.54	0.40	0.66
m_2	B	0.80	0.10	0.69	0.54	0.83
Monetary policy rule						
r_π	N	0.60	0.10	0.59	0.42	0.75
r_R	N	0.80	0.10	0.91	0.89	0.93
Persistence of shocks						
ρ_u	B	0.70	0.10	0.67	0.50	0.84
ρ_j	B	0.70	0.10	0.88	0.83	0.93
ρ_A	B	0.70	0.10	0.67	0.51	0.83
ρ_i	B	0.70	0.10	0.66	0.56	0.75
Volatility of shocks						
σ_R	IG	0.10	inf	0.002	0.001	0.002
σ_u	IG	0.10	inf	0.007	0.004	0.010
σ_j	IG	0.10	inf	0.454	0.278	0.622
σ_A	IG	0.10	inf	0.012	0.005	0.019
σ_i	IG	0.10	inf	0.031	0.027	0.036
σ_Y	IG	0.001	inf	0.051	0.037	0.065

Table 1 Prior a posterior distribution

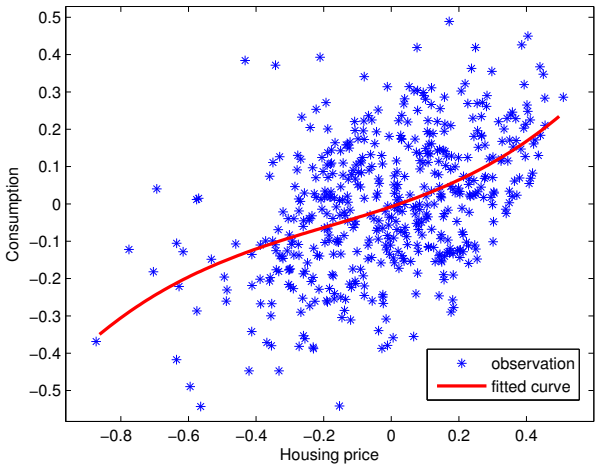


Figure 4 Simulation of the model with always binding constraint

The parameter α expresses labour share of patient households and its posterior mean is 0.55. It implies that the share of borrowing constrained households is 0.45. This proportion is in reverse order compared to Hloušek [4] who made the analysis on older data. Posterior mean of parameter ψ is 0.2 which shows very low capital adjustment cost. Parameters m_1 and m_2 are loan-to-value ratios for entrepreneurs and impatient households, respectively. It indicates that houses owned by impatient households are more easily collateralizable than entrepreneurial real estates, $m_2 = 0.69 > m_1 = 0.54$. Parameters of the monetary policy rule show high degree of interest rate smoothing ($r_R = 0.91$) and the quite standard weight to inflation ($1+r_\pi = 1.59$) which is comparable with other empirical studies for the Czech economy. Regarding the shocks, the most persistent shock is the housing preference shock. This shock is also the most volatile. This result is similar to Hloušek [4] and is caused by unusual boom and bust in the house

prices in recent years.

6 Results of simulation

This model assumes that the borrowing constraints (for both impatient households and entrepreneurs) are binding near the steady state and they are binding for decrease as well as increase in housing prices. This benchmark model was simulated in reaction to series of housing price shocks (five hundred replications). Result of this simulation together with fitted polynomial curve is depicted in Figure 4. We observe that there is not asymmetry that was found in the data and estimated parameters of equation (1) are not statistically significant (not shown here).

As a next step we introduce occasionally binding constraint mechanism that was illustrated in Section 3 and is modelled using *Occbin* toolbox developed by Guerrieri a Iacoviello [3]. This toolbox uses a piecewise linear perturbation method that is able to solve dynamic models which switches between two regimes: binding and not binding constraint. This method provides a good approximation of a dynamic programming solution that is often use for such non-linearities but is much faster even for models with many state variables.

First, we look at reaction of the model variables to positive and negative housing price shocks of magnitude of one standard deviation. These impulse responses are depicted in Figure 5. The dashed line represents linear solution, the solid line represents piecewise linear solution. In reaction to the fall of house price, the value of houses which serves as collateral decreases, therefore the mortgage loan decreases which lowers consumption and output. Both linear and piecewise linear solution coincides as the constraint is binding. In reaction to the rise of house price the size of mortgage loan increases, but to much less extent in the case of piecewise linear solution. The credit constraint is not binding here as consumers spread out consumption in time. This reaction is reflected in smaller increase of consumption and output. On the other hand, linear solution (with binding constraint) is mirror image of the previous case when house price fell.

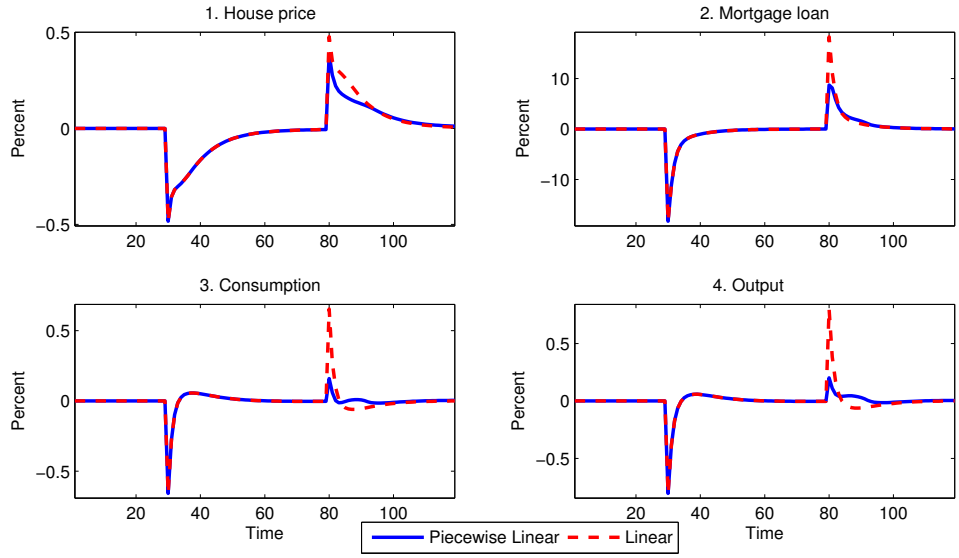


Figure 5 Impulse responses to housing price shock

The results of model simulation to the sequence of housing price shocks for the piecewise linear solution is depicted in Figure 6. The fitted curve already resembles the curve estimated for the data. The rise in house prices is accompanied by smaller increase in consumption, but for falling house prices the consumption drops almost proportionally. The estimated polynomial equation is as follows:

$$\Delta c = \underset{(0.006)}{0.026} + \underset{(0.032)}{0.269}\Delta q - \underset{(0.130)}{0.488}\Delta q^2 - \underset{(0.226)}{0.510}\Delta q^3 + \epsilon, \tag{2}$$

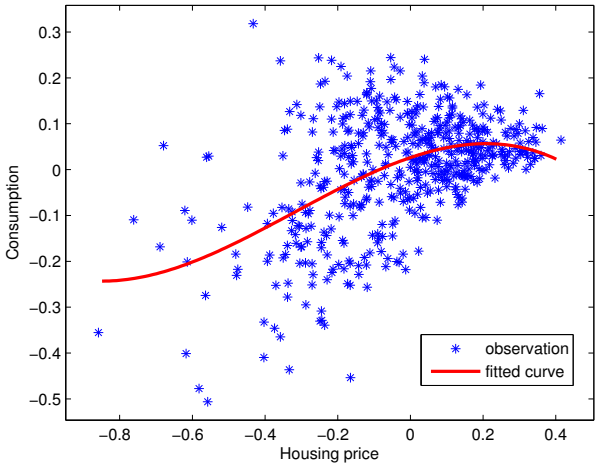


Figure 6 Simulation of the model with occasionally binding constraint

All the parameters are statistical significant, have corresponding signs, but partly differ in size from estimation made on data. Nevertheless, we are able to capture asymmetrical pattern in behaviour or consumption and housing prices.

7 Conclusion

In this paper we tried to explain asymmetrical relationship between housing prices and consumption by occasionally binding constraint on housing wealth. This mechanism was applied on DSGE model with collateral constraint. The model was estimated on Czech data and then it was subject to series of simulations in reaction to house price shocks. The credit constraint was binding for decrease but relaxed for increase in the housing prices. This mechanism produced asymmetrical behaviour of consumption that was not present in the benchmark model with always binding constraint.

The topic for further research is to extend the model for open economy that will be more suitable for the Czech economy conditions.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/1121/2014 at Masaryk University.

References

[1] Iacoviello, M.: House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle. In: *American Economic Review*, 2005, 739-764.

[2] Guerrieri, L., and Iacoviello, M.: Collateral Constraints and Macroeconomic Asymmetries. In: *Working paper*, July 2014.

[3] Guerrieri, L., and Iacoviello, M.: Occbin: A Toolkit to Solve Models with Occasionally Binding Constraints Easily. In: *Journal of Monetary Economics*, 2015, 22-38.

[4] Hloušek, M. DSGE model with collateral constraint: estimation on Czech data, In Ramík, J., Stavárek, D. *Proceedings of 30th International Conference Mathematical Methods in Economics.* Karviná: Silesian University, School of Business Administration, pp. 296-301, 2012.

[5] Hodrick, R. J., and Prescott, E. C.: Postwar US business cycles: an empirical investigation, *Journal of Money, Credit, and Banking*, 1997, 1-16.

Heterogeneity and efficiency of meat processing companies in the Czech Republic

Tamara Rudinskaya[1]

Abstract In this paper different ways of considering heterogeneity in stochastic frontier models based on data about Czech meat processing farms are studied. Recent studies proved that food processing firms' heterogeneity can influence the result of efficiency estimation. It is possible to take heterogeneity into account by including those effects in the mean and/or variance of distribution of inefficiency (observed heterogeneity) or by randomizing some parameters of the model (unobserved heterogeneity). For empirical study panel data set from the years 2005–2012 contained 757 observations of 130 meat processing companies was used. All models were estimated in the form of Translogarithmic production function. The models with observed heterogeneity (heterogeneity in variance and Equality Constrained Scaling Model) and unobserved heterogeneity (True Random Effects and Random Parameters Model) were compared to Aigner, Lowell and Schmidt (ALS) model without heterogeneity by statistical tests. Our results indicate that all estimated technical efficiency coefficients are higher than in the case of ALS model.

Keywords: Stochastic Frontier Analysis, heterogeneity, Czech Republic, meat processing companies

JEL Classification: D24
AMS Classification: 90B30

1 Introduction

The literature concerning different methods of efficiency measuring has increased rapidly during the last decade. There are two main ways to model frontier efficiency, namely the Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). In this paper we deal with modelling firm specific heterogeneity using SFA.

For capturing heterogeneity traditional models of SFA have been extended. Initially, when the heterogeneity accounting studies started to develop it was assumed that time invariant parts in the models represent inefficiency, and time variant parts can be seen as firm or unit specific heterogeneity. However, this opinion has entirely changed (see e.g. [5]). In recent papers, not changing in time part, are mainly assumed as firm specific heterogeneity while the time variant part is considered as inefficiency. Which one of these interpretations is right is not a simple question. It is obvious that there are firm specific heterogeneity factors which do not change in time and which are beyond the managerial control. These should of course be interpreted as time invariant heterogeneity [7].

Both approaches (DEA and SFA) assume that firms are not heterogeneous but inefficient, since all inefficiency scores are estimated by assuming a homogeneous technology available to all producers. This suggests that the impact of inefficiency in the agriculture often is overestimated [6].

Technical efficiency analysis in view of heterogeneity of food processing industry in the Czech Republic was conducted by Čechura and Hockmann [2,3]. In these research papers authors dealt with unobserved heterogeneity using production function and True Random effects and Random Parameters Model (in 2011) and output distance function and True Random effects model (in 2014). Authors concluded that intersectoral heterogeneity and heterogeneity among firms are an important characteristic of EU food processing industry. Kroupová [8] analysed technical efficiency of ecologic farm in the Czech Republic using Pitt and Lee Random effects model with heteroscedasticity and heterogeneity.

The purpose of this paper is to study the different ways how the firm specific observed and unobserved heterogeneity can be taken into account using SFA.

[1] CULS, Department of Economics, e-mail: rudinskayats@mail.ru

2 Materials and methods

2.1 SFA and heterogenity

It is possible to take heterogeneity into account by including those effects in the mean and/or variance of the distribution of inefficiency (observed heterogeneity) or by randomizing some parameters of the stochastic frontier model (unobserved heterogeneity).

Observed heterogeneity. Observed heterogeneity can be introduced into the specification by several methods. A common approach is incorporating of vector of variables $\boldsymbol{z_i}$ that contains the information about heterogeneity directly into a model. In this case $\boldsymbol{z_i}$ appears to be a goal function itself.

$$y_i = \beta' x_i + \alpha' z_i + v_i - u_i \tag{1}$$

Two other methods of introducing observed heterogeneity in the frontier model are in the variance parameter and in the mean of the underlying inefficiency.

The first estimated model is the normal- truncated normal model with heterogeneity in variance $Var[u_i] = \sigma_u^2 \exp(\delta' z_i)$[5]. The second model is Equality Constrained Model [1] was used for accounting for heterogeneity in variance and mean. The mean and standard deviation of the underlying truncated normal variable $\boldsymbol{u_i}$ are scaled by the same exponential function of the data[2].

$$\mu_i = \mu * \exp(\delta' z_i) \text{ and } \sigma_{u,i} = \sigma_u * \exp(\gamma' z_i) \tag{2}$$

Unobserved heterogeneity. During the past two decades there was a development of various forms of econometric methods, which can, especially on panel data, identify the unobserved heterogeneity. Unobserved heterogeneity can be taken into account by randomising some of the parameters of the model in which case it is assumed that this randomisation captures all time invariant unobserved heterogeneity. Among the models that are able to introduce unobserved heterogeneity belong: True Fixed and Random Effects Model [4], Random Parameters Model [4], Fixed-Management Model [1]. The True Random and Random Parameter Model will be estimated in this paper.

"True" random effects model. Green [4] argued that the random effects model has three significant weaknesses. The first is its implicit assumption that the effects are not correlated with the included variables. The second problem with the random effects is its implicit assumption that the inefficiency is the same in every period. For a long time series data, this is likely to be an undesirable assumption. The third shortcoming of this model is that in this model $\boldsymbol{u_i}$ carries both the inefficiency and, in addition, any time invariant firm specific heterogeneity.

To avoid the former limitations Green [4] proposed "True" random effects model that is as follows:

$$y_{it} = \alpha + \beta' x_{it} + w_i + v_{it} \pm u_{it} \tag{3}$$

where w_i is the random firm specific effect and $\boldsymbol{v_{it}}$and $\boldsymbol{u_{it}}$ are the symmetric and one sided components specified earlier [4].

Random Parameters Model. In this model, heterogeneity is captured by an unobservable variable that is simulated by suitable estimation procedures. A general form of the random parameters stochastic frontier formulation may be written as [4]:

$$y_{it} = \boldsymbol{\beta_i' x_{it}} + v_{it} - u_{it}, \qquad u_i = |N[\mu_{it}, \sigma_{uit}^2]|, \quad \mu_{it} = \boldsymbol{\delta_i' m_{it,}} \quad \sigma_{uit}^2 = \sigma_u^2 * exp(\gamma_i' w_{it}) \tag{4}$$

The model allows parameter values to vary across the population according to some pre-specified distribution. The model allows, all at once, half normal and truncated normal distribution for $\boldsymbol{u_i}$ and firmwise and/or timewise heteroscedasticity in $\boldsymbol{u_{it}}$.

2.2 Data set

The panel data set was collected from the Albertina database. The database contains all registered companies and organisations in the Czech Republic. The analysis uses information from the final accounts of companies whose main activity is food processing in the period from 2005 till 2012. After the cleaning process (removing companies with missing observations and negative values of the variables), the unbalanced panel data set contains 757 observations of 130 meat processing companies. The data set represents 14 regions of the CR.

[2] Originally Alvarez, Amsler, Orea, and Schmidt proposed scaling of the standard deviation and mean of $\boldsymbol{u_i}$ by linear function of the data. Being skeptical of the linear scaling of the variance, Greene proposed exponential form instead.

In the analysis were used following variables: Output, Labour, Capital and Material. Output is represented by the total sales of goods, products and services of the food processing company. In order to avoid inflation changes, Output was deflated by the price index of food processing companies according to the branch. (2005=100). The Labour input is used in the form of total personnel costs per company, divided by the average annual regional wage. The Capital variable is represented by the value of tangible assets. Material is total costs of material and energy consumption per company. Capital and Material were deflated by the price index of industrial sector (2005=100).

For estimation maximum simulated likelihood method in the computer program NLOGIT version 4.0 – LIMDEP version 9.0 was used.

3 Results and discussion

For estimation production function in the Translog (transcendental logarithmic) form was used.

The three factor translog production function is:

$$\ln(Y) = \ln(A) + \alpha_K \ln(K) + \alpha_L \ln L + \alpha_M \ln M + \beta_{KK} \ln(K) \ln(K) + \beta_{LL} \ln(L) \ln(L) + \beta_{MM} \ln(M) \ln(M) + B_{KL} \ln(K) \ln(L) + \beta_{KM} \ln(K) \ln(M) + \beta_{LM} \ln(L) \ln(M) = f(K, L, M) \quad (5)$$

where A = total factor productivity, L = labor, K = capital, M = material, and Y = output.

Parameters		**Default**[3]		**Heterogeneity in variance of u**[4]**(HV)**		**Equality Constrained Scaling Model (ES)**	
		coefficient	z-value	coefficient	z-value	coefficient	z-value
First-order effects	Const.	0.1968***	10.163	0.1808***	5.793	0.1476***	5.989
	T	0.0043	0.683	0.0134*	1.861	0.0021	0.316
	K	0.0671***	5.835	0.0816***	4.657	0.0275**	2.397
	L	0.1872***	8.129	0.1236***	3.997	0.2684***	12.578
	M	0.7202***	48.791	0.7207***	41.730	0.6292***	51.069
Second – order effects	TT	-0.0087**	-2.003	-0.0113*	-1.772	-0.0101**	-2.356
	KK	0.0203*	1.771	0.0183	1.134	0.0419***	3.549
	LL	0.0751***	2.793	0.0623*	1.707	0.2322***	6.592
	MM	0.1531***	16.129	0.1645***	12.105	0.2301***	21.598
	KT	-0.0029	-0.543	0.0004	0.048	0.0021	0.361
	LT	0.0035	0.412	0.0038	0.292	-0.0073	-0.720
	MT	0.0053	1.026	-0.0021	-0.280	0.0096	1.561
	KL	0.0017	0.111	0.0184	0.868	-0.0601***	-3.556
	KM	0.0458***	-5.617	-0.0478***	-4.395	0.0056	0.742
	LM	-0.0669***	-4.305	-0.0742***	-3.855	-0.1789***	-11.815
Parameters of variance (mean) in $\boldsymbol{u_{it}}$[5]	Const.			-0.4432***	-3.728	11.7392	-0.211
	D2[6]			-1.6053***	-12.407	-1.0821***	-8.717
	D3			-2.3297***	-7.836	-1.4968***	-7.623
	D4			-1.8249***	-6.018	-1.0674***	-5.390
	D5			-1.9631***	-7.084	-1.2147***	-6.997
λ		3.1634***	15.481	2.5892***		37.6958***	
Sigma		0.4718***	911.410	0.4607		11.7399	
Log likelihood		-131.5504		-93.4344		18.4944	

Table 1 Parameters estimation of model with observed heterogeneity

Note: ***, **, * denote significance at the 1%, 5% and 10% level.
Source: own calculations

[3] The default specification is Aigner, Lovell and Shmidt's canonical normal-half-normal model in the form $y = \beta' x + v + u, u = |U|$

[4] The variance of $\boldsymbol{u}$ is exponential

[5] For Equality Constrained Scaling Model heterogeneity effects are included into the variance and the mean of inefficiency component $\boldsymbol{u_{it}}$, and besides the estimated parameters (except constant parameter) are equal.

[6] D1, D2, D3, D4, D5 are dummy variables that represent firm's size according to the number of employees. D1: less than 10 employees, D2: 11-50, D3: 51-150, D4: 151-250, D5: more than 251 employees. Variable D1 was not included into the model to avoid multicollinearity.

In the estimation, we normalized all variables in logarithm by their sample mean. This has the advantage that the first-order parameters can be interpreted as elasticity of outputs at sample mean.

The estimated production elasticities (Table 1) satisfy the assumption of monotonicity and quasiconcavity, i.e., the elasticities are positive and the diminishing marginal productivity for each input was estimated ($\beta_{rr} + \beta_r^2 - \beta_r < 0$), for *r = K, L, M)* for *Default* and *HV* model. *ES* model does not satisfy the assumption of quasiconcavity for production factor *Capital* and *Labour*.

Since the values of production factors were normalised by their arithmetic means, in translogarithmic model these coefficients denote the variation or possible percentage change in aggregate output as a result of one percent change in the input that is production, or output, elasticity. The highest elasticity in all models displays production factor Material. The parameter $\boldsymbol{\lambda}$ is the relation between the variance of $\boldsymbol{u_{it}}$ and $\boldsymbol{v_{it}}$. Thus, the parameter indicates the significance of technical inefficiency in the residual variation. A value larger than one suggests that variation in $\boldsymbol{u_{it}}$ prevails the variation in the random component $\boldsymbol{v_{it}}$.

Technical change has positive impact on production in the case of HV model (the variable *Time* is positive and significant at 10% level of significance).

Heterogeneity was introduced into the models as a group size dummy variable. The estimated parameters of these variables are significant in the case of both models, what confirms the presence of heterogeneity in the data.

Parameters		**Default (D)**		**True Random Effects Model (TRE)**		**Random Parameter Model (RPM)**	
		Coefficient	Z-value	Coefficient	Z-value	Coefficient	Z-value
First-order effects	Const.	0.1968***	10.163	0.0286*	1.772	0.1169***	7.967
	T	0.0043	0.683	0.0182***	3.826	0.0203***	5.469
	K	0.0671***	5.835	0.0764***	8.710	0.0385***	5.603
	L	0.1872***	8.129	0.1207***	9.395	0.2586***	24.351
	M	0.7202***	48.791	0.8074***	90.374	0.7284***	98.622
Second – order effects	TT	-0.0087**	-2.003	-0.0088	-1.573	-0.0091***	-2.908
	KK	0.0203*	1.771	-0.0029	-0.330	0.0219***	3.177
	LL	0.0751***	2.793	0.1006***	6.225	0.1823***	13.750
	MM	0.1531***	16.129	0.1732***	25.042	0.2056***	40.258
	KT	-0.0029	-0.543	-0.0008	-0.164	-0.0048	-1.301
	LT	0.0035	0.412	0.0017	0.171	0.0065	1.061
	MT	0.0053	1.026	-0.0039	-0.755	-0.0034	-0.990
	KL	0.0017	0.111	0.0303***	3.590	-0.0097	-1.335
	KM	0.0458***	-5.617	-0.0535***	-15.949	-0.0112***	-3.087
	LM	-0.0669***	-4.305	-0.0902***	-8.965	-0.1881***	-23.161
λ		3.1634***	15.481	1.9628***	25.279	3.8765***	9.520
Sigma		0.4718***	911.410	0.3227***	74.092	0.3096***	49.959
Log likelihood		-131.5504		-50.6689		-13.1201	

Table 2 Parameters estimation of model with observed heterogeneity

Note: ***, **, * denote significance at the 1%, 5% and 10% level.
Source: own calculations

The first-order estimated parameters *Capital, Labour, Material* are significant under z-test at 1% level of significance in both cases. Signs of the coefficients are positive that is consistent with economic theory (the assumption of monotonicity is fulfilled). The curvature condition of quasiconcavity in inputs is achieved in the case of all production factors. That is, the estimates are consistent with the economic theory (at least at the sample mean).

The highest elasticity displays production factor Material. The parameter λ is more than one indicates presence of inefficiency in the data. Technical change has positive impact on production (the variable *Time* is positive and significant at 1% level of significance in both cases). It is characterised by Labour-intensive, and Capital- and Material-saving behaviour.

Indicator	ALS	HV	ES	TRE	RPM
Mean	0.7606	0.7752	0.8746	0.8206	0.8119
St.dev.	0.1178	0.1234	0.1366	0.0866	0.0971
Min	0.4528	0.3765	0.4666	0.1505	0.2237
Max	0.9793	0.9763	0.9801	0.9791	0.9773
Return to scale	0.9745	0.9259	0.9251	1.0045	1.0255

Table 3 Statistics of efficiency scores

Source: own calculations

All estimated technical efficiency coefficients are higher than in the case of Aigner, Lowell and Schmidt model (ALS). The highest technical efficiency was estimated using ES model, the lowest displays HV model. Models with observed heterogeneity demonstrate estimation of slightly diminishing return to scale, whereas the models with unobserved heterogeneity represent increasing return to scale.

Indicator	ALS	HV	ES	TRE	RPM
ALS	1				
HV	0.8793	1			
ES	0.6736	0.8860	1		
TRE	0.5326	0.4850	0.3713	1	
RPM	0.8335	0.7459	0.5850	0.4834	1

Table 4 The Spearman correlations of the inefficiency rankings

Source: own calculations

The correlation matrixes based on Spearman's correlation test between the ranks obtained by the inefficiency results from different models are presented in Table 5. The correlation among the RE model and its extended form HV are positive and close to 1 which indicates that these models rank the firms very similarly. The high correlation, besides that, display models HV and ES, ALS and RPM. The correlation of inefficiency intimates of TRE and other models is low, especially in the case of ES model.

Testing of the models (goodness-of-fit tests)

For alternative model selection, Akaike (AIC) and Bayesian Information Criterion (BIC) can be used. The model with least information loss (least AIC and BIC) is assumed to be the null model. The results show that the Equality Constrained Scaling Model (ES) and Random Parameters model (RPM) have lower AIC and BIC criterion, therefore fit the data better than other models.

Criterion	ALS	HV	ES	TRE	RPM
AIC	0.3909	0.3011	0.00923	0.1838	0.0928
BIC	0.4946	0.4292	0.14335	0.2938	0.2273

Table 5 Testing parameters of the model

Source: own calculations

According to AIC and BIC criteria the Equality Constrained model and Random Parameters model fit the data best.

4 Conclusions

The main interest of this paper was to study different ways how the firm specific heterogeneity can be taken into account in the SFA. We look at the potential advantages of heterogeneity extended stochastic frontier models over conventional models in technical efficiency measurement.

Our data consists of 130 meat processing companies which are assumed to vary by their size. Observed heterogeneity was taken into account by dummy variables representing firm's size. For empirical purpose model with heterogeneity in variance and Equality Constrained Scaling Model (heterogeneity in mean and variance parameter) were estimated. As the models representing unobserved heterogeneity "True" Random Effects Model and Random Parameters Model were estimated.

Our basic result is that all estimated technical efficiency coefficients in the case of observed and unobserved heterogeneity are higher than in the case of Aigner, Lowell and Schmidt model (ALS). The highest technical efficiency was estimated using Equality Constrained Scaling Model, the lowest displays the model with heterogeneity in variance parameter. Models with observed heterogeneity demonstrate estimation of slightly diminishing return to scale, whereas the models with unobserved heterogeneity represent increasing return to scale.

The correlation among the Aigner, Lovell and Shmidt model and the model with heterogeneity in variance parameter are positive and close to 1 which indicates that these models rank the firms very similarly. The high correlation, besides that, display models HV and ES, ALS and RPM. The correlation of inefficiency intimates of TRE and other models is low, especially in the case of ES model.

References

[1] Alvarez, A., Amsler, C., Orea, L., Schmidt, P.: Interpereting and testing the scaling property in models where inefficicency depends on firm characteristics. *Journal of productivity analysis* **25**, 3 (2006), 201-212.

[2] Čechura, L., Hockmann, H.: Heterogeneity in Technology and Efficiency – Specifics of the Food Processing Industry in the Visegrád Countries. In: *Paper prepared for presentation at the EAAE 2014 Congress 'Agri-Food and Rural Innovations for Healthier Societies'* August, 26-29, 2014 Ljubljana, Slovenia.

[3] Čechura L., Hockmann H.: Efficiency and heterogeneity in Czech food processing industry, In: *EAAE 2011 Congress, Change and Uncertainty Challenges for Agriculture*, Zurich, Switzerland.

[4] Greene, W., *Fixed and Random Effects in Stochastic Frontier Models*. Journal of Productivity Analysis, 23, 2005, pp.7–32

[5] Greene, W.: *Econometric Analysis*, 7th edition. Boston: Pearson Education, Inc., 2011, p.297. ISBN 10: 0-13-139538-6.

[6] Hockmann H., Pieniadz A., *Farm Heterogeneity and Efficiency in Polish Agriculture: A Stochastic Frontier Analysis.* 12th Congress of the European Association of Agricultural Economists – EAAE 2008.

[7] Kopsakangas-Savolainen, M., Svento, R.: Observed and unobserved heterogeneity in stochastic frontier models. Working Paper No. 0805 August, 2008, ISSN 1796-7961.

[8] Kroupová, Z.: Technická efektivnost ekologického zemědělství České republiky. *Ekonomická revue ER-CEREI* 13 (2010), 63–75.

Regional Input-Output Models: Assessment of the Impact of Investment in Infrastructure on the Regional Economy

Jaroslav Sixta[1], Jakub Fischer[2]

Abstract. Regional input-output tables provide detailed description of economic linkages between producers. Usually, official statistical authorities do not publish regional input-output tables with some exceptions (like the U.S. Bureau of Economic Analysis). It means that pure regional input-output tables are very scarce and re-gional input-output tables are usually compiled for groups of countries. We have been working on the construction of regional input-output tables for the regions ("kraje") of the Czech Republic since 2013. These tables are based on national supply and use tables from the Czech national accounts for 2011. Computation of regional tables has to be based on many simplifications. Models based on regional input-output tables provide improved source of information in comparison with standard models based on national input-output tables. The main principles are based on Leontief's approach with regional multipliers. The paper is focused on a common situation of the impact of increased investments into infrastructure on the regional economy. Besides, the case of dependency between the regions is discussed and main issues are tackled and explained.

Keywords: Input-Output, GDP, regional, models.

JEL classification: C67, R10, O11
AMS classification: 665C20

1 Introduction

Input-Output Analysis (IOA) represents reborn scientific discipline that is going to be widely used among different group of users. IOA has been used for many years in the West but after the boom of affordable computers in 1990s it became more common. There can be found many analyses focused on economy, environment, energy, etc. Input-Output became also popular in the Czech Republic after 2000. Input-output analysis lost its unearned stigma of a simple planning tool used mainly in the centrally planned economy. One of the specific usages of Input-Output Analysis is the regional analysis, [10]. Usually, regional data are very hard to obtain especially when some detailed structures are needed, [3].

Regional Input-Output Tables (RIOTs) compiled for groups of regions within the country are very scarce, [6] and [9]. The basis for the Regional Input-Output Analysis is formed by specific technical coefficients, see Miller and Blair [5]. These coefficients are available for the US, see [1]. Usually, RIOTs can be obtained for group of countries of some blocks. They are also available for China or Spain. These tables were being compiled for Finland but the process was stopped, see [12]. Regional Input-Output Tables for the Czech Republic were prepared within the 3 years long project that is going to be finished by December 2015. These tables are prepared for all 14 regions of the Czech Republic ("Kraje"), NUTS 3. Since RIOTs should reflect regional differences, they are more usable for specific regional models like the impact of investments in the region, organisation of different events, etc.

Our paper illustrates the difference in the use of regional and national tables. It is illustrated on the investments into public infrastructure in the regions Praha and Moravskoslezsky kraj. The impact

[1]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, sixta@vse.cz
[2]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, fischerj@vse.cz

on regional output, value added and import is estimated.

1.1 Methodology of Experimental Regional Input-Output Tables

Before the process of computation of RIOTs is started, main methodological issues have to be solved in line with [8]. Among all of them, the issue of consistency between official national Input-Output Tables has to be solved. The problem lies in the identification of local kind-of-activity units (LKAU), see [2] and [4]. Official national accounts published by the Czech Statistical Office do not distinguish LKAU for RIOTs correctly[1]. Finally, it has be carefully assessed if users appreciate rather methodological purity than consistency with main macroeconomic aggregates. With respect to this, we decided to design our RIOTs fully compatible with official data and research done in [7]. Since the principles of LKAU are not fully in line with the methodology, it has its own consequences on the interpretation of results. Subsequently, the model of RIOT has to be selected, as well. Our RIOTs are designed as balanced product by product input-output tables. Alternative approach lies in the computation of regional multipliers only. Finally, RIOTs are more valuable when they are split between imported and regionally produced products. Our RIOTs have this breakdown and therefore the import dependency of the region can be estimated, [11].

Regional Input-Output Tables were prepared for the Czech Republic for 2011 at basic prices. These 14 sets of symmetric product-by-product tables are broken down by 82 products of CZ-CPA. The sum of all components except imports and exports of all 14 tables is equal to national aggregates. Imports and exports are not consolidated since on the regional level these indicators consist from two parts. The first part is given by the international trade with goods and services. This refers to the national foreign trade. The second part is given by the inter-regional trade. Therefore export and import within regional input-output tables is inflated in comparison with national symmetric input-output tables.

2 Regional Input-Output Analysis

There are lots of applications of Input-Output Analysis ranging from simple static models to advanced dynamic models or models using some data or principles like CGE or DSG. Very often the input-output analysis is used for the evaluation of a particular event with specific respect to a target region. We illustrate the advantages of using regional input-output multipliers resulting from RIOTs. For simplicity, we selected two regions, Praha and Moravskoslezsky kraj. The comparison is provided with the country average. Regional specifics can be described by technical coefficients (input coefficients) defined as share of intermediate consumption on output, see following table 1.

Product	Czech Republic	Praha	Moravskoslezsky kraj
Agriculture	1.5%	0.9%	1.1%
Mining	2.6%	1.9%	3.1%
Manufacturing	26.2%	13.8%	32.4%
Energy	3.6%	4.1%	4.2%
Construction	5.1%	6.3%	4.2%
Services	24.7%	36.2%	20.2%

Table 1 Regional technical coefficients, %

Figures in table 1 are aggregated for illustrative purposes. It is clear that different technical coefficients refer to the input structure of output. The difference between the regions of Praha and Moravskoslezsky kraj is quite significant. The problem lies mainly in the difference between Prague and all other regions. It seems that capital city influences the analysis.

Suppose the increase of investments into public infrastructure by 15 000 mil. CZK. It is adequate to a 15-20 km of new highway or significant reconstruction of existing one. For the illustration, simple static input-output model is used as following:

[1]The problem is described in http://apl.czso.cz/pll/rocenka/rocenkaout.dod_uziti?mylang=EN

$$x = (I - A)^{-1} y, \tag{1}$$

where

x	output vector,
y	vectors of final use,
$(I - A)^{-1}$	Leontief inverse.

The impact of the increase of gross fixed capital formation is measured by the change of output. For these purposes, input-output tables for domestic production are used.

$$\Delta x = (I - A)^{-1} \Delta y, \tag{2}$$

Investment shock is measured by tables for domestic production where $y = 15\,000$ CZK mil, according to formula 2. The results are computed at the full level of 82 products. The resulting impacts on account of goods and services are illustrated in table 2.

Indicator	**Czech Republic**	**Praha**	**Moravskoslezsky kraj**
Output	0.3%	1.3%	3.1%
net taxes on product	0.1%	0.4%	0.9%
Imports	0.1%	1.0%	1.2%
Resources	0.3%	1.3%	2.5%
Intermediate consumption	0.4%	1.6%	3.3%
Final consumption expenditures	0.0%	0.0%	0.0%
of which households	0.0%	0.0%	0.0%
Gross capital formation	1.6%	6.5%	16.1%
of which fixed capital	1.6%	6.4%	16.3%
changes in inventories	0.0%	0.0%	0.0%
Exports	0.0%	0.0%	0.0%
Uses	0.3%	1.3%	2.5%

Table 2 Results of regional input-output analysis, %

The impact of the increase of gross fixed capital formation of 15 000 CZK mil is different between the regions. It means that investments into public infrastructure have different consequences depending on the regional specifics. When using average multipliers for the Czech Republic, the overall output will increase by the 32.5 billion CZK (0.3%). In Moravskoslezsky kraj, the increase of output is the most significant with about 32.3 billion CZK (3.3%) in comparison with Praha (32.1 billion CZK, 1.3%). The different impact is also on the imports. The most significant increase of imports is observed for Moravskoslezsky kraj. It refers to the availability of the production of necessary inputs to the construction industry. The simple static input-output analysis provides only the first step after the investment shock. In the reality, the shock will be split into several periods with many multiplication effects. More advanced models including also impacts on household consumption and further investments generated by additional operating surplus are used in practice.

Effects of investment shock should be analysed also from the perspective of employment. This will be available by December 2015. The investment shock has different impact on the structure of output and on the requirements for imports. Table 3 illustrates regional demands for imports when investing 15 billion CZK into public infrastructure.

3 Regional Composition of Value Added

Even simple static Input-Output Analysis cannot provide fully adequate estimates of gross value added, the structure of affected industries shows specifics of regional links. It seems that the impacts of the investments shocks should be identical in all regions but the results have to be influenced by the input

Product	Czech Republic		Praha		Moravskoslezsky kraj	
Agriculture	7.4	0.0%	5.4	0.0%	9.9	0.1%
Mining	56.6	0.4%	74.6	0.5%	11.3	0.1%
Manufacturing	542.5	3.6%	368.2	2.5%	658.9	4.4%
Energy	94.7	0.6%	143.6	1.0%	92.1	0.6%
Construction	3 724.0	24.8%	4 668.6	31.1%	3 520.1	23.5%
Services	880.2	5.9%	1 001.9	6.7%	853.3	5.7%
Total	5 305.5	35.4%	6 262.3	41.7%	5 145.5	34.3%

Table 3 Regional demand for inputs, CZK billion, %

structure of regional output. The increase of investments by 15 billion CZK will be differently reflected in Czech regions. Figure 1 illustrates different impact on Praha, Moravskoslezsky kraj and the Czech Republic. Results for Moravskoslezsky kraj are close to the country average but the structure of the region Praha is very different. In Praha, the most significant difference is found for the output of services.

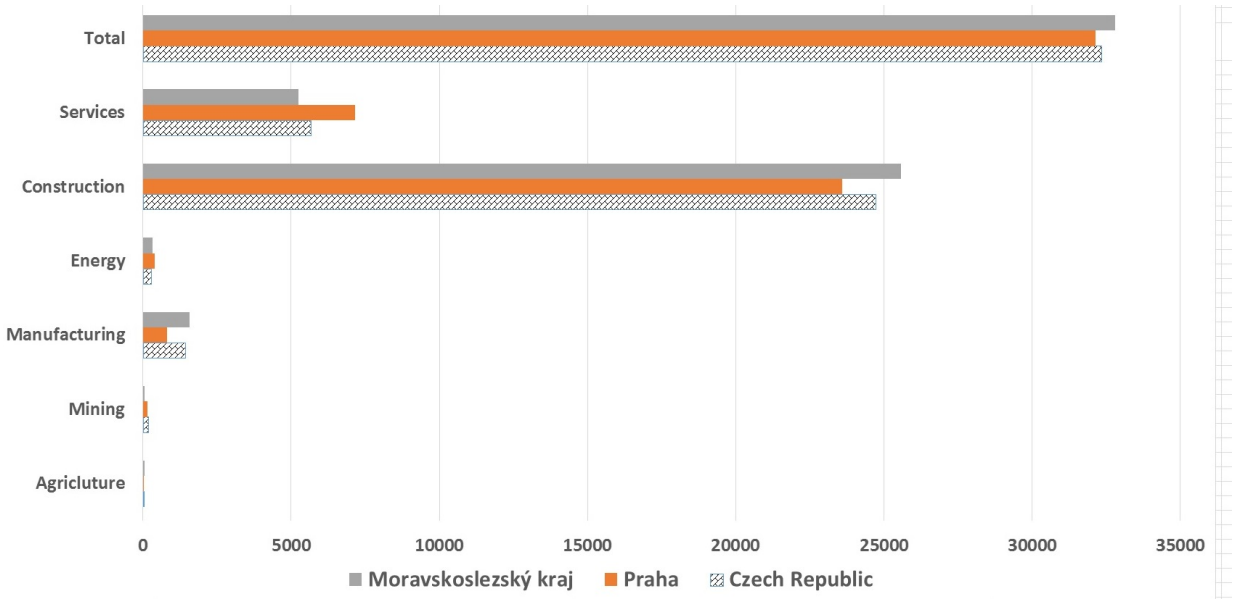


Figure 1 Change of the output structure

Beside output structure and imports, the size and structure of gross value added is different region by region. It is given mainly by the availability of regional producers to be able to satisfy demands for specific inputs of construction industry. If the region is not available to satisfy such demands, the inputs (both materials and services) will be imported. The assumptin that is behind the simple static input-output analysis is the stability of a price level, at least for a short period. When analysing the investments, the highest increase of gross value added is in Moravskoslezsky kraj (9 470.3 CZK mil), the increase in Praha (8 470.3 CZK mil) is lower than in average (9 404.1 CZK mil), see table 4.

The relative impact of the external investment shock is different region by region. It means that regional economy can be stimulated by mix of policies on demand. There are two effects. The first is connected with the advantages of regional economy and the second is connected with incentives for goods and services that are not available and have to be imported. Detailed economic analysis is not a target of our paper. The mix of economic policies working in the same way as infrastructure construction can contain promotion of small business, education, regional taxes, etc.

Anyway, 1 CZK invested into infrastructure immediately increase regional gross value added by 0.627 CZK in the Czech Republic. The highest value is observed in Moravskoslezsky kraj where it reaches 0.631 CZK, the lowest value is in Praha, 0.565 CZK. The composition of additional gross value added created by the 1 CZK investment is illustrated in figure 2.

Product	Czech Republic		Praha		Moravskoslezsky kraj	
Agriculture	17.5	0.02%	4.9	0.27%	12.1	0.27%
Mining	72.9	0.16%	1.9	−0.22%	28.2	0.12%
Manufacturing	405.9	0.05%	169.0	0.39%	368.2	0.36%
Energy	88.4	0.05%	83.9	0.26%	110.8	0.56%
Construction	6 502.7	2.48%	5 538.1	10.93%	6 662.9	25.80%
Services	2 316.7	0.11%	2 672.4	0.37%	2 288.1	1.29%
Total	9 404.1	0.27%	8 470.3	0.99%	9 470.3	2.68%

Table 4 Regional gross value added, CZK billion, %

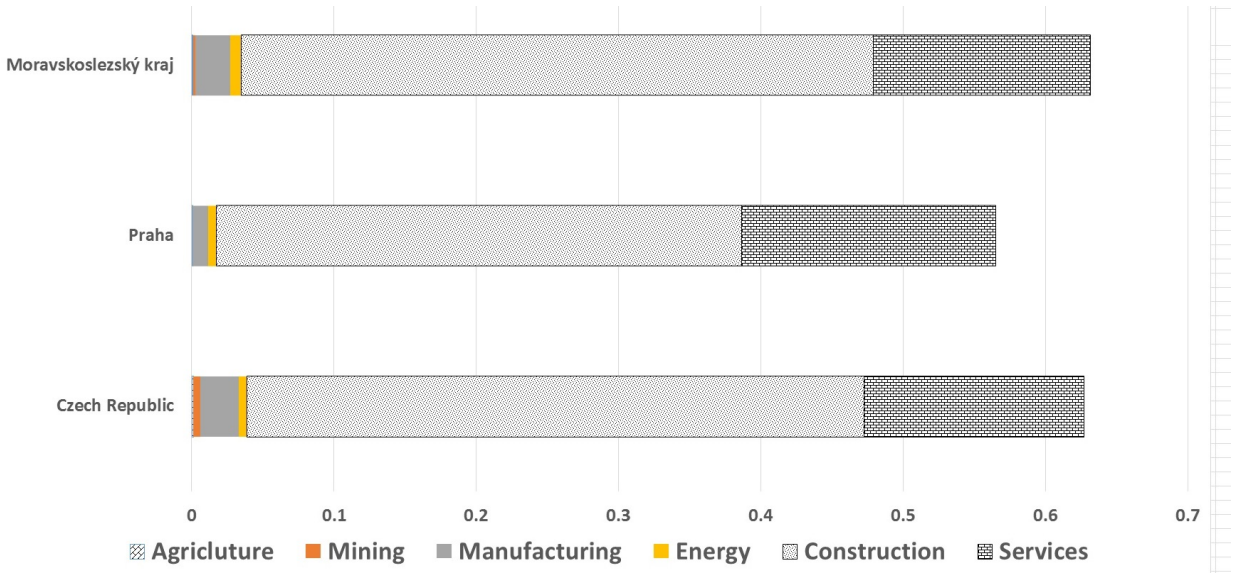


Figure 2 Relative impact on the regional value added

4 Conclusion

Regional input-output tables provide lots of information about regional economy that can be easily used for detailed regional analyses. The quality of statistical information hidden in regional Input-Output Tables is important but for practical applications, the expert knowledge of the region is crucial. Application of either simple Input-Output Analysis or mode developed modes provides results that have to be correctly interpreted. It is necessary to have in mind that Regional Input-Output Tables represent a model of regional economy with many minor imperfections. The most important obstacle is a poor quality of identification of local kind-of-activity unit.

An illustrative example of investment shock was presented on Czech Regional Input-Output tables for 2011. These tables come from the results of our research project. They are derived from officially published indicators from national accounts. These tables were compiled for 14 regions at basic prices in dimension of 82 x 82 products. Tables are split between the use of domestic (regional) production and imported products.

Simple static Input-Output Analysis was used for illustration of the regional difference. It is shown that at least in the short period, the regional impact of increased investments is different region to region. It means that when analysing regional impacts of different events, it is suitable to use rather regional tables than countries averages. This was illustrated on the case of regions Praha and Moravskoslezsky kraj. Since our project finishes by the end of 2015, Regional Input-Output Tables will be freely available for download in January or February 2016.

Acknowledgements

This paper is prepared under the support of the project "Regional estimates of gross domestic product based on the expenditure approach" of the Czech Science Foundation, project No 13-15771S.

References

[1] Bureau of Economic Analyses. [online]. URL: https://www.bea.gov/regional/rims/rimsii/

[2] European Union: *Manual on regional accounts methods*. Luxembourg, 2013.

[3] European Union: *Eurostat Manual of Supply, Use and Input-Output Tables*. [online], 2008. URL: http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-013-EN.pdf

[4] Eurostat: *European System of Accounts - ESA 1995*. Luxembourg, Office for Official Publications of the European Communities, 1996.

[5] Miller, R. E. and Blair, P. D.: *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, New York, 2009. ISBN-13 978-0-511-65103-8.

[6] Kahoun, J. and Sixta J.: Regional GDP Compilation: Production, Income and Expenditure Approach, *Statistika* **4** (2013), 24–36.

[7] Kramulová, J. and Musil, P.: Experimentální odhad složek výdajové metody regionálního HDP v ČR, *Politická ekonomie* **6** (2013), 814–833.

[8] Leontief, W.: *Input-Output Economics, second edition*. Oxford University Press, New York, 1984. ISBN-13 978-0-19-503527-8.

[9] Sixta J. and Fischer J. and Zbranek J.: Regional Input-Output Tables. In: *Mathematical Methods in Economics 2014*, Olomouc, 2014, 896–902.[online]. URL: http://mme2014.upol.cz/conference-proceedings

[10] Sixta J.: Development of Input-Output Tables in the Czech Republic, *Statistika* **2** (2013), 4–14.

[11] Sixta J. and Vltavská K. and Fischer, J.: Compilation of Regional Input-Output Tables. In: *Applications of Mathematics and Statistics in Economy (AMSE 2013)*, Banská Bystrica, 2013, s. 23.

[12] Statistics Finland. Online. [online]. URL: https://www.stat.fi/til/apt/index_en.html

Optimisation of University Timetable via Complex Goal Programming Model

Veronika Skocdopolova[1]

Abstract. Goal programming is a widely used approach for solving not only multi-objective decision problems. It is 60 years since the goal programming has been formulated for the first time. Since that time it becomes very popular technique that has many real-world applications. Optimisation of a timetable leads to cost reducing or avoiding time wasting. Problem of university timetable construction is NP-hard problem that each university deals each term with. Some of them use time-tested scheduling board, the other use more or less sophisticated computer-based system. The University of Economics, Prague, belongs to the first group. The problem of this process is that the timetable from the previous year is only modified a little bit. Therefore the timetable does not react on specific changes between the two academic years. This paper presents a complex model, which utilizes goal programming with multistage penalisation function for solving the timetabling problem. The complex model takes into consideration the specificity of timetable of the department of econometrics at the University of Economics, Prague. It also deals with teachers' time preferences.

Keywords: timetabling, goal programming, multistage penalization function, teachers' preferences.

JEL Classification: C44
AMS Classification: 90C29

1 Introduction

Timetabling is a widely used technique. It is utilized in transportation (bus schedules), healthcare (planning the shifts of the staff) or in education. It helps avoiding time or resource wasting and therefore it reduces cost. University timetabling is a problem that each university deals each term with. There are two main approaches how to construct a timetable using mathematical modelling. One of them is creating a complex model usually with integer variables ([3], [5]). However, solving integer programming models is, in general, NP-hard problem [4], so it leads to utilizing various heuristic or metaheuristic methods ([2], [6]). Heuristic and metaheuristic methods give us solutions that are relatively close to optimal solution in relatively reasonable time. The other approach consists in decomposition of the problem into several interrelated stages ([1], [8], [9]). This means outputs of one stage are inputs in the next stage. This paper deals with complex model designed for timetabling at the department of econometrics, University of Economics, Prague. The specific points of timetabling at that department were described in [9] and [10]. The complex model is formulated using goal programming with multistage penalization function (for more see [7]). The model also deals with teachers' time preferences. This results in greater satisfaction of teachers with the timetable.

1.1 Goal Programming with Multistage Penalization Function

The generic goal programming model can be formulated as follows:
Minimise

$$z = f(\mathbf{d}^-, \mathbf{d}^+) \tag{1}$$

subject to

$$\mathbf{x} \in X\,, \tag{2}$$

$$f_k(\mathbf{x}) + d_k^- - d_k^+ = g_k\,,\ k = 1, 2, \ldots, K, \tag{3}$$

$$d_k^-, d_k^+ \geq 0\,, \qquad k = 1, 2, \ldots, K,$$

where (1) is a general penalisation function, X is the set of feasible solution satisfying all of the constraints including non-negativity constraints (2), $f_k(\mathbf{x})$ is an objective function that represents the k-th criterion, g_k is the k-th goal,

[1] University of Economics, Prague, Faculty of Informatics and Statistics, department of econometrics, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, skov01@vse.cz.

the decision maker wants to meet, d_k^- is negative deviation from the *k*-th goal, which represents the underachievement of the *k*-th goal, d_k^+ is positive deviation from the *k*-th goal, which shows the overachievement of the *k*-th goal; and (3) are the goal constraints. It happens in practice that certain small deviation from the goal acceptable with just a little penalization and greater deviation is allowed but with higher penalization. To preserve the linear form of the model we can use the multistage penalization function. In this case we have to change the goal constraints (3) to the following one

$$f_k(\mathbf{x}) + d_k^- - d_{k1}^+ - d_{k2}^+ = g_k\,, \qquad k = 1, 2, \ldots, K,$$

$$0 \le d_{k1}^+ \le h_{k1}^+, 0 \le d_{k2}^+ \le h_{k2}^+, d_k^- \ge 0\,, \qquad k = 1, 2, \ldots, K,$$

where d_{k1}^+ is the 1st stage positive deviation that is non-negative and has upper bound h_{k1}^+, and d_{k2}^+ is the 2nd stage positive deviation that is also non-negative with upper bound h_{k2}^+. The multistage penalization function can be formulated as follows:

$$z = \sum_{k=1}^{K} \left(u_k d_k^- + v_{k1} d_{k1}^+ + v_{k2} d_{k2}^+ \right),$$

where v_{k1} and v_{k2} are penalty constants of the 1st an 2nd stage positive deviation and u_k is penalty constant of the negative deviation.

2 Mathematical model

In this part there is presented a model for timetabling at the department of econometrics at the University of Economics, Prague. The model takes into consideration teachers' time preferences. It also deals with assigning of classrooms in two distant campuses. Also the problem of assigning computer classrooms is solved within this model.

In the mathematical model there are defined the following index sets:

- i – teachers that participate in teaching,
- j – courses that have to be scheduled; this set has several subsets:
  - $PC1$ – courses that need computer classroom every week,
  - $PC2$ – courses that need computer classroom every other week,
  - $PC3$ – courses that don't need computer classroom,
  - R – courses that have to be placed into particular classroom,
  - $BL1$ – courses that have to be placed into block, 1st course of the block,
  - $BL2$ – courses that have to be placed into block, 2nd course of the block,
- k – time windows,
- l – classrooms that are at department's disposal, this set has two subsets:
  - JM – classrooms located in the campus Jižní Město,
  - ZI – classrooms located in the campus Žižkov,
- m – working days.

Variables of the model are as follows:

- x_{ijkl}, $\forall\, i, j, k, l$ – binary variable that equals 1 if the teacher i teaches the course j in the time window k in the classroom l, and 0 otherwise,
- y_{im}, $\forall\, i, m$ – binary variable that equals 1 if the teacher i teaches on the working day m, and 0 otherwise,
- a_{im}, $\forall\, i, m$ – binary variable that equals 1 if the teacher i teaches on the working day m in the campus Jižní Město, and 0 otherwise,
- b_{im}, $\forall\, i, m$ – binary variable that equals 1 if the teacher i teaches on the working day m in the campus Žižkov, and 0 otherwise,
- δ_1^-, δ_1^+ – negative and positive deviation from total maximal subject preference,
- δ_2^-, δ_2^+ – negative and positive deviation from total maximal time window preference,
- δ_{3i}^-, δ_{3i}^+, $\forall\, i$ – 1st level negative and positive deviation from number of courses, that should teach the teacher i,

- δ_{4i}^{-}, δ_{4i}^{+}, $\forall\, i$ – 2nd level negative and positive deviation from number of courses, that should teach the teacher i,
- δ_{5l}^{-}, δ_{5l}^{+}, $\forall\, l$ – 1st level negative and positive deviation expressing the difference between the capacity of the course assigned to the classroom l and the capacity of classroom l,
- δ_{6l}^{-}, $\forall\, l$ – 2nd level negative deviation expressing the difference between the capacity of the course assigned to the classroom l and the capacity of classroom l,
- δ_{7i}^{-}, δ_{7i}^{+}, $\forall\, i$ – negative and positive deviation from the maximal number of days per week when the teacher i would like to teach,
- δ_{8im}^{-}, δ_{8im}^{+}, $\forall\, i, m$ – negative and positive deviation from the maximal number of courses per day that the teacher i would like to teach on the day m.

Input parameters of the model are denoted as follows:

- RN_l, $\forall\, l$ – numeric code that replaces label of classroom l,
- CPT_l, $\forall\, l$ – capacity of classroom l,
- PC_l, $\forall\, l$ – indicates whether classroom l is equipped with student computers (value 1) or not (value 0),
- TW_l, $\forall\, l$ – numeric code of time window when the classroom l is at disposal,
- AR_l, $\forall\, l$ – indicates whether the classroom l is located in the campus Jižní Město (value 1) or Žižkov (value 0),
- NC_i, $\forall\, i$ – number of courses that should be thought by the teacher i (goal value),
- DPW_i, $\forall\, i$ – maximal number of days per week that the teacher i should teach (goal value),
- CPD_i, $\forall\, i$ – maximal number of courses per day that the teacher i should (goal value),
- PS_{ij}, $\forall\, i, j$ – teacher i preference of course j at the scale 0–5, where 0 means, that the teacher i cannot teach the course j,
- PT_{ik}, $\forall\, i, k$ – teacher i preference of time window k at the scale 0–5, where 0 means, that the teacher i cannot teach at the time window k,
- CAP_j, $\forall\, j$ – required capacity of the course j,
- $COMP_j$, $\forall\, j$ – indicates, whether for course j needs to be placed in classroom with student computers every week (value 1), every other week (value 0,5), or never (value 0),
- $ROOM_j$, $j \in R$ – numeric code of a classroom where the course j have to be assigned,
- WIN_k, $\forall\, k$ – numeric code of the time window k,
- D – number of time windows per day,
- $PREF\_S$ – maximal total subject preference (goal value),
- $PREF\_T$ – maximal total time window preference (goal value),
- $p_1, p_2, p_{31}, p_{32}, p_4, p_{51}, p_{52}, p_6, p_7, p_8$ – penalty constants of each deviation included in the penalty function.

Mathematical model can be formulated as follows:

Minimize

$$z = p_1\delta_1^- + p_2\delta_2^- + p_{31}\sum_i \delta_{3i}^- + p_{32}\sum_i \delta_{3i}^+ + p_4\sum_i (\delta_{4i}^- + \delta_{4i}^+) +$$
$$+ p_{51}\sum_l \delta_{5l}^- + p_{52}\sum_l \delta_{5l}^+ + p_6\sum_l \delta_{6l}^- + p_7\sum_i \delta_{7i}^+ + p_8\sum_i\sum_m \delta_{8im}^+$$

subject to the following constraints:

The teacher i can teach in the time window k at most one course

$$\sum_j\sum_l x_{ijkl} \le 1, \quad \forall i,k\,,$$

every course j has to be assigned exactly to one teacher i at the time window k in the classroom l

$$\sum_i\sum_k\sum_l x_{ijkl} = 1, \quad \forall j\,,$$

and to every classroom l can be assigned at most one course at the particular time

$$\sum_i\sum_j\sum_k x_{ijkl} \le 1, \quad \forall l\,.$$

If the teacher *i* set the preference of course *j* to 0, he/she mustn't teach the course *j*

$$\sum_k \sum_l x_{ijkl} \leq PS_{ij}, \quad \forall i, j,$$

as well as with the zero preference of time window *k*

$$\sum_j \sum_l x_{ijkl} \leq PT_{ik}, \quad \forall i, k.$$

At the time window *k*, which the course *j* is assigned to, the classroom *l* has to be available

$$\sum_i \sum_k \sum_l WIN_k x_{ijkl} = \sum_i \sum_k \sum_l TW_l x_{ijkl}, \quad \forall j.$$

The following constraints assure the assignment of computer classrooms. All the courses $j \in PC1$ that have to be thought in a computer classroom every week have to be assigned to a computer classroom at the time window *k*

$$\sum_i \sum_{j \in PC1} \sum_l PC_l x_{ijkl} = \sum_i \sum_{j \in PC1} \sum_l COMP_j x_{ijkl}, \quad \forall k,$$

while the number of computer classrooms used in the time window *k* have to be greater or equal than a half of number of courses that have to be thought in a computer classroom every other week ($COMP_j = 0{,}5$, $j \in PC2$)

$$\sum_i \sum_{j \in PC2} \sum_l PC_l x_{ijkl} \geq \sum_i \sum_{j \in PC2} \sum_l COMP_j x_{ijkl}, \quad \forall k.$$

This assures rotation of computer classrooms and non-computer classrooms. Courses for those computer are not needed have to be assigned to non-computer classrooms

$$\sum_i \sum_{j \in PC3} \sum_l PC_l x_{ijkl} = \sum_i \sum_{j \in PC3} \sum_l COMP_j x_{ijkl}, \quad \forall k.$$

If the teacher *i* teaches on the day *m* in the campus Jižní Město, the value of the binary variable a_{im} has to be 1

$$a_{im} \leq \sum_j \sum_{k=7m-6}^{7m} \sum_{l \in JM} x_{ijkl}, \quad \forall i, m,$$

$$D \cdot a_{im} \geq \sum_j \sum_{k=7m-6}^{7m} \sum_{l \in JM} x_{ijkl}, \quad \forall i, m.$$

Accordingly if the teacher *i* teaches on the day *m* in the campus Žižkov, the value of the binary variable b_{im} has to be 1

$$b_{im} \leq \sum_j \sum_{k=7m-6}^{7m} \sum_{l \in ZI} x_{ijkl}, \quad \forall i, m,$$

$$D \cdot b_{im} \geq \sum_j \sum_{k=7m-6}^{7m} \sum_{l \in ZI} x_{ijkl}, \quad \forall i, m.$$

To avoid wasting of teachers' time the transfer from one campus to another is prohibited during one day. Each day *m* can the teacher *i* teaches at most in one of the campuses

$$a_{im} + b_{im} \leq 1, \quad \forall i, m.$$

If the teacher *i* teaches on the day *m* in one of the campuses, it means that he/she teaches on the day *m*

$$y_{im} = a_{im} + b_{im}, \quad \forall i, m.$$

The following constraints assure the courses thought in block. This means that two courses of the same subject that should be thought in block are assigned to the following time

$$\sum_i \sum_{j \in BL1} \sum_k \sum_l TW_l x_{ijkl} - \sum_i \sum_{j \in BL2} \sum_k \sum_l TW_l x_{ijkl} = 1,$$

in the same classroom

$$\sum_i \sum_{j \in BL1} \sum_k \sum_l RN_l x_{ijkl} - \sum_i \sum_{j \in BL2} \sum_k \sum_l RN_l x_{ijkl} = 0,$$

with the same teacher

$$\sum_{j \in BL1} \sum_k \sum_l x_{ijkl} - \sum_{j \in BL2} \sum_k \sum_l x_{ijkl} = 0, \quad \forall i .$$

Some of the courses ($j \in R$) have to be assigned to a particular classroom

$$\sum_i \sum_k \sum_l RN_l x_{ijkl} = ROOM_j, \quad j \in R .$$

The following constraints are soft or goal constraints. With this constraints the deviations are allowed. One of the goal of this model is the maximization of the total subject preference

$$\sum_i \sum_j \sum_k \sum_l PS_{ij} x_{ijkl} + \delta_1^- - \delta_1^+ = PREF\_S ,$$

and maximization of the total time window preference

$$\sum_i \sum_j \sum_k \sum_l PT_{ik} x_{ijkl} + \delta_2^- - \delta_2^+ = PREF\_T .$$

The penalty function includes only negative deviations from this two goals.

Another goal is to follow the course load for each teacher *i*

$$\sum_j \sum_k \sum_l x_{ijkl} + \delta_{3i}^- + \delta_{4i}^- - \delta_{3i}^+ - \delta_{4i}^+ = NC_i, \quad \forall i ,$$

$$\delta_{3i}^- \le 1, \quad \forall i ,$$

$$\delta_{3i}^+ \le 1, \quad \forall i .$$

In this case the principle of multistage penalization is utilized. The teacher *i* should teach the load of courses NC_i. The positive or negative deviation of one course is possible with small penalization. This means that the teacher can teach one course more or less than NC_i. Greater deviations (positive or negative) are also allowed but with higher penalization. The penalization function includes all four deviations.

The fourth goal of the model is to utilize the capacity of classrooms

$$\sum_i \sum_j \sum_k CAP_j x_{ijkl} + \delta_{5l}^- + \delta_{6l}^- - \delta_{5l}^+ = \sum_i \sum_j \sum_k KAP_l x_{ijkl}, \quad \forall l ,$$

$$\delta_{5l}^- \le 0{,}5 \cdot KAP_l \quad \forall l ,$$

$$\delta_{5l}^+ \le 0{,}1 \cdot KAP_l, \quad \forall l .$$

With a small penalization it is possible to exceed the capacity of the classroom *l* up to 10 % (It is assumed, that not all of the registered students will attend the course). It is also possible not to utilize the capacity of the classroom up to 50 % of its capacity. If necessary course *j* can be assigned to a classroom with capacity more than twice exceeding the requirement of the course. Nevertheless this results in higher penalization. On the other hand the exceeding of the room capacity for more than 10 % is not allowed. The penalization function cover all three allowed deviations.

The fifth goal ensures, that the teacher *i* should teach only DPW_i days per week

$$\sum_m y_{im} + \delta_{7i}^- - \delta_{7i}^+ = DPW_i, \quad \forall i .$$

Both deviations are possible, but penalized is only the positive deviation.

The last goal assure the number of courses each teacher has per day

$$\sum_j \sum_{k=7m-6}^{7m} \sum_l x_{ijkl} + \delta_{8im}^- - \delta_{8im}^+ = CPD_i, \quad \forall i,m.$$

Again, both deviations are allowed, but only the positive one is penalized.

All the variables are binary except the deviations that are just non-negative:

$$x_{ijkl} \in \{0,1\}, \quad \forall i,j,k,l\ ,\ y_{im} \in \{0,1\}, \quad \forall i,m\ ,$$

$$A_{im} \in \{0,1\}, \quad \forall i,m\ ,\ B_{im} \in \{0,1\}, \quad \forall i,m\ ,$$

$$\delta_1^-, \delta_1^+, \delta_2^-, \delta_2^+ \geq 0\ ,$$

$$\delta_{3i}^-, \delta_{3i}^+, \delta_{4i}^-, \delta_{4i}^+, \delta_{7i}^-, \delta_{7i}^+ \geq 0, \quad \forall i\ ,$$

$$\delta_{5l}^-, \delta_{5l}^+, \delta_{6l}^-, \delta_{6l}^+ \geq 0, \quad \forall l\ ,$$

$$\delta_{8im}^-, \delta_{8im}^+ \geq 0, \quad \forall i,m\ .$$

3 Conclusion

Timetabling is used in many segments of economy. It helps saving both resources and time. In this paper the complex model for timetabling at the University of Economics, Prague was introduced. The model was formulated as goal program with six goals and multistage penalization function. Next step of this research is implementation of this model into an application that may utilize the cooperation between Microsoft Excel and Gurobi solver via MPL modelling system.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics, Prague (grant no. F4/62/2015) and by institutional support for long-term conceptual development of science and research at the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Al-Husain, R., Hasan, M. K. and Al-Qaheri, H.: A Sequential Three-Stage Integer Goal Programming (IGP) Model for Faculty-Course-Time-Classroom Assignments. *Informatica* **35** (2011), 157–164.
[2] Aladag, C., Hocaoglu, G. and Basaran, M.: The effect of neighbourhood structures on tabu search algorithm in solving course timetabling problem. *Expert Systems with Applications* **36** (2009), 12349–12356.
[3] Bakir, M. A. and Aksop, C.: A 0-1 Integer Programming Approach to a University Timetabling Problem. *Hacettepe Journal of Mathematics and Statistics* **37** (2008), 41–55.
[4] Cerny, M.: *Computations, Volume III* (In Czech). Professional Publishing, Prague, 2012.
[5] Daskalaki, S., Birbas, T. and Housos, E.: An integer programming formulation for a case study in university timetabling. *European Journal of Operational Research* **153** (2004), 117–135.
[6] De Causmaecker, P., Demeester, P. and Berghe, G.: A decomposed metaheuristic approach for a real-world university timetabling problem. *European Journal of Operational Research* **195** (2009), 307–318.
[7] Jones, D., Tamiz, M.: *Practical Goal Programming*. International Series in Operations Research & Management Science 141. Springer, New York, 2010.
[8] Mirrazavi, S. K., Mardle, S. J. and Tamiz, M.: A Two-Phase Multiple Objective Approach to University Timetabling Utilising Optimisation and Evolutionary Solution Methodologies. *The Journal of the Operational Research Society* **54** (2003), 1155–1166.
[9] Skocdopolova, V.: University timetable construction – computational approach. Mathematical Methods in Economics 2013, Jihlava, 2013, 808–813.
[10] Skocdopolova, V.: Construction of a university timetable via goal programming. Efficiency and Responsibility in Education 2013. Prague: Czech University of Life Sciences in Prague, 2013, 565–572.

Second Order Optimality in Transient and Discounted Markov Decision Chains

Karel Sladký[1]

Abstract. The article is devoted to second order optimality in Markov decision processes. Attention is primarily focused on the reward variance for discounted models and undiscounted transient models (i.e. where the spectral radius of the transition probability matrix is less then unity). Considering the second order optimality criteria means that in the class of policies maximizing (or minimizing) total expected discounted reward (or undiscounted reward for the transient model) we choose the policy minimizing the total variance. Explicit formulae for calculating the variances for transient and discounted models are reported along with sketches of algorithmic procedures for finding second order optimal policies.

Keywords: dynamic programming, transient Markov reward chains, discounted Markov reward chains, reward-variance optimality.

JEL classification: C44, C61, C63
AMS classification: 90C40, 60J10, 93E20

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect variability-risk features of the problem. Perhaps the best known approaches stem from the classical work of Markowitz on mean variance selection rules, i.e. we optimize the weighted sum of the expected total (or average) reward and its variance. In the present paper we restrict attention on transient and discounted models with finite state space.

2 Notation and Preliminaries

In this note, we consider at discrete time points Markov decision process $X = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$, and compact set $\mathcal{A}_i = [0, K_i]$ of possible decisions (actions) in state $i \in \mathcal{I}$. Supposing that in state $i \in \mathcal{I}$ action $a \in \mathcal{A}_i$ is chosen, then state j is reached in the next transition with a given probability $p_{ij}(a)$ and one-stage transition reward r_{ij} will be accrued to such transition.

A (Markovian) policy controlling the decision process, $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$ for every $n = 0, 1, 2, \ldots$, and $f_i^n \in \mathcal{A}_i$ is the decision (or action) taken at the nth transition if the chain X is in state i. Let $\pi^m = (f^m, f^{m+1}, \ldots)$, hence $\pi = (f^0, f^1, \ldots, f^{m-1}, \pi^m)$, in particular $\pi = (f^0, \pi^1)$. The symbol E_i^π denotes the expectation if $X_0 = i$ and policy $\pi = (f^n)$ is followed, in particular, $\mathsf{E}_i^\pi(X_m = j) = \sum_{i_j \in \mathcal{I}} p_{i,i_1}(f_i^0) \ldots p_{i_{m-1},j}(f_{m-1}^{m-1})$; $\mathrm{P}(X_m = j)$ is the probability that X is in state j at time m.

Policy π which selects at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary, hence following policy $\pi \sim (f)$ X is a homogeneous Markov chain with transition probability matrix $P(f)$ whose ij-th element is $p_{ij}(f_i)$. Then $r_i^{(1)}(f_i) := \sum_{j \in \mathcal{I}} p_{ij}(f_i) r_{ij}$ is the expected one-stage reward obtained in

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: sladky@utia.cas.cz

state i. Similarly, $r^{(1)}(f)$ is an N-column vector of one-stage rewards whose i-the elements equals $r_i^{(1)}(f_i)$. Observe that $\mathsf{E}_i^\pi(X_m = j) = [P^m(f)]_{ij}$ ($[A]_{ij}$ denotes the ij-th element of the matrix A, $A \geq B$, resp. $A > B$ iff for each i, j $[A]_{ij} \geq [B]_{ij}$ resp. $[A]_{ij} \geq [B]_{ij}$ and $[A]_{ij} > [B]_{ij}$ for some i, j). The symbol I denotes an identity matrix and e is reserved for a unit column vector.

Considering the standard probability matrix $P(f)$ the spectral radius of $P(f)$ is equal to one. Recall that (the Cesaro limit of $P(f)$) $P^*(f) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=}^{n-1} P^k(f)$ (with elements $p_{ij}^*(f)$) exists, and if $P(f)$ is aperiodic then even $P^*(f) = \lim_{k\to\infty} P^k(f)$ and the convergence is geometrical. The (row) vector $g^{(1)}(f) = P^*(f)\, r^{(1)}(f)$ is the vector of average rewards, its $i$the entry $g_i^{(1)}(f)$ denotes the average reward if the process starts in state i. Moreover, if $P(f)$ is unichain, i.e. $P(f)$ contains a single class of recurrent states, then $p_{ij}^*(f) = p_j^*(f)$, i.e. limiting distribution is independent of the starting state and $g^{(1)}(f)$ is a constant vector with elements $\bar{g}^{(1)}(f)$. It is well-known (cf. e.g. [3]) that also $Z(f)$ (fundamental matrix of $P(f)$), and $H(f)$ (the deviation matrix) exist, where $Z(f) := [I - P(f) + P^*(f)]^{-1}, H(f) := Z(f)\,(I - P^*(f))$.

Transition probability matrix $\tilde{P}(f)$ is called *transient* if the spectral radius of $\tilde{P}(f)$ is less than unity, i.e. it at least some row sums of $\tilde{P}(f)$ are less than one. Then $\lim_{n\to\infty}[\tilde{P}(f)]^n = 0$, $\tilde{P}^*(f) = 0$ $g^{(1)}(f) = \tilde{P}^*(f)\, r^{(1)}(f) = 0$ and $\tilde{Z}(f) = \tilde{H}(f) = [I - \tilde{P}(f)]^{-1}$. Observe that if $P(f)$ is stochastic and $\alpha \in (0, 1)$ then $\tilde{P}(f) := \alpha P(f)$ is transient, however, if $\tilde{P}(f)$ is transient it may happen that some row sums may be even greater than unity.

3 Reward Variance for Finite and Infinite Time Horizon

Let $\xi_n(\pi) = \sum_{k=0}^{n-1} r_{X_k,X_{k+1}}$ resp. $\xi_n^\alpha(\pi) = \sum_{k=0}^{n-1} \alpha^k r_{X_k,X_{k+1}}$ with $\alpha \in (0, 1)$, be the stream of undiscounted, resp. α-discounted rewards, received in the n next transitions of the considered Markov chain X if policy $\pi = (f^n)$ is followed. Supposing that $X_0 = i$, on taking expectation we get for the first and second moments of $\xi_n(\pi)$, resp. $\xi_n^\alpha(\pi)$,

$$v_i^{(1)}(\pi, n) := \mathsf{E}_i^\pi(\xi_n(\pi)) = \mathsf{E}_i^\pi \sum_{k=0}^{n-1} r_{X_k,X_{k+1}}, \quad v_i^{(2)}(\pi, n) := \mathsf{E}_i^\pi(\xi_n(\pi))^2 = \mathsf{E}_i^\pi(\sum_{k=0}^{n-1} r_{X_k,X_{k+1}})^2, \tag{1}$$

resp.

$$v_i^{\alpha(1)}(\pi, n) := \mathsf{E}_i^\pi(\xi_n^\alpha(\pi)) = \mathsf{E}_i^\pi \sum_{k=0}^{n-1} \alpha^k r_{X_k,X_{k+1}}, \quad v_i^{\alpha(2)}(\pi, n) := \mathsf{E}_i^\pi(\xi_n^\alpha(\pi))^2 = \mathsf{E}_i^\pi(\sum_{k=0}^{n-1} \alpha^k r_{X_k,X_{k+1}})^2. \tag{2}$$

It is well known from the literature (cf. e.g. [2],[3],[7]) that for the time horizon n tending to infinity policies maximizing or minimizing the values $v_i^{(1)}(\pi, n)$ and $v_i^{\alpha(1)}(\pi, n)$ can be found in the class of stationary policies, i.e. there exist $\hat{f}, \hat{f}^\alpha, \bar{f}, \bar{f}^\alpha \in \mathcal{F}$ such that for all $i \in \mathcal{I}$ and any policy $\pi = (f^n)$

$$v_i^{(1)}(\hat{f}) := \lim_{n\to\infty} v_i^{(1)}(\hat{f}, n) \geq \limsup_{n\to\infty} v_i^{(1)}(\pi, n), \quad \lim_{n\to\infty} v_i^{(1)}(\bar{f}, n) \leq \liminf_{n\to\infty} v_i^{(1)}(\pi, n), \tag{3}$$

$$v_i^{\alpha(1)}(\hat{f}^\alpha) := \lim_{n\to\infty} v_i^{\alpha(1)}(\hat{f}^\alpha, n) \geq \limsup_{n\to\infty} v_i^{\alpha(1)}(\pi, n), \quad \lim_{n\to\infty} v_i^{\alpha(1)}(\bar{f}^\alpha, n) \leq \liminf_{n\to\infty} v_i^{\alpha(1)}(\pi, n). \tag{4}$$

3.1 Finite Time Horizon

If policy $\pi \sim (f)$ is stationary, the process X is time homogeneous and for $m < n$ we write for undiscounted random reward $\xi_n = \xi_m + \xi_{n-m}$ (here we delete the symbol π and tacitly assume that $\mathrm{P}(X_m = j)$ and ξ_{n-m} starts in state j). Hence $[\xi_n]^2 = [\xi_m]^2 + [\xi_{n-m}]^2 + 2 \cdot \xi_m \cdot \xi_{n-m}$. Then for $n > m$ we can conclude that

$$\mathsf{E}_i^\pi[\xi_n] = \mathsf{E}_i^\pi[\xi_m] + \mathsf{E}_i^\pi\big\{\sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}_j^\pi[\xi_{n-m}]\big\}. \tag{5}$$

$$\mathsf{E}_i^\pi[\xi_n]^2 = \mathsf{E}_i^\pi[\xi_m]^2 + \mathsf{E}_i^\pi\big\{\sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}_j^\pi[\xi_{n-m}]^2\big\} + 2 \cdot \mathsf{E}_i^\pi[\xi_m] \sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j) \cdot \mathsf{E}_j^\pi[\xi_{n-m}]. \tag{6}$$

In particular, from (3), (5) and (6) we conclude for $m = 1$

$$v_i^{(1)}(f, n+1) = r_i^{(1)}(f_i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i) \cdot v_j^{(1)}(f, n) \tag{7}$$

$$v_i^{\alpha(2)}(f, n+1) = r_i^{(2)}(f_i) + 2 \cdot \sum_{j\in\mathcal{I}} p_{ij}(f_i) \cdot\ r_{ij} \cdot v_j^{(1)}(f, n) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\, v_j^{(2)}(f, n) \tag{8}$$

where $r_i^{(1)}(f_i) := \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)\ r_{ij}, \quad r_i^{(2)}(f_i) := \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)[r_{ij}]^2$.

Since the variance $\sigma_i(f,n) = v_i^{(2)}(f,n) - [v_i^{(1)}(f,n)]^2$ from (7),(8) we get

$$\begin{aligned}\sigma_i(f,n+1) &= r_i^{(2)}(f_i) + \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)\cdot\sigma_j(f,n) + 2\sum_{j\in\mathcal{I}}\ p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{(1)}(f,n)\\ &\quad -[v_i^{(1)}(f,n+1)]^2 + \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)[v_j^{(1)}(f,n)]^2 \qquad (9)\\ &= \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)[r_{ij} + \ v_j^{(1)}(f,n)]^2 - [v_i^{(1)}(f,n+1)]^2 + \sum_{j\in\mathcal{I}}\ p_{ij}(f_i)\cdot\sigma_j(f,n). \qquad (10)\end{aligned}$$

Using matrix notations equations (7),(8) can be written as:

$$v^{(1)}(f,n+1) = r^{(1)}(f) + P(f)\cdot v^{(1)}(f,n) \qquad (11)$$

$$v^{(2)}(f,n+1) = r^{(2)}(f) + 2\cdot P(f)\circ R\cdot v^{(1)}(f,n) + P(f)\ \cdot v^{(2)}(f,n) \qquad (12)$$

where $R = [r_{ij}]_{i,j}$ is an $N\times N$-matrix, and
$r^{(2)}(f) = [\ r_i^{(2)}(f_i)], \quad v^{(2)}(f,n) = [v_i^{(2)}(f,n)]$, $v^{(1)}(f,n) = [(v_i^{(1)}(f,n)]$ are column vectors.
The symbol $\circ$ is used for Hadamard (entrywise) product of matrices. Observe that
$r^{(1)}(f) = (P(f)\circ R)\cdot e, \quad r^{(2)}(f) = [P(f)\circ(R\circ R)]\cdot e$.

3.2 Transient Models

In this subsection we focus attention on transient models, i.e. we assume that the transition probability matrix $\tilde{P}(f)$ with elements $p_{ij}(f_i)$ is substochastic and $\rho(f)$, the spectral radius of $\tilde{P}(f)$, is less than unity. Then $\tilde{P}^*(f) = \lim_{n\to\infty}[\tilde{P}(f)]^n = 0$ and for the fundamental and deviation matrices we get $\tilde{Z}(f) = \tilde{H}(f) = [I - \tilde{P}(f)]^{-1}$.

Then on iterating (11) we easily conclude that there exists $v^{(1)}(f) := \lim_{n\to\infty} v^{(1)}(f,n)$ such that

$$v^{(1)}(f) = r^{(1)}(f) + \tilde{P}(f)\cdot v^{(1)}(f) \Longleftrightarrow v^{(1)}(f) = [I - \tilde{P}(f)]^{-1} r^{(1)}(f). \qquad (13)$$

Similarly, from (12) (since the term $2\cdot P(f)\circ R\cdot v^{(1)}(f,n)$ must be bounded) on letting $n\to\infty$ we can also verify existence $v^{(2)}(f) = \lim_{n\to\infty} v^{(2)}(f,n)$ such that

$$v^{(2)}(f) = \ r^{(2)}(f) + 2\cdot\tilde{P}(f)\circ R\cdot v^{(1)}(f) + \tilde{P}(f)\, v^{(2)}(f) \qquad (14)$$

hence

$$v^{(2)}(f) = [I - \tilde{P}(f)]^{-1}\ \left\{\ r^{(2)}(f) + 2\cdot\tilde{P}(f)\circ R\cdot v^{(1)}(f)\right\}. \qquad (15)$$

On letting $n\to\infty$ from (9), (10) we get for $\sigma_i(f) := \lim_{n\to\infty}\sigma_i(f,n)$

$$\begin{aligned}\sigma_i(f) &= r_i^{(2)}(f_i) + \sum_{j\in\mathcal{I}}\ \tilde{p}_{ij}(f_i)\cdot\sigma_j(f) + 2\sum_{j\in\mathcal{I}}\ \tilde{p}_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{(1)}(f)\\ &\quad -[v_i^{(1)}(f)]^2 + \sum_{j\in\mathcal{I}}\ \tilde{p}_{ij}(f_i)[v_j^{(1)}(f)]^2 \qquad (16)\\ &= \sum_{j\in\mathcal{I}}\ \tilde{p}_{ij}(f_i)[r_{ij} + v_j^{(1)}(f)]^2 - [v_i^{(1)}(f)]^2 + \sum_{j\in\mathcal{I}}\ \tilde{p}_{ij}(f_i)\cdot\sigma_j(f). \qquad (17)\end{aligned}$$

Hence in matrix notation

$$\sigma(f) = \ r^{(2)}(f) + \tilde{P}(f)\cdot\sigma(f) + 2\cdot\tilde{P}(f)\circ R\cdot v^{(1)}(f) - [v^{(1)}(f)]^2 + \tilde{P}(f)\cdot[v^{(1)}(f)]^2. \qquad (18)$$

After some algebra (18) can be also written as

$$\sigma(f) = [I - \tilde{P}(f)]^{-1}\cdot\{\ r^{(2)}(f) + 2\cdot\tilde{P}(f)\circ R\cdot v^{\alpha(1)}(f) - [v^{(1)}(f)]^2\}. \qquad (19)$$

In particular, if $\tilde{P}(f) := \alpha P(f)$ then (19) reads

$$\sigma(f) = [I - \alpha P(f)]^{-1}\cdot\{\ r^{(2)}(f) + 2\cdot\alpha P(f)\circ R\cdot v^{(1)}(f)\} - [v^{(1)}(f)]^2. \qquad (20)$$

where $r^{(2)}(f) = [P(f)\circ(R\circ R)]\cdot e$.

3.3 Discounted Case

From (4) similarly to subsection 3.1 for $n > m$ we can conclude that

$$\mathsf{E}_i^{\pi}[\xi_n^{\alpha}] = \mathsf{E}_i^{\pi}[\xi_m^{\alpha}] + \alpha^m \mathsf{E}_i^{\pi}\{\sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j)\cdot \mathsf{E}_j^{\pi}[\xi_{n-m}^{\alpha}]\}. \tag{21}$$

$$\begin{aligned}\mathsf{E}_i^{\pi}[\xi_n^{\alpha}]^2 &= \mathsf{E}_i^{\pi}[\xi_m^{\alpha}]^2 + \alpha^{2m}\mathsf{E}_i^{\pi}\{\sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j)\cdot \mathsf{E}_j^{\pi}[\xi_{n-m}^{\alpha}]^2\}\\ &\quad +2\cdot\alpha^m\cdot \mathsf{E}_i^{\pi}[\xi_m^{\alpha}]\sum_{j\in\mathcal{I}} \mathrm{P}(X_m = j)\cdot \mathsf{E}_j^{\pi}[\xi_{n-m}^{\alpha}].\end{aligned} \tag{22}$$

In particular, from (2), (21) and (22) we conclude for $m = 1$

$$v_i^{\alpha(1)}(f, n+1) = r_i^{(1)}(f_i) + \alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot v_j^{\alpha(1)}(f,n) \tag{23}$$

$$v_i^{\alpha(2)}(f, n+1) = r_i^{(2)}(f_i) + 2\cdot\alpha\cdot\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f,n) + \alpha^2\cdot\sum_{j\in\mathcal{I}} p_{ij}(f_i)\, v_j^{\alpha(2)}(f,n) \tag{24}$$

and from (23),(24), for the variance $\sigma_i^{\alpha}(f,n) := v_i^{\alpha(2)}(f,n) - [v_i^{\alpha(1)}(f,n)]^2$ we get

$$\begin{aligned}\sigma_i^{\alpha}(f,n+1) &= r_i^{(2)}(f_i) + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{\alpha}(f,n) + 2\cdot\alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f,n)\\ &\quad -[v_i^{\alpha(1)}(f,n+1)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)[v_j^{\alpha(1)}(f,n)]^2 \qquad (25)\\ &= \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij} + \alpha\cdot v_j^{\alpha(1)}(f,n)]^2 - [v_i^{\alpha(1)}(f,n+1)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{\alpha}(f,n). \qquad (26)\end{aligned}$$

Using matrix notations equations (23), (24) can be written as:

$$v^{\alpha(1)}(f, n+1) = r^{(1)}(f) + \alpha\cdot P(f)\cdot v^{\alpha(1)}(f,n) \tag{27}$$

$$v^{\alpha(2)}(f, n+1) = r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f,n) + \alpha^2\cdot P(f)\, v^{\alpha(2)}(f,n) \tag{28}$$

recall that $R = [r_{ij}]_{i,j}$ is an $N\times N$-matrix, and $\circ$ is used for Hadamard (entrywise) product of matrices.

On iterating (27) we easily conclude that there exists $v^{\alpha(1)}(f) := \lim_{n\to\infty} v^{\alpha(1)}(f,n)$ such that

$$v^{\alpha(1)}(f) = r^{(1)}(f) + \alpha P(f)\cdot v^{\alpha(1)}(f) \Longleftrightarrow v^{\alpha(1)}(f) = [I - \alpha P(f)]^{-1} r^{(1)}(f). \tag{29}$$

Similarly to the transient case on letting $n\to\infty$ for discounted models also exists $v^{\alpha(2)}(f) = \lim_{n\to\infty} v^{\alpha(2)}(f,n)$ and by (28)

$$v^{\alpha(2)}(f) = r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f) + \alpha^2\cdot P(f)\, v^{\alpha(2)}(f), \tag{30}$$

so

$$v^{\alpha(2)}(f) = [I - \alpha^2\cdot P(f)]^{-1}\left\{\ r^{(2)}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f)\right\}. \tag{31}$$

On letting $n\to\infty$ from (25), (26) we get for $\sigma_i^{\alpha}(f) := \lim_{n\to\infty}\sigma_i^{\alpha}(f,n)$

$$\begin{aligned}\sigma_i^{\alpha}(f) &= r_i^{(2)}(f_i) + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{\alpha}(f) + 2\cdot\alpha\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\ r_{ij}\cdot v_j^{\alpha(1)}(f)\\ &\quad -[v_i^{\alpha(1)}(f)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)[v_j^{\alpha(1)}(f)]^2 \qquad (32)\\ &= \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij} + \alpha\cdot v_j^{\alpha(1)}(f)]^2 - [v_i^{\alpha(1)}(f)]^2 + \alpha^2\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot\sigma_j^{\alpha}(f). \qquad (33)\end{aligned}$$

Hence in matrix notation

$$\sigma^{\alpha}(f) = r^{(2)}(f) + \alpha^2\cdot P(f)\cdot\sigma^{\alpha}(f) + 2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f) - [v^{\alpha(1)}(f)]^2 + \alpha^2\cdot P(f)\cdot[v^{\alpha(1)}(f)]^2. \tag{34}$$

After some algebra (34) can be also written as

$$\sigma^{\alpha}(f) \quad = \quad [I-\alpha^2\cdot P(f)]^{-1}\cdot\{\ r^{(2)}(f)+2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f)\}-[v^{\alpha(1)}(f)]^2. \tag{35}$$

where $r^{(2)}(f)=[P(f)\circ(R\circ R)]\cdot e$.

(35) is similar to the formula for the variance of discounted rewards obtained by Sobel [6] by different methods (see also [5]).

3.4 Comparison of Transient and Discounted Models

Let us consider transient model where the transient transition probability matrix $\tilde{P}(f):=\alpha P(f)$, called the α-transient model. Then by (20) the variance

$$\sigma(f) \quad = \quad [I-\alpha P(f)]^{-1}\cdot\{\ r^{(2)}(f)+2\cdot\alpha P(f)\circ R\cdot v^{\alpha(1)}(f)\}-[v^{\alpha(1)}(f)]^2.$$

On the other hand for the α-discounted model we get by (35)

$$\sigma^{\alpha}(f) \quad = \quad [I-\alpha^2\cdot P(f)]^{-1}\cdot\{\ r^{(2)}(f)+2\cdot\alpha\cdot P(f)\circ R\cdot v^{\alpha(1)}(f)\}-[v^{\alpha(1)}(f)]^2.$$

Since $[I-\alpha P(f)]^{-1}>[I-\alpha^2P(f)]^{-1}$ from (15),(31) and from (19),(20) we immediately conclude that

$$v^{\alpha(2)}(f)-\tilde{v}^{\alpha(2)}(f) \ = \ \{[I-\alpha P(f)]^{-1}-[I-\alpha^2P(f)]^{-1}\}\cdot\{\ r^{(2)}(f)+2\alpha P(f)\circ R\cdot\tilde{v}^{\alpha(1)}(f)\}>0, \tag{36}$$

$$\tilde{\sigma}^{\alpha}(f)-\sigma^{\alpha}(f) \ = \ \{[I-\alpha P(f)]^{-1}-[I-\alpha^2P(f)]^{-1}\}\cdot\{\ r^{(2)}(f)+2\alpha P(f)\circ R\cdot\tilde{v}^{\alpha 1)}(f)>0. \tag{37}$$

Observe that

$$[I-\alpha P(f)]^{-1}-[I-\alpha^2P(f)]^{-1}=\sum_{k=0}^{\infty}\{(\alpha P(f))^k-(\alpha^2P(f))^k\}=\sum_{k=0}^{\infty}\{(1-\alpha^k)\alpha^k(P(f))^k\}>0.$$

In words:
Total reward of α-transient Markov reward chain is greater the that of α-discounted Markov reward chain (this fact was also mentioned in [1] and in [4] using approached different to ours).

4 Finding Optimal Policies

For finding second order optimal policies, at first it is necessary to construct the set of optimal transient or optimal α-discounted policies. Recalling (3),(4) for the both optimality criteria mentioned above optimal policies can be found in the class of stationary policies, i.e. there exists $\hat{f},\hat{f}^{\alpha}\in\mathcal{F}$ such that

$$v^{(1)}(\hat{f})\geq v^{(1)}(\pi) \quad \text{resp.} \quad v^{\alpha(1)}(\hat{f})\geq v^{\alpha(1)}(\pi) \quad \text{for every policy} \quad \pi=(f^n). \tag{38}$$

Let $\mathcal{F}^{(0)}\subset\mathcal{F}$ be the set of all transient optimal stationary policies, $\mathcal{F}^{\alpha(0)}\subset\mathcal{F}$ be the set of all α-discounted optimal stationary policies.

Obviously, following policy $\hat{f}\in\mathcal{F}^{(0)}$, resp. $\hat{f}\in\mathcal{F}^{\alpha(0)}$, the corresponding action $\hat{f}_i\in\mathcal{A}_i^{(0)}\subset\mathcal{A}_i$, resp. $\hat{f}_i^{\alpha}\in\mathcal{A}_i^{\alpha(0)}\subset\mathcal{A}_i$. For finding the sets $\mathcal{F}^{(0)}$, $\mathcal{F}^{\alpha(0)}$ of optimal policies well-known policy iteration algorithms, value iteration algorithms or modified value iteration algorithms can be used.

Assume that for $\hat{f}^{(1)},\hat{f}^{(2)}\in\mathcal{F}^{(0)}$ resp. $\hat{f}^{\alpha(1)},\hat{f}^{\alpha(2)}\in\mathcal{F}^{\alpha(0)}$

$$\hat{v}^{(1)}:=v^{(1)}(\hat{f}^{(1)})=v^{(1)}(\hat{f}^{(2)}) \ \text{ resp. } \ \hat{v}^{\alpha(1)}:=v^{\alpha(1)}(\hat{f}^{\alpha(1)})=v^{\alpha(1)}(\hat{f}^{\alpha(2)}),$$

however for the corresponding variances $\sigma_i(\hat{f}^{(1)})\neq\sigma_i(\hat{f}^{(2)})$, $\sigma_i(\hat{f}^{\alpha})\neq\sigma_i(\hat{f}^{\alpha})$.

In what follows we show existence of $f^*\in\mathcal{F}^{(0)}$, $f^{\alpha*}\in\mathcal{F}^{\alpha(0)}$, such that

$$\sigma(f^*)\leq\sigma(\pi) \quad \text{resp.} \quad \sigma^{\alpha}(f^*\alpha)\leq\sigma^{\alpha}(\pi) \tag{39}$$

for every policy $\pi=(f^n), f^n\in\mathcal{F}^{(0)}$ resp. $\pi=(f^n), f^n\in\mathcal{F}^{\alpha(0)}$.

To this end from (18), resp. from (34), we have for transient, resp. discounted, case

$$\sigma(f) = \tilde{P}(f) \circ R \circ R \cdot e + 2 \cdot \tilde{P}(f) \circ R \cdot v^{(1)}(f) - [v^{(1)}(f)]^2 + \tilde{P}(f) \cdot [v^{(1)}(f)]^2 + \tilde{P}(f) \cdot \sigma(f), \qquad (40)$$

resp. for every $f \in \mathcal{F}^{\alpha(0)}$

$$\sigma^\alpha(f) = P(f) \circ R \circ R \cdot e + 2 \cdot \alpha \cdot P(f) \circ R \cdot v^{\alpha(1)}(f) - [v^{\alpha(1)}(f)]^2 + \alpha^2 \cdot P(f) \cdot [v^{\alpha(1)}(f)]^2 + \alpha^2 \cdot P(f) \cdot \sigma^\alpha(f). \quad (41)$$

Now let $h(f), h^\alpha(f)$ equal the first three terms on the RHS of (40) and (41), i.e.

$$h(f) := \tilde{P}(f) \circ R \circ R \cdot e + 2 \cdot \tilde{P}(f) \circ R \cdot v^{(1)}(f) - [v^{(1)}(f)]^2 + \tilde{P}(f) \cdot [v^{(1)}(f)]^2,$$

$$h^\alpha(f) := P(f) \circ R \circ R \cdot e + 2 \cdot \alpha \cdot P(f) \circ R \cdot v^{\alpha(1)}(f) - [v^{\alpha(1)}(f)]^2 + \alpha^2 \cdot P(f) \cdot [v^{\alpha(1)}(f)]^2.$$

Observe that $h(f), h^\alpha(f)$ well correspond to one-stage rewards in (13) and in (29).

On comparing (40) and (13), resp. (41) and (29), we can conclude that the minimal values $\sigma(f^*)$, resp. $\sigma^\alpha(f^*)$ with $f \in \mathcal{F}^{(0)}$, resp. with $f \in \mathcal{F}^{\alpha(0)}$, exist.

5 Conclusions

We have received formulas for the variance of total rewards for transient and discounted Markov reward chains. This enables in the class of optimal policies select policies minimizing variance of total expected reward.

Acknowledgements

This research was supported by the Czech Science Foundation under Grants 13-14445S, 15-10331S and by CONACyT (Mexico) and AS CR (Czech Republic) under Project 171396.

References

[1] Feinberg, E.A. and Fei, J.: *Inequalities for variances of total discounted costs.* J. Applied Probability **46** (2009), 1209–1212.

[2] Mandl, P.: *On the variance in controlled Markov chains.* Kybernetika **7** (1971), 1–12.

[3] Puterman, M.L.: *Markov Decision Processes – Discrete Stochastic Dynamic Programming.* Wiley, New York 1994.

[4] Righter, R.: *Stochastic comparison of discounted rewards.* J. Applied Probability **48** (2011), 293–294.

[5] Sladký, K.: *The variance of discounted rewards in Markov decision processes: Laurent expansion and sensitive optimality.* In: Proceedings of the 32th International Conference Mathematical Methods in Economics 2014, Palacký University, Olomouc, pp. 908–913.

[6] Sobel, M.: *The variance of discounted Markov decision processes.* J. Applied Probability **19** (1982), 794–802.

[7] Veinott, A.F.Jr.: *Discrete dynamic programming with sensitive discount optimality criteria.* Annals Math. Statistics **40** (1969), 1635–1660.

Constant Market Share Analysis of EU Exports: Tool for Analyzing the EU Competitiveness in Trade

Michaela Staníčková[1], Bohdan Vahalík[2], Lenka Fojtíková[3]

Abstract. Understanding a country's relative performance on the aspects of trade should provide a guide to its competitiveness in global markets. It offers limited insight into the main determinants of competitiveness, or of the policy levers that might be pulled to improve it what requires an analytical approach. Constant market share (CMS) analysis is frequently used in applied studies of export. CMS analysis is used to decompose changes in the European Union's (EU) share of the global market for exports into competitiveness and structural effects over 2004-2013. CMSA is used to identify and quantify competitiveness from the export data and breaks down a country´s export performance into a pure competitiveness effect and a structural effect within the EU countries. The EU has experienced competitive pressures from China and other emerging economies that have resulted in a loss of global market share. Changes of a country's market share in world exports result from many interrelated factors. Therefore, the analysis of the export performance of a country should be put in perspective by analyzing long periods in order to identify trends and comparing the EU Member States.

Keywords: CMSA, competitiveness effect, export shares, structural effect, trade.

JEL Classification: F14, F43, L6, O52
AMS Classification: 62P20, 93D25

1 Introduction

Changes in a country's share of global export trade are a useful indicator of how well the economy can generate income to pay for imports. Changes in a country's market share in world exports can be influenced by many interrelated factors, such as competitiveness and specialization. Domestic and external macroeconomic features and developments can influence the relative competitiveness of exports. Structural factors, such as the endowment of productive factors or technology and geographical linkages, also condition the sectoral specialization of exports and its distribution among different trading partners [3]. If a country specializes in exports of goods (or towards areas) where demand is particularly buoyant, its aggregate market share will increase even if competitiveness does not improve. At the same time, even if a country maintains its share in individual markets, it can still have a decrease in its aggregate market share if the country is specialized in markets that grow more slowly than world exports or in products for which demand is growing more slowly than average [6]. Export analysis of the EU selected countries via constant market share analysis is the main focus of the paper.

2 Methodology and Data

The paper uses data of export from the *UNCTAD statistic database of international trade* which contains detailed annual nominal exports of goods data for all countries by commodity and partner country expressed in US dollars. Dataset is based on the Standard International Trade Classification (SITC) revision 3 at the 3-digit level which contains 255 product groups. Database also provides bilateral trade flows of goods data among all countries, i.e. bilateral trade exports of selected countries with other 225 states. Annual export shares and export growth of all countries and products were calculated for period 1995-2013 from dataset. This is the longest time period generated from UNCTAD therefore growth rates are calculated only for period 1996-2013.

The paper provides the Constant Market Share (CMS) analysis for Visegrad countries (Czech Republic, Hungary, Poland and Slovakia). Czech Republic is observed country in the following part to understand each component of CMS calculations. CMS presents the arithmetic breakdown on the countries' market share growth into two

[1] VŠB – TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, michaela.stanickova@vsb.cz.
[2] VŠB – TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, bohdan.vahalik@vsb.cz.
[3] VŠB – TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava 1, Czech Republic, lenka.fojtikova@vsb.cz.

components. The market (or competitiveness) effect contains factors reflecting changes in individual market share. The structural component reflects impact of export product specialization and geographical orientation. The total effect (TE) of countries' market share growth is result of difference between a country's export growth and export growth of the world in the given year. As it is expressed in (1), the total effect is difference between a country´s export growth and world´s export growth. If a country´s export growth (g) is higher than world´s export growth (g^*), country is increasing its share on world market and vice versa. For example, following [4], [1] or [5], TE, i.e. total change of Czech's exports worldwide is proxied by the difference between growth of total Czech export of goods and growth of total world export of goods.

$$TE = g - g^* = \sum_i \sum_j \theta_{ij} g_{ij} - \sum_i \sum_j \theta^*_{ij} g^*_{ij}, \tag{1}$$

where i refers product group, j refers country of trade partner, g_{ij} refers to annual export growth of product i to destination j, θ_{ij} represents market share of product i in total export of observed country to destination j, and θ_{ij} can be calculated as follows (2) and it is is share of product i in total Czech exports to destination j in time t-1:

$$\theta_{ij} = \frac{X_{ij,t-1}}{\sum_i \sum_j X_{ij,t-1}}. \tag{2}$$

Variable g_{ij}, expressed by (3), is percentage change of Czech exports of product i to country j in time t. It means that the $\theta_{ij}g_{ij}$ is e.g. calculated as share of product i in total Czech exports to destination j in 1995 multiplied by percentage change of Czech exports of product i to country j between 1995-1996.

$$g_{ij} = \frac{X_{ij,t} - X_{ij,t-1}}{X_{ij,t-1}}. \tag{3}$$

The same approach is used for second part of (1). Variable θ^*_{ij} is share of product i in total world´s exports to destination j in time t-1, see (4). It is important to note that in this case, total Czech exports must be excluded:

$$\theta^*_{ij} = \frac{X^*_{ij,t-1}}{\sum_i \sum_j X^*_{ij,t-1}}. \tag{4}$$

Variable g^*_{ij} represents percentage change of world´s export growth of product i to country j, see (5). Also in this case must be applied that growth of total world export is reduced by total Czech exports:

$$g^*_{ij} = \frac{X^*_{ij,t} - X^*_{ij,t-1}}{X^*_{ij,t-1}}. \tag{5}$$

If growth of Czech´s export is higher than growth of world´s export, TE will be positive and Czech Republic gains total market share and vice versa. TE, following [1], can be calculated as a sum of two effects. The first effect is the Market Share Effect (MSE) interpreted as indicator of competitiveness and it is result of effective changes in market shares of observed country. MSE is difference between growth rate of Czech exports and growth rate of world exports, see (6). Comparison is made between growth rates of Czech Republic and world exports for each product i to each destination country j multiplied by given product and geographical structure of Czech exports as the sum over i (j). As [5] mentioned, MSE captures extent of changes in shares that have been only due to changes in price and non-price competitiveness by abstracting from changes in product and geographical structures. That is why MSE is commonly referred as a competitiveness effect.

$$MSE = \sum_i \sum_j \theta_{ij} \quad g_{ij} - g^*_{ij} \quad . \tag{6}$$

The second effect is the Combined Structure Effect (CSE) and it is result of country´s product specialization and geographical concentration. CSE consist of the Product Structure Effect (PSE), the Geographical Structure Effect (GSE) and residual term called the Mixed Structure Effect (MIX), i.e. TE = MSE + CSE = MSE + PSE + GSE + MIX. CSE determines extent of changes in total effect of market share caused by relative product or geographical specialization of observed country. It captures relative evolution of each destination market in the world which is defined as difference between world export growth (g^*_{ij}) in product i to destination j and total world export growth (g^*) in the given year. This relative evolution is weighted by relative importance of destination j for observed country. This importance is expressed as difference between share of destination j for each product i in total country´s export (θ_{ij}) and total world export (θ^*_{ij}) to that destination j in product i. For example, CSE is positive if

Czech Republic is relatively more (less) specialized in individual markets that grow above (below) world´s average, or negative if Czech Republic is relatively less (more) specialized in individual markets that grow above (below) the world´s average, as it is expressed by (7):

$$CSE = \sum_i \sum_j \left(\theta_{ij} - \theta_{ij}^* \right) \left(g_{ij}^* - g^* \right) . \tag{7}$$

CSE takes into account product as well as geographical specialization of exports and it can be broken down further (CSE = PSE + GSE + MIX). PSE measures the contribution of Czech´s product composition to total effect of market share, see (8). It measures how relative product specialization of Czech Republic fits to more dynamic industries of the world demand. PSE is positive if Czech´s composition of exports is more concentrated in products which are demanded above the world average and vice versa.

$$PSE = \sum_i \left(\theta_i - \theta_i^* \right) \left(g_i^* - g^* \right) . \tag{8}$$

GSE measures contribution of Czech´s geographical export market composition to the total effect of market share, see (9). It means that it measures how relative trade partner specialization of Czech Republic is orientated to more dynamic markets of the world. GSE is positive if Czech´s composition of exports is more concentrated in trade partners which growth above world´s average and vice versa.

$$GSE = \sum_j \left(\theta_j - \theta_j^* \right) \left(g_j^* - g^* \right) . \tag{9}$$

MIX represents a residual term because product and geographical structures of each observed country are not independent, therefore the sum of PSE and GSE does not equals CSE. MIX is defined as follows (10), where the first element measures relative specialization of the Czech Republic by comparing Czech´s export composition with world export composition. As mentioned in [5], this is equivalent to the Revealed Comparative Advantage composed by [2]. The second element expresses relative product specialization of observed country. The third element expresses relative trade partner orientation.

$$MIX = \sum_i \sum_j \left[\left(\theta_{ij} - \theta_{ij}^* \right) - \left(\theta_i - \theta_i^* \right) \frac{\theta_{ij}^*}{\theta_i^*} - \left(\theta_j - \theta_j^* \right) \frac{\theta_{ij}^*}{\theta_j^*} \right] g_{ij}^* . \tag{10}$$

MIX components are explained in (11) – (16) using the example of Czech Republic again. Product composition of the Czech Republic is calculated as share of product *i* in Czech exports to all trading partners (11).

$$\theta_i = \sum_j \theta_{ij} \tag{11}$$

Product composition of the world as share of product *i* in world exports to all trading partners (12):

$$\theta_i^* = \sum_j \theta_{ij}^* . \tag{12}$$

Trading partner composition of the Czech Republic as share of each trading partner *j* in Czech exports in all commodities (13):

$$\theta_j = \sum_i \theta_{ij} . \tag{13}$$

Trading partner composition of the world as share of each trading partner *j* in world exports in all commodities (14):

$$\theta_j^* = \sum_i \theta_{ij}^* . \tag{14}$$

Growth rate of world exports of product *i* is calculated as ratio of share of product *i* in world exports to all trading partners multiplied by its growth rate and product composition of the world (15):

$$g_i^* = \frac{\sum_j \theta_{ij}^* g_{ij}^*}{\theta_i^*} . \tag{15}$$

Growth rate of world exports to destination *j* is calculated as ratio of share of destination *j* in world exports in all products multiplied by its growth rate and trading partner composition of the world (16):

$$g_j^* = \frac{\sum_i \theta_{ij}^* g_{ij}^*}{\theta_j^*}. \tag{16}$$

3 Results and discussion

Results show that Visegrad countries exhibit many common characteristics related to exports, which is obviously linked to their common economic development and subsequently level of competitiveness, especially within the EU. Visegrad countries have showed positive total effect in most years because their growth rate of export was higher than world´s average. Also in the most cases, Visegrad countries have a common trend in total effect of CMS as it can be seen in Figure 1, but their share of total export growth is different. Total effect of CMS was negative in 1996, 1999, 2010 and 2012 because of similar economic development. Negative total effect in 1996 can be caused by fading process in period when selected countries transitioned to the market type of economy. Negative total effect in other years was usually caused by economic crisis. Visegrad countries have reached also similar trend in export growth rate during the entire period. All countries show high positive effect during the 90´s as well as in period of preparation to the EU accession. During this period, total effect has already begun to decline. After 2008, all Visegrad countries show very low positive or negative total effect, i.e. their export growth have fluctuated around the world export growth. Although Visegrad countries show similar trend, there are some differences, e.g. Hungarian economy reached significantly higher export growth rates than other economies during the 90´s, but in other periods showed the lowest export growth rate. Slovakia managed to achieve considerably higher growth rate of exports than world´s average in pre-crisis period as well as in post-crisis period. The highest overlap of export growth rate of Visegrad countries was reached in pre-crisis period when it reached almost 10% over the world´s average. On the other hand, there was no overlap in post-crisis period.

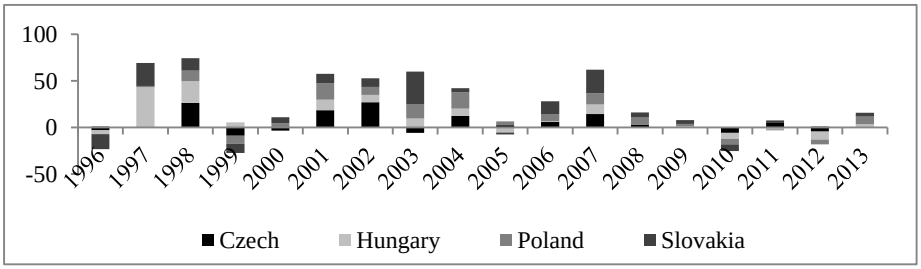


Figure 1 Total effect of the CMS Analysis

Comparison of main components of CMS analysis among all Visegrad countries is presented in Figure 2. It reveals that their export have had a common trend in total effect as well as in its single components. Visegrad countries export performance showed significant gains of market share during the entire period (top left picture). The fastest onset was recorded by Hungary which was leading until the advent of economic crisis. Hungarian cumulative total effect gains of total market share about 110%. Since that, Hungarian stagnated and lost part of its market share. Rest of Visegrad countries have reached closely related development of market share achievement until 2006 when their cumulative gain ranged 70-90%. Before the economic crisis, the cumulative gains started to diverged. Whereas the Czech Republic and Poland growth of market share stagnated in range of 90-100%, Slovakia have continued to achieve higher market share of reached almost 130% in 2013 in comparison with 1996. Export competitiveness, measured by market share effect, is pictured in top right figure. There can be seen, that major part of total effect of CMS Analysis is caused by market share (competitiveness) effect resulting from influence of relative specialization of Visegrad countries and thus it had major impact on result of total CMS effect. This is consistent with findings [6] where major contribution of competitiveness effect is mentioned as well. Growth of this indicator provides information about ability of economy to gain higher market share due to changes in price and non-price competitiveness factors. It is very positive results for Visegrad countries. Their growth of competitiveness was the main factor of their market share increase. Detailed view on top right figure shows that Hungary achieved fast growth of competitiveness, but it gains of competitiveness stagnated since 2004. The same happened to the Czech Republic and Poland since 2008. Slovakia has increased its market share due to increasing competitiveness during the entire period, even in the post-crisis period.

Bottom parts of Figure 2 compare contributions of product (bottom left) and geographical (bottom right) structure effect of CMS Analysis. As it was already said, PSE is positive if country´s composition of exports is more concentrated in products which are demanded above the world average and vice versa. Bottom left figure shows that all Visegrad countries had negative contribution to export performance. Cumulative loss of product market share reached of about 10-25%. It means that their export product structure did not correspond to products where world´s export growth was more dynamic during the entire period. Product effect was positive only in four years

during period 1996-2013. The most negative product effect was reached by the Czech Republic, the least by Slovakia. The biggest losses of export share of products were recorded in years of economic crises (1998-1999, 2008. 2010-2011) in all Visegrad countries. These results show very bad progress in the shift of product export structure towards more dynamic products. It is strongly connected with industrial structure of Visegrad countries which are strongly specialized on automotive industry and less focused on more dynamic sectors, especially high-tech industry. Geographical structure effect contributed the most to Slovakia because it was positive almost during the entire period. At Figure 2 can be seen that Slovakia was changing its geographical export composition during 90´s, as well as other Visegrad countries, towards more dynamic markets. It subsequently caused growth of geographical effect during following years and its contribution to total effect of market share. There can be also seen that since beginning of the economic crisis in 2008, all countries lost significantly their geographical structure effect because their export is focused on the EU market which lost its dynamics. Geographical structure effect of Slovakian export growth reached the same value at the end of selected period mainly due to the economic crisis. Other Visegrad countries decreased their geographical structure effect by 5-15% since 2008 in comparison with 1996.

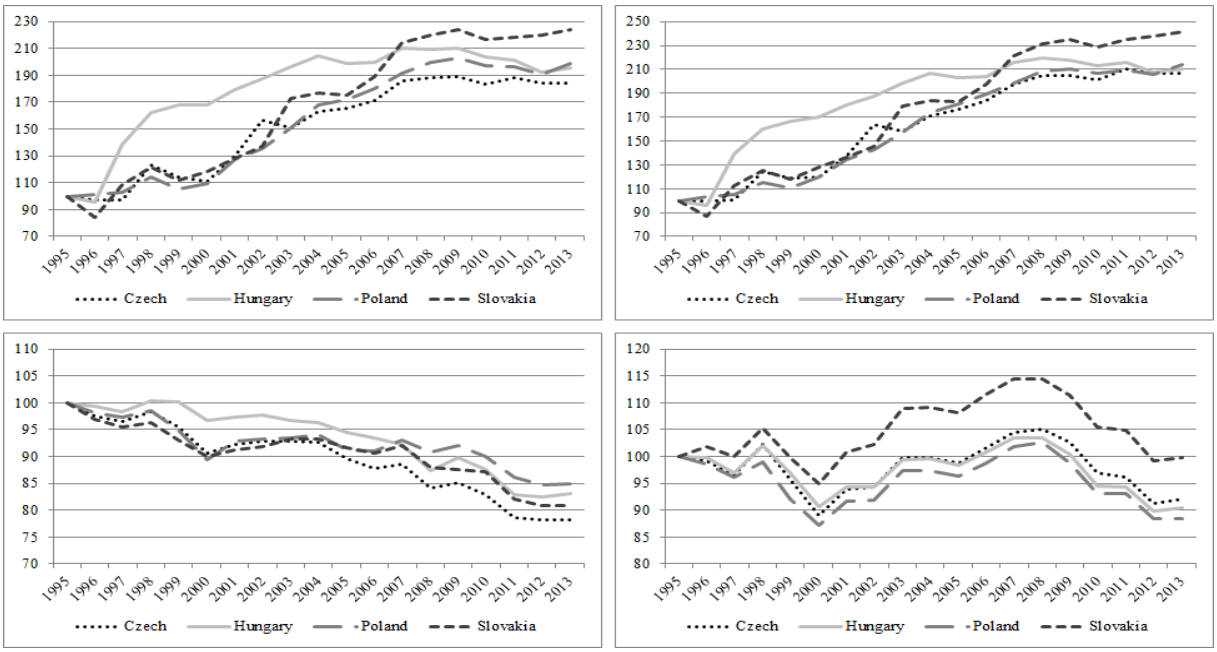


Figure 2 Cross-country CMS Analysis results; index 1996 = 100, cumulative results

4 Conclusion

Based on CMS analysis was found out that all countries gained high profit from market (competitiveness) structure effect due to their transformation and increasing of export competitiveness (see Figures in Annex). Share of combined structure effect is much lower, but there can be seen significant differences among Visegrad countries in decomposition of market share total effect. While Hungary gained main profit from competitiveness structure effect during the second half of 90´s and its contribution was significantly lower in following period, Poland gained major competitiveness effect before and after the EU accession. Evolution of total effect was considerably changeable in the Czech Republic and Slovakia during the entire period, but also in this case competeveness structure effect has dominated. Combined structure effect is composed of product, geographical and mixed effect. Visegrad countries have similar trend in all parts of CSE because nearly all components show positive and negative effect in each year. Product structure effect of all countries has very low positive values in some years but mostly it was negative. That means that Visegrad countries export product composition is more concentrated in products whose demand is under the world average. Geographical structure effect development is very similar. Visegrad countries have got common trend during the entire period. Also in this case, Visegrad countries exhibit big changes between positive and negative values but average for whole period is gently negative, i.e. Visegrad countries had more orientated export on less dynamic markets then it was world average.

Acknowledgements

Supported by the SGS project (SP2015/106) of Faculty of Economics, VŠB-Technical University of Ostrava and the Operational Programme Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Amador, J. and Cabral S.: The Portuguese Export performance in Perspective: A Constant market Share Analysis. In: *Banco de Portugal Economic Bulletin*, Banco de Portugal, Lisbon, 2008, 201-221.

[2] Balassa, B.: Trade Liberalization and Revealed Comparative Advantage, The Manchester School of Economic and Social Studies **33(2)** (1965), 99-123.

[3] Melecký, L.: Use of DEA Approach to Measuring Efficiency Trend in Old EU Member States. In: *Proceedings of XI. International Scientific Conference Liberec Economic Forum 2013*. Technical University of Liberec, Liberec, 2013, 381-390.

[4] Nyssens, A. and Poullet G.: *Parts de marché des producteurs de lÚEBL sur les marches extérieurs et intérieur, Cahier 7*, Banque Nationale de Belgique.

[5] Pandiella, A.G.: A Constant Market Share Analysis of Spanish Goods Exports, *OECD Economics Department Working Papers* **1186** (2015).

[6] Pavličková V.: The Competitiveness of Slovak Foreign Trade in the European Market. *Economic Annals* **196** (2013), 7-49.

Annex

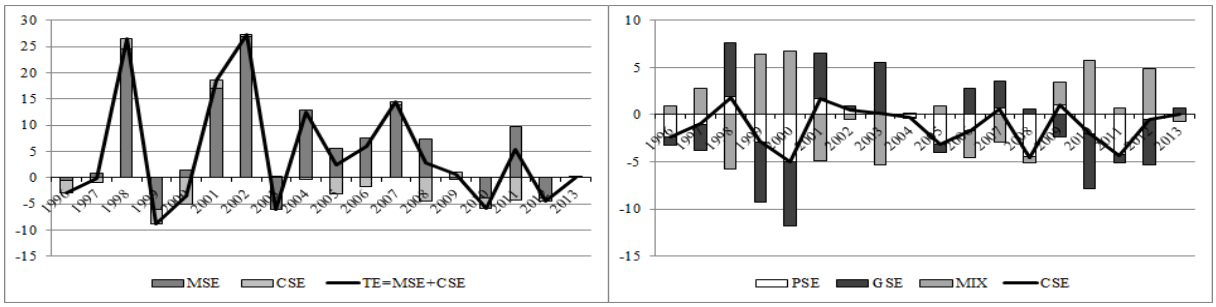


Figure 1 Main result of a CMS Analysis for Czech Republic

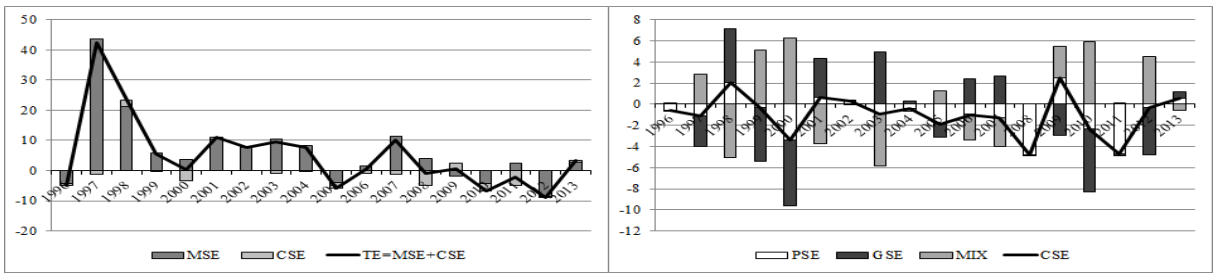


Figure 2 Main result of a CMS Analysis for Hungary

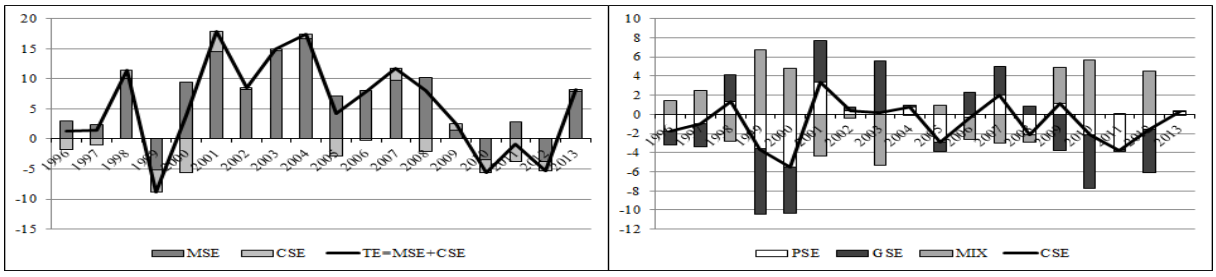


Figure 3 Main result of a CMS Analysis for Poland

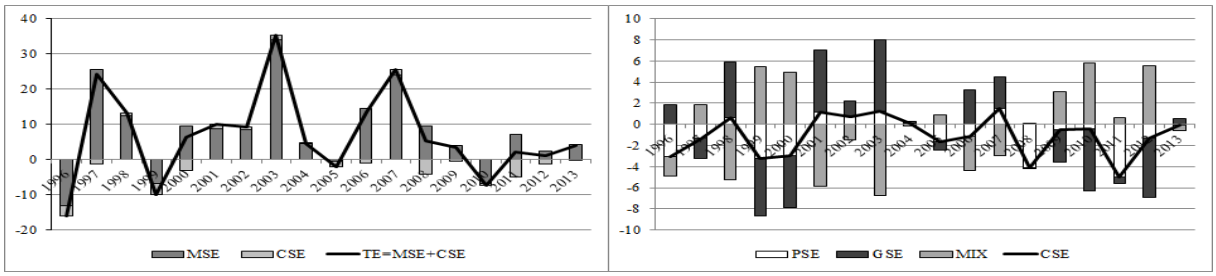


Figure 4 Main result of a CMS Analysis for Slovakia

Soft consensus model under linguistically labelled evaluations

Jan Stoklasa [1], Vojtěch Sukač[2], Tomáš Talášek[3], Jana Talašová[4]

Abstract. In group decision making and evaluation, the ability to reach an overall consensus of a majority of most relevant experts (decision makers) is often an important issue. Mathematical models capable of providing insights into the process of consensus reaching and also the resulting consensual evaluations are therefore studied in detail in the literature. In this paper we recall the soft consensus model proposed in 1986 by Kacprzyk and Fedrizzi and further developed by these authors in the following years. The idea of consensus is reflected in the concept of a majority of experts that are important enough agreeing on the evaluation (e.g. recommendation of a given alternative). In the multiple criteria group decision making model proposed in this paper we assume that from a given group of experts each expert provides evaluations of all the alternatives with respect to all the criteria in linguistic terms (with meanings modelled by fuzzy numbers). An overall fuzzy evaluation of each alternative by each expert is computed using fuzzy weighted average. The competences of the experts are assessed using predefined linguistic labels (with meanings modelled by fuzzy numbers) and a soft consensus is sought. The model finds a subset of the set of alternatives containing those alternatives that are evaluated "good enough" by a "sufficient amount" of "important" experts. The group evaluations of the alternatives from this set are computed from their overall fuzzy evaluations provided by the individual experts by fuzzy weighted average; fuzzy numbers representing the competences of the experts are used as weights. The best-evaluated alternative is then selected using e.g. a defuzzification by the center of gravity method.

Keywords: Soft consensus, fuzzy, linguistic modelling, evaluation.

JEL classification: C44
AMS classification: 90B50, 91B74

1 Introduction

Decision making is an integral part of our everyday lives and in the management science (and management practice in general) also a key aspect in the well being of companies. In situations, where many alternatives and many criteria need to be considered, when the description of the problem and its structure is very complex or when time constraints exist, mathematical models for decision support provide a useful tool to reach good decisions. When there is much at stake, when objectivity is required or when more points of view need to be considered, multiple people (so called experts) are involved in the decision making process [1]. This allows the problem to be approached in a complex way, but also introduces possible conflicts of interests, distortions of evaluations or generally speaking dissensus in the decision making process. Also reluctance of the experts to change their originally declared opinions in the decision making progress can affect the outcome of the decision making. The path to the group consensus can be seen as an iterative process (see e.g. [6] for the Delphi method) where in each step all the experts consider the evaluations suggested by all their fellow decision-makers, hear their arguments and decide whether to modify the evaluation they suggest themselves. Even in such a situation the importance of the experts (e.g. their

[1]Palacký University, Olomouc, Faculty of Science, Dept. of Math. Analysis and Applications of Mathematics, 17. listopadu 1192/12, 77146, Olomouc, Czech Republic, jan.stoklasa@upol.cz
[2]dtto, vojta.sukac@gmail.com
[3]dtto, tomas.talasek@upol.cz
[4]dtto, jana.talasova@upol.cz

education, expertise, power, ...) can be taken into account. Ideally this process converges to a situation, where no expert is willing to change his/her opinion/evaluation any more. We can not expect that an overall agreement is found this way. As suggested by Kacprzyk and Fedrizzi [2] and further elaborated by e.g. Kacprzyk, Fedrizzi and Nurmi [4] a softer version of agreement - the so called soft consensus can be found as an evaluation or alternative that is agreed upon by a sufficient amount of sufficiently important experts. This approach is adopted in the presented paper.

This paper deals with the following multiple criteria group decision-making problem - we assume that a group of experts is to evaluate a set of alternatives against a set of criteria and together choose the most suitable alternative (in the soft consensus sense outlined in the previous paragraph). Within the spirit of linguistic fuzzy modelling (see e.g. [8] for more details) the evaluations are provided using linguistic terms (values of a linguistic variable), each expert is assigned a competence level represented by a fuzzy number and the quantifiers describing the relative size of the subgroup of experts with close evaluations are also defined using linguistic variables. At first, each expert evaluates each alternative with respect to each criterion by a linguistic term represented in the model by a fuzzy number. An overall evaluation is computed for each expert and each alternative using the fuzzy weighted average with fuzzy weights (see e.g. [7]). The proposed approach allows for different sets of weights of criteria to be used by different experts. The experts are assigned various decision competences expressed again linguistically with a fuzzy-number representation in the mathematical model. Consensus reaching in this model is based on the concept of finding the best alternative not among all alternatives, but among those that are sufficiently good according to a sufficient amount of relevant experts.

To introduce the novel multiple criteria fuzzy group decision making model, we first recall the most relevant concepts of fuzzy set theory and linguistic fuzzy modelling. Then we describe the proposed model step by step. A short discussion concerning the proposed model and the results that can be provided by it follows and the courses of further development of the model, its validation and related issues are summarized in the end.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $\mu_A : U \to [0,1]$. For each $x \in U$ the value $\mu_A(x)$ is called a *membership degree* of the element x in the fuzzy set A and μ_A is called a *membership function* of the fuzzy set A. The symbol $\mathcal{F}(U)$ denotes the family of all fuzzy sets on U. The *height of* a fuzzy set A is a real number $\mathrm{hgt}(A) = \sup_{x \in U}\{\mu_A(x)\}$. In the case when a fuzzy set A on U is used to represent a meaning of a linguistic term we require the fuzzy set A to be normal, that is $\mathrm{hgt}(A) = 1$. This means that there exists at least one such $x \in U$ that is fully compatible with the respective linguistic term (it is well described by it). Other important concepts related to fuzzy sets are: a) the *kernel* of A, $\mathrm{Ker}(A) = \{x \in U \mid \mu_A(x) = 1\}$; b) the *support* of A, $\mathrm{Supp}(A) = \{x \in U \mid \mu_A(x) > 0\}$; and c) for $\alpha \in [0,1]$ the *α-cut* of A, $A_\alpha = \{x \in U \mid \mu_A(x) \geq \alpha\}$. If the support of A is a finite set, $\mathrm{Supp}(A) = \{x_1, \ldots, x_k\}$, then the fuzzy set A will be described as $A = \{{}^{\mu_A(x_1)}\!/_{x_1}, \ldots, {}^{\mu_A(x_k)}\!/_{x_k}\}$. The cardinality of such a fuzzy set A is given by $Card(A) = \sum_{i=1}^{k} A(x_i)$. A *union* of fuzzy sets A and B on U is a fuzzy set $A \cup B$ on U with the membership function for all $x \in U$ given by $\mu_{(A \cup B)}(x) = \max\{\mu_A(x), \mu_B(x)\}$. An *intersection* of fuzzy sets A and B on U is a fuzzy set $A \cap B$ on U with the membership function for all $x \in U$ given by $\mu_{(A \cap B)}(x) = \min\{\mu_A(x), \mu_B(x)\}$.

Let V be a nonempty subset of $\mathbb{R}$. *A fuzzy number* N is fuzzy set on the universe V which fulfills the following conditions: i) $\mathrm{Ker}(N) \neq \emptyset$, ii) N_α are closed intervals for all $\alpha \in (0,1]$ and iii) $\mathrm{Supp}(N)$ is bounded. The family of all fuzzy numbers on V is denoted by $\mathcal{F}_N(V)$. Each fuzzy number N is determined as $N = \left\{[\underline{N}(\alpha), \overline{N}(\alpha)]\right\}_{\alpha \in [0,1]}$, where $\underline{N}(\alpha)$ and $\overline{N}(\alpha)$ is lower and upper bound of the α-cut of fuzzy number N respectively, for $\alpha \in (0,1]$, and the closure of the support of N, $\mathrm{Cl}(\mathrm{Supp}(N)) = [\underline{N}(0), \overline{N}(0)]$. A *trapezoidal fuzzy number* N is determined by an ordered quadruple $(n^1, n^2, n^3, n^4) \subset V^4$ of significant values of N satisfying $(n^1, n^4) = \mathrm{Supp}(N)$ and $[n^2, n^3] = \mathrm{Ker}(N)$. The membership function

of trapezoidal fuzzy number N is defined by (1). N is called a triangular fuzzy number if $n^2 = n^3$.

$$\mu_N(x) = \begin{cases} 0 & \text{if } x < n^1, \\ \frac{x-n^1}{n^2-n^1} & \text{if } n^1 \leq x < n^2, \\ 1 & \text{if } n^2 \leq x \leq n^3, \\ \frac{n^4-x}{n^4-n^3} & \text{if } n^3 < x \leq n^4, \\ 0 & \text{if } x > n^4. \end{cases} \tag{1}$$

To obtain a real-number-representation of a fuzzy number $N \in \mathcal{F}_N([a,b])$ the *center of gravity* of the fuzzy number can be computed using (2).

$$T_N = \frac{\int_a^b x\mu_N(x)\mathrm{d}x}{\int_a^b \mu_N(x)\mathrm{d}x} \tag{2}$$

This paper utilizes the linguistic fuzzy modelling approach (see e.g. [8]) to deal with group decision-making problems. The key term of linguistic modelling is the *linguistic variable* introduced by Zadeh [9], which is defined by the quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), V, G, M)$, where $\mathcal{V}$ is the name of the linguistic variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic values, $V \subset \mathbb{R}$ is the universe on which the fuzzy numbers representing meanings of these linguistic terms are defined, G is a syntactic rule for generating the linguistic values of $\mathcal{V}$, M is a mapping that assigns to each linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ its meaning $A = M(\mathcal{A}) \in \mathcal{F}_N(V)$.

3 Proposed multiple criteria group decision-making model

In the following text we consider the following multiple-criteria group decision making situation: we suppose a set $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ of $n \geq 2$ alternatives is given, each alternative X_i, $i = 1, \ldots, n$ is evaluated by p experts $\boldsymbol{E} = \{E_1, \ldots, E_p\}$, $p \geq 2$, according to a set of criteria $\boldsymbol{C} = \{C_1, \ldots, C_m\}$, $m \geq 2$. We also assume that the competences of experts are given by fuzzy numbers $L^k \in \mathcal{F}_N([0,1])$, $k = 1, \ldots, p$ (in this case 0 can be interpreted as a completely incompetent expert and 1 means a fully competent expert). A linguistic variables with a linguistic term set {*extremely low competence, very low competence, low competence, average competence, high competence, very high competence, extremely high competence*} can be used to assign competences to experts; an example of the meanings of these linguistic terms is provided in Figure 1 a).

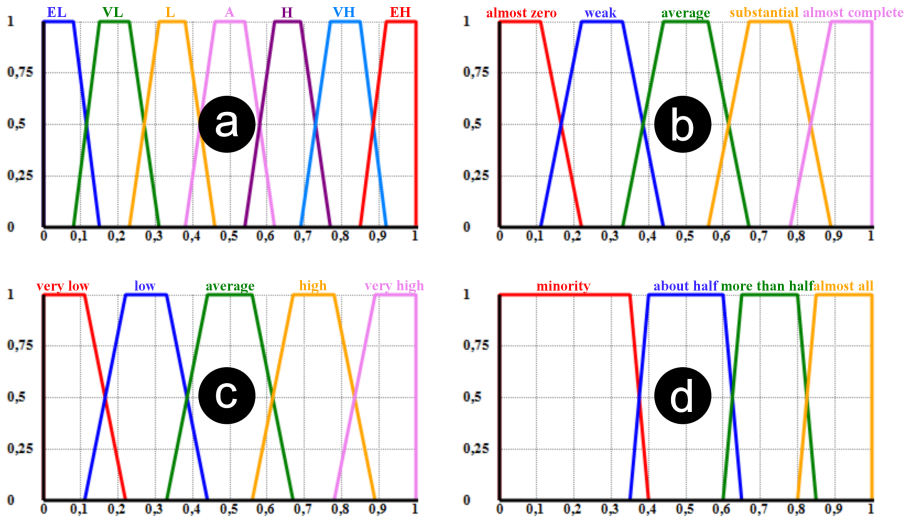


Figure 1 Meanings of the linguistic terms expressing: a) the competence of experts, b) the fulfillment of partial goals, c) the importances of criteria, d) the meanings of the linguistic quantifiers.

Expert E_k, $k = 1, \ldots, p$ evaluates each alternative X_i, $i = 1, \ldots, n$ according to each criterion C_j, $j = 1, \ldots, m$ by a fuzzy number $H_{ij}^k \in \mathcal{F}_N([0,1])$ which expresses the level of fulfillment of the partial goal

corresponding to the given criterion C_j. A linguistic variable with the linguistic term set {*almost zero fulfillment, weak fulfillment, average fulfillment, substantial fulfillment, almost complete fulfillment*} can be used by the experts to express the fulfillment of the partial goals by the alternatives. The meanings of these linguistic terms can be e.g. represented by the fuzzy numbers in Figure 1 b).

Each expert E_k, $k = 1, \ldots, p$, assigns a weight $W_j^k \in \mathcal{F}_N([0,1])$ to each criterion C_j, $j = 1, \ldots, m$. This weight expresses the importance of criterion C_j perceived by the given expert. The sets of weights of the criteria can be different for each expert. To facilitate the decision making process, a linguistic variable describing the importance of the criteria with the linguistic term set {*very low importance, low importance, average importance, high importance, very high importance*} can be provided to all the experts (possible meanings of the respective linguistic terms are depicted in Figure 1 c)).

3.1 Evaluations by individual experts

When an expert E_k completes the assessment of an alternative X_i with respect to all the criteria, an overall evaluation of the alternative by the given expert H_i^k needs to be determined. These overall evaluations $H_i^k \in \mathcal{F}_N([0,1])$ of alternative X_i, $i = 1, \ldots, n$ according to expert E_k, $k = 1, \ldots, p$ are computed in the form of (3), where the lower and upper bounds of the α-cuts $\underline{H}_i^k(\alpha)$ and $\overline{H}_i^k(\alpha)$ are computed using (4) and (5) respectively.

$$H_i^k = \left\{ \left[\underline{H}_i^k(\alpha), \overline{H}_i^k(\alpha) \right] \right\}_{\alpha \in [0,1]} \tag{3}$$

$$\underline{H}_i^k(\alpha) = \min_{w_j^k \in [\underline{W}_j^k(\alpha), \overline{W}_j^k(\alpha)], j=1,\ldots,m} \frac{\sum_{j=1}^m w_j^k \cdot \underline{H}_{ij}^k(\alpha)}{\sum_{j=1}^m w_j^k} \tag{4}$$

$$\overline{H}_i^k(\alpha) = \max_{w_j^k \in [\underline{W}_j^k(\alpha), \overline{W}_j^k(\alpha)], j=1,\ldots,m} \frac{\sum_{j=1}^m w_j^k \cdot \overline{H}_{ij}^k(\alpha)}{\sum_{j=1}^m w_j^k} \tag{5}$$

The overall evaluations H_i^k express the level of fulfillment of the overall goal by alternative X_i according to expert E_k. The final decision - the selection of the collectively preferred alternative - is reached based on the idea that the "optimal" alternative should be such that is good enough according to a sufficient amount of important experts. The experts specify the acceptability levels of the alternatives (acceptability is assessed based on the overall fuzzy evaluation of an alternative). This is done by introducing a linguistic variable with the linguistic term set $\mathcal{T}(Acc)$ = {*excellent, good, acceptable, borderline, unacceptable*} = $\{\widehat{\mathcal{A}}_1, \ldots, \widehat{\mathcal{A}}_5\}$ with meanings represented by trapezoidal fuzzy numbers $\widehat{A}_r = (\widehat{a}_r^1, \widehat{a}_r^2, \widehat{a}_r^3, \widehat{a}_r^4)$, $r = 1, \ldots, 5$, on the evaluation universe $[0,1]$. The terms $\mathcal{A}_1, \ldots, \mathcal{A}_5$ expressing *"at least* $\widehat{\mathcal{A}}_r$*"*, for all $r = 1, \ldots, 5$, are subsequently added into $\mathcal{T}(Acc)$, with meanings defined by (6).

$$\mu_{A_r}(x) = \begin{cases} 0 & \text{if } x < \widehat{a}_r^1, \\ \frac{x - \widehat{a}_r^1}{\widehat{a}_r^2 - \widehat{a}_r^1} & \text{if } \widehat{a}_r^1 \le x < \widehat{a}_r^2, \\ 1 & \text{if } x \ge \widehat{a}_r^2, \end{cases} \qquad r = 1, \ldots, 5 \tag{6}$$

Levels of acceptance θ_{ri}^k of alternative X_i, by expert E_k are calculated for each overall evaluation H_i^k using (7). These numbers can be interpreted as truth values of the statement: *"the acceptance of alternative* X_i *by expert* E_k *is* $\mathcal{A}_r$*"*.

$$\theta_{ri}^k = \mathrm{hgt}\big(A_r \cap H_i^k\big) \tag{7}$$

3.2 Aggregation of the individual overall evaluations - a soft consensus perspective

Since we are dealing with a multi-expert (group) decision making problem, neither the individual overall evaluations nor the levels of acceptance of each alternative by each expert are a sufficient output. The group perspective now needs to be taken into account. We require a promising alternative to be evaluated well enough according to a sufficient amount of important experts. First let us define linguistic quantifiers to assess the quantity of the experts. We will consider the linguistic quantifier set $\{\widehat{\mathcal{Q}}_1, \ldots, \widehat{\mathcal{Q}}_4\}$={*almost*

all, more than half, about half, minority}; see Figure 1 d) for their possible meanings. We also need to specify what the linguistic expression *important expert* means. This can be done by defining the meaning of this linguistic term by a trapezoidal fuzzy number $M(\textit{important expert}) = IE = (0.3, 0.8, 1, 1)$, $IE \in \mathcal{F}_N([0,1])$.

For the group aggregation, it is necessary to determine the importance level ζ_k of each expert, $k = 1, \ldots, p$. If the competence of the expert E_k is L^k, then ζ_k can be computed by (8). A fuzzy set of important experts can then be defined as $B = \{{}^{\zeta_1}\!/_{E_1}, \ldots, {}^{\zeta_p}\!/_{E_p}\}$.

$$\zeta_k = \mathrm{hgt}\big(IE \cap L^k\big),\ k = 1, \ldots, p \tag{8}$$

The meanings of the linguistic quantifiers $\{\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4\}$ expressing *at least* $\widehat{\mathcal{Q}}_s$, $M(\textit{at least } \widehat{\mathcal{Q}_s}) = M(\mathcal{Q}_s) = Q_s$, $Q_s \in \mathcal{F}_N([0,1])$, $s = 1, \ldots, 4$, can be defined analogously to the meanings of $\mathcal{A}_r$ (see (6)).

In accordance with [4] we now need to find the truth value of the statement "*A given quantity Q_s of important experts B suggest a given evaluation A_r of the alternative X_i*"for all $s = 1, \ldots, 4$, $r = 1, \ldots, 5$ and $k = 1, \ldots, p$. That is we need to determine $\xi_i^{r,s} = \mathrm{truth}\big(Q_s\ B \text{ experts suggest } A_r\big)$, which can be rewritten in the form of $\xi_i^{r,s} = \mathrm{truth}\big(Q_s\ (B \text{ and suggests } A_r) \text{ experts are } B\big)$. Suppose that at least one of the experts has nonzero importance level. Then $\xi_i^{r,s}$ can be computed by (9).

$$\xi_i^{r,s} = \mu_{Q_s}\left(\frac{\sum_{k=1}^{p} \min\{\zeta_k, \theta_{ri}^k\}}{\sum_{k=1}^{p} \zeta_k}\right),\ r = 1, \ldots, 5,\ s = 1, \ldots, 4 \tag{9}$$

For all r, s the set $\Upsilon_{r,s}$ is defined by (10), which includes such alternatives which are $\mathcal{A}_r$ according to $\mathcal{Q}_s$ of important experts, that is a set of alternatives for which the statement "*The quantity Q_s of important experts B suggest the evaluation A_r for the alternative*"is true.

$$\Upsilon_{r,s} = \Big\{X_i \in \boldsymbol{X} \,\Big|\, \xi_i^{r,s} = 1\Big\}. \tag{10}$$

The decision makers then go through these sets $\Upsilon_{r,s}$ in a predefined order $\mathcal{K}$ until the first nonempty set is found. The order $\mathcal{K}$ is a given sequence of pairs $\big\{(r,s)_t\big\}$ such that the pair $(r,s)_t$ should be checked for nonemptyness before the pair $(r,s)_{t'}$, for $t < t'$. The order $\mathcal{K}$ can be specified e.g. in the following way $\mathcal{K} = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(1,3),(2,3),\ldots\big\}$. This order can be rewritten using linguistic terms as " *check first (excellent by almost all) then (excellent by majority), (good or better by almost all), (good or better by majority), (acceptable or better by almost all), ...".* The first non-empty set $\Upsilon_{r,s}$ according to $\mathcal{K}$ (denoted Υ^*) includes those alternatives, among which the optimal one should be chosen.

The sequence $\mathcal{K}$ does not need to include all possible combinations of r and s. Combinations that are deemed uninteresting from the decision making point of view can be omitted - it is sufficient to consider only relevant combinations (e.g. (borderline by majority) can be omitted, as such alternatives are definitely not good ones to consider). It is therefore possible that there is no nonempty $\Upsilon_{r,s}$ to be found. In this case, no alternative is recommended (none of the alternatives is considered good enough – in the consensus sense – by the experts).

In case of a nonempty Υ^*, the fuzzy weighted average with fuzzy weights L^k, $k = 1, \ldots, p$ can be applied to aggregate experts' evaluations of alternatives H_i^k, $i : X_i \in \Upsilon^*$. This way group evaluations of the alternatives from Υ^* can be computed in the form of fuzzy numbers on $[0,1]$. As all these alternatives are sufficiently good according to a large enough number of important experts, an easy selection based e.g. on the centers of gravity of the resulting group evaluations can be performed.

4 Conclusion

In this paper we have considered multiple criteria decision making problems under group decision making in the framework of consensus reaching. We have proposed a new linguistic fuzzy modelling approach that uses linguistic evaluations of the alternatives against criteria and adopts the soft-consensus approach. This way a consensual solution of the multiple-criteria group decision making problem is an alternative, that is sufficiently well evaluated by a sufficiently large subgroup of the experts, moreover the importance

of the experts is taken into account. As such the proposed model first performs a pre-selection of the alternatives and finds those, that comply with this definition of soft consensus. In the next step the overall group of the evaluations of these alternatives are computed from individual evaluations provided by the experts and using the experts' competences as weights (the fuzzy weighted average aggregation is applied).

The consensus approach obviously is not the only one to multiple-criteria group decision making problems. The group evaluations could e.g. be computed first and the alternative with the highest group evaluation chosen. The soft consensus perspective provides not only means for the reflection of expert competences, but also requires agreement among the experts on the evaluations (particularly when the competences/weights of the experts are similar). Further research will include the performance assessment of the proposed group MCDM model on practical problems and a further study of the dynamics of consensus reaching among experts and its reflection in the model.

Acknowledgements

Supported by the grant IGA PrF 2015 013 Mathematical models of the Internal Grant Agency of Palacký University, Olomouc and by the grant GA14-02424S Methods of operations research for decision support under uncertainty of the Grant Agency of the Czech Republic.

References

[1] Eklund, P., Rusinowska, A., De Swart, H.: Consensus reaching in committees. *European Journal of Operational Research*, 178, 2007, 185–193.

[2] Kacprzyk, J. and Fedrizzi, M.: 'Soft'consensus measures for monitoring real consensus reaching processes under fuzzy preferences. *Control and Cybernetics* **15** (1986), 309–323.

[3] Kacprzyk, J. and Roubens, M., Eds.: *Non-Conventional Preference Relations in Decision Making*, Springer-Verlag, 1988.

[4] Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1997): 'Soft' degrees of consensus under fuzzy preferences and fuzzy majorities. In J. Kacprzyk et al. (eds.): *Consensus under Fuzziness*, Springer Science+Business Media, New York, 1997, 55–83.

[5] Klamler, C.: The Dodgson ranking and the Borda count: a binary comparison. *Mathematical Social Sciences*, **48** (2004), 103–108.

[6] Linstone H. A., Turoff, M. (Eds.): *The Delphi Method: Techniques and Applications.* Addison-Wesley Pub. Co, 1975.

[7] Pavlačka, O.: Modeling uncertain variables of the weighted average operation by fuzzy vectors. *Information Sciences* **181** (2011), 4969–4992.

[8] Stoklasa, J.: *Linguistic models for decision support.* Lappeenranta: Lappeenranta University of Technology, 2014.

[9] Zadeh, L., A.: Fuzzy Sets. *Information and control*, **8**(3), (1965), 338–353.

[10] Zadeh, L., A.: The concept of a linguistic variable and its application to approximate reasonning-I. *Information Sciences*, **8**(3), (1975), pp. 199–249.

On Robust Testing for Normality of Error Terms in Regression Models

Luboš Střelec[1]

Abstract. Testing for normality of error terms constitutes one of the most important steps of regression model verification and validation, because failure to assess non-normality of the regression residuals may lead to incorrect results since significant deviations from normality can substantially affect the performance of usual statistical inference techniques. Thus, in the majority of cases of relevant regression analysis normality of error terms is expected. However, this is not true in many practical situations. While OLS estimator is known to be very sensitive to outliers, the robust regression estimator (e.g. Least Trimmed Squares, LTS) is known not to be unduly affected by the presence of outliers.

The aim of this paper is to present and discuss the trade-off between power and robustness of selected classical and robust normality tests of error terms in regression models. For this purpose we use OLS and LTS residuals from linear regression models with various distributed dependent variable including models with presence of outliers. To explore the power and robustness of selected classical tests and robust tests of general RT class tests for normality, we perform simulation study.

Keywords: Error terms, Least Trimmed Squares, linear regression model, Ordinary Least Squares, robustness, RT class tests, testing for normality.

JEL classification: C12, C15, C52, C63
AMS classification: 11K45, 62J05, 62J20, 62F03

1 Introduction

As it is generally known, the regression analysis is a very popular tool in econometrics. Within the framework of verification and validation of regression models, testing for normality constitutes one of the most important issues. In the majority of cases of relevant analysis, it is expected that residuals represent Gaussian distributed random variable with constant expected value and constant variance – however, this is not true in many practical situations of economic and financial issues, where presence of outliers is typical phenomena. Presence of outliers in data set as well as the other significant deviations from normality may lead to incorrect results of usual statistical inference techniques used in linear regression models. Thus, the normality assumption of the regression residuals is very important. Moreover, the Ordinary Least Squares (OLS) estimator is known to be very sensitive to outliers and non-normal errors. In contrast, the robust least squares (e.g. Least Trimmed Squares, LTS) estimator is known not to be unduly affected by the presence of outliers [8].

Hence, the aim of this contribution is to present and discuss the trade-off between power and robustness of selected classical and robust normality tests of error terms in regression models. For this purpose we take advantages of p-location outlier models introduced by [2]. The main advantage of the p-location outlier model is that this type of models simulates presence of outliers, typical phenomena for economic and financial data sets. The reason is also the fact that in presence of outliers, a mean may reflect a misleading center of the data. Subsequently, the variance and standard deviation are then strongly affected by outliers since their computation is based on the average of the squared differences from the mean [2]. As a consequence, normally distributed stochastic errors are necessary in order to make a not misleading inferences which explains a necessity and importance of robust tests of normality [14].

[1]Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, lubos.strelec@mendelu.cz

The rest of paper is organized as follows. In the second part, the general RT class of tests for normality is introduced. In the third part, we present and discuss some interesting results of power of selected classical and robust tests for normality of error terms in regression models. In the fourth part, illustrative example using regression data with outlier is presented. Finally, the last section is Summary.

2 General RT class tests for normality

The general RT class is based on robustification of the classical Jarque-Bera test introduced by [6]. The general RT class test statistic is defined by [13] for purpose of robust testing for normality against Pareto tails and has the following form

$$RT = \frac{n}{C_1}\left(\frac{M_{j_1}^{\alpha_1}(r_1, T_{(i_1)}(s_1))}{M_{j_2}^{\alpha_2}(r_2, T_{(i_2)}(s_2))} - K_1\right)^2 + \frac{n}{C_2}\left(\frac{M_{j_3}^{\alpha_3}(r_3, T_{(i_3)}(s_3))}{M_{j_4}^{\alpha_4}(r_4, T_{(i_4)}(s_4))} - K_2\right)^2,$$

where M_j are jth theoretical central moment estimators of the random variable defined as $M_j(r, T(F_n, s)) = \frac{1}{n-2r}\sum_{m=r+1}^{n-r}\varphi_j(X_{(m)} - T(F_n, s))$ for $j \in \{0, 1, 2, 3, 4\}$, where φ_j is a tractable and continuous function where $\varphi_0(x) = \sqrt{\pi/2}|x|$ and $\varphi_j(x) = x^j$ for $j \in \{1, 2, 3, 4\}$, $X_{(m)}$ is the order statistic, $T(F_n, s)$ is a location functional applied to the sample $X_1, X_2, \ldots, X_n$, r and s are the trimming constants for moments and location, respectively, K_1 and K_2 are small-sample variants of mean corrections, C_1 and C_2 are asymptotic constants, and finally, α_1, α_2, α_3 and α_4 are exponents.

For the construction of the general RT class test statistic it was used location functional approach, which has been introduced by P. E. Bickel and E. L. Lehmann in a series of papers (e.g. [3]) and as was shown in [13] and [14] it looks to be playing a crucial role also by robust testing for normality. In the general RT class we used the following four different location estimators: mean ($T_{(0)} = \frac{1}{n}\sum_{i=1}^{n} X_i$), median ($T_{(1)} = F^{-1}(1/2)$), trimmed mean ($T_{(2)}(s) = \frac{1}{n-2s}\sum_{i=s+1}^{n-s} X_{(i)}$, where $X_{(i)}$ is the order statistic of the sample and s is the trimming constant for location), and pseudo-median ($T_{(3)} = \text{median}_{i\leq j}(X_i + X_j)/2$, i.e. the median of the set $\{(X_1 + X_1)/2, (X_1 + X_2)/2, (X_1 + X_3)/2, \ldots, (X_1 + X_n)/2, (X_2 + X_2)/2, (X_2 + X_3)/2, \ldots, (X_2 + X_n)/2, \ldots, (X_{n-1} + X_n)/2, (X_n + X_n)/2\}$).

Note, that some theoretical results on consistency and asymptotic distribution χ^2 of the general RT class test statistic can be found in [13] and [14]. As it was also shown in [13], the power of RT class test mimics the effectiveness of location estimator in typical cases. Thus, trade-off between power and robustness is a typical issue here.

3 Power simulation study

3.1 Simulation setup

To compare the power and robustness of several tests for normality, we perform the Monte Carlo simulation study using R [9]. In this contribution we consider the classical linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is $(n\times 1)$-vector, $\mathbf{X}$ is a nonstochastic $(n\times k)$-matrix of rank k, $\boldsymbol{\beta}$ is the $(k\times 1)$-vector of unknown parameters, and $\boldsymbol{\epsilon}$ is the disturbance $(n \times 1)$-vector. Therefore, our simulation study is based on sample OLS residuals $\hat{\epsilon}$ which are given by

$$\hat{\boldsymbol{\epsilon}} = \left(\mathbf{I_n} - \mathbf{X}\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\right)\mathbf{y},$$

where $\mathbf{I_n}$ is identity matrix of size n, and on sample LTS residuals $r(\beta)$ based on the LTS estimator

$$\hat{\beta}_n^{(LTS)} = \underset{\beta\in\mathbb{R}^k}{\text{argmin}}\sum_{i=1}^{h} r_{(i)}^2(\beta),$$

where $r_{(i)}(\beta)$ represents ith order statistics of squared residuals $r_1^2(\beta), \ldots, r_n^2(\beta)$ with $r_i(\beta) = y_i - x_i\boldsymbol{\beta}$, and h is trimming constant which have to satisfy $\frac{n}{2} < h \leq n$ (for more details of LTS see [10] and [11]).

The approach using p-location outlier models of the dependent variable y and uniformly distributed independent variables was already studied in [15], but only for the OLS residuals. Furthermore, in [17] normality of the OLS residuals of various alternative distributed dependent variable y including distributions with heavy, light and very short tails was studied. However, in present contribution we assume p-location outlier models of the dependent variable y and the OLS as well as LTS residuals for the purpose of comparison. Generally, we assume the p-location outlier model as used in [2], i.e. we consider $y_1, \ldots, y_{n-p}$ to be *iid* from $N(0,1)$ and $y_{n-p+1}, \ldots, y_n$ to be *iid* from $N(\lambda, 1)$. For our simulation we consider the following parameters setup: $p \in \{1, 3, 5\}$, $\lambda \in \{3, 10\}$ and $n \in \{20, 100\}$. As regressor X we assume one uniformly distributed variable. Consequently, for analyzed p-location outlier models of the dependent variable y we estimate OLS as well as LTS residuals which are then used to test of normality. The number of replications of the Monte Carlo simulations is 100,000.

Note that two typical extremal behaviors occur in robust testing: the tests which are more robust have smaller power (since they are not affected by single outliers) and tests with higher power are typically less robust (because they are affected by single outliers) – for more details see e.g. [13]. Thus, in our power study we consider some classical non-robust tests of normality with higher power against the broad scale of alternative distributions – the Shapiro-Wilk test (SW, [12]) as the most popular omnibus test of normality for a general use, the Jarque-Bera test (JB, [6]) as the most widely adopted omnibus test of normality in econometrics, finance and related fields, and the Anderson-Darling test (AD, [1]) and the Lilliefors test (LT, [7]) as the most famous tests of normality based on the empirical distribution function – accompanied with several new tests for normality based on robust characteristics, in particular, the medcouple test (MC_{LR}, [4]), the robust Jarque-Bera test (RJB, [5]), and the selected robust tests from the RT class, namely $MMRT1$, $MMRT2$, $TTRT1$ and $TTRT2$ – for more details of these tests see [14].

3.2 Results of the Monte Carlo study

The results of the Monte Carlo simulations are summarized in Table 1. Based on this table we can conclude that the most commonly used tests for normality, such as the SW, LT and JB tests, are too strict in rejecting normality in the case of a small number of outliers, even for large sample size $n = 100$. In contrast, the MC_{LR} and $TTRT1$ are more robust. To illustrate this general framework, we suppose the p-outlier model for $p = 1$, $\lambda = 3$ and $n = 100$ – for this alternative the SW and JB tests reject normality of OLS as well as LTS residuals in 19% and 27% cases, respectively. In contrast, the MC_{LR} and $TTRT1$ tests reject normality only in 5% cases, what corresponds with significance level.

	$\lambda = 3$						$\lambda = 10$					
	$p = 1$		$p = 3$		$p = 5$		$p = 1$		$p = 3$		$p = 5$	
	OLS	LTS	OLS	LTS	OLS	LTS	OLS	LTS	OLS	LTS	OLS	LTS
AD	0.158	0.152	0.271	0.265	**0.228**	**0.204**	0.998	**1.000**	0.975	**1.000**	**0.946**	**1.000**
JB	**0.296**	**0.293**	0.231	0.200	0.057	0.044	**1.000**	**1.000**	0.954	**1.000**	0.557	0.001
LT	0.108	0.108	0.217	0.215	0.196	0.191	0.958	0.982	0.952	**1.000**	0.922	**1.000**
MC_{LR}	0.050	0.050	0.081	0.090	0.105	0.111	0.058	0.054	0.465	0.518	0.646	0.890
RJB	0.284	0.270	0.241	0.210	0.058	0.048	**1.000**	**1.000**	0.958	**1.000**	0.566	0.346
SW	0.209	0.207	0.273	**0.270**	0.203	0.179	**1.000**	**1.000**	0.974	**1.000**	0.935	**1.000**
$MMRT1$	0.255	0.260	0.217	0.201	0.147	0.107	**1.000**	**1.000**	0.930	0.999	0.859	0.991
$MMRT2$	0.246	0.264	0.186	0.175	0.147	0.099	**1.000**	**1.000**	0.921	**1.000**	0.868	0.997
$TTRT1$	0.049	0.045	**0.274**	0.269	0.216	0.199	0.000	0.000	**0.978**	**1.000**	0.918	**1.000**
$TTRT2$	0.250	0.257	0.244	0.180	0.084	0.034	**1.000**	**1.000**	0.970	**1.000**	0.825	0.202
AD	0.086	0.088	0.292	0.293	0.526	0.531	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
JB	**0.265**	**0.265**	**0.586**	**0.586**	0.740	0.737	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
LT	0.066	0.066	0.175	0.177	0.336	0.337	0.964	0.964	**1.000**	**1.000**	**1.000**	**1.000**
MC_{LR}	0.048	0.046	0.056	0.057	0.073	0.072	0.046	0.049	0.061	0.059	0.096	0.094
RJB	0.248	0.247	0.573	0.568	**0.744**	**0.739**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
SW	0.191	0.193	0.490	0.494	0.677	0.684	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$MMRT1$	0.242	0.246	0.530	0.535	0.679	0.683	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$MMRT2$	0.238	0.243	0.523	0.531	0.672	0.680	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$TTRT1$	0.051	0.056	0.122	0.132	0.174	0.191	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
$TTRT2$	0.212	0.217	0.478	0.480	0.624	0.628	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Table 1 Power of the selected normality tests for $n = 20$ (upper block) and $n = 100$ (lower block)

The differences in powers for OLS and LTS residuals are in most cases negligible. However, very interesting result we can find in the case of p-outlier model for $p = 5$, $\lambda = 10$ and $n = 20$ where the power of the classical JB test for OLS residuals is 0.557, while for LTS residuals is biased – power is 0.001. Similarly, the power of the $TTRT2$ test for OLS and LTS residuals is 0.825 and 0.202, respectively. It is caused by construction of the test statistics which are based on empirical moments.

4 Illustrative example

For illustration purposes, we consider the data set of HDI and GDP of the EU countries. Note that the data as well as correlation and cluster analysis for this data set were performed in [16]. In this contribution we focus on normality of the OLS and LTS residuals of the linear model $HDI_i = \beta_0 + \beta_1 GDP_i + \epsilon_i$ for $i = 1, 2, \ldots, n$. The scatter plot for this data set is presented in Figure 1. Subsequently, the histograms, boxplots and QQ plots of the OLS and LTS residuals are presented in Figure 2. From this figures we can state that one outlier (Luxembourg) affects especially OLS estimator – in contrast, the LTS estimate is not affected by this outlier.

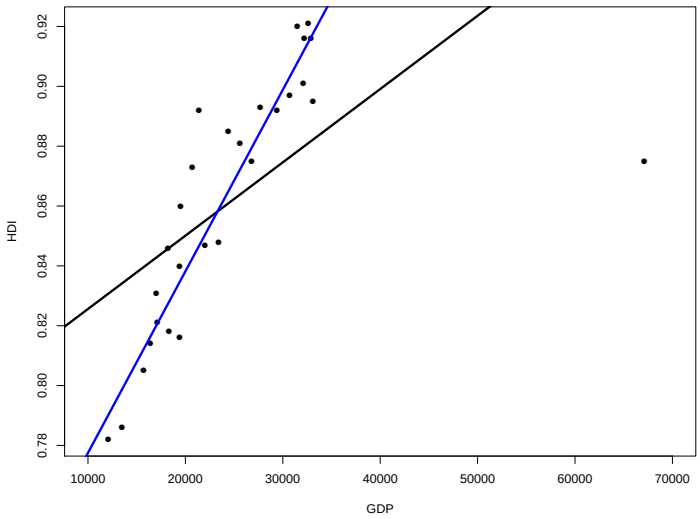


Figure 1 Scatter plot – OLS estimate (black), LTS estimate (blue)

For mentioned data set, in Table 2 the test statistics and p-values of analyzed normality tests are presented. The results show, that normality of the OLS residuals is rejected only by $MMRT1$, at 5% significance level. In the case of the LTS residuals the normality is rejected by the majority of analyzed tests – only very robust MC_{LR} and $TTRT1$ tests do not reject the normality of LTS residuals, at 5% significance level. Thus, these tests can be recommended in cases involving a small number of outliers.

	OLS		LTS	
	statistic	p-value	statistic	p-value
AD	0.586	0.116	4.746	< 0.001
JB	3.316	0.073	485.039	< 0.001
LT	0.136	0.192	0.341	< 0.001
MC_{LR}	3.409	0.202	1.941	0.483
RJB	2.879	0.132	44651.617	< 0.001
SW	0.929	0.058	0.466	< 0.001
$MMRT1$	4.988	0.038	432.783	< 0.001
$MMRT2$	3.812	0.056	414.401	< 0.001
$TTRT1$	1.895	0.789	14.544	0.159
$TTRT2$	1.287	0.479	1031397.3	< 0.001

Table 2 Test statistics and p-values of the selected normality tests for OLS and LTS residuals

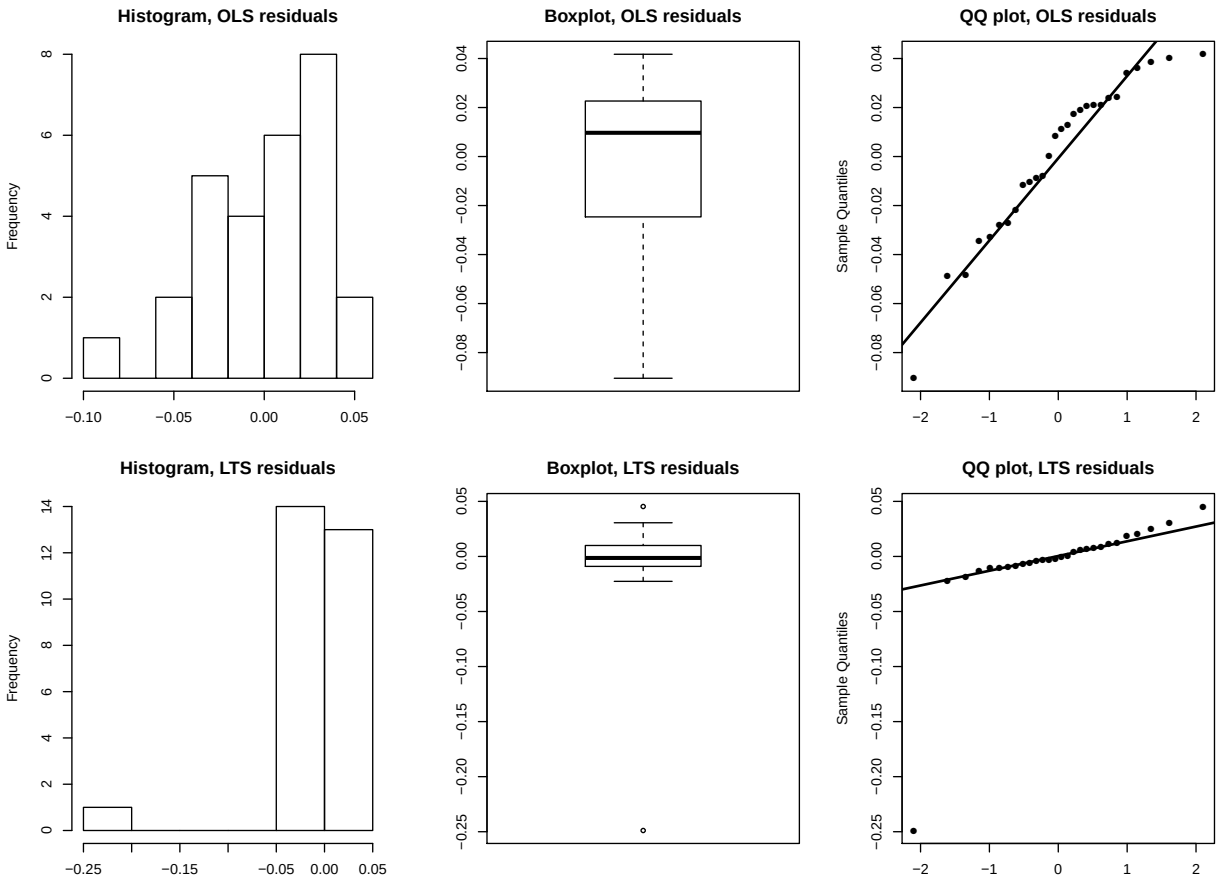


Figure 2 Histograms, boxplots and QQ plots – OLS residuals (upper line), LTS residuals (lower line)

5 Summary

In this contribution we present and discuss some interesting results of power and robustness of selected normality tests for error terms of linear regression model using simulation of p-location outlier models. Based on our simulation study we can conclude that analyzed RT class tests ($MMRT1$, $MMRT2$, $TTRT1$ and $TTRT2$) are more robust than the classical JB in the cases of small number of outliers ($p = 3, 5$) in comparison with the sample size. In contrast, if we suppose a largest number of outliers in small sample size (e.g. $p = 5$ outliers in small sample size $n = 20$) we need the test with high power, because the hypothesis of normality is not sustainable. In this case analyzed RT class tests have higher power than the classical JB as well as than RJB, while power of analyzed RT class tests is in the most cases comparable with the SW test. Finally, as was also presented in illustrative example, the most robust tests for normality are MC_{LR} and $TTRT1$ tests.

Acknowledgements

Special Thanks to Assoc. Professor Milan Stehlík for cooperation on research of robust testing for normality.

References

[1] Anderson, T. W., and Darling, D. A.: Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **23** (1962), 193–212.

[2] Balakrishnan, N.: Permanents, order statistics, outliers, and robustness. *Revista Matemática Complutense*, **20** (2007), 7–107.

[3] Bickel, P. J., and Lehmann, E. L.: Descriptive statistics for nonparametric models II. location. *The Annals of Statistics*, **3** (1975), 1045–1069.

[4] Brys, G., Hubert, M., and Struyf, A.: Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, **23** (2008), 429–442.

[5] Gel, Y. R., and Gastwirth, J. L.: A robust modification of the Jarque-Bera test of normality. *Economics Letters*, **99** (2008), 30–32.

[6] Jarque, C. M., and Bera, A. K.: A test for normality of observations and regression residuals. *International Statistical Review*, **55** (1987), 163–172.

[7] Lilliefors, H. W.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62** (1967), 399–402.

[8] Önder, A. Ö., and Zaman, A.: Robust tests for normality of errors in regression models. *Economics Letters* **86**, 1 (2005), 63–68.

[9] R Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org/.

[10] Rousseeuw, P. J.: Least median squares regression. *Journal of the American Statistical Association* **79**, 388 (1984), 871–880.

[11] Rousseeuw, P. J., and Leroy, A. M.: *Robust Regression and Outlier Detection.* Wiley, New York, 1987.

[12] Shapiro, S. S., and Wilk, M. B.: An analysis of variance test for normality. *Biometrika*, **52** (1965), 591–611.

[13] Stehlík, M., Fabián, Z., and Střelec, L.: Small sample robust testing for normality against Pareto tails. *Communications in Statistics – Simulation and Computation* **41**, 7 (2012), 1167–1194.

[14] Stehlík, M., Střelec, L., and Thulin, M.: On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **130**. (2014), 98–108.

[15] Střelec, L., Grochová, L, and Kolman, P.: On power comparison of normality tests of error terms in regression models. In: *11th International Conference of Numerical Analysis and Applied Mathematics 2013: ICNAAM 2013*, edited by T. E. Simos, et al., AIP Conference Proceedings 1558, American Institute of Physics, New York, 2013, 1847–1850. ISBN 978-0-7354-1184-5.

[16] Střelec, L., Hampel, D., Grochová, L., Janová, J.: On the measurement of sustainable economic performance. In: *Proceedings of the 32nd International Conference Mathematical Methods in Economics 2014*. Palacký University, Olomouc, 2014, 985–990. ISBN 978-80-244-4209-9.

[17] Střelec, L.: Notes on power of normality tests of error terms in regression models. In: *12th International Conference of Numerical Analysis and Applied Mathematics 2014: ICNAAM 2014*, edited by T. E. Simos, et al., AIP Conference Proceedings 1648, American Institute of Physics, New York, 2015, 720013(1)–720013(4). ISBN 978-0-7354-1287-3.

A Data Processing Method for the Purpose of Optimizing Production of a Certain Type of Car

Petr Suchánek[1], Robert Bucki[2], Franciszek Marecki[3]

Abstract. Knowledge of the market environment is the major key to the success of companies. Good market knowledge greatly helps to optimize production which results in minimizing manufacturing costs and maximizing profits. The fundamental issue remains the knowledge of behavior as well as decision-making processes of customers who tend to buy high quality products. Therefore, it is necessary to pay great attention to the parameters of specific products in relation to their marketability as parameters of each product are input data for customers' decision-making processes. The paper highlights the problem of practical demonstration of a specific market product analysis - a car in order to determine the optimal value of production. On the basis of the survey data it is necessary to specify the car production size of different brands to achieve the maximal financial efficiency. Data on customer preferences may be obtained by questionnaire methods. Bayesian formulas and principles of information diagnostic systems and self-learning systems were used for data processing and creation of the mathematical model.

Keywords: decision-making processes, questionnaire methods, Bayesian formulas, information diagnostic system, concluding generator, self-learning system.

JEL Classification: C11
AMS Classification: 90B50, 62C10

1 Introduction

Manufacturing optimization is the key issue covering all range of companies. The requirement for the efficiency of production creates the need for good management which focuses on production and management processes, new technologies and, of course, optimization methods [1]. Production optimization can be viewed from two perspectives. One of them is the actual optimization of operations and production logistics (e.g. for production lines) and supply chain [8], [11], then the second one is production optimization in terms of its content, in other words, optimization of production in relation to the required specifications of each manufactured product. The article is focused on the second of these two aspects. A typical example might be, for example, automobile production, where customers can choose different types of equipment by means of configurators. This example falls into the category of multi-objective optimization problems which depend on a probability measure corresponding to many economic, financial, and production activities [9]. Let us consider the problem of production M car brands (e.g. FIAT, Volkswagen, Renault, etc.) by different companies belonging to one group (e.g. Volkswagen Group, etc.). Customers evaluate these cars from the point of view of N parameters e.g.: price, power, fuel type, fuel consumption, maximum speed, silhouette, comfort, running costs, warranty, insurance costs, quality, etc. The problem is to determine the volume of production of each car brand. Too large production causes losses (sales at a lower price). If the production amount is too low, the gain is lower than expected. Moreover, some customers purchase competing cars which do not have all their preferred parameters. In a market economy it is essential to analyze the market - that is, the analysis of customer preferences. Data on customer preferences may be obtained by classic or web-based questionnaire methods [4], [12], [16], [17]. At the same time, the following are of a vital importance: interviewers, clients, survey questions, surveys periods, etc. On the basis of the survey data it is necessary to specify the car production size of different brands to achieve the maximal financial efficiency. The decision problem formulated in this way belongs to the operational research area. It is characterized by its massive character and data uncertainty. To solve this problem, diagnostic methods can be used such as the Bayesian formulas [2], [5], [10]. However, the classical diagnostic methods can be modified so that the market analysis is adjusted to changes in time [14]. Therefore, the customer preferences analysis [3] model should be recursive. In this particular case a self-learning computer system can be developed, i.e. the one that generates

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 73340 Karviná, Czech Republic, suchanek@opf.slu.cz.
[2] Institute of Management and Information Technology in Bielsko-Biała, ul. Legionów 81, 43-300 Bielsko-Biała, Poland, rbucki@wsi.net.pl
[3]Academy of Business in Dąbrowa Górnicza, ul. Cieplaka 1C, 41-300 Dąbrowa Górnicza, Poland, fmarecki@gmail.com.

results matching current data [15]. Creating a computer system becomes the most important issue for processing large data bases in the shortest real time and is of a competitive advantage in today's market environment [6].

Let us assume that *K* customers are surveyed. Data obtained from questionnaires filled out by *K* customers are stored in the database consisting of *K* records. Each record consists of *M* + *N* fields. These records distinguish the following:

- *M* fields matching car branches;
- *N* fields matching parameters of cars.

Let us assume that the elements of the database are binary values:

1) $b_m = 1$ if the *l*-th customer chooses the *m*-th brand of car;
2) $b_m = 0$ otherwise, $m = 1,\ldots,M$.

At the same time we assume that a customer prefers only one brand of car.

3) $b_m = 1$ if the *l*-th customer prefers (chooses) the *n*-th parameter;
4) $b_m = 0$, otherwise, $n = 1,\ldots,N$.

At the same time we assume that a customer prefers more than one parameter. The problem consists of determining:

1) The percentage share of cars of various brands *m*, $m = 1,\ldots,M$ in the car market sales.
2) The probability that customers choose the *m*-th brand car, $m = 1,\ldots,M$ if the car manufacturer produces it with determined parameters. So, the parameters of the manufactured car are treated as conjunction.

On the basis of the survey (among *K* customers) the problem formulated above can be solved. The expert system of market analysis consists of:

1) database;
2) base of knowledge;
3) concluding generator.

2 Database

The database consisting of *K* rows (surveys) and $M + N$ columns (record fields) is made on the basis of the carried out surveys. The elements of this base are binary numbers (0 or 1) and they mean that a customer chooses the *m*-th brand car and its *n*-th parameter. It is assumed that a customer chooses only a car of a specific brand with a few determined parameters.

This is the reason of originating the decision diagnostic problem concerning buying a car. The problem consists of various customers' preferences and the fact that a company manufactures cars of only one brand with pre-defined parameters (which do not include all parameters preferred by customers). So, the company should produce a car with preferences which are close to its customers' preferences. Otherwise, customers may turn to a rival company to buy another car.

The Bayesian method can be implemented to diagnose the problem formulated previously. This method lets us create the information diagnostic system based on artificial intelligence methods [13]. Such a system could be a self-learning one i.e. it could adjust itself to changeable conditions resulting from the process of changes of customers' preferences in time. Such preferences can be revealed in the subsequent survey sessions. Let us assume that, after the analysis of fields *m*, $m = 1,\ldots,M$, there are *K* questionnaires in which customers chose one of *M* cars which is illustrated in the formula (1):

$$\sum_{m=1}^{M} K_m = K \tag{1}$$

where K_m - the number of customers who chose the *m*-th brand of car, $m = 1,\ldots,M$.

Let us assume that the analysis of a field *n*, $n=1,\ldots,N$ of the database shows that the *n*-th parameter was chosen (preferred) by L_n customers, $n = 1,\ldots,N$ and at the same time the following becomes valid (2):

$$L_n \le K \tag{2}$$

So, the *n*-th parameter can be chosen (preferred for each car). Moreover (3):

$$\sum_{n=1}^{N} L_n < K \cdot N \tag{3}$$

which means that all parameters are not preferred for all cars. Further, let us assume that, after the database analysis, the number $L_{m,n}$ is determined where: $L_{m,n}$ is the number of customers who choose the *n*-th parameter and the *m*-th car. At the same time it is assumed that (4):

$$L_{m,n} < K\ ,\ m = 1,...,M \tag{4}$$

which is understood that not all customers choose the *n*-th parameter for the *m*-th car brand, $m = 1,\ldots,M$. Moreover (5):

$$L_{m,n} < N\ ,\ n = 1,...,N \tag{5}$$

which means that not all customers choose the *m*-th band of car for the *n*-th parameter, $n = 1,\ldots,N$.

3 Base of knowledge

The base of knowledge can be made on the basis of the database. Knowledge is expressed by probabilities characterizing the market of car buyers. On the basis of data from the database it is possible to calculate the probability p_m representing a statistical customer who is expected to buy a car of the *m*-th brand (with any parameters) (6):

$$p_m = \frac{K_m}{K}\ ,\ m = 1,...,M \tag{6}$$

The probability q_n representing a statistical customer who is expected to buy a car of any brand which is calculated as shown in (7):

$$q_n = \frac{L_n}{K}\ ,\ n = 1,...,N \tag{7}$$

Moreover, it is possible to determine the conditional probability $q_{m,n}$ representing customer's preferences concerning the *n*-th parameter shown by (8):

$$q_{m,n} = \frac{L_{m,n}}{K_m}\ ,\ m = 1,...,M\ ,\ n = 1,...,N \tag{8}$$

Probabilities determined by formulas (6), (7) and (8) form the knowledge base of the required diagnostic system. Such a base consists of:

1) Vectors of absolute probabilities: p_m, $m = 1,\ldots,M$ and q_m, $n = 1,...,N$
2) Matrixes of relative probabilities: $q_{m,n}$, $m = 1,\ldots,M$, $n = 1,\ldots,N$

Elements of the knowledge base (probabilities) change if elements of the database are subject to the change. So, after each survey session the database and the knowledge base change. Therefore, the database and knowledge system is dynamic (in time).

4 Concluding generator

The so-called concept of a car brand profile produced by the *m*-th company, $m = 1,\ldots,M$ is introduced in the discussed mathematical model of diagnosing the market state of the car buyers. Such a profile is presented in the parameter vector form of the brand manufactured car. It must be emphasized that the company manufactures cars which have the same parameters. On the other hand, preferences concern various parameters of a car. If a company decides to manufacture the *m*-th brand cars with certain parameters, a customer will buy it with a probability which can be determined. Let us assume that a company decided to produce its brand cars with parameters written in the vector below (9):

$$X = [x_n],\ n = 1,...,N \tag{9}$$

The elements of the vector are defined as follows:

i. $x_n = 1$ if the *n*-th parameter is included in the car;

ii. $x_n = 0$ if the *n*-th parameter is not included.

Let *X* be called the company profile in short in the next part of the article. The problem consists in determining the probability $P_m(X)$, $m = 1,\ldots,M$ which means that a client will buy a car produced by the company which is characterized by profile *X*. The problem formulated in the above way is to be solved by means of the Bayesian statistics on condition that a client buys only one car.

5 Bayesian statistic formulas

Based on [7] it is assumed that there are *M* events A_m, $m = 1,\ldots,M$ (purchasing the *m*-th brand car) which exclude each other and are mutually related. It is also assumed that events create the total system (probabilities): $p(A_m) > 0$, $m = 1,\ldots,M$ meeting the condition (10):

$$p(A_1) + \ldots + p(A_m) + \ldots + p(A_M) = 1 \tag{10}$$

Event A_m is understood as purchasing the *m*-th brand car (manufactured by the *m*-th company). It assumed that a customer buys only one car, which is justified.

Let us assume that there are *N* independent and non-exclusive events B_n, $n = 1,\ldots,N$. These events are interpreted as the choice in the survey of the *n*th parameter. Let us assume that each of *N* parameters of the *m*-th type is chosen (which was made in the *m*-th company) e.g. the colour of the body, upholstery, etc. Let us assume there are probabilities of these events $p(B_n^m) > 0$, where $m = 1,\ldots,M$ and $n = 1,\ldots,N$. These probabilities can be calculated on the basis of data from the database.

The profile of the *m*-th company defines conjunction B^m of events B_n^m, $n = 1,\ldots,N$. This assumption means that parameters can be chosen simultaneously. The probability of the conjunction is determined as (11):

$$p(B^m) = p(B_1^m) \cdot \ldots \cdot p(B_n^m) \cdot \ldots \cdot p(B_N^m) > 0 \tag{11}$$

Events A_m and B_n are dependent because the choice of the *m*-th car is associated with the specified parameters B_n, $n = 1,\ldots,N$. For the conjunction of dependent events the known Bayesian formula (12) is implemented as follows:

$$p(A_m \wedge B^m) = p(A_m) \cdot p(B^m / A_m) = p(B^m) \cdot p(A_m / B^m) \tag{12}$$

The formula (12) enables us to obtain the conditional probability $p(B / A_m)$ as expressed in (13):

$$p(A_m / B^m) = \frac{p(A_m) \cdot p(B^m / A_m)}{p(B^m)}, \; m = 1,\ldots,M \tag{13}$$

As events B_n are assumed to be independent, then (14):

$$p(B^m / A_m) = \prod_{n=1}^{N} p(B_n^m / A_m) \tag{14}$$

and (15)

$$p(B) = \prod_{n=1}^{N} p(B_n) \tag{15}$$

Substituting equations (14) and (15) into the formula (13), finally, we are able to obtain (16):

$$p(A_m / B^m) = \frac{p(A_m) \cdot \prod_{n=1}^{N} p(B_n^m / A_m)}{\prod_{n=1}^{N} p(B_n^m)} \tag{16}$$

The assumption that the customer buys only one car, comes down to accepting that events B_n are mutually exclusive. In this case the formula (16) takes the form (17):

$$p\left(A_m / B_n^m\right) = \frac{p\left(A_m\right) \cdot \sum_{n=1}^{N} p(B_n^m / A_m)}{\sum_{n=1}^{N} p(B_n^m)} \tag{17}$$

The probability of choosing the *m*-th car brand manufactured by the *m*-th company (events A_m) on condition that this car possesses defined parameters (conjunction of events B_n) can be determined from Bayes formula (16) and (17). If the *n*-th parameter does not occur in a car, then such an event is not taken into account in the presented model.

6 Information diagnostic system

The problem formulated hereby consists of determining probabilities $P_m(X)$, $m=1,...,M$ of market buying preferences concerning a car produced by a company characterized by profile *X*. Especially, the most probable buy can be determined, i.e. a car of the company which probability $P_m(X)$ is the highest (in the set *M* of cars/companies). The probability $P_m(X)$ of buying the *m*-th brand car depends on the profile of the company *X*. This profile is known by people filling the survey form. However, they can prefer other parameters of the *m*-th brand car (e.g. price, fuel consumption, etc.). This is the reason why it can be analyzed "*What would happen if* " a company changes its profile (some parameters). The change of a company profile *X* changes the probability $P_m(X)$ which represents the total amount of sales. Adjusting the company's profile to meet the expectations of the market/customers increases the sale incomes. However, it is connected with additional costs. So, the analysis of incomes/profits requires comparing increasing incomes and costs. Incomes can be evaluated by means of the Bayesian method on the basis of survey data.

7 The self-learning system

The Bayesian statistical method of car market analysis presented in the paper is based on survey data. The data is stored in the database on the basis of which the base of knowledge is prepared. The knowledge system lets us come to certain conclusions about sales of cars characterized by determined parameters.

The presented diagnostic system can be treated as a self-learning one as:

1) The company can adjust its profile to market conditions on the basis of survey data (the database and the base of knowledge).
2) Survey data can be stored and removed from the database. Especially, the out-of-date data should be removed. This lets us conclude that the conclusion system can adjust itself to customers' preferences (probabilities of the database change).

The market of other objects can be analyzed analogously to the car market (materials, products, goods, services, etc.).

8 Conclusion

The success of today's companies is increasingly dependent on precise knowledge of the market environment. The market environment can be monitored and evaluated from different perspectives and with the use of a variety of methods and tools. The primary requirement is to define the purpose of the evaluation, obtain relevant input data and use appropriate methods of assessment. One of the ways of obtaining information for research, education and other professional activities is the appropriate questionnaire research. It can be used for various purposes. In this paper the questionnaire research was considered as a source of data for the model used for production optimization of certain type of car. A car is one of the products which the customer can choose from a variety of configurations described by different sets of parameters. To assess the impact of various factors on the industrial area it is important to pay increasing attention to searching for dependencies between parameters and marketability of individual products. It turns out that knowledge of the parameters of individual products and their dependencies can be very effectively used e.g. among others, to optimize the size of products. This ultimately leads to a reduction in waste and maximizes profits of companies.

Acknowledgements

This paper was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2015.

References

[1] Adacher, L., and Cassandras, C.G.: Lot size optimization in manufacturing systems: The surrogate method. International Journal of Production Economics **155**(SI) (2014), 418-426.
[2] Billingsley, P.: Probability and Measure. Wiley, New Jersey, 2012.
[3] Bodnarova, A., Gavalec, M., and Mls, K.: Multi-criteria models with trend evaluation in decision making support. In: Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011. Janska Dolina, 2011, 47-52.
[4] Davis, R. N.: Web-based administration of a personality questionnaire: Comparison with traditional methods. Behavior Research Methods Instruments & Computers **31**(4) (1999), 572-577.
[5] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian Data Analysis. Chapman and Hall/CRC, London, 2013.
[6] Janáková, M.: Active Initiatives to ICT Innovations for Support of Competitive Advantage. In: ICT Support for Complex Systems, IDIMT 2012. Jindrichuv Hradec, 2012.
[7] Jensen, F. V.: An Introduction to Bayesian Networks. Springer-Verlag, New York, 1996.
[8] Josvai, J.: Optimization methods of sequence planning using simulation. Simulation in Produktion und Logistk, Paderborn, 2013.
[9] Kankova, V.: Economic and Financial Problems via Multiobjective Stochastic Optimization. In: Proceedings of the 31st International Conference on Mathematical Methods in Economics 2013. Jihlava, 2013, 362-367.
[10] Marin, J. M., Pillai, N. S., Robert, C. P., and Rousseau, J.: Relevant statistics for Bayesian model choice. Journal of the Royal Statistic al Society Series B-Statistical Methodology **76**(5) (2014), 833-859.
[11] Modrak, V., and Marton, D.: Configuration complexity assessment of convergent supply chain systems. International Journal of General Systems **43**(5) (2014), 508-520.
[12] Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., and Singer, E.: Methods for Testing and Evaluating Survey Questionnaires. Wiley-Interscience, New York, 2004.
[13] Price, Ch.: Computer-Based Diagnostic Systems. Springer, 1999.
[14] Rupp, A. A., Templin, J., and Henson, R. A.: Diagnostic Measurement: Theory, Methods, and Applications (Methodology in the Social Sciences). The Guilford Press, New York, 2010.
[15] Sher, E., Chronis, A., and Glynn, R.: Adaptive behavior of structural systems in unpredictable changing environments by using self-learning algorithms: A case study. Simulation-Transactions of the Society for Modeling and Simulation International **90**(8) (2014), 991-1006.
[16] Suchanek, P., Bucki, R., Litavcova, E., et al.: Modeling Customer Response by the Questionnaire Evaluation Method. In: Proceedings of the 31st International Conference on Mathematical Methods in Economics 2013. Jihlava, 2013, 875-880.
[17] Thomas, S. J.: Using Web and Paper Questionnaires for Data-Based Decision Making: From Design to Interpretation of the Results. Corwin, Thousand Oaks, 2004.

Non–linear restrictions in DSGE models: The effect of foreign exchange interventions in Czech economy

Karolína Súkupová[1], Osvald Vašíček[2]

Abstract. In this paper we deal with zero lower bound on interest rates and implementing this restriction into the DSGE models. As an example, we apply the zero lower bound into the relation for interest rates in the estimated DSGE model based on Czech data. Besides that, we study the foreign exchange interventions used by Czech National Bank as an unconventional monetary policy tool, when interest rates cannot be used. We implement the foreign exchange intervention function into the model and study impact of intervention on the behaviour of economy represented by model.

In the first section, we describe the method and its application into the model. In the second section, we verify its robustness and consistence simulating the impulse responses to model innovations and paths of model variables for stochastically generated sets of model parameters. In the third section, we discuss results of these simulations and compare behaviour of economy represented by model with economic reality.

Keywords: unconventional monetary policy, exchange rate intervention, DSGE models, Zero Lower Bound.

JEL classification: C11, C15, E43, E52
AMS classification: 91B51, 91B64

1 Introduction

Using the interest rates as the monetary policy tool became widely used approach to monetary policy. During the Great Recession, monetary authorities in Europe lowered the interest rates to stimulate economies during the decline. As the recession ended, monetary authorities were challenged by the deflationary risks and low economic growth combined with interest rates on their minimal values. Both ECB and CNB lowered interest rates to the technical zero in the end of 2012 and considered further steps. This situation restored old theoretical discussions among central bankers and academia about Zero Lower Bound on interest rates.

Since the third quarter of 2012, the policy interest rates in Czech economy were on technical zero, while the economic growth was near or below zero and inflation rate was falling down. Czech National Bank decided to use foreign exchange rate intervention as an unconventional monetary policy tool in November of 2013. Their primary goal was to avoid deflation, as the inflation rate was decreasing and risk of the deflationary spiral became real. Their intention was to renew economic growth, stimulate domestic demand and help exporters. First economic data measured after the depreciation were published in the second half of 2014. The economic growth in the first quarter of 2014 was one of the strongest in Europe, domestic demand increased as well as exports. Czech national bank explained this positive change as the result of FX intervention, critics of intervention emphasized positive foreign demand shocks.

In the first section of this paper, we introduce the model and its features, that enable us to model FX intervention. In the second section, we describe the method for treating the non–linear restrictions in DSGE model. In the third section, we are testing the robustness of this algorithm and the fourth

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, 324246@mail.muni.cz
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, osvald@mail.muni.cz

section consists of economic analysis of the impact of foreign exchange rate intervention in a model that represents real economy. Our analysis of exchange rate intervention may be divided into two parts. Both regimes will be analysed under the condition of Zero Lower Bound on interest rates. In the first part, we will analyse the negative intervention shock, i. e. unexpected depreciation of the foreign exchange rate and its effects in economy. The second analysed regime is the exchange rate targeting. We are focused especially on the negative foreign inflation shock, that struck economy in 2014.

2 Model

Our work is based on the diploma thesis from Malovaná [4]. The model used for our analysis is a model of a small open economy with nominal rigidities, which follows standard framework derived in [3]. The model consists of households, two stages of domestic producers, importers and monetary authority. The foreign economy is modelled as a closed version of the domestic economy. The intervention of a central bank is modelled folowing Montoro, Ortiz in [5]. The continuum of dealers is operating in a domestic economy. They obtain sale and purchase orders of domestic bonds ϖ_t^d and ϖ_t^{cb} from households and central bank and sale orders of foreign bonds ϖ_t^* and ϖ_t^{cb*} from foreign investors and domestic central bank. They maximize an objective

$$-E_t^d e^{-\gamma \Omega_{t+1}^d},$$

where E_t^d is the expectation operator, γ is the the coefficient of absolute risk aversion and Ω_{t+1}^d is total investment after returns. The result of this optimization is the UIP condition

$$E_t s_{t+1} - s_t = i_t - i_t^* + \gamma \sigma^2 (\varpi_t^* + \varpi_t^{cb*}).$$

Central bank may intervene into the foreign exchange rate by selling or purchasing bonds on market. The amount of bonds in this operation is determined by the equation

$$\varpi_t^{cb*} = \chi_T (e_t - e_T) + \chi_e (e_t - e_{t-1}) + \chi_q q_t + \epsilon_t^{cb}$$

Monetary authority may intervene under the different regimes. The combination of the parameters $(\chi_T = 1, \chi_e = 0, \chi_q = 0)$ implies exchange rate targeting, the combination $(\chi_T = 0, \chi_e = 1, \chi_q = 0)$ means so called "leaning against the wind", i.e. smoothing of exchange rate. The combination $(\chi_T = 0, \chi_e = 0, \chi_q = 1)$ describes the case of intervention to the real exchange rate. The calibration $(\chi_T = 0, \chi_e = 0, \chi_q = 0)$ implies that only unexpected intervention shocks influence the foreign exchange rate.

The model is estimated using Bayesian techniques in Matlab under toolbox Dynare. We used 11 observed time series, namely domestic real output, domestic consumption, domestic exports, domestic imports, imports inflation, domestic CPI, domestic nominal interest rate (PRIBOR), foreign real output, foreign CPI, foreign nominal interest rate and exchange rate (CZK/EUR). Except the domestic and foreign nominal interest rates, we transformed the original series by taking logarithms and the first differences.

3 Restrictions in DSGE models

There are more ways to treat restrictions in the DSGE models. The piecewise linear algorithms belong to the earliest approach and the complete implementation into the Dynare was provided by Guerierri and Iacoviello in [1]. The next generation of algorithms was based on the nonlinear solution of a model. The approach used in this paper belongs to this kind of algorithms. The method is introduced and derived by Holden and Paetz in [2] and it is based on so called shadow price shocks. The constrained model equation takes a form

$$x_t = \mu_x + \phi_{-1} y_{t-1} + \phi_0 y_t + \phi_{+1} E_t y_{t+1} - \mu_y (\phi_{-1} + \phi_0 + \phi_{+1}) + \epsilon_{SP}.$$

The shadow price shock ϵ_{SP} is incorporated into the equation of the constrained variable and it drives this variable from the negative values back to the zero level. The algorithm saves values of ϵ_{SP} and impulse responses of the model variables to all values of this shock. The impulse response to the model's shock ϵ under constraint may be written as

$$\text{irf}_\epsilon^x = \mu_x + v + \alpha M,$$

where v denotes the impulse response to this shock without constraint, M is the matrix of the impulse response functions to the values of the shadow price shock and α is the magnitude of these shocks. As the magnitude α is not known, we need to solve quadratic programming problem

$$\alpha^* = \underset{\substack{\alpha \geq 0 \\ \mu^* + v^* + M^* \alpha \geq 0}}{\operatorname{argmin}} \{\alpha'(\mu^* + v^*) + \frac{1}{2}\alpha'(M^* + M^{*'})\alpha\}$$

4 Simulation

The non-linear optimization requires specific conditions to be satisfied by model parameters. We would like to know, how sensitive to change in parameters is the algorithm. We simulated paths of the model variables and impulse response functions for different sets of model parameters. These sets are generated from the posterior distributions with the estimated characteristics, i.e. posterior mean and standard deviation. All samples were generated within the Matlab, using functions from the statistical toolbox. We needed to generate samples from Beta distribution, Gamma distribution, and normally distributed samples. As the Matlab functions for generating samples from Gamma and Beta distribution take as input scale parameters and shape parameters instead of mean (μ) and standard deviation (σ), we needed to calculate them. Then, the shape parameters for Beta distribution are

$$\alpha = \frac{(1-\mu)\mu^2 - \mu\sigma^2}{\sigma^2} \qquad \beta = \frac{(1-\mu)^2\mu - (1-\mu)\sigma^2}{\sigma^2}$$

and the shape parameter and the scale parameter for Gamma distribution are

$$k = \left(\frac{\mu}{\sigma}\right)^2 \qquad \theta = \frac{\sigma^2}{\mu}.$$

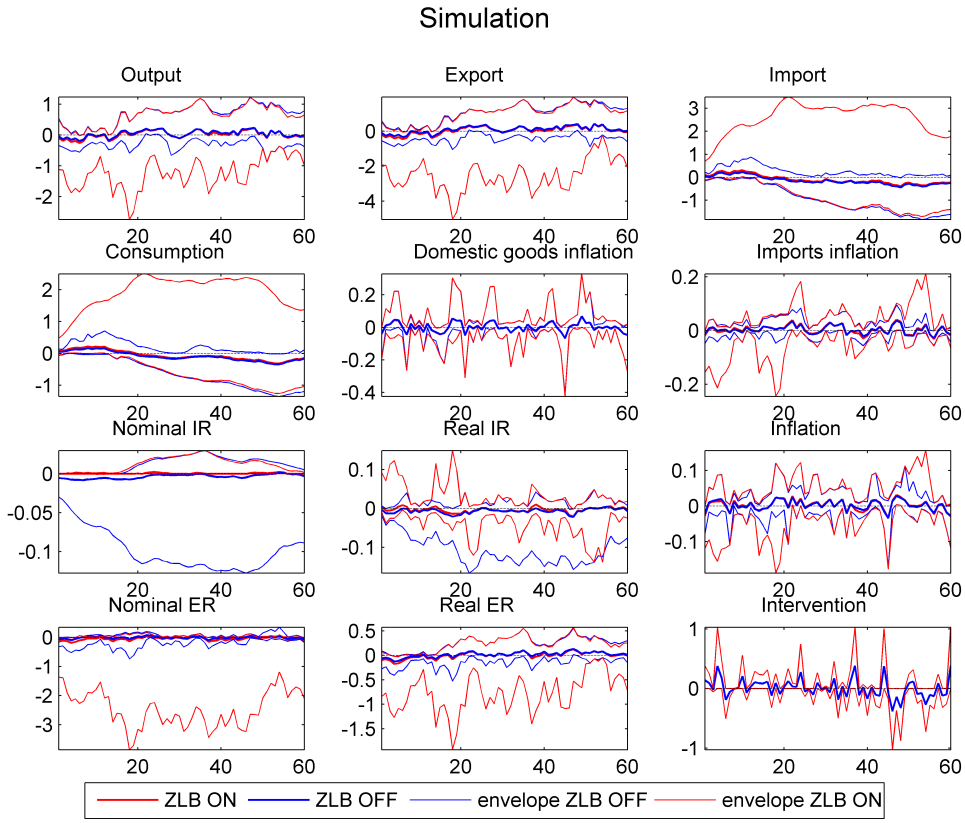


Figure 1 Simulation of paths of selected model variables. Number of generated samples of parameters $N = 200$.

In the extreme case, we generated 2000 samples and then executed simulation of our model with each set of parameters. In all runs that were executed, the algorithm worked properly. In the case pictured in the Figure 1 we generated 200 samples of each parameter. There are pictured simulated paths of selected model variables. The thick line captures the case of 'mean' parameters, the thin lines are envelopes, i. e. the minimum and the maximum paths. These simulated paths have not any straightforward economic interpretation. They exemplify the most adverse case of parameters and shocks.

5 Results

We simulated two modifications of the model. The first version of the model represents the economy, where Zero Lower Bound on interest rates is binding. The central bank reached limits of the standard monetary policy tools, but there has not been applied any form of intervention yet. The central bank decides to depreciate the currency and sells foreign assets to dealers. This decision may be represented by the negative intervention shock that strikes the economy through the UIP condition. This scenario correspond to the situation in the Czech economy in the second half of 2013.

The depreciation of the foreign exchange rate leads to increase of the interest rates as the foreign investors want the same real returns. This is the reason why the cases 'ZLB ON' and 'ZLB OFF' are the same. Ceteris paribus, nominal interest rate is higher and Zero Lower Bound no longer binds in economy. The imports became more expensive and the import inflation is higher. Also, the demand for the imports decreases. On the other hand, exports are increased as same as domestic output. In the short term, domestic consumption decreases, but in the long term increases as same as the demand for imports. That may be explained by the higher interest rates that make current consumption more expensive compared to future and by short term influence of intervention shock. Both higher imports inflation and domestic goods inflation lead in the short time to higher domestic inflation rate.

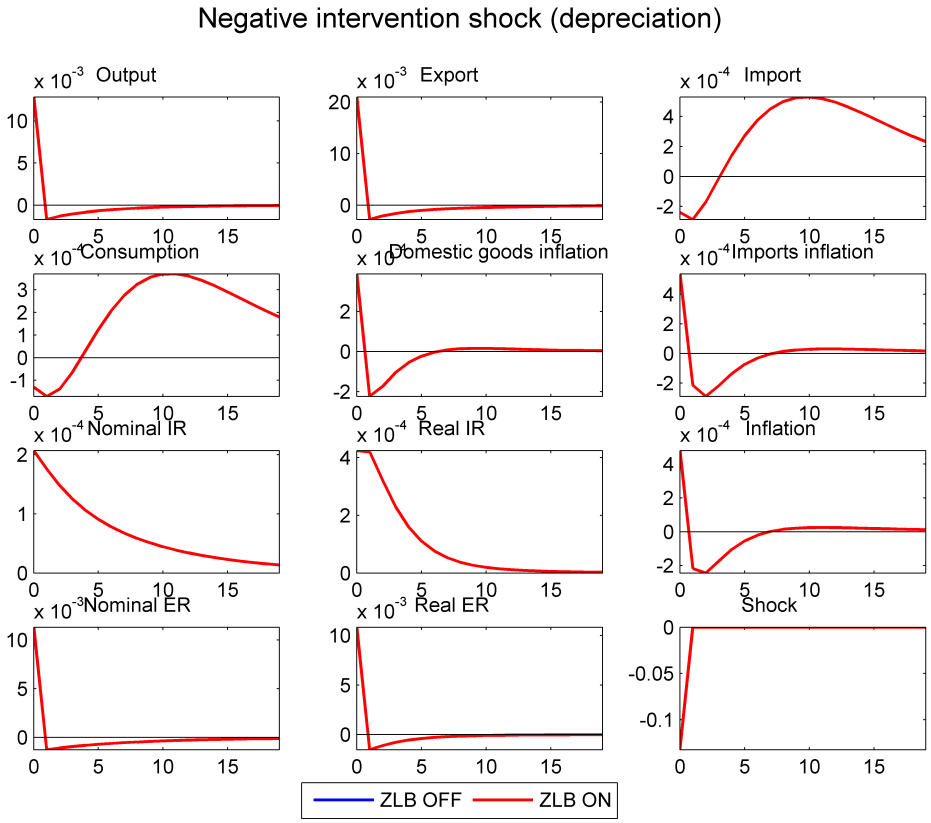


Figure 2 The impulse response functions of selected model variables to intervention shock.

The depreciation of the foreign exchange rate tends to move the economy to the desired direction, nevertheless the short time intervention shock cannot significantly influence the economy in the longer

term. To increase the inflation expectations, Czech National Bank declared the foreign exchange rate commitment. They are prepared to 'intervene on the foreign exchange market if needed to weaken the koruna so that the exchange rate of the koruna is kept close to CZK 27 to the euro', as long as the inflation rate will be under the inflation target. Recently, Czech National Bank will not discontinue the use of this monetary policy instrument until the second half of 2016.

The second version of a model includes a foreign exchange rate targeting. Central bank depreciated currency and try to fix the new level of foreign exchange rate. That refers to the current situation in the Czech economy. Even that the growth of domestic product in the first quarter of 2014 became one of the strongest in Europe and the domestic consumption and expectations showed positive development, the inflation rate decreased. Representatives of the central bank admitted that the inflation rate decreases and moves away from the inflation target. They explained it by the negative foreign inflation shock, especially by low oil prices. In our model, negative inflation shock is equivalent to the positive shock in technology.

Foreign positive technology shock causes decrease of the prices in the foreign economy and the subsequent decrease of the prices of imports. The demand for the imports in the domestic economy as same as consumption increase and domestic consumers substitute some ratio of domestic goods to imports. Cheaper imports decrease the domestic inflation rate. In the case, when ZLB does not bind on the interest rates, central bank would decrease the nominal interest rates. The combination of the fixed nominal interest rates and the lower inflation rate increases the real returns and that attracts the foreign investors. Their demand for domestic currency makes pressure on the appreciation of the foreign exchange rate. Comparison of the classic and constrained impulse responses shows that the fixed nominal interest rate increases responses of consumption and imports above the unrestricted level.

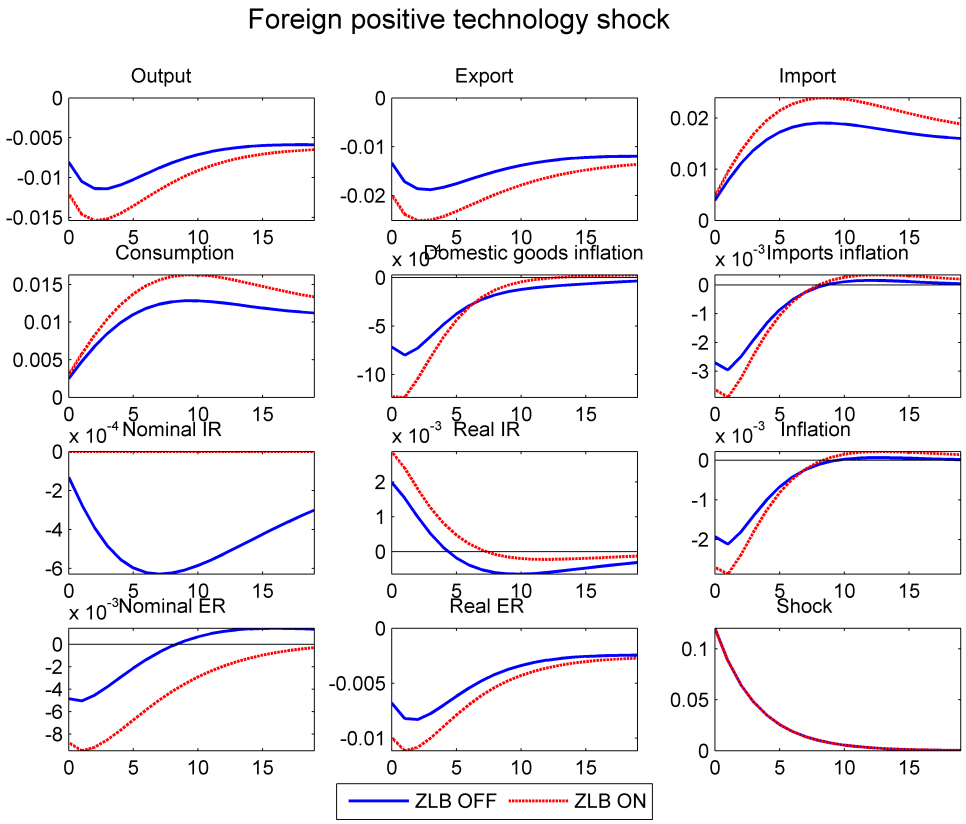


Figure 3 The impulse response functions of selected model variables to foreign technology shock.

These results suggest that the negative foreign inflation shock tends to neutralise the intervention policy and delays the return of the inflation rate to target as well as keeps the interest rates on the zero values. The Czech National Bank declared that its bank board is ready to consider the second depreciation of the foreign exchange rate in case of permanent deflationary risk, but the impact of that may be questionable. However, the recent situation suggests that this intervention may not be necessary.

As the ECB announced the quantitative easing and the price of oil is slightly increasing, the foreign negative inflation shock will be weaker and the foreign exchange rate targeting will have intended effect on economy.

6 Conclusion

In this paper, we dealt with the implementation of non–linear constraints into the DSGE models. We were using the shadow price algorithm that enables us to simulate the model that represents the economy, where the values of the nominal interest rates reached the Zero Lower Bound. We were analysing the robustness of this algorithm considering the changes of parameters within the estimated intervals. We generated samples of the parameters from their posterior distributions and executed the simulation of the model for each generated set of parameters.

The main goal of our work was to analyse the foreign exchange rate intervention, used by the Czech National Bank as an unconventional monetary policy tool. We analysed two modifications of the model. The first was a model without foreign exchange rate targeting and the second was a model with inflation targeting. We analysed the impulse responses that represent the behaviour of economy after the central bank depreciates the foreign exchange rate. Our results suggested that the depreciation tends to have the desired impact on economy, but the short term intervention shock cannot influence economy in the long term. In the second case, we studied the impact of the negative foreign inflation shock (positive technology shock) on the economy under the inflation targeting. Our results showed that the foreign positive technology shock tends to neutralise the intervention policy and may extend the period of low inflation rates and the extremely low interest rates.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1235/2014. This support is gratefully acknowledged.

References

[1] GUERRIERI, Luca and Matteo IACOVIELLO. *OccBin: A Toolkit for Solving Dynamic Models With Occasionally Binding Constraints Easily* 2014

[2] HOLDEN, Tom. and Michael PAETZ. *Efficient simulation of DSGE models with inequality constraints* Discussion Paper 1612, School of Economics, University of Surrey. 2012

[3] JUSTINIANO, Alejandro. PRESTON, Bruce. *Monetary policy and uncertainty in an empirical small open economy model.* Working Paper Series WP-09-21, Federal Reserve Bank of Chicago, 2010.

[4] MALOVANÁ, Simona. *The Effectiveness of Unconventional Monetary Policy Tools at the Zero Lower Bound: A DSGE Approach.* Diploma thesis, Charles University, Prague. 2014.

[5] MONTORO, Carlo and Marco ORTIZ. *Foreign Exchange Intervention and Monetary Policy Design: A Market Microstructure Analysis.* Third draft. 2013.

Estimation and Analysis of Consumption Function in Slovakia

Kvetoslava Surmanová[1], Andrea Furková[2], Marian Reiff[3]

Abstract. This paper presents an econometrics model of consumption expenditure in the Slovak republic. The consumer spending is often discussed topic in economics. The original consumption function authored by Keynes was subjected to criticism since its inception on the theoretical and empirical level too. In our paper we will first focus on the application of econometric Friedman´s permanent income hypothesis. This research is based on quarterly data covering the period from 1995 to 2014. The reference period is characterized by forming economy to a market economy. This process was accompanied by a number of one-off measures for example tax reform, privatization and change currency. In addition, consumer behaviour was affected by the global financial and economic crisis. Due to adverse developments in the labour market worsened income situation of the population which led us to the observed indicator. Therefore, in the second part we devote stability testing structure of the analysed indicator.

Keywords: consumption, permanent income, partial adjustment, rational expectations

JEL Classification: E21
AMS Classification: 62P20

1 A brief history of modern consumption theory

Consumption is the main component of aggregate demand and, therefore is important in the economic analysis and in particular in macroeconomics. Among the most common and best-known approaches to the analysis of consumption function can be included Keynesian theory. The main well-known basic characteristics of Keynes´ analysis are that the marginal propensity to consume (MPC) falls with income, as does the average propensity to consume (APC).

In response to the Keynes´ theory Milton Friedman proposed in 1956 his permanent income hypothesis. The main idea of the permanent income hypothesis of Friedman arises from the fact that households are faithful to their consumption patterns, regardless of short-term income and thus consumption is much more stable than expected. According to Friedman, population realizes its consumption with respect to the amount of permanent income[4].

At the same times was developing theory of lifecycle model. Authors were Modigliani and Brumberg in 1955. According to lifecycle theory residences maximize their lifetime utility subject to their lifetime budget constraint. It is most often formulated at the macro level, but we can also see it applied in microeconomics.

Both theories (permanent-income hypothesis and life-cycle hypothesis) were developed at least partly in response to the discrepancy between Keynes's conjecture and the empirical evidence.

Davidson, Hendry, Srba and Yeo [4] published in 1978 theory based on Friedman´s hypothesis. It was the first econometric model of consumption function incorporating an error correction mechanism (ECM).

Many studies have used these models or a combination of the variables suggested by these approaches, in order to explain consumption (see Byrne and Davis [3], Bover [2], Senaj [6]).

[1] University of Economics/Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemska 1, kvetoslava.surmanova@euba.sk.
[2] University of Economics/Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemska 1, andrea.furkova@euba.sk
[3] University of Economics/Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemska 1, marian.reiff@euba.sk.
[4] Permanent mean to be oriented long-term increase in income.

1.1 Household consumption in the years 1995 to 2014

Household consumption is recognized as one of the components of GDP expenditure. This indicator captures the total consumption of goods and services in households based on classification COICOP according to ESA 2010. On the reference period of the years 1995–2014 the final consumption shows relatively fast and steady upward trend. This course was disturbed only by the financial crisis in the fourth quarter of 2008 (see Figure 1).

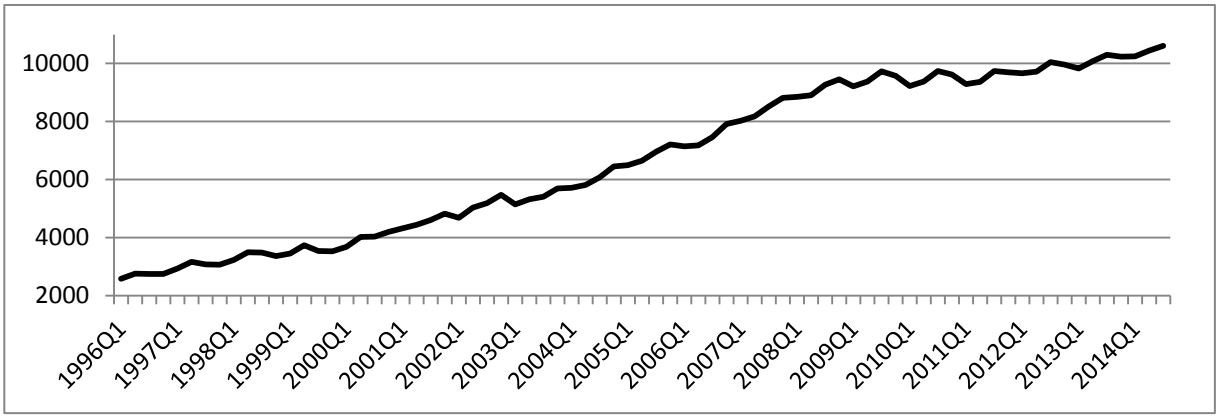


Figure 1: Final household consumption over time (in mil. EUR and constant prices)

About half of aggregate household consumption is formed by expenditures on food and soft drinks, along with spending on housing and related services. It is interesting that while spending on food, by comparing the years 1996 and 2014 have dropped from about 28% to 18% of the total population consumption, expenditure on housing and related services have increased in the period by about 7% (from 19% to 26%).

Figure 2 demonstrates the share of consumption in total GDP. Household final consumption bears to the total GDP of more than 50 percent for the whole period.

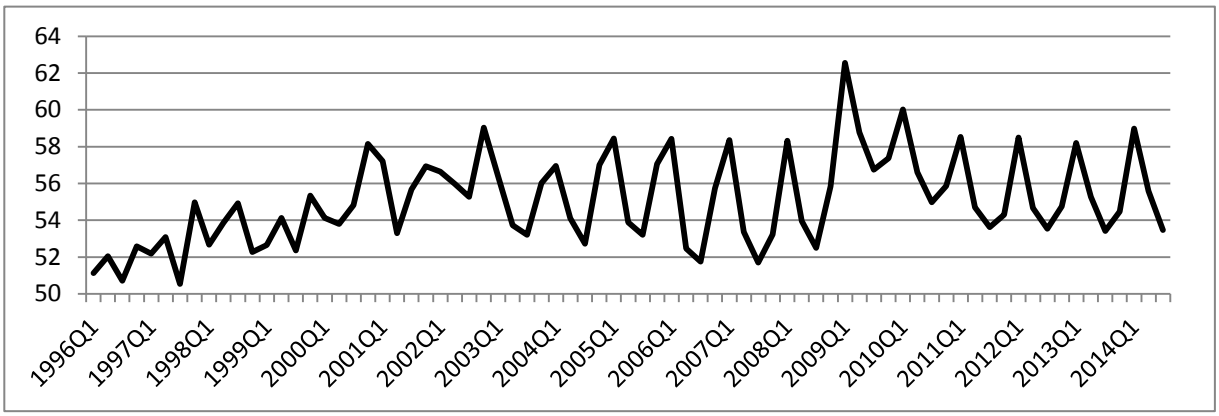


Figure 2 Percentage ratio of final household consumption to GDP (in constant prices)

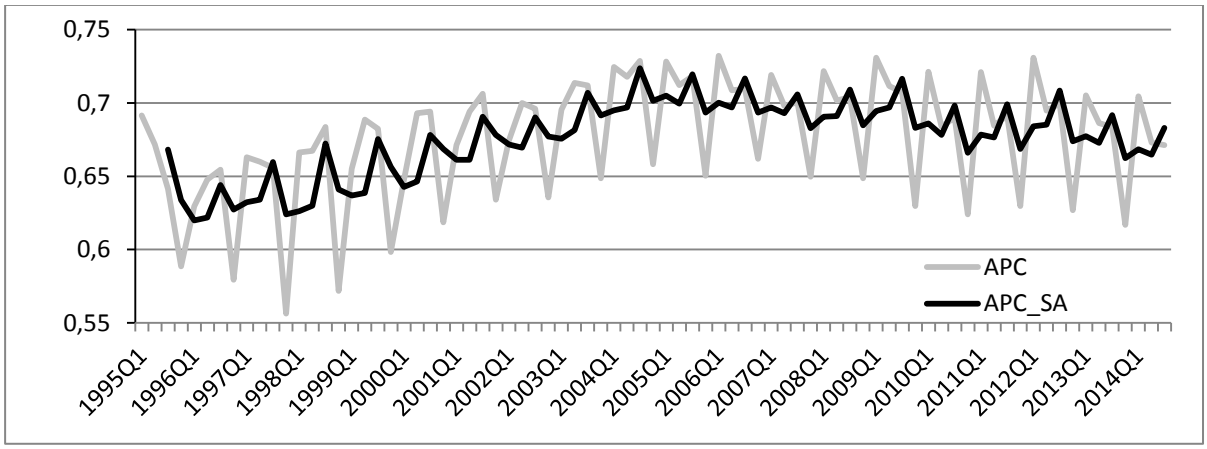


Figure 3 Average propensity to consume (APC) and seasonal adjusted APC

The economic development at the end of the year 2008 was marked by the crisis, resulting there were problems in the labor market. These changes led to stagnation in labor income, negative expectations of households, etc. These factors have led to a reduction in the average propensity to consume. The change that has occurred has not been large, which can be observed in Figure 3. The slight change in APC captured in crisis period was, in our opinion dampened the already low level of disposable income in the economy. Due to the seasonal component of time series of population consumption curve we present the average propensity to consume, even after adjusting for seasonal component[5].

1.2 Concepts of consumption function econometric modeling

Consumption function based on the principle rational expectations

As a reaction to Keynes's theory, further advanced theories of consumption and savings have been developed. One of the theories is Friedman's permanent income theory. In 1957 Friedman expanded Keynes's findings by dividing consumption and income for permanent and temporary part. The inquiry found that households are faithful to their consumption patterns, regardless of short-term income and thus consumption is much more stable than expected. Emphasis was given on forming expectations of future income. This theory can be written as follows:

$$C_t = \beta_0 + \beta_1 Y_t^p + u_t \tag{1}$$

Where β_1 denotes the marginal propensity to consume out of permanent income Y_t^p. Based on the formation of expectations regarding the future income

$$Y_t^p = Y_{t-1}^p + \lambda\left(Y_t - Y_{t-1}^p\right), \tag{2}$$

while for coefficient of adaptive expectations λ apply $0 < \lambda \leq 1$, equation (1) can be modify to autoregression model of adaptive expectations:

$$C_t = \beta_0\lambda + \beta_1\lambda Y_t + (1-\lambda)C_{t-1} + v_t. \tag{3}$$

Stochastic term $v_t = u_t - (1-\lambda)u_{t-1}$ is composed stochastic term and to estimate parameters (3) we can no longer apply Ordinary least squares method. Suitable method to estimate parameters is Instrumental variables method.

Consumption function based on the principle partial adjustment

From the point of view of econometric model parameters estimation is easier Friedman's hypothesis defined as a partial adjustment model.
Starting from the model:

$$C_t^p = \beta_0 + \beta_1 Y_t + u_t. \tag{4}$$

For immeasurable, permanent consumption C_t^p it is necessary to adopt a presumption of partial adaptation:

$$C_t - C_{t-1} = \varphi\left(C_t^p - C_{t-1}\right). \tag{5}$$

where φ is coefficient of partial adjustment and applies $0 < \varphi \leq 1$.
By substitution of (4) dividing into (5) and after adjustments it is received again autoregression model:

$$C_t = \beta_0\lambda + \beta_1\varphi Y_t + (1-\varphi)C_{t-1} + v_t. \tag{6}$$

The comparison of random components in the relations (3) and (6) is a random component in relation (6) as a simpler $v_t = \varphi u_t$. Therefore, to estimate parameters for the equation (6) it is possible to use Ordinary least squares method.

Consumption function based on the principle of Error correction model (ECM)

Concept of ECM model is developed by Davidson, Hendry, Srba and Yeo [4]. Based on a long-term relationship (1) conditioned by autonomous consumption value $\beta_0 = 0$. Following equation (7) is received by logarithm of (1) and utilization of the model Autoregressive distributed lag - ADL(1, 1):

$$c_t = m + \alpha_1 c_{t-1} + \beta_0 y_t^d + \beta_1 y_{t-1}^d + u_t. \tag{7}$$

[5] The time series APC was adjusted by the method of moving averages.

In the Error correction model have been used the first differences of the variables c_t a y_t^d to ensure their stationarity because all the selected time series were nonstationary and integrated fist order. As the integration of both time series is the same, the cointegration concept can be applied. In [5], we can see that construction of the ECM took into account cointegration vector and residual values from the cointegration equation with one period lag:

$$\Delta c_t = \beta_0 + \beta_1 \Delta y_t^d + \lambda(c_{t-1} - \alpha_0 - \alpha_1 y_{t-1}^d) + v_t \quad (8)$$

ECM models are currently widely used models. Their advantage is the ability to separate the short-term effects from long-term impact on the determinants of explained variable.

1.3 A Test Friedman´s permanent income hypothesis in conditions of Slovakia

Analysis has been carried out on quarterly data for the period 1995–2014. The starting model for implemented analysis is model (6). Variables of final consumption of household (C) and Gross domestic product (Y^d) are in mil. EUR, in current prices.

In Table 1 we present the results of estimated models. Two variants of models were estimated: model of partial adjustment and model of rational expectations. For each alternative, firstly basic models (M1 and M5) were estimated.

Chow test was used to test the structural stability of parameters. Value of Chow statistic greater than the critical value of the Fisher distribution at level of significance 5% confirmed the rejection of the null hypothesis. Null hypothesis states that regression of structural change before and after are the same. The break point for structural change of parameters is the period 2008Q4, when in the Slovak Republic showed the effects of the financial crisis. Chow test itself doesn't show which parameters are changing. Therefore, the additional models with dummy variable crisis (M2 and M6) were estimated. Dummy variable c*risis* = 1 for $t \in (2008Q4, 2014Q3)$ and *crisis* = 0 during the period before of the structural change.

The time series used for model estimation are without seasonal adjustment. Therefore, the basic model extended to include dummy variable of seasonal character (M3 and M7) was estimated. Models M3 and M7 were tested with Chow test too. Parameters of both models are unstable again. Estimated models M4 and M8 show a statistically significant change of intercept in the period from 2008Q4.

Models of partial adjustment (M1 – M4) were estimated with least square method and models M5 – M8 with instrumental variables method.

	Model of partial adjustment				**Model of rational expectations**			
	M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**
cons	80.94	-120.94	140.4	-84.85	-117.94	-146.02	250.43	78.05
Y^d	0.13***	0.22***	0.19***	0.29***	0.48***	0.48**	-0.07	-0.04
$C_{(-1)}$	0.81***	0.73***	0.72***	0.62***	0.31	0.33	1.09***	1.03***
crisis	-	1 448.89**		-180.05***		-51.55		-207.31***
crisis*Y^d	-	-0.11***						
crisis*S4				-360.90***				
S1			-158.42***	-109.33***			-304.76***	-252.42**
S2			-239.24***	-216.11***				
R^2	0.996	0.99	0.997	0.999	0.989	0.99	0.996	0.997
D-W	1.994	2.17	1.95	2.16	2.388	2.46	1.99	2.37
Chow	7.689		19.17		4.705		22.54	

Table 1 Results of estimated consumption function, own calculations
*** Parameter is significant on 1% level of significance
** Parameter is significant on 5% level of significance

Conclusion

The purpose of this paper was to specify and estimate a consumption function as a Friedman´s permanent-income hypothesis. We focused on testing the structural stability of parameters after the period 2008Q4.

From the results of the different model estimation summary (Table 1) we can conclude that under the Slovak Republic conditions, it is more appropriate to apply the model of partial adjustment in order to estimate consumption function. All parameters in M1 to M4 are statistically significant and R2 is at least 99%. Chow test has shown that the incorporation of dummy variable, crisis into the model is necessary, leading to an improvement of the original basic model. Based on our opinion, model M4 describes the behavior of household final consumption best. This claim, we could demonstrate as the predictive power of the model, which may be a subject of further research of presented problem.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0444/15 "Econometric Analysis of Production Possibilities of the Economy and the Labour Market in Slovakia"

References

[1] Arlt, J., Čutková, J.and Radkovský, Š. 2001. *Analýza spotřební funkce v podmínkach ČR*. WP **34**, Praha 2001.

[2] Bover, O.: *Wealth effects on consumption: Microeconometric estimates from the Spanish survey of household finances*. Bank of Spain working paper, Documentos de Trabajo, No. **0522**, 2005.

[3] Byrne, J. P. and Davis, P. E.: Disaggregate wealth and aggregate consumption: an investigation of empirical relationships for the G7. *Oxford Bulletin of Economics and Statistics* **65** (2), 2003, 197 – 220.

[4] Davidson, J. E. H., Hendry, D. F., Srba, F.and Yeo, S.: Econometric Modelling of the Aggregate Time-Series Relationship Between consumers´ Expenditure and Income in the United Kingdom. *The Economic Journal*, **88**, 1978, 661 – 692.

[5] Lukáčik, M., Lukáčiková, A., and Szomolányi, K.: The Long-run Determinants of Money Demand in Slovakia. In: *Quantitative Methods in Economics: Multiple Criteria Decision Making XIII*. IURA Edition, Bratislava, 2006, 92–100.

[6] Senaj, M.: *Odhad spotrebnej funkcie pre Slovensko a prognóza spotreby*. WP 1/2007, 2007, online www.nbs.sk/PUBLIK.

[7] www.statistics.sk

Efficient Market Theory from the point of view of Markov chains analysis

Milan Svoboda[1], Pavla Říhová[2]

Abstract. This paper follows our previous studies which were dealing with applicability of Markov chains analysis (MCA) for prediction of the Czech Stock Market. These studies indicated that the trading strategies based on MCA could beat the market. However, that is in contradiction to the Efficient Market Theory. Now we focus on the question whether share markets perform random walk. When analysing we proceed from the following idea. In the case that a share rate makes a random walk, other step should not be dependent on the previous step. Provided that, the probability of the transition from any state *i* to the state *j* should be equal to the relative frequency of the state *j* and should be equal to the fixed vector components. The transition probability matrix should be identical to the limit matrix. The study is carried out on the day prices changes of three share titles ČEZ, KB, O2. In order to verify the hypothesis of concordances of individual rows in the transition probability matrix with the fixed vector, The Chi-square test was used. The test statistics values do not support the statement of completely random walk of share rates in the Czech Stock Market.

Keywords: Markov chain analysis, Efficient market theory, transition probability matrix, stock market, technical analysis.

JEL Classification: C02, C13, G14, G19
AMS Classification: 90C40, 91B82

1 Introduction

One of the characteristics of share markets is their permanent variation. There are periods of different lengths either of share price growth or of share price decrease. All investors' dream is to be able to predict this variation and therefore to know the future share price. One of the employed approaches to prediction is technical analysis (TA). According to TA the successful prediction of share price development is possible. TA is represented by a wide range of methods which can deduce the future prices from the past prices and the trade volumes. The main aim of TA is the prognosis of short term price movements, important being not the price level but the price estimation. TA works on two assumptions:

- Share prices are moving in trends. Rates have their clearly defined trend (movement direction). Trends have a certain degree of inertia. The share price moves in a certain direction and the movement in this direction becomes weaker after a period of time. Afterwards the price moves in the opposite direction. These trend changes depend on each other.
- Development formations and cycles repeat themselves. This assumption is based on the human nature to behave in a similar way under the same circumstances.

TA methods were ignored by academicians for a long time. The reason is Efficient market theory. According to this theory the share price performs a random walk and that is why the share price is not predictable. The idea of a random walk was probably first published by mathematician L Bachalier [1] in his thesis. Later the theory was developed and with empirical studies supported by E. Fama [3], [4]. According to this theory all expected information is already calculated in the share price. As the purpose of the rate change is considered to be unexpected information. As effective is reckoned such market, which absorbs very quickly all this unexpected information. Since it is not possible to predict unexpected information in advance because it comes accidentally, the price change is accidental as well. If this theory was valid, TA would not be applicable and in the long term all investors would reach comparable results. With the massive introduction of computer technology and with electronic databases development there are more and more studies appearing, e.g. Sweeney[7] and Brock et al.[2], showing that TA methods are able to overcome the market and therefore calling Efficient market theory into question.

[1] University of West Bohemia, Husova 11, 306 14 Pilsen, Czech Republic, e-mail: svobodm@kem.zcu.cz.
[2] University of West Bohemia, Husova 11, 306 14 Pilsen, Czech Republic, e-mail: divisova@kem.zcu.cz.

This paper follows our previous studies Svoboda and Lukáš [5], [6], where we modelled the share price variation by Markov chains. Now we are going to concentrate on the question whether these models support the theory of share markets random walk.

2 Data and Methodology

2.1 Markov chains

Markov chains are used for modelling of processes which can occur in one of finite (countable) number of states in discrete time moments. Markov chain (MC) is understood as a sequence of discrete random variables X_1, X_2, X_3,.... with the Markov property which can be formally described as follows:

$$P(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, .., X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

In words, MC is a random process with a discrete set of states, discrete time and of that kind that the probability $p_i(n)$ that at the time moment t_n the process will be in state i, is stochastically dependent only on the state at the previous time moment, i.e. on the state at the time t_{n-1}. Particular realizations x_i are elements of a countable set $S = \{s_i\}$, $i = 1, 2, ..., N$ which is called a state space. Behaviour of the described process is determined by:

- vector of unconditional probabilities $\mathbf{p}^T(n) = [p_1(n), p_2(n), ..., p_N(n)]$, $n = 0, 1, 2, ...$ where ^{T} denotes transposition,
- transition probability matrix **P** whose elements give conditional probabilities $p_{ij} = P(X_n = s_j \mid X_{n-1} = s_i)$, $i,j = 1, 2, .., N$, where p_{ij} can depend on n. In the case that p_{ij} does not depend on n at all, we speak about homogenous MC, in the opposite case we speak about non-homogenous MC.

If we know the probability of particular states appearance at the time moment when the process starts, we can describe behaviour of the process using the following relations:

$$\mathbf{p}^T(n) = \mathbf{p}^T(n-1)\mathbf{P} \tag{1}$$

and by successive substitution we get:

$$\mathbf{p}^T(n) = \mathbf{p}^T(0)\mathbf{P}^n \tag{2}$$

Behaviour of homogenous MC after n time moments is determined by the initial vector and n^{th} power of transition probability matrix. When analysing stochastic process behaviour we are interested in transition probability between each state. These probabilities are given by elements of matrix $\mathbf{P}^n$ (called p_{ij}^n).

A MC is said to be a regular chain if some power of the transition matrix has only positive elements. From theory it is known that if MC is regular then:

$$\mathbf{U} = \lim_{n\to\infty} \mathbf{P}^n \ , \tag{3}$$

where **U** is an N by N matrix, whose rows are the same and consist of a vector $\mathbf{u} = [u_1, u_2, ..., u_N]$ which we call fixed vector (stationary distribution). By limit transition of relation (1) we get to the conclusion that **u** is a solution of homogenous system of algebraic equations when meeting the condition that the sum of probabilities in each row equals 1. We can describe it as follows:

$$\mathbf{u} = \mathbf{uP} \ \text{ and } \sum_{i=1}^{N} u_i = 1 \ , \tag{4}$$

Fixed vector components can be interpreted as parts of the total time which the process spends in particular states during a rather long period. But that implies that fixed vector components limitly equal relative frequency of particular states.

If Efficient Market Theory is valid, the share price performs a random walk. It means that the following "step" should not be conditioned by the previous step. Probability of transition from any state i to the state j should be the same and equal relative frequency of state j. Transition probability matrix P should be therefore also identical to the limit matrix U, i.e. every row should be the same as fixed vector.

2.2 Data

The study is based on the data from the Prague Stock Exchange in the period of 7 years from 2 January 2006 until 2 January 2013, specifically for the close daily prices of the following shares: ČEZ, Komerční Banka and Telefonica O2. In the monitored period of time the mentioned companies regularly paid out the dividends. For this reason the close prices on the Ex date were increased by a dividend. In this way the price drop, caused by the loss of dividend claim, was eliminated. Out of these modified close prices we have calculated the share price daily change y_t. The length of the time period has been determined by the number of growing or decreasing close prices in sequence according to the relations:

$$y_t = \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{5}$$

Table 1 illustrates the calculation y_t of closing prices (denoted Pt) of ČEZ shares.

t	1	2	3	4	5	6	7
P_t	801,2	809,0	813,7	807,5	802,1	819,0	826,0
y_t		0,0097	0,0058	-0,0076	-0,0067	0,0211	0,0085

Table 1 Example - calculation of y_t

2.3 State space

We have defined state space model according to the following principle. The state space has eight states. Four states are defined for the cases when the share price decreases $\{D_4, D_3, D_2, D_1\}$. D_4 denotes the highest decrease and D_1 the lowest one. Four states are defined for the cases when the share price grows $\{G_1, G_2, G_3, G_4\}$. G_1 denotes the lowest growth and G_4 the highest one. The size of growth or decrease will be defined for each share according to standard deviation (denoted σ), which is calculated out of daily price changes y_t. For particular titles we have calculated the following figures: σ_{O2} = 0,01442, σ_{CEZ} = 0,01996, σ_{KB} = 0,02355. We define particular states as follows:

$$D_4: y_t < -1{,}5\sigma, \quad D_3: -1{,}5\sigma \le y_t < -1\sigma, \quad D_2: -1\sigma \le y_t < -0{,}5\sigma, \quad D_1: -0{,}5\sigma \le y_t < 0,$$
$$G_1: 0 \le y_t < 0{,}5\sigma, \quad G_2: 0{,}5\sigma \le y_t < 1\sigma, \quad G_3: 1\sigma \le y_t < 1{,}5\sigma, \quad G_4: 1{,}5\sigma \le y_t .$$

For thus defined model, we calculated conditional transition probability between particular states p_{ij}, the number of particular states appearance n_i, and conditional probability of decrease $d_i = \sum_{j=1}^{4} p_{ij}$ and growth $g_i = \sum_{j=5}^{8} p_{ij}$.

3 Empirical results

We have implemented an application for VBA in MS Excel for our calculations. All results are presented in Table 2. Calculated probabilities p_{ij} are shown with three decimal places, therefore the row sum of probabilities does not have to be equal to 1.

n-1 \ n		D_4	D_3	D_2	D_1	G_1	G_2	G_3	G_4	d_i	g_i	n_i
D_4	O2	0,139	0,101	0,139	0,127	0,101	0,038	0,152	0,203	0,506	0,494	79
	CEZ	0,167	0,051	0,128	0,154	0,179	0,064	0,077	0,179	0,500	0,500	78
	KB	0,139	0,089	0,101	0,190	0,127	0,165	0,051	0,139	0,519	0,481	79
D_3	O2	0,079	0,059	0,119	0,168	0,327	0,139	0,059	0,050	0,426	0,574	101
	CEZ	0,105	0,074	0,095	0,147	0,232	0,116	0,126	0,105	0,421	0,579	95
	KB	0,089	0,069	0,188	0,168	0,218	0,149	0,059	0,059	0,515	0,485	101

D_2	O2	0,041	0,041	0,165	0,230	0,284	0,132	0,074	0,033	0,477	0,523	243
	CEZ	0,049	0,061	0,147	0,208	0,282	0,184	0,045	0,024	0,465	0,535	245
	KB	0,053	0,061	0,132	0,268	0,246	0,132	0,075	0,035	0,513	0,487	228
D_1	O2	0,037	0,068	0,141	0,239	0,312	0,129	0,041	0,032	0,485	0,515	410
	CEZ	0,024	0,034	0,147	0,265	0,313	0,142	0,058	0,017	0,470	0,530	415
	KB	0,039	0,059	0,129	0,249	0,308	0,120	0,048	0,048	0,476	0,524	441
G_1	O2	0,033	0,049	0,124	0,274	0,329	0,118	0,037	0,035	0,480	0,520	508
	CEZ	0,029	0,062	0,153	0,235	0,297	0,163	0,045	0,016	0,478	0,522	485
	KB	0,022	0,043	0,110	0,300	0,337	0,116	0,049	0,024	0,475	0,525	493
G_2	O2	0,034	0,060	0,149	0,226	0,285	0,153	0,068	0,026	0,468	0,532	235
	CEZ	0,034	0,034	0,123	0,299	0,272	0,160	0,045	0,034	0,489	0,511	268
	KB	0,036	0,063	0,129	0,263	0,277	0,129	0,058	0,045	0,491	0,509	224
G_3	O2	0,039	0,049	0,098	0,225	0,255	0,206	0,098	0,029	0,412	0,588	102
	CEZ	0,030	0,040	0,150	0,240	0,250	0,140	0,080	0,070	0,460	0,540	100
	KB	0,029	0,049	0,167	0,176	0,245	0,147	0,098	0,088	0,422	0,578	102
G_4	O2	0,074	0,062	0,173	0,173	0,123	0,185	0,049	0,160	0,481	0,519	81
	CEZ	0,096	0,164	0,096	0,137	0,123	0,151	0,068	0,164	0,493	0,507	73
	KB	0,088	0,077	0,154	0,143	0,176	0,121	0,088	0,154	0,462	0,538	91

Table 2 transition probabilities p_{ij}

Now we will determine fixed vector components for particular shares. We will get them by solving the system of equation (4). The solution is seen in Table 3.

	u_1 (D_4)	u_2 (D_3)	u_3 (D_2)	u_4 (D_1)	u_5 (G_1)	u_6 (G_2)	u_7 (G_3)	u_8 (G_4)
O2	0,0449	0,0574	0,1382	0,2330	0,2887	0,1331	0,0580	0,0467
CEZ	0,0443	0,0540	0,1393	0,2359	0,2763	0,1518	0,0569	0,0415
KB	0,0449	0,0574	0,1296	0,2507	0,2802	0,1268	0,0586	0,0518

Table 3 elements of fixed vector

As it has been said before, if Efficient market theory was valid, individual rows of the matrix P should be identical to the fixed vector. Now we will verify whether the differences between the fixed vector and individual rows of the matrix P could be caused by a coincidence or whether the differences are significant enough to declare that a difference exists.

To verify the hypothesis of the concordance between the fixed vector and individual rows of the matrix P we will use The Chi-square test:

$$G_j = \sum_{i=1}^{k} \frac{\left(n_i - n\pi_i\right)^2}{n\pi_i} \tag{6}$$

where G_j is value of statistic for j^{th} state,
n_i are numbers of realised transitions from the j^{th} state to the i^{th} state,
n is the number of j^{th} states,
π_i is relative frequency of i^{th} state (theoretical frequency).

Statistics G has, provided sufficiently large number of elements, approximately χ^2 distribution with $v = k - 1$ degrees of freedom, which means that in our case v=7. The test will be carried out for $\alpha = 0{,}05$. When $\alpha = 0{,}05$ and v=7, null hypothesis is denied if value of statistic $G > 14{,}1$.

state	D_4	D_3	D_2	D_1	G_1	G_2	G_3	G_4
G_{O2}	91,27	5,30	4,45	6,16	15,05	4,13	9,23	36,00
$G_{ČEZ}$	71,62	32,02	5,19	17,13	13,36	8,94	4,15	57,76
G_{KB}	37,96	11,93	3,94	2,80	27,84	0,90	10,21	32,33

Table 4 values G_j for particular shares

4 Conclusion

As it is seen from the results, the critical values of all shares were considerably exceeded in both side states, i.e. in case of strong daily growth or decrease of share price. On the contrary in case of small changes of share price the values of test statistics are low. When we observe the probability of transition between particular states in Table 2, we can see that there is a higher probability of transition from the state D_4 to the state D_4 than from other states. A similar situation applies for the state G_4. These findings do not support one of the basic assumptions of Efficient market theory that the share price reacts immediately to unexpected information. The results thus indicate that the Czech stock market is not able to absorb new information quickly and exactly and therefore it does not behave effectively.

Other research will be developed in the following courses:

- By prolonging the observed time rows,
- By a differently defined state space,
- By applying the method to other stock markets,
- By creation of business strategies which make use of MC analysis.

Acknowledgements

The research project was supported by the grant no. GA15-20405S of the Grant Agency, Prague, Czech Republic and by the grant SGS-2015-026 of the Faculty of Economics, University of West Bohemia in Pilsen..

References

[1] Bachelier, L.: Theorie de la spéculation. 1900. In: Bachelier, L., Samuelson, P. A., Davis, M., Etheridge, A.: *Louis Bachelier's Theory of Speculation: the Origins of Modern Finance*, Princeton NJ: Princeton University Press, 2006, ISBN 978-0-691-11752-2

[2] Brock, W., Lakonishok, J. and LeBaron, B.: Simple Technical Trading Rules and The Stochastic Properties of Stock Returns, Journal of Finance, 1992, Vol. 47, No. 5, pp. 1731 – 1764.

[3] Fama, E., F. Efficient Capital Markets: A Review of Theory and Empirical Work, Journal of Finance, 1970, Vol. 25, No.2, pp. 383 – 417.

[4] Fama, E., F. and Blume, M., E. Filter Rules and Stock Market Trading, Journal of Business, 1966, Vol. 39, pp. 283 – 306.

[5] Svoboda, M., Lukáš, L. Application of Markov chain analysis to trend prediction of stock indices, Proceedings of 30th International Conference Mathematical Methods in Economics, 2012, pp. 848 – 853

[6] Svoboda, M., Lukáš, L. Application of non-homogeneous Markov chain analysis to trend prediction of stock indices, Proceedings of 31th International Conference Mathematical Methods in Economics, 2013, pp. 881 – 886.

[7] Sweeny, R., J. Some new filter rule tests: methods and results and Quantitative Analysis. Journal of Financial 1988, Vol. 23, pp. 285–300.

Long-Run Elasticity of Substitution

Karol Szomolányi, Martin Lukáčik, Adriana Lukáčiková[1]

Abstract. Actual approaches dealing with measuring supply side of the economy focus on concept of constant elasticity of substitution production function normalization. There is empirical evidence that aggregate production function is better characterized by a non-unitary elasticity of substitution. However, there are theoretical and empirical suggestions for dual elasticity of substitution (short-run and long-run or local and global) supported by the non-stationarity of the studied data as well.
In the paper we show, how to estimate long-run elasticity of substitution using labour income co-integration estimation form. We use production function normalization where normalization values are expressing long-run dynamics to derive the estimation form. For estimation we use the U.S. data with huge range and find an indication that U.S. long-run production function is in Cobb-Douglas form.

Keywords: constant elasticity substitution and Cobb-Douglas production function, inputs substitution elasticity, labour share, capital share.

JEL Classification: C13, E23
AMS Classification: 62H12, 91B64

1 Introduction

Actual approaches dealing with measuring supply side of the economy focus on concept of constant elasticity of substitution (CES) production function normalization suggested by De la Grandville [9]. There is empirical evidence that aggregate production function is better characterized by a non-unitary elasticity of substitution, e.g. Chirinko et al. [8], Klump et al. [19] and León-Ledesma et al. [23]. Chirinko [7] survey suggests that most evidence favors elasticities ranges of 0.4 – 0.6 for the US. Jones [15], [16] argued that capital shares exhibit swings and trends in many countries as to be inconsistent with Cobb-Douglas or CES with Harrod-neutral technological progress needed for neoclassical balanced growth.[2]

Recent theoretical papers distinguish long-run and short-run elasticity of substitution of inputs. Jones [15] and [16] consider Cobb-Douglas production function from in the long-run view and CES production function with an elasticity of substitution less than one from the short-run view, because input innovations are Pareto distributed from the long-run. Jürgen [17] suggests that transition state economies have less substitution elasticity than the steady state economies. Jürgen [18] predicts that competitive equilibrium can be realized only under Cobb-Douglas technology.

Empirical papers estimating the elasticity of substitution use estimation forms with output, labour, capital and capital and labour incomes[3]. Obviously output and labour income are non-stationary and co-integration analysis is a suitable method for a labour income specification. In fact we should consider that many empirical papers estimates short-run elasticity of substitution.[4]

León-Ledesma et al. [22] verified the supply side system of equations by Monte Carlo simulation. Even if capital and labour were generated by a non-stationary process they estimates suggest good results using given specifications and generalized least square method. However, they did not consider changes of elasticity of substitution in time. More over their simulated supply side is independent on demand shocks, while it is known that money is not neutral in the short-run.

In the paper we will attempt to estimate long-run elasticity of substitution using normalized CES production function where normalization values are expressing long-run dynamics. We will derive labour income co-integration estimation form from the Jürgen's [18] and Jones's [15] production function forms. We will estimate the parameters using American annual labour income and intensively expressed output data series. We choose the labour income specification because of the enormous range of the data series (1929 – 2012) suitable for co-integration analysis. Moreover, in respect to Jones's [16] theoretical suggestion we suppose Harrod-neutral non-stationary technical progress innovating labour. Therefore logs of labour income and average product of the labour datasets are non-stationary and co-integrated, while logs of capital income and average product of capital

[1] University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovakia, karol.szomolanyi@euba.sk.
[2] See Barro and Sala-i-Martin [4]
[3] See Klump et. al [20]
[4] See Chirinko, Fazzari and Meyer [8], Klump et al. [19], León-Ledesma et al. [23] and Chirinko [7].

datasets may be stationary. Therefore it is not possible to analyse long-run relationship between these two variables using co-integration analysis.

2 Specification

Jürgen [17] derived normalized CES production function with baseline values of production corresponding to a situation where the economy is in a long-run (optimal) position. His version of production function in extensive form is:

$$Y_t = \left(Y_t^*\right)^{1-\frac{\sigma^*-1}{\sigma^*}\frac{\sigma}{\sigma-1}} \left[\alpha \left(\frac{K_t}{K_t^*} \right)^{\frac{\sigma-1}{\sigma}} \left(K_t^*\right)^{\frac{\sigma^*-1}{\sigma^*}} + (1-\alpha) \left(\frac{A_t N_t}{A_t^* N_t^*} \right)^{\frac{\sigma-1}{\sigma}} \left(A_t^* N_t^*\right)^{\frac{\sigma^*-1}{\sigma^*}} \right]^{\frac{\sigma}{\sigma-1}}; \quad t \in T \tag{1}$$

where long-run production function expresses baseline point of normalization:

$$Y_t^* = \left[\alpha \left(K_t^*\right)^{\frac{\sigma^*-1}{\sigma^*}} + (1-\alpha) \left(A_t^* N_t^*\right)^{\frac{\sigma^*-1}{\sigma^*}} \right]^{\frac{\sigma^*}{\sigma^*-1}}; \; t \in T \tag{2}$$

Values denoted with asterisks are long-run. Inputs are capital K and labour N, output is Y, A is technology, α is capital share, time is defined in the set T. We suppose that a long-run elasticity of substitution is higher than a short-run elasticity $\sigma^* > \sigma$.

If inputs are optimally used long-run (2) and short-run (1) production function forms coincide. It follows that if $K = K^*$ and $N = N^*$ than $\sigma = \sigma^*$. Special case is Jones's [15] modification, where long-run production function (2) is in a Cobb-Douglas form.

We consider Harrod-neutral technological progress with constant growth rate γ in the form:

$$A_t = A_0 e^{\gamma t}; \; t \in T \tag{3}$$

Using first order condition of firm maximizing profit we can derive a long-run estimation form from the long-run production function (2):

$$\ln\left(w_t^*\right) = \eta_0^* + \eta_1^* \ln\left(\frac{Y_t^*}{N_t^*}\right) + \eta_2^* t + v_t^* \tag{4}$$

where w is labour income rate and coefficients are:

$$\eta_0^* = \ln(1-\alpha)$$
$$\eta_1^* = \frac{1}{\sigma^*}$$
$$\eta_2^* = \frac{\sigma^*-1}{\sigma^*}\gamma$$

Analogically we can derive a short-run estimation form from the short-run production function (1):

$$\ln\left(w_t\right) = \eta_0^* + \eta_1 \ln\left(\frac{Y_t}{N_t}\right) + \eta_0 t + v_t \tag{5}$$

where coefficients are:

$$\eta_0^* = \log(1-\alpha)$$
$$\eta_1 = \frac{1}{\sigma}$$
$$\eta_0 = \frac{\sigma-1}{\sigma}\gamma_N$$

The stochastic term v captures deviations of the short-run non-neutral technology from the long-run Harrod-neutral technology growth $A - A^*$. Combing short-run and long-run specifications we will gain co-integration relation in the form:

$$\Delta\left[\ln\left(w_t\right)\right] = \eta_0 + \eta_1\Delta\left[\ln\left(\frac{Y_t}{N_t}\right)\right] + \lambda\left[\ln w_{t-1} - \eta_0^* - \eta_1^*\ln\left(\frac{Y_{t-1}}{N_{t-1}}\right) + \eta_2^* t\right] + u_t \tag{6}$$

We do not use procedure of Engel and Granger [10] to test co-integration because of the impossibility to test long-run coefficients directly. Square bracket expression multiplying co-integration coefficient λ is a long-run equilibrium residual which we do not want to estimate independently. The expression is balanced if every its term is stationary including the long-run equilibrium deviation. If this is not true, long-run equilibrium deviation do not participate in explaining of the differences of dependent variable, it means that $\lambda = 0$. We prefer dynamic ordinary least square methods suggested by Stock and Watson [26] to Johansen [14] procedure, because we consider only one co-integrating relation. According to the procedure of Stock and Watson we can write specification (6) in the form:

$$\Delta\left[\ln\left(w_t\right)\right] = \beta_0 + \beta_1\Delta\left[\ln\left(\frac{Y_t}{N_t}\right)\right] + \beta_2\ln w_{t-1} + \beta_3\ln\left(\frac{Y_{t-1}}{N_{t-1}}\right) + \beta_4 t + u_t \tag{7}$$

By the estimation of (7) we focus on test of co-integration – significance test of parameter $\beta_2 = \lambda$ with critical values of Banerjee, Dolado and Mestre [3]. Their critical values vary depending on whether the deterministic trend is included into test equation or no. Long-run production function is Cobb-Douglas, if $\sigma^* = 1$, i.e. if $-\beta_2/\beta_3 = 1$ and $-\beta_4/\beta_2 = 0$.

3 Data

There are various specifications suitable for the production function estimate. In order to focus on long run production function parameters we need to use as long as possible dataset. American data series of labour income and output with the range since 1929 till 2012, i.e. 84 observations are satisfactorily long. This is one of reasons, why we focus on labour income specification. We gathered annual data from the American national acoount system (NIPA)[5]. To derive dataset we followed Gollin [12] and Klump et al [19].

Gollin [12] refers an inconsistency between a theory and observed values of labour share. This inconsistency comes from incorrect calculation of labour share. Compensation to employees is not suitable indicator for labour income because they exclude proprietors (self-employed) labour income. It is it is unclear how the income of self-employed workers should be categorized in the labour-capital dichotomy.

We consider two approaches. Following Krueger [21] and Antràs [1] we add two thirds of self-employed workers' income to the compensations of employees. We denote this approach by the symbol (*a*). Blanchard's [6], Gollin's [12] and Bentolila's and Saint-Paul's [5] approach (*b*) is to use compensation per employee as a shadow price of labor of self-employed workers, i.e. labour income in extensive form *Nw* is:

$$\text{labour income} = \left(1 + \frac{\text{self employed}}{\text{total employment}}\right)\cdot \text{compensation to employees} \tag{8}$$

Gollin [12] introduced two more ways to modify data for correct labour share calculation, but as he stated these two ways are not suitable for the U.S. economy.

We consider GDP for output. We can use employment or number of hours worked as labour indicator. For a long-run analysis we consider employment to be satisfactory measure of the labour.

We used augmented Dicky-Fuller test, Phillips-Perron test and correlogram to test unit root of the variables (see Lukáčik and Lukáčiková [24]). We preferred Phillips-Peron test, if many autoregressive terms solved autocorrelation is the test specification. We state that all variables are integrated of order 1.

We could not reject unit root hypothesis of non-differentiated logs of labour incomes computed by both approaches (*a*) and (*b*). However, using correlograms we state non-stationarity of the variables. Autocorrelation values are slowly decreasing with time. We consider co-integration analysis as a suitable statistical tool for labour income-output relationship analysis.

4 Estimates

We estimated coefficients of (7) specification using both data combinations (*a*) and (*b*). We tested for autocorrelation using LM test and correlogram (see Ivaničová et al [13]). In the (*b*) case we needed to include one period

[5] http://www.bea.doc.gov/bea/dn/nipaweb/index.asp

lagged first difference of the log of labour income to solve autocorrelation. The estimation stability was verified by recursive methods.

Co-integration test was realized by testing of significance of λ coefficient using critical values suggested by Banerjee et al [3]. The coefficient is statistically significant at 10 % level in (*a*) and at 1 % level in (*b*). Estimates of long-run elasticity of substitution are close to one with values of 0.997 (*a*) and 0.976 (*b*). The long-run trend estimates are close to 0 with values of -0.001 (*a*) and -0.003 (*b*).

5 Conclusion

We found co-integration relation between U.S. labour income and output in intensive forms. Using two data combinations, one relation is statistically significant at 10 percent level and other at 1 percent level. In both combinations the long-run elasticity of substitution is close to 1 suggesting Cobb-Douglas long-run production function in U.S. economy.

The results correspond with the results of Szomolányi et al [28] who used same estimation form and Visegrad Group of 4 countries (V4) data. Using quarterly data in period from the 1st quarter 1995 to the 4th quarter 2012 the estimates of the elasticity of substitution were close to 1 in V4 economies. The result also corresponds with theoretical framework of Jones [15], [16] and Jürgen [17], [18].

The results could be a contribution to the world discussion focusing on production function parameters estimation. Non-stationarity of data series should be considered in estimation and co-integration analysis should be used. If we consider non-neutrality of demand side, hopeful production function parameters estimation method is implementing the labour income co-integration estimation form into the vector-autoregressive model according to Garratt et al [11], Assenmacher-Wesche and Pesaran [2] and Lukáčik et al [25]. This approach provides an estimation of both short-run and long-run coefficients in one step. The results may also involve improvements in methods for output gap detection based on production function (see more in Surmanová and Furková [27]).

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0444/15 "Econometric Analysis of Production Possibilities of the Economy and the Labour Market in Slovakia".

References

[1] Antràs, P.: Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution. *The B.E. Journal of Macroeconomics* **4** (2004), April, 1–36.
[2] Assenmacher-Wesche, K., and Pesaran, M.H.: *A VECX Model of the Swiss Economy*. CESifo Working Paper 2281, 2009.
[3] Banerjee, A., Dolado, J.J., and Mestre, R.: Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework. *Journal of Time Series Analysis* **19** (1998), 267–283.
[4] Barro, R.J., and Sala-I-Martin, X.: *Economic Growth.* 2nd ed. The MIT Press, 2003.
[5] Bentolila, S., and Saint-Paul, G.: Explaining Movements in the Labor Share. *The B.E. Journal of Macroeconomics* **3** (2003), October, 1–36.
[6] Blanchard, O.J.: The Medium Run. *Brookings Papers on Economic Activity* **28** (1997), issue 2, 89–158.
[7] Chirinko, R.S.: σ: The Long and Short of It. *Journal of Macroeconomics* **30** (2008), June, 671–686.
[8] Chirinko, R.S., Fazzari, S.M., and Meyer, A.P.: How Responsive is Business Capital Formation to its User Cost? *Journal of Public Economics* **74** (1999), October, 53–80.
[9] De La Grandville, O.: In Quest of the Slutzky Diamond. *American Economic Review* **79** (1989), 468–481.
[10] Engle, R.F., and Granger, C.W.J.: Co-integration and error correction representation, estimation and testing. *Econometrica* **55** (1987), March, 251–276.
[11] Garratt, A., Lee, K., Pesaran, M.H., and Shin, Y.: *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*. Oxford University Press, New York, 2006.
[12] Gollin, D.: Getting Income Shares Right. *Journal of Political Economy* **110** (2002), issue 2, 458–474.
[13] Ivaničová, Z., Chocholatá, M. and Surmanová, K.: *Ekonometrické modelovanie*. Ekonóm, Bratislava, 2012.
[14] Johansen, S.: Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration. *Journal of Econometrics* **69** (1995), September, 111–132.
[15] Jones, C.I.: Growth, Capital Shares, and a New Perspective on Production Functions. *Proceedings* (2003), Federal Reserve Bank of San Francisco, issue November.
[16] Jones, C.I.: The Shape of Production Functions and the Direction of Technical Change. *The Quarterly Journal of Economics* **120** (2005), May, 517–549.
[17] Jürgen, A.: A Dual Elasticity of Substitution Production Function with an Application to Cross Country Inequality. *Economics Letters* **102** (2009), January, 10–12.

[18] Jürgen A.: On the Dynamic Implications of the Cobb-Douglas Production Function. *The Selected Works of Jürgen Antony* (2010), available from: http://works.bepress.com/antony/7
[19] Klump, R., McAdam, P., and Willman, A.: Factor Substitution and Factor Augmenting Technical Progress in the US. *Review of Economics and Statistics* **89** (2007), 183–192.
[20] Klump, R., McAdam, P., and Willman, A.: The Normalized CES Production Function: Theory and Empirics. *Journal of Economic Surveys* **26** (2012), 769–799.
[21] Krueger, A.B.: Measuring Labor's Share. American Economic Review **89** (1999), May, 45–51.
[22] León-Ledesma, M.A., McAdam, P., and Willman, A.: *In Dubio pro CES: Supply Estimation with Mis-Specified Technical Change*. European Central Bank, Working Paper No. 1175, 2010.
[23] León-Ledesma, M.A., McAdam, P., and Willman, A.: Identifying the Elasticity of Substitution with Biased Technical Change. *American Economic Review* **100** (2010), 1330–1357.
[24] Lukáčik, M., and Lukáčiková, A.: Význam testovania stacionarity v ekonometrii. *Ekonomika a informatika* **6** (2008), issue 1, 146–157.
[25] Lukáčik, M., Szomolányi, K., Lukáčiková, A., and Hančlová, J.: Long-run Structural Macroeconometric Models of the Slovak and Czech Economies. In: *Proceedings of the Quantitative Methods in Economics: Multiple Criteria Decision Making XV*. IURA Edition, Bratislava, 2010, 107–116.
[26] Stock, J.H., and Watson, M.W.: A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* **61** (1993), 783–820.
[27] Surmanová, K., and Furková, A.: Methods for Output Gap Detection. In: *Proceedings of Mathematical Methods in Economics 2011 (part II)*. Janská Dolina, Slovakia, 2011, 665–670.
[28] Szomolányi, K., Lukáčik,M., and Lukáčiková, A.: Estimation of the Production Function Parameters in V4 Economies. In: *Proceedings of Mathematical methods in economics 2013*. College of Polytechnics, Jihlava, 2013, 898–902.

Moving Average influence on the prediction of the EUR/USD exchange rate using the Radial Basis Function Network

Pavel Šedek[1]

Abstract. The objective of this paper is to analyze the impact of a time series data processing on a prediction accuracy using artificial neural network. According with authors' previous work, the Radial Basis Function network model was utilized for prediction of the EUR/USD hourly spot exchange rate. As this time series is one of the most volatile financial instruments, conventional prediction methods obviously fails and so the prediction using the radial basis function network without appropriate data processing.

In this paper, the hourly close rates were smoothened using a moving average and its exponential alternative. Resulting data were split in two parts, first one for training the neural network and the consecutive one for testing. Training of the neural network was performed using a set of vectors generated with a time window. Prediction quality of the results computed by the artificial neural network was measured by the mean square error function.

Keywords: EURUSD, foreing exchange, time series prediction, radial basis function network, aritificial neural network, moving average

1 Introduction

The subject of this article is to further analyze the effect of smoothening the EUR/USD exchange spot rate via simple moving average and it's exponential alternative before forecasting it using radial basis funcion network.

Smoothing the data using the moving averages is one of the data preprocessing/preparation technique connected with the data prediction. Using such a method could have strong influence on the result computed using other methods such as artificial neural networks.

In the authors latest work [8] was done some optimization of the radial basis network parameters. More detailed optimization is described in this paper.

2 Review of related work

The detailed literature review about RBF networks could be studied in [8]. The conclusion of that review was that the radial basis function network as a prediction method worths further research. Detailed review is not in the scope of this article as was done in previous work.

2.1 Authors previous work

This paper examines in further detail the use of radial basis funcion network on the EUR/USD spot exchange rate, following the [8]. The radial basis function network have been already examined for timeseries prediction in papers [2] and [3]. The prediction of EUR/USD exchange rate have also been focused in [7].

[1]Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Economics, Management and Humanities, Technicka 2, 166 27 Prague, Czech Republic, sedekpav@fel.cvut.cz

3 Brief model description

This section was previously published in [8].

The radial basis network model (Fig. 1) is an artificial neural network model with the multilayer feed-forward network topology. As stated in [6, p. 2], despite of the identical model, the node transfer function and the training algorithms differs from multi-layer perceptron or other common types of feed-forward neural networks.

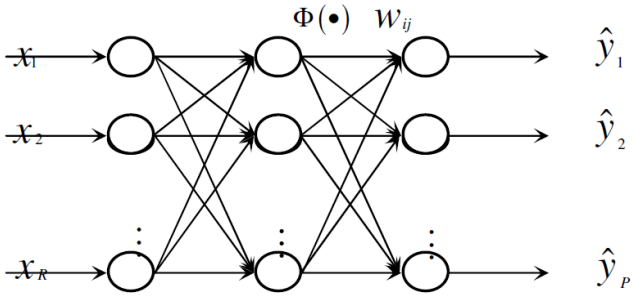


Figure 1 Structure of RBF neural network [1, p. 170]

Instead of the step or sigmoidal functions obviously used in multilayer perceptron or feed-forward network, the radial basis function also called Gaussian function is utilized as the transfer function. This function is defined by the following formula:

$$f(x) = e^{-x^2}$$

In the netowrk context, output value on each network node is calculated by the following expression [1, p. 171]:

$$\hat{y}_j = \sum_{i=1}^{M} w_{ij}\Phi_i(x) = \sum_{i=1}^{M} w_{ij} e^{-\frac{|x-c_i|^2}{2\sigma_i^2}}$$

As stated in previous work [2], literature shows many different approaches of training radial basis networks as Orthogonal forward regression [6, p. 10], Hierarchically Self-Organizin Learning [6, p. 11], Resource Allocating Network [6, p. 14] or Heuristic Called Forward Selection [9, p. 115].

For the purpose of this paper, the calculations are computed in the MathWorks Matlab software. The implementation of the radial basis network is extensively described in the MathWorks Matlab documentation [5]. Brief schema of the network topology is shown on figure 2.

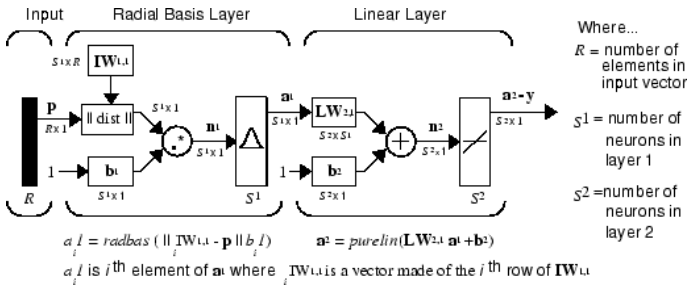


Figure 2 Radial basis network (Matlab) [5]

As previously stated in [2], the training algorithm is described in the software documentation[4] as follows:

Initially the radbas layer has no neurons. The following steps are repeated until the network's mean squared error falls below goal.

- *The network is simulated.*
- *The input vector with the greatest error is found.*
- *A radbas neuron is added with weights equal to that vector.*
- *The purelin layer weights are redesigned to minimize error.*

The network performance is measured using an error function, which is calculated as for other types of feed-forward neural network using the mean square error. The error function is defined by the following expression [1, p. 171]:

$$E = \sum_{j=1}^{P} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{P} (y_j - \sum_{i=1}^{M} w_{ij} \Phi_i(x))^2$$

4 Implementation and simulation

The implementation and simulation was realized using MathWorks Matlab software.

4.1 Dataset

The same dataset source as in [8] was used (EUR/USD H1 foreign exchange rate close price downloaded from Dukascopy [1]). Further processing was done using moving average and it's exponential alternative, both with 24 hour period. The difference among the moving averages could be seen on figure 3.

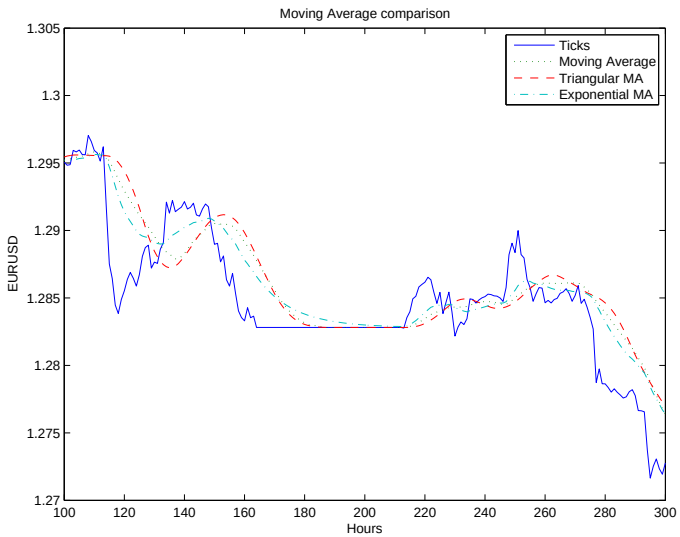


Figure 3 Moving average effect on the source data

The equation for the simple moving average state as follows.

$$SMA_i = \frac{\sum_{j=i}^{i+d-1} ((d+i-j) * X_j)}{\sum_{j=i}^{i+d-1} (d+i-j)}$$

And for the exponential moving average as follows.

$$EMA_i = \frac{\sum_{j=i}^{i+d-1} (S^j * X_j)}{\sum_{j=i}^{i+d-1} S^j}$$

[1] https://www.dukascopy.com/plugins/fxMarketWatch/?historical_data

For the purpose of network training training, the 24 hour samples were transformed to the matrix as a prediction source and a 1 hour sample as a prediction target. From each matrix first 1000 vectors are used for training and 3333 vectors for validation and testing.

4.2 Training

The training of the neural network was performed in loops to analyze the influence of maximal neurons constraint and the spread parameter of the radial basis function. This optimization was performed with integer steps from values 1 to 100 for neurons and 1 to 50 for spread. This process was performed for both, the plain moving average and it's exponetnial alternative.

On the following figures (and) is visualized the mean square error of both optimizations. Note that just the areas where the optimization performs best is shown on the figures.

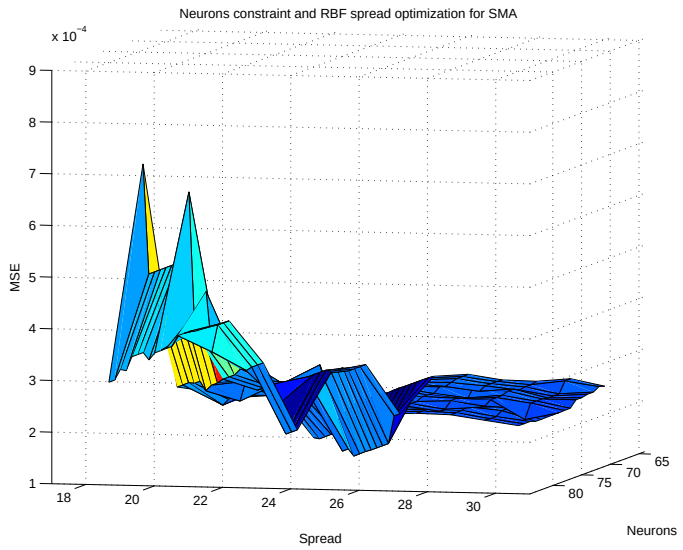


Figure 4 Neurons constraint and RBF spread optimization for SMA

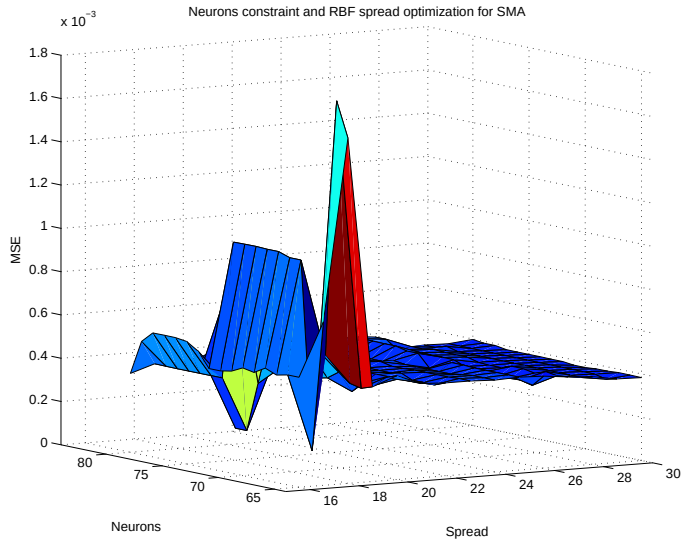


Figure 5 Neurons constraint and RBF spread optimization for EMA

On both figures, it could be seen, that some values of spread have strong influence on the mean square error. The resulting values from optimization are listed in the table 1

averaging method	MSE
simple	0.000118054
exponential	0.000147514

Table 1 Optimization results

Detailed optimization of the raw data prediction was not performed since the results of samples were much worse then after smoothing with the moving average.

5 Conclusion

In contrast with previous expectations, the smoothing using the simple moving average gives better prediction results then the exponential one. However, using radial basis function networks for prediction of pure data without smoothing it moves the accuracy to much worse values.

Further testing of this prediction on the other financial instruments should be performed. Also, the real-time testing should be done to verify this optimization.

References

[1] Guo, C., Hou, Z.-g., and Zeng, Z.: "RBF Neural Network Based on Particle Swarm Optimization". In: *LNCS 6063 - Advances In Neural Networks - ISNN2010* 7.June (2010), pp. 169–176.

[2] Kříž, R. and Šedek, P.: "RECURRENCE PLOTS OF ELECTRICITY PRICE SERIES AND THE PREDICTION OF ELECTRICITY PRICES USING ARTIFICIAL NEURAL NETWORKS". In: *POSTER 2014 - 18th International Student Conference on Electrical Engineering*. Vol. 18. 2014. ISBN: 978-80-01-05499-4.

[3] Kříž, R. and Šedek, P.: "The Prediction of Tropospheric Ozone Using a Radial Basis Function Network". English. In: *ISCS 2014: Interdisciplinary Symposium on Complex Systems SE - 13*. Ed. by Sanayei, A., E. Rössler, O., and Zelinka, I. Vol. 14. Emergence, Complexity and Computation. Springer International Publishing, 2015, pp. 115–123. ISBN: 978-3-319-10758-5. DOI: 10.1007/978-3-319-10759-2_13.

[4] MathWorks: *MATLAB Functions in Neural Network Toolbox*.

[5] MathWorks: *Radial Basis Neural Networks - MATLAB & Simulink*.

[6] Sundararajan, N., Saratchandran, P., and Lu, Y. W.: *Radial Basis Function Neural Networks with Sequential Learning: MRAN and Its Applications*. Progress in neural processing. World Scientific, 1999. ISBN: 9789810237714.

[7] Šedek, P.: "Automatické obchodování pomocí umělých neuronových sítí". Master Thesis. Czech Technical University in Prague, 2013.

[8] Šedek, P.: "EUR / USD H1 Spot Exchange Rate Prediction via Radial Basis Function Network". In: *Proceedings of the 19th International Scientific Student Conferenece POSTER 2015*. Vol. 19. 2015. ISBN: 978-80-01-05499-4.

[9] Tykierko, M.: "Automatic change detection in dynamical system with chaos based on model, fractal dimension and recurrence plot". In: *Computer Aided Systems Theory - EUROCAST 2007*. 2007, pp. 113–120. ISBN: 978-3-540-75867-9. DOI: 10.1007/978-3-540-75867-9_15.

Term structure of interest rates: comparison of the Czech Republic and Germany

Blanka Šedivá [1], Patrice Marek[2]

Abstract. The aim of the paper is to demonstrate the possibility of using parametric and nonparametric methods for estimating the term structure of interest rates. First, the principles of several yield curves construction are described and later, advantages and disadvantages of parametric models (Nelson-Siegel and Svensson class model) and nonparametric models are presented. The nonparametric methods based on a kernel estimation are emphasized and closely described, demonstrated and analysed. The both methods (parametric and nonparametric) are used to estimate the term structure of interest rates based on the market price of the Czech government bonds and German government bonds.

Keywords: Yield curves construction, Nelson-Siegel model, Svensson model, Nonparametric models, Czech government bonds, German government bonds

JEL classification: C51, G12
AMS classification: 62J02, 91B55

1 Introduction

The relationship between risk-free rate zero-coupon asset and its maturity is denoted as a time structure of interest rates and a graphical interpretation of such relationship is called a yield curve. Many authors are interested in the time structure of interest rates mainly from economical point of view. They are focused on questions: What information is hidden in the yield curve and what can we find out from the shape of the yield curve in the present or future economical state. There are many hypotheses developed during years which explain the shape of a yield curve, detailed overview can be found in [4]. The idea about the time structure of interest rates provides very important information especially to investors, who use information about interest rates for the risk management, pricing the underlying assets and derivates. It is also important to central banks who are interested in historical yield curves and these curves are the source of information for setting up the correct monetary politics. Many studies dedicated to estimation of the interest rate time structure come from the environment of central banks and other central financial institutions. For example, European central bank *https://www.ecb.europa.eu/stats/money/yc* publishes actual spot, forward and par yield curves estimated by Svensson model [15]. This model is used for the estimation of American yield curves published in [1]. Historical information about the interest rates time structure for the US is also available as par yield curves denoted by the method of quasi-cubic splines which can be found on http://*www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/Historic-Yield-Data-Visualization.aspx*.

In the last couple of years the estimations of interest rates time structure is one of the classical topics solved in economical, econometric and financial studies. There are two possible approaches for these estimations from the methodical point of view – nonparametrical and parametrical methods. Nonparametrical models derived by using spline principles were used in nowadays classical text [6]. In this text the estimation of discount function is derived by using polynomial splines. In [14] the exponential splines are used for the estimation of discount function and authors show that this approach provides better yield curve estimation than the polynomial approach. Besides the type of spline functions for nonparametrical estimations also the choice of minimizing criterion is important as it also contains the element defining smoothness of the function.

The parametrical approaches are represented by classical methods as Nelson-Siegel [7] and Swenson

[1]University of West Bohemia, Plzeň, Czech Republic, sediva@kma.zcu.cz
[2]University of West Bohemia, Plzeň, Czech Republic, patrke@kma.zcu.cz

[13] and its different versions [11], [1]. Recently, dynamical models derived from mentioned models are becoming more and more popular. Since parametric models are non-linear and we use methods based on a minimization of non-linear criterion functions we usually need to deal with the problems of sensitivity of these algorithms on the initial conditions and also with other problems connected to numerical stability of estimations [5]. The modern theoretical concepts of interest rate time structures assume that yield curves and recent economical activities expressed for example by macroeconomic indicators (inflation, GDP, ..) are strongly connected. The expected future inflation and expected future real economy progression might be important factors that affect the shape of the yield curve. On the other hand the yield curve can be used as predictor. The strong relationship between yield curves and macroeconomic indicators is a topic of many theoretical and empirical studies. One of the main approaches used for mutual relationship modelling uses dynamical linear and non-linear models.

The foregoing empirical studies show that it is impossible to classify which approach to modelling is better than the others. Every method used has its advantages and disadvantages and the model choice mainly depends on the author's subjective opinion. Particular application of methods which are used to estimate the interest rate time structure in the Czech environment based on actual prices of Czech assets was possible to use after sufficient development of the market with these assets. Similarly, as in other countries, numerous modelling approaches are used. The most often estimation is probably parametrical approach based on Nelson-Siegel model or Svensson model [9]. The complex estimate of daily coefficients for Czech yield curves modelled by Nelson-Siegel approach is available in [2]. In [4] three parametrical functions are chosen for time structure models. These functions combine polynomial and exponential functions. In [8] the estimation of interest rate time structure is based on interest swaps and for its estimation the bootstrap method is used. Modern approaches to interest rate time structures modelling which use dynamical Nelson-Siegel model and that include latent regional factors are mentioned in [12]. In this article, the regional dynamics of yield curves for the Czech Republic, Hungary, Poland and the Slovak Republic is investigated.

2 Yield curve relationship in continous time

The term structure of interest rates is a central concept in monetary and financial economies. The estimation of a yield curves is based on an assumed functional relationship between either par yields, spot rates y, forward rates f or discount factor d on the one hand and maturities on the other. The key term in construction of yield curves is discount factor that is defined as the quantities used at a given point in time to obtain the present value of future cash flow. A discount function $d(t_1, t_2)$ is the collection of discount factor at the time t_1 for all maturities t_2.

The forward rates f are defined as the instantaneous rates that are derived from the concept of the discount factor.

$$f(t) = \lim_{h \to 0} \left(\frac{d(t)/d(t+h) - 1}{h} \right) = \frac{-d(t)'}{d(t)}. \tag{1}$$

On the other hand the discount rate can be computed as $d(t) = \exp\left(-\int_0^t f(\tau)\, \mathrm{d}\tau \right)$. The yield of the spot rates is formulated as mean forward rates in interval $(0, t)$ and the spot yield is computed as

$$y(t) = \frac{1}{t} \int_0^t f(\tau)\, \mathrm{d}\tau = \frac{1}{t} \int_0^t \frac{-d(\tau)'}{d(\tau)}\, \mathrm{d}\tau = \frac{-\ln(d(t))}{t}. \tag{2}$$

These relations can be also inverted to express forward and spot rate $f(t) = t\, y(t)' + y(t)$ and $d(t) = \exp(-t\, y(t))$, with initial condition $f(0) = y(0) = 0$. From the above relationships, it is clear that relations between the discount rate of d, the forward rate of f and the spot rate of return y are unambiguous and knowledge in the discount rate d is equivalent knowledge of f or y.

The construction of these curves poses several problems for applied research. Many governments do not issue longer term (i.e. greater then 1–2 years) zero-coupon bonds. Hence the yield curve must be inferred from other instruments.

One possibility is the estimate of the time structure from bond prices, we assume that there is absence of arbitrary and financial flows associated with the bond issue are known: coupons c in time $t_1, t_2, \ldots, t_J$

and payment of the nominal value of R in time t_J. Arbitrage in the bond markets will cause the price P of the bond is equal to the present value of expected cash flows. The second problem is that, in practice, small pricing errors perhaps trading, taxation, illiquidity spreads necessitate adding an error term to present value of expected cash flows. The statistical problem is how estimate the function d from a sample of coupon paying bonds.

3 Yield curve models and estimation methods

The aim of the estimates in the term structure of interest rates is to find a model curves that give us the most reliable and useful estimates, not just on a specific day, but also in a longer time frame. It is necessary to take into consideration several requests and their corresponding criteria, when estimating the yield curve. The basic requirements for "good" estimates of yield curves include

- the requirement to the smoothness: the estimated yield curves should be relatively smooth, because the yield can be also affected by other random factors;
- the requirement for flexibility of curves: the estimation technique should provide sufficient flexibility at shorter maturities;
- the requirement for stability: the estimates of the yield curve should remain stable in the sense that small changes in data in a single maturity have a disproportionate impact on forward rates at other maturities.

3.1 Methods of estimations

The constructions of yield curves are usually based on the market prices of assets that have the same risk and liquidity. The selection of appropriate assets for the estimation the time structure is therefore usually restricted to the actual market price of government bonds.

Given the discount function we can price any coupon bond by summing the price of its individual cash flows. Let suppose we have an information on K bonds, every bond is represented by a vector $B_k = [\mathbf{b}_k, \mathbf{t}_k, p_k], k = 1, 2, 3, ..., K$, where vector $\mathbf{b}_k = (b_{1k}, b_{2k}, \ldots, b_{nk})$ denotes the payments returned to the owner of the bond k at times $\mathbf{t}_k = (t_{1k}, t_{2k}, \ldots, t_{nk})$ and p_k is the market price of the bond. Knowledge of discount rate can be used to express price of bonds as a sum of discounted cash flows $p_k^M = \sum_{j=1}^{n_k} b_{jk}\left(t_{jk}\right) d\left(t_{jk}\right)$.

The yield rates curve can be estimated by minimising the sum of differences between the model prices and observed prices of a set of bonds: $\min \sum_k \left(p_k - p_k^M\right)^2$, where the superscript M stands for "model" price, or the model can be estimated by minimising the sum of differences between the model rates and observed rates of a set of bonds: $\min \sum_k \left(y_k - y_k^M\right)^2$.

3.2 Parametric models

Various parametric models can be used to determine the yield curve. The method developed by Nelson and Siegel [7] assumes explicitly the following function form for the spot rates

$$y(t) = \beta_1 + \beta_2 \left(\frac{1 - \exp\left(-\frac{t}{\lambda}\right)}{\frac{t}{\lambda}} \right) + \beta_3 \left(\frac{1 - \exp\left(-\frac{t}{\lambda}\right)}{\frac{t}{\lambda}} - \exp\left(-\frac{t}{\lambda}\right) \right). \tag{3}$$

The popularity and practical usability of the model is mainly based on the fact that this model is able to describe the time structure with a small number of parameters, while the three components, which are included into the model, allow to describe various theoretical and empirical observed shapes of the yield curve, starting from a monotonic curves through S-shaped curve to curve with one local maximum.

To improve the flexibility of the curves and fit, Svensson [13] extended Nelson and Siegel's function by adding a further term that allows for a second "jump". The extra precision in achieved at the cost of adding two parameters β_4 and λ_2 which have to be estimated. The spot forward rates curve thus becomes

$$y(t) = \beta_1 + \beta_2 \left(\frac{1 - \exp\left(-\frac{t}{\lambda_1}\right)}{\frac{t}{\lambda_1}} \right) + \beta_3 \left(\frac{1 - \exp\left(-\frac{t}{\lambda_1}\right)}{\frac{t}{\lambda_1}} - \exp\left(-\frac{t}{\lambda_1}\right) \right) + \\ + \beta_4 \left(\frac{1 - \exp\left(-\frac{t}{\lambda_2}\right)}{\frac{t}{\lambda_2}} - \exp\left(-\frac{t}{\lambda_2}\right) \right). \quad (4)$$

But the model calibration and validation, i.e. obtaining parameters values such that the model yields accord with the market yields is difficult, many authors have reported "numerical difficulties". There are two reasons for these difficulties. First, the objective function is non convex and exhibits several local minima, so the optimization algorithm is also sensitive to the initial parameters value. The second problem is collinearity and this stems from the model specification. Both models can be interpreted as a factor models but factors in model are not independent. For example the Nelson-Siegel's model includes three factors, which can be interpreted as level (β_1), steepness (β_2) and curvature (β_3).

3.3 Nonparametric models

Rather than specifying a single functional form over the entire maturity range, nonparametric methods can be used. Over a closed interval, a given continuous function can be approximated by selection an arbitrary function, however quite frequently it displays insufficient smoothing properties.

In this article we focused on nonparametric estimations based on kernel regression smoothing. The inspiration for this article was the work [3], which included also asymptotic properties of kernel estimates of the yield curve and this paper follows on our previous published paper [10].

Assume the following regression model $Y_i = m(X_i) + \varepsilon_i$, $\varepsilon_i \sim iid, E(\varepsilon_i) = 0, i = 1, 2, \ldots, n,$. The Nadaraya–Watson kernel regression algorithm for estimation of the conditional expectation of a random variable $E(Y|X = x) = m(x)$ is based on local weighting mean

$$\widehat{m}(x) = \sum_i w_i(x)\, Y_i, \qquad \text{with} \quad w_i(x) = \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_i K\left(\frac{x - X_i}{h}\right)} I_{\left[\sum_i K\left(\frac{x - X_i}{h}\right) \neq 0\right]}, \quad (5)$$

where $K(\cdot)$ is a kernel as a weighting function and h is a bandwidth. Basic calculus shows that $\widehat{m}_n(x)$ is the solution to a weighted least square problem, being the minimizer $\sum_i (Y_i - \beta_0)^2 K\left(\frac{x - X_i}{h}\right)$.

It is known, that the significant effect to the quality of kernel estimation have the choices of kernel function and bandwidth. In our study, we used three types of kernel function: Gaussian $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$, exponential $K(u) = \frac{1}{2} \exp(-|u|)$ and quartic (biweight) $K(u) = \frac{15}{16}\left(1 - u^2\right)^2 I_{[|u| \leq 1]}$ form of kernel function and the optimal window width choice was founded by cross validation algorithm. These three types were selected because of their qualities and behaviour in the simulation studies.

The estimation of term structure with using nonparametric regression was based on local weighted minimization differences between model price and market price which is calculated as sum of discounted cash flows

$$\min_d \sum_k w_k(t) \left(p_k - \sum_{j=1}^{J_k} b_k(t_{jk}) d(t_{jk}) \right)^2. \quad (6)$$

4 Data and empirical findings

The parametric and nonparametric approaches, described in Section 3, were applied to real data from the Czech Republic and Germany. The final data set for estimating the German term structure of interest rates comprises standard bonds issued by the Federal Republic of Germany, five-year special federal bonds and Federal Treasury notes. The final data for estimating the Czech term structure contains government bonds. For short end of yield curve we used PRIBOR and PRIBID, resp. EURIBOR. The data were obtained on May 10, 2015 and these data can be obtained at *http://www.boerse-frankfurt.de/en/bonds*

and *http://www.patria.cz/kurzy/online/dluhopisy.html*. We adopt the standard approach and consider all bonds with a remaining time to maturity above six month. The yield of bonds with residual maturity below six month are excluded because they appear to be more liquid. In running the programming we choose several initial values for each model and then we used the best fitting model. That approach can eliminat problems with several local minimum of criterion function and dependence on choosing initial value. The quality of models were compared by mean square error $MSE_P = 1/K \sum_k \left(p_k - p_K^M\right)^2$ and $MSE_y = 1/K \sum_k \left(ytm_k - ytm_K^M\right)^2$, where ytm_k is the calculated yield to maturity of bonds.

The estimated parameters of the Nelson-Siegel's model (NS model) formulated in Eq. (3), Svensson model (S model) formulated in Eq. (4) and nonparametric model formulated in Eq. (6) are summarized in Table 1

Parametric models								
	β_1	β_2	β_3	β_4	λ_1	λ_2	MSE_p	MSE_y
NS model CZ	0.0167	-0.0155	0.0687		39.9977		1.4764	0.0457
S model CZ	0.0059	-0.0044	-0.0044	20.9269	11.3985	12896.2185	1.3589	0.0461
NS model GE	0.0117	-0.0120	0.0568		39.9999		2.6407	0.0259
S model GE	0.0000	-0.0001	0.0043	15.0001	10.3986	12721.9560	3.4148	0.0266
Nonparametric models								
			h				MSE_p	MSE_y
Gaussian kernel CZ			0.1594				4.9881	0.1014
Exponential kernel CZ			0.1000				5.1373	0.1033
Quartic kernel CZ			0.9187				5.1686	0.0847
Gaussian kernel GE			0.0913				5.9345	0.0520
Exponential kernel GE			0.2255				5.9334	0.0515
Quartic kernel GE			0.1841				5.9106	0.0530

Table 1 Estimated parameters of the Nelson-Siegel model, Svensson model and nonparametric model for the Czech Republic (CZ) and Germany (GE)

Our results demonstrate the problem with numerical solutions of optimization problem with several parameters. Although the Svensson model is extension form of Nelson Siegel model, the obtained estimates of parameters are completely different. Also the interpretation of value can be problematic especially for parameters lambda in both models. There is the same problem with estimation of bandwidth for nonparametric estimation but the values of criterion MSE are close. An analysis of the choice of kernel function is one of the possible directions for our future work.

The estimated yield curves based on the Nelson-Siegel model, Svensson model and nonparametric kernel approach for the Czech Republic and Germany are in Figure 1.

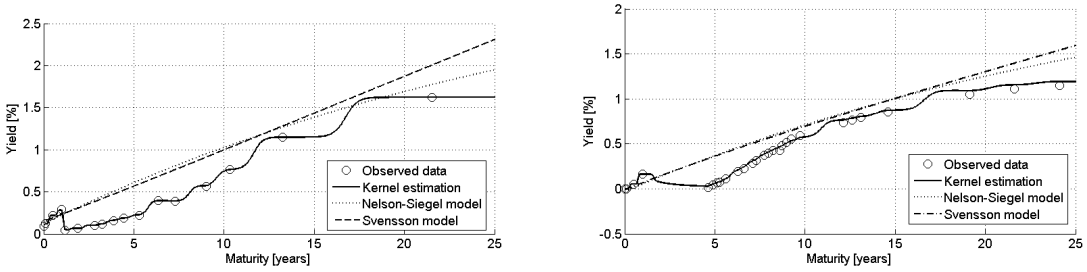


Figure 1 Estimated curves for the Czech Republic (left) and Germany (right)

The observed and modelled prices for bonds are shown in Figure 2. It shows more precise estimates for short end but there are not conspicuous differences between parametric and nonparametric approach.

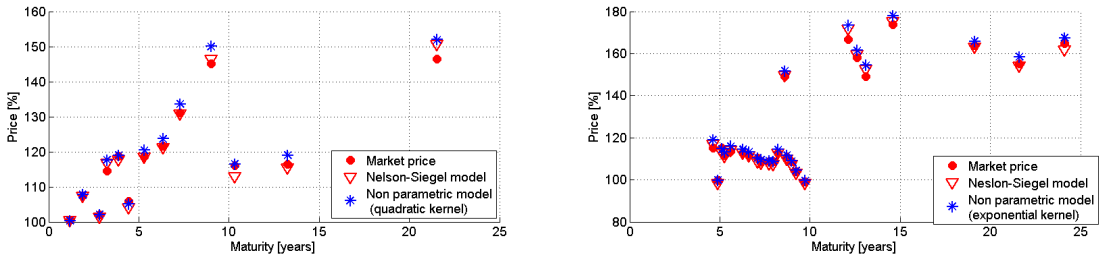


Figure 2 Market and estimated prices of bonds for the Czech Republic (left) and Germany (right)

5 Conclusions

The Nelson–Siegel and Svensson model are parametric models that have four or six parameters. Within these models, it is very difficult to estimate parameters. A different approach for estimating the time structure of interest rates is based on nonparametric methods. Our studies shown that the results obtained from parametric and nonparametric methods are comparable and both can be used for estimation of the market bond prices. The advantage of nonparametric methods are quickness of algorithm and the fact, that the results are more robust and do not depend on initial values. The disadvantage can be seen in its huge flexibility and sensitivity to data set, especially in the case there are only small number of market bond prices available. The choice of optimal kernel function is also the open problem.

References

[1] Gurkaynak, R. S., Sack, B. P., and Wright, J. H.: *The U.S. Treasury Yield Curve: 1961 to the Present.* Working paper no. 2006-28, FEDS Working Paper No. 2006-28, 2006.

[2] Kladívko, K.: The Czech Treasury Yield Curve from 1999 to the Present. *Czech Journal of Economics and Finance (Finance a úvěr) 60*, **4** (2010), 307–335.

[3] Linton, O.B. and Mammen, E. and Nielsen, J. P. and Tanggaard, C.: Yield Curve Estimation by Kernel Smoothing Methods. *Available at SSRN: http://ssrn.com/abstract=222369.*

[4] Málek, J., Radová, J., and Štěrba, F.: Konstrukce výnosové křivky pomocí vládních dluhopisů v České republice. *Politická ekonomie* **6** (2007), 792–808.

[5] Manousopoulos, P. and Michalis Michalopoulos, M.: Comparison of non-linear optimization algorithms for yield curve estimation *European Journal of Operational Research. 192* **2** (2009), 594–602.

[6] McCulloch, J. H.: Measuring the term structure of interest rates. *Journal of Business* (1971), 19–31.

[7] Nelson, C. R., and Siegel, A. F.: Parsimonious modeling of yield curves. *Journal of Business* (1987).

[8] Plánička, P.: Stanovení bezrizikové výnosové křivky ve výnosovém oceňování podniků za použití IRS. *Finanční řízení podniků a finančních institucí, Ostrava* (2011), 349–357.

[9] Slavík, M.: Odhad časové struktury úrokových sazeb z cen domácích dluhopis. *Finance a úvěr* **11** (2001), 591–606.

[10] Šedivá, B., and Ťoupal, T.: Odhady časové struktury úrokových sazeb s využitím neparametrických jádrových metod. *Řízení a modelování finančních rizik*, (2014), 753–761.

[11] Söderlind, P., and Svensson, L.: New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics 40*, **2** (1997), 383–429.

[12] Šopov, B., and Seidler, J.: Yield curve dynamics: Regional common factor model. *Prague Economic Papers* **2** (2011), 140–156.

[13] Svensson, L. E.: *Estimating and interpreting forward interest rates: Sweden 1992-1994.* Tech. rep., National Bureau of Economic Research, 1994.

[14] Vasicek, O. A., and Fong, G. H.: Term structure modeling using exponential splines. *The Journal of Finance 37*, **2** (1982), 339–348.

[15] Webpages of European Central Bank (ECB).: Available at https://www.ecb.europa.eu, April 2013.

Pareto efficiency of consumers' allocations in oligopolistic general equilibrium with innovation

Ľubica Šimková[1], Milan Horniaček[2]

Abstract. We analyze Pareto efficiency of consumers' allocations (collections of sequences of their expected consumption and provision of labor services) in an infinite horizon general equilibrium model of an oligopolistic economy with innovation. Firms, and banks can change their single period production possibility sets by technological and/or product innovation. The latter is a result of successful research and development, outcome of which is stochastic. Consumers maximize average discounted expected utility. We analyze consumers' allocations when firms, labor unions, and banks use strategies that maximize average discounted expected real wealth of consumers. Under such strategies, it is not possible to increase sum of average discounted expected consumed quantities of each consumption good and decrease sum of average discounted expected supplies of each labor service to each firm and bank that is not yet zero. For discount factor close enough to one, under such strategies, consumers' allocation is strictly Pareto efficient.

Keywords: consumers' allocation, general equilibrium, innovation, maximization of consumers' real wealth, Pareto efficiency.

JEL Classification: D43, D59, D69
AMS Classification: 91B15, 91B51, 91B70

1 Introduction

Standard general equilibrium models follow the paradigm of perfect competition. Equilibrium consumers' allocation in them is strictly Pareto efficient (see [3]). This property of general competitive equilibrium is used as justification for competition/antitrust policy despite the fact that most industries in developed countries have oligopolistic character. Therefore, welfare analysis of general equilibrium in an oligopolistic economy is needed. The current paper is a contribution to this analysis.

We analyze a dynamic stochastic general equilibrium model, in which firms, (commercial) banks, and labor unions behave strategically. Consumers do not behave strategically. They make their decisions on consumption, labor supply, and bank operations on the basis of maximization of average discounted expected utility. Firms and banks can change their single period production possibility sets by innovation. The latter is a result of successful research and development (henceforth, R&D), outcome of which is stochastic. R&D is the only source of uncertainty in our model.

The model can be viewed as a stochastic game with firms, banks, and labor unions as players. The players maximize average discounted expected real wealth of their stakeholders (i.e., shareholders or members). This is a generalization of the objective of an imperfectly competitive firm developed by Dierker and Grodal [1] to our infinite horizon stochastic model. Coalitions of players use the same objective. A profile of strategies of members of a coalition maximizes average discounted real wealth of its stakeholders if there does not exist another profile of strategies that (with unchanged strategies of the players outside the coalition) allows its stakeholders to increase sum of average discounted expected consumed quantities of each consumption good and to decrease sum of average discounted expected supplies of each labor service to each firm and bank (that is not yet zero).

We have shown in [2] that a strong Markov perfect general equilibrium (henceforth, SMPGE) exists in our model for any discount factor. In each SMPGE the real wealth of all consumers (who are stakeholders of the grand coalition) is maximized. In the present paper, we use this property of an SMPGE to show that, for a discount factor close enough to one, the equilibrium consumers' allocation (i.e., the equilibrium collection of sequences of their expected consumption and provision of labor services) is strictly Pareto efficient with respect to the set of all feasible consumers' allocations.

[1] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 82105 Bratislava, lubica.simkova@fses.uniba.sk.
[2]Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, 82105 Bratislava, milan.horniacek@fses.uniba.sk.

2 Model and its analysis

2.1 Formulation of the model

The model has the form of an infinite horizon stochastic game with discount factor $\delta \in (0,1)$. We denote it by $\Gamma(\delta)$. It is described in our accompanying paper [2]. Here we describe only its features that are necessary to follow the arguments in the next subsection.

$J \cup B \cup L$ is the finite set of players in $\Gamma(\delta)$. J is the set of firms, B is the set of banks, and L is the set of labor unions. I is the finite set of infinitely living consumers. They are endowed by labor services. Their endowments in each period depend on the set of goods that can be produced. They maximize average discounted expected utility, using single period vonNeumann – Morgenstern utility function, which is the same for each period. We denote single period utility function of consumer $i \in I$ by u_i. Its argument is a vector containing consumption of non-durable consumption goods, available quantities of durable consumption goods of all vintages, and provision of each labor service to each firm and bank. We assume that it is generic, continuous, strictly concave, strictly increasing in consumption of each non-durable consumption good, strictly increasing in available quantity of each durable consumption good of each vintage, and strictly decreasing in provision of each type of labor service to each firm and bank.

Since product innovation is possible in our model, the set of goods in it, $G = G_C \cup G_J \cup G_B$, is countable infinite. Nevertheless, only a finite subset of them can be produced in any period. G_C (G_J, G_B) is the countable infinite set of consumption goods (producer goods, banking services). The maximum feasible output of good $g \in G$ in any period is $y_g^{\max} > 0$, the upper bound on its price is $p_g^{\max} > 0$ and the lower bound of its price is $p_g^{\min} \in (0, p_g^{\max})$. We assume that $\sup_{g \in G} y_g^{\max}$ and $\sup_{g \in G} p_g^{\max}$ are finite and $\inf_{g \in G} p_g^{\min} > 0$.

The maximal duration of any R&D project is $\overline{T} \in N$ periods.

In $\Gamma(\delta)$ we restrict attention to pure Markov strategies. We denote the set of pure Markov strategies of player $k \in J \cup B \cup L$ by $\overline{S}_k$ and set $\overline{S} = \prod_{k \in J \cup B \cup L} \overline{S}_k$. A pure Markov strategy of a player depends only on payoff relevant elements of a non-terminal history. We denote by S the subset of $\overline{S}$ that contains only pairs of contract proposals of a buyer and a seller that coincide. S^+ is the subset of S containing only profiles of continuous strategies (i.e., strategies that are continuous in payoff relevant elements of a non-terminal history). Since the outcome of a strategy profile (i.e., inputs, outputs, prices, and R&D activities) depends only on concluded contracts with positive traded quantity to which it leads and on its prescriptions after histories consistent with it, we can restrict attention to S^+. An SMPGE on S^+ is immune to deviations by any coalition to any profile of its strategies (even non-Markov one).

We denote by $\Im$ the set of states. Each $\sigma \in \Im$ contains all non-terminal histories that have the same payoff relevant elements. $\Im_1$ is the set of states at the beginning of a period. We denote the initial state by σ_0. $\Gamma_{(\sigma)}(\delta)$ is the class of subgames following histories belonging to σ. Restriction of any set or function defined for $\Gamma(\delta)$ to $\Gamma_{(\sigma)}(\delta)$ is indicated by subscript "(σ)". For each $D \in 2^{J \cup B \cup L}$ function γ_D assigns to each $s \in con(S^+)$ the vector of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks multiplied by minus one by stakeholders of players in D. It is continuous on S^+. For $D \in 2^{J \cup B \cup L}$ and $s \in \overline{S}$, $A_D(s)$ is the set of all vectors of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks (multiplied by minus one) that are feasible for stakeholders of players in D when the players follow s.

For each $s \in \overline{S}$, $\Im_1(s)$ is the subset of $\Im_1$ containing states that occur with a positive probability when the players follow s. For each $s \in \overline{S}$, each $\sigma \in \Im_1(s)$, and each $t \in N$, $\pi(s, \sigma, t)$ is the probability of occurrence of state σ in period t when the players follow s. For each $\sigma \in \Im_1(s)$, let $N(\sigma) \subset G$ be the finite set of goods that can be produced (in the case of labor services provided) at state σ.

For the purposes of the argument in the following subsection, properties of single period production possibility sets and abilities of firms to utilize them by their strategies are summarized by set $\overline{\Psi}$. It expresses relations between average discounted expected used quantities of labor services with minus sign in each firm and bank and average discounted expected outputs of consumption goods at each $\sigma \in \Im_1$. It contains only vectors of the described form that are generated by some strategy profile belonging to S^+.Thus, elements of $\overline{\Psi}$ are feasible not only from technological but also from economic point of view, i.e. from the point of view of ability of firms to finance investment and R&D leading to them. (Note that if $\psi \in \overline{\Psi}$ is generated by $s \in \overline{S} \setminus S^+$, then it is generated also by some $s' \in S^+$.) We denote by Ψ the upper boundary of $\overline{\Psi}$. We assume that Ψ is convex. This corresponds to non-decreasing returns to scale in production. Function γ_ψ assigns to each $\psi \in \Psi$ the vector of average discounted expected aggregate consumed quantities of new consumption goods and average discounted expected aggregate labor supplies to individual firms and banks multiplied by minus one by all consumers.

For each $\psi \in \Psi$, denote by $\overline{S}(\psi)$ the set of Markov strategy profiles that generate ψ.

Denote by $\wp(\psi)$ the set of price vectors

$$p = (p(\sigma))_{\sigma \in \Im_1} = \left(\left((p_{kg}(\sigma))_{g \in G_L \cap N(\sigma)} \right)_{k \in J \cup B}, (p_g(\sigma))_{g \in G_C \cap N(\sigma)} \right)_{\sigma \in \Im_1}$$
$$\in \prod_{\sigma \in \Im_1} \left(\prod_{g \in G_L \cap N(\sigma)} \left[p_g^{\min}, p_g^{\max} \right]^{\#(J \cup B)} \times \prod_{g \in G_C \cap N(\sigma)} \left[p_g^{\min}, p_g^{\max} \right] \right)$$

for which there exist $s \in \overline{S}(\psi)$ and $K(\psi, p)$ such that the subspace of

$$\bigcup_{s \in \overline{S}(\psi)} \left(\prod_{\sigma \in \Im_1(s)} \begin{pmatrix} \prod_{g \in G_L \cap N(\sigma)} \left[-\sum_{t \in N} \delta^{t-1} \pi(s, \sigma, t) y_g^{\max}, 0 \right]^{\#(J \cup B)} \\ \times \prod_{g \in G_C \cap N(\sigma)} \left[0, \sum_{t \in N} \delta^{t-1} \pi(s, \sigma, t) y_g^{\max} \right] \end{pmatrix} \right)$$

defined by $\sum_{\sigma \in \Im_1(s)} p(\sigma) z(\sigma) = K(\psi, p)$ is tangent to Ψ at ψ. We assume that $\wp_{(\sigma)}(\psi)$ is non-empty for each $\sigma \in \Im_1$, each $\kappa \in (0,1]$, and each $\psi \in \Psi_{(\sigma)}$.

Definition 1. *A strategy profile $s^* \in \overline{S}$ is an SMPGE of $\Gamma(\delta)$ if there do not exist $\sigma \in \Im$, $D \in 2^{J \cup B \cup L} \setminus \{\varnothing\}$, $s_D \in S_{D(\sigma)}$, $x_D \in A_{D(\sigma)}(s^*_{-D(\sigma)}, s_D)$,[3] and $z_D \in \prod_{g \in G_C} (0, \infty) \times \prod_{k \in J \cup B} \prod_{g \in G_L} (0, \infty)$ such that $x_{Dg} = \gamma_{Dg}(s^*_{(\sigma)}) + z_{Dg}$ for each $g \in G_C$ and $x_{Dkg} = \gamma_{Dkg}(s^*_{(\sigma)}) + z_{Dkg}$ for each $(k, g) \in (J \cup B) \times G_L$ with $\gamma_{Dkg}(s^*_{(\sigma)}) > 0$.*

In the next subsection we will use the following result proven in [2].

Theorem 1. *For each $\delta \in (0,1)$ and each $\psi^* \in \Psi$ there exists $s^* \in S^+$ that (i) is an SMPGE of $\Gamma(\delta)$, (ii) it generates a vector of prices belonging to $\wp(\psi^*)$, and (iii) $\gamma_{J \cup B \cup L}(s^*) = \gamma_\psi(\psi^*)$.*

2.2 Pareto efficiency of consumers' allocations

A consumers' allocation generated by $s \in \overline{S}$ is a collection of consumption – labor supply sequences of all consumers that occurs when the players follow s, i.e., $x(s) = \left(\left\{ (x_i(s, \sigma, t))_{\sigma \in \Im_1 : \pi(s, \sigma, t) > 0} \right\}_{t \in N} \right)_{i \in I}$, where $x_i(s, \sigma, t)$ is the vector of consumed quantities of consumption goods and supplied quantities of labor services to firms and banks by consumer i at state σ in period t when the players follow s. We identify the set of feasible consumers' allocations with the set of consumers' allocations generated by strategy profiles belonging to $\overline{S}$. This again takes into account not only technological but also economic feasibility of a consumers' allocation (firms have to be able to finance investments and R&D leading to use of labor services and production of consumption goods and they have to be able to sell produced consumption goods).

A consumers' allocation $x(s)$ is strictly Pareto efficient if and only if there does not exist $s' \in \overline{S}$ such that $x(s')$ satisfies

$$(1-\delta) \sum_{t \in N} \left(\delta^{t-1} \sum_{\sigma \in \Im_1(s')} \pi(s', \sigma, t) u_i(x_i(s', \sigma, t)) \right) \geq (1-\delta) \sum_{t \in N} \left(\delta^{t-1} \sum_{\sigma \in \Im_1(s)} \pi(s, \sigma, t) u_i(x_i(s, \sigma, t)) \right) \quad (1)$$

for each $i \in I$ with strict inequality for at least one $i \in I$.

[3] We replace $A_{D(\sigma)}(s^*_{-D(\sigma)}, s_D)$ by $A_{J \cup B \cup L(\sigma)}(s)$ if $D = J \cup B \cup L$.

For each $k \in J \cup B$ we denote by $Y_k^{\max}$ k's maximal (in terms of inclusion) single period production possibility set . It cannot be further enlarged by any innovation. We assume that it exhibits constant returns to scale, i.e., its upper boundary is flat and it has no infinite portion parallel with some coordinate axis. (It is flat in the sense that its projection on any finite dimensional vector space of dimension $n \in N$ is a subset of a vector space of dimension $k < n$.) In the proof of our result we will use the following two assumptions.

Assumption 1. There exist $s^{(0)} \in S^+$ and $\tau^{(0)} \in N$ such that for each $\delta \in (0,1)$ (a) if $\sigma \in \Im_1$ occurs with a positive probability when the players follow $s^{(0)}$ and a single period production possibility set of at least one $k \in J \cup B$ at σ is a strict subset of $Y_k^{\max}$, then $s^{(0)}$ prescribes in $\Gamma_{(\sigma)}(\delta)$ R&D project that aims at achieving single period production possibility set $Y_k^{\max}$ for each $k \in J \cup B$, the probability that this R&D project succeeds in $\tau^{(0)}$ periods is positive, and for each $\sigma' \in \Im_1$, at which single period production possibility set of each $k \in J \cup B$ equals $Y_k^{\max}$, $s^{(0)}$ prescribes in $\Gamma_{(\sigma')}(\delta)$ for each $k \in J \cup B$ only use of netput[4] vectors lying in the upper boundary of $Y_k^{\max}$, which generate an element of $\Psi_{(\sigma)}$,[5] (b) there does not exist $s' \in \overline{S}$ that achieves the same in less than $\tau^{(0)}$ periods with a positive probability.

Assumption 2. (a) The use of producer goods and labor services as inputs in R&D is associated with non-decreasing returns to scale, i.e., for any R&D project the probability of is success in a given number of periods is convex in the vector of inputs. (b) The probability of success of an R&D project in a given number $\tau \in \{1,\ldots,\overline{T}\}$ periods is strictly increasing in the used quantity of any input into R&D in any period $t \in \{1,\ldots,\tau\}$ of its duration.

Proposition 2. *There exists $\underline{\delta} \in (0,1)$ such that for each $\delta \in [\underline{\delta},1)$ and each SMPGE $s^* \in \overline{S}$ of $\Gamma(\delta)$, $x(s^*)$ is strictly Pareto efficient.*

Proof. Taking into account part (b) of Assumption 1 and part (b) of Assumption 2, if the grand coalition wants to increase the average discounted expected real wealth of stakeholders of its members by increasing consumption or decreasing use of labor services at the expense of innovation in some period, then the best chance of doing so is by increasing consumption or decreasing use of labor services in the first period of $\Gamma_{(\sigma)}(\delta)$ for some $\sigma \in \Im_1$ described in part (a) of Assumption 1. It follows from part (a) of Assumption 2 that the optimal increases in consumed quantities (and, hence, also optimal decreases in inputs to R&D) in the first period of $\Gamma_{(\sigma)}(\delta)$ are bounded away from zero. Hence, the set of changes that need to be considered is compact. Each such change fails to increase the average discounted expected real wealth of stakeholders of members of the grand coalition for a discount factor close enough to one. Since the set of changes to be considered is compact, there exists $\underline{\delta} \in (0,1)$ such that for each $\delta \in [\underline{\delta},1)$ each of the changes mentioned above fails to increase the average discounted expected real wealth of stakeholders of members of the grand coalition.

It follows from the above considerations and part (b) of Assumption 2 that, for each $\delta \in [\underline{\delta},1)$ $\Psi_{(\sigma_0)}$ coincides with the set of vectors ψ generated by strategy profiles described in part (a) of Assumption 1. Thus, it is flat and convex. Moreover, for each $\delta \in [\underline{\delta},1)$ each SMPGE of $\Gamma(\delta)$ has the properties described in part (a) of Assumption 1. Take (arbitrary) such $\delta \in [\underline{\delta},1)$ and s^*. Let ψ^* be an element of $\Psi_{(\sigma_0)}$ generated by s^*.

Denote the vector of prices of labor services and consumption goods at states belonging to $\Im_1(s^*)$ by $p(s^*)$. Thus,

$$p(s^*) = \left(\left(\left(p_{kg}(s^*,\sigma)\right)_{g \in G_L \cap N(\sigma)}\right)_{k \in J \cup B}, \left(p_g(s^*,\sigma)\right)_{g \in G_C \cap N(\sigma)}\right)_{\sigma \in \Im_1(s^*)}.$$

Suppose that $x(s^*)$ is not strictly Pareto efficient. Then there exists $s' \in \overline{S}$ such that $x(s')$ satisfies

$$(1-\delta)\sum_{t \in N}\left(\delta^{t-1}\sum_{\sigma \in \Im_1(s')}\pi(s',\sigma,t)u_i\left(x_i(s',\sigma,t)\right)\right) \geq (1-\delta)\sum_{t \in N}\left(\delta^{t-1}\sum_{\sigma \in \Im_1(s^*)}\pi(s^*,\sigma,t)u_i\left(x_i(s^*,\sigma,t)\right)\right) \qquad (2)$$

[4] In a netput vector net outputs are recorded with a positive sign and net inputs with a negative sign. Each element of a single period production possibility set of any firm is a netput vector.

[5] The last subpart of part (a) of Assumption 1 is innocuous and it can be achieved in any subgame following a history belonging to $\Im_1$ at which the single period production possibility set of each $k \in J \cup B$ equals $Y_k^{\max}$. We include it only in order to shorten arguments in the proof of Proposition 2.

for each $i \in I$ with strict inequality for at least one $i \in I$. Taking into account properties of $\Psi_{(\sigma_0)}$, we can assume without loss of generality that there is $T \in N$ such that, starting from it, components of $x(s')$ are the same in each period (and it is obtained by the convex combination of single period components of $x(s')$ with coefficients equal to $(1-\delta)\delta^{t-1}$).

For each $i \in I$ for whom (2) holds as a strict inequality, the difference between average discounted expected expenditures on quantities of consumption goods specified by $x_i(s')$ and average discounted expected wage income from labor supplies specified by $x_i(s')$ exceeds *i*'s average discounted expected income from dividends and net interest payments (i.e., difference between interest paid to him and interest paid by him) generated by s^*. Otherwise, (using the fact that prices are bounded from above) there would be $t_i \in N$ such that starting from this period consumer i could buy components of $x_i(s')$ without borrowing of money beyond that generated by s^*. Thus, he could buy components of $x_i(s^*)$ prior to period t_i and components of $x_i(s')$ starting from period t_i. With respect to strict concavity of his single period utility function, this would increase his average discounted expected utility in comparison with that obtained form $x_i(s^*)$, contradicting the fact that he demanded $x_i(s^*)$ under s^*.

Also, taking into account that consumers' single period utility functions are strictly increasing in consumed quantity of any consumption good and strictly decreasing in supply of each labor service to any firm or bank, for each $i \in I$ for whom (2) holds as equality, the difference between average discounted expected expenditures on quantities of consumption goods specified by $x_i(s')$ and average discounted expected wage income from labor supplies specified by $x_i(s')$ equals or exceeds *i*'s average discounted expected income from dividends and net interest payments generated by s^*. (The argument showing this is analogous to the one in the preceding paragraph.)

Taking into account considerations in the preceding paragraph, there exists $K(\psi^*, p(s^*))$ such that the subspace of

$$\Pi_{\sigma \in \Im_1(s^*)} \left(\Pi_{g \in G_L \cap N(\sigma)} \left[-\sum_{t \in N} \delta^{t-1} \pi(s^*, \sigma, t) y_g^{\max}, 0 \right]^{\#(J \cup B)} \times \Pi_{g \in G_C \cap N(\sigma)} \left[0, \sum_{t \in N} \delta^{t-1} \pi(s^*, \sigma, t) y_g^{\max} \right] \right)$$

defined by

$$\sum_{\sigma \in \Im_1(s^*)} p(\sigma) z(\sigma) = K(\psi^*, p(s^*)) \tag{3}$$

cuts $\Psi_{(\sigma_0)}$ at ψ^*. Thus, in a neighborhood of ψ^* in $\Psi_{(\sigma_0)}$ there are points that lie below the subspace (3). Take (arbitrary) such $\psi^{**} \in \Psi_{(\sigma_0)}$. According to Theorem 1, there exists an SMPGE s^{**} of $\Gamma(\delta)$ satisfying $\gamma_{J \cup B \cup L}(s^{**}) = \gamma_{\psi}(\psi^{**})$. Thus, the grand coalition can increase average discounted expected real wealth of stakeholders of its members in $\Gamma(\delta)$ at s^{**} by switching to s^*. This contradiction with requirement of Definition 1 for the grand coalition and $\sigma = \sigma_0$ completes the proof. *Q.E.D.*

The proof of Proposition 2 has the same point of departure (after establishing that $\Psi_{(\sigma_0)}$ is flat) as the standard proof of strict Pareto efficiency of an equilibrium consumers' allocation in a general competitive equilibrium (see [3], subsection 5.5.1, p. 117-119). If an equilibrium consumers' allocation x^* is weakly Pareto dominated by another feasible consumers' allocation x', then we can draw two conclusions from it. First, each consumer, who is strictly better off under x' than under x^*, has income that is not sufficient to buy his part of x'. Second, each consumer, who is equally well off under x' than under x^*, has (with respect to local non-satiation of his preferences) income that just suffices or is insufficient to buy his part of x'. In the standard general equilibrium theory these two conclusions are used (through summation of consumers' budget inequalities) to show that x' is not feasible or that the sum of profits is not maximized in the competitive equilibrium generating x^* (see [3], subsection 5.5.1 for the latter approach). This is possible thanks to the fact that firms cannot affect prices. In our oligopolistic model firms can affect prices, so the approach in Proposition 2 has to be different. Summation of consumers' budget inequalities shows – using the fact that $\Psi_{(\sigma_0)}$ is flat – that aggregate "budget line" cuts $\Psi_{(\sigma_0)}$. (There must be points in it that lie bellow x'. Since $\Psi_{(\sigma_0)}$ is flat, this implies that a part of aggregate "budget

line" lies above $\Psi_{(\sigma_0)}$.) And now the multiplicity of equilibria implied by Theorem 1 enters the scene. Since aggregate "budget line" cuts $\Psi_{(\sigma_0)}$, it lies above some elements of $\Psi_{(\sigma_0)}$ that are generated by consumers' allocation in SMPGE. That is, the aggregate "budget line" enabling consumers to buy x^* enables them also to buy more of each consumption good and provide less of each labor service to each firm and bank (unless it is already zero) than in some SMPGE. This contradicts the definition of an SMPGE and completes the proof that x^* is strictly Pareto efficient.

As it is clear from considerations in the above paragraph, the fact that $\Psi_{(\sigma_0)}$ is flat plays a crucial role in the proof of an equilibrium consumers' allocation in an SMPGE. This fact stems from two assumptions: possibility of innovation (and its features) and discount factor close to one. The former can ultimately lead to single period production possibility sets of firms and banks with flat upper boundaries. The latter creates a motivation for firms and banks to aim at ultimate single period production possibility sets.

3 Conclusions

We have analyzed a dynamic general equilibrium model with non-decreasing returns to scale in production as well as in R&D. This includes strictly increasing returns to scale in production – a typical situation in which perfect competition (that competition policy authorities in the European Union and antitrust authorities in the USA try to impose) cannot work. We have shown that an equilibrium immune to deviations by all coalitions in all subgames generates a strictly Pareto efficient consumers' allocation. Thus, it has also the property for which a general competitive equilibrium is celebrated and which is one of the main motivations for competition/antitrust policy. This suggests that the motivation for competition/antitrust policy is questionable. Therefore, further research on interactions between oligopolistic firms in general equilibrium that would respect their natural technological characteristics (in particular increasing returns to scale in both production and R&D) and would be free of any politically motivated bias against cooperation between firms and banks, is needed. We hope that our step in this direction will soon have followers.

A driving factor for our result is attractiveness of R&D for patient firms and banks (i.e., when the discount factor is close enough to one). Firms and banks can cooperate in R&D as well as in selling innovated products. Nevertheless, they have to do it in the way that is immune to deviation by any coalition. Since successful R&D cannot be undone, those who innovate cannot be punished by those who fail to innovate. Therefore, by cooperating in R&D firms generate at least as much innovations as they would without cooperation, in fierce competition. Thus, our paper contributes also to the literature showing that dynamic competition based on innovation is much more important than static competition protected by competition/antitrust laws, which originated by [4].

Acknowledgements

The research reported in the present paper was financially supported by grant APVV-14-0020.

References

[1] Dierker, E., and Grodal, B.: The price normalization problem in imperfect competition and the objective of the firm. *Economic Theory* **14** (1999), 257-284.

[2] Horniaček, M., and Šimková, Ľ: Strong Markov perfect general equilibrium with innovation. In: *Proceedings of the Mathematical Methods in Economics 2015.* University of West Bohemia, Cheb, 2015.

[3] Mukherji, A.: *Walrasian and non-Walrasian equilibria.* Clarendon Press, Oxford, 1990.

[4] Schumpeter, J. A.: *Capitalism, socialism, and democracy.* 2nd edition. Harper and Brothers, New York, 1947.

Application of the Modern Approach to Age-specific Fertility Rates Stochastic Modelling in the Czech Republic

Ondrej Simpach[1]

Abstract. Mortality and fertility represent two basic components of natural population replacement. While the mortality can be modelled based on known conditions (decrease of the probability of death in time and increase of life expectancy, which can be applied to the most of advanced populations in general), in the case of fertility the approach must be slightly different. Changes in the level and distribution of fertility are characterized by social conditions, lifestyle and also by decision makings of individuals. There are studies that try (using the principal components method and stochastic trend) to find a model that explains and predicts the future development of age-specific fertility rates. Our study utilize also this approach.

The aim of this paper is to identify appropriate form of the Lee-Carter model based on the data set of Czech females' age-specific fertility rates. It will be used for modelling a matrix of age-specific fertility rates of Czech females. The predictions of these rates up to the year 2050 will be constructed. It is required that the shape of the distribution of predicted fertility rates follows the current data (which is not respected in some studies). The model also has to respect the trend that arises from past data. Predictions which are output of this paper provide lower values than expectations published by Eurostat. As our projections of age-specific fertility rates of Czech females are predicted by modern and sophisticated system, it is possible to claim that the future would be more pessimistic than Eurostat expects.

Keywords: fertility, Lee-Carter model, demography, stochastic modelling.

JEL classification: C61, C63
AMS classification: 62H12

1 Introduction

The population development and improving the living standard in the country is closely related to postponement of first childbirth to the later age and together the decline of live births in total (see e.g. the paper from Rueda, Rodriguez [13] or Simpach, Pechrova [16]). In many populations this decrease is below the level of simple reproduction of the population (2.08 children per 1 female within the reproduction period). Czech Republic have good database that will allow us to obtain the age-specific fertility rates $f_{x,t}$ of Czech females from 1925 to 2012. During this period, the development of these rates was affected by a wide range of social changes, which also brought the consequences of changes in fertility of Czech females. Difficult to explain and difficult to predict is the behaviour as the result of individual decisions in family planning (Peristera, Kostaki [11]). The level of fertility cannot permanently decrease in the future, because there is its logical value below which the fertility never decreased before, (e.g. the past development of the Czech time series showed that in 1999 there was the total fertility rate 1.13 live birth child per 1 female during her reproductive period and the range of values 1.13–1.18 was in many other cases during the 90s of the last century). Neither this value can permanently grow in the future, because of health point of view there is a maximum possible value that a female cannot exceed (see e.g. Caputo, Nicotra, Gloria-Bottini [3] or Myrskylae, Goldstein, Cheng [10]). The level of fertility varies between

[1]University of Economics Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, Winston Churchill sq. 4, 130 67 Prague 3 - Zizkov, Czech Republic, ondrej.simpach@vse.cz

its logical lower and upper limits in time, and just as well depends on the shape of the distribution of age-specific rates (Rueda, Rodruguez [13]).

In our paper we estimate the principal components that explain fertility from the multidimensional matrix of age-specific fertility rates (see Hyndman, Booth [4]). This will be performed using the singular value decomposition (SVD) and by Lee-Carter model (Lee, Carter [7]). Estimates are made on the basis of the shortened length of the analysed matrix for the period 1998-2012 (see also Simpach, Langhamrova [15]), because in the case of fertility modelling it makes sense to use only the current and stable database, which is not deflected by past influences. (The shorter is the analysed database of fertility, the more realistic results for the Czech population can be expected). We also calculate the predictions of age-specific fertility rates for the period 2013-2050 on the basis of this model and in the conclusion we compare our results with low variant of fertility projection according to Eurostat. It will be obvious that Eurostat projections are initially overestimated above the level of 2012, on the other hand the projections by Lee-Carter model are initially slightly underestimated bellow the level 2012. We can use these estimates as a lower and upper limit of the potential future fertility in the Czech Republic.

2 Methodology and Data

We use the empirical data from Czech Statistical Office (CZSO) about the number of live-born persons to x-year old mothers in time t ($N_{x,t}$) and the number of mid-year female population x-year old in time t ($S_{x,t}$) for our calculations, where x = 15–49 completed years of life and t = 1925–2012. This allows us to calculate the age-specific fertility rates as

$$f_{x,t} = \frac{N_{x,t}}{S_{x,t}}, \tag{1}$$

and after ($\times$ 1,000) we interpret the result as the number of live births per 1,000 x-year old females in time t. In order to be clear, which changes in age-specific fertility rates occurred in the past, their development is shown in Figure 1 as "rainbow" chart. (This technology uses Hyndman, Booth [4] and Hyndman [5]). We can see the changing of maximum values of the age-specific fertility rates in the past and especially since the beginning of 90s of the last century become the trend of postponing childbirth to the later ages and in these days the modal age have exceeded the value of 30 years.

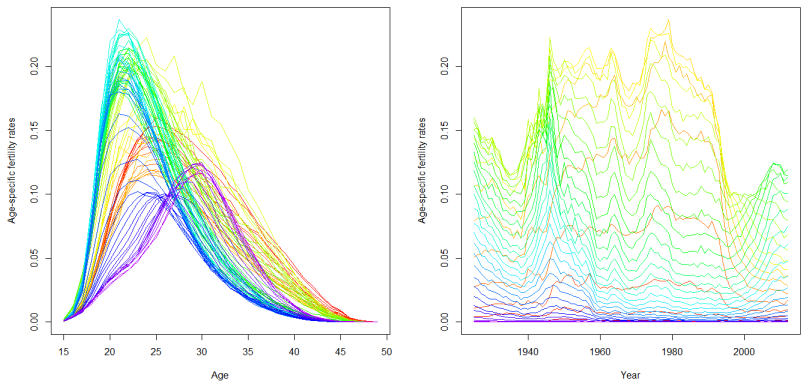


Figure 1 Empirical values of $f_{x,t}$ of the Czech females for the period 1925–2012 in rainbow charts. The left chart represents changes of $f_{x,t}$ over time, the right one represents changes of $f_{x,t}$ within the age groups. *Source: Data CZSO, author's calculations and illustration based on Hyndman [5] approach.*

Used approach based on principal component is, that the empirical values of age-specific fertility rates of Czech females can be decomposed (see Lee, Carter [7] or Lee, Tuljapurkar [8]) as

$$f_{x,t} = a_x + b_x \times k_t + \varepsilon_{x,t} \tag{2}$$

where x = 15–49 years of life, t = 1,2, ..., T, a_x are the age-specific fertility profiles independent of time, b_x are the additional age-specific components determine how much each age group changes when parameter k_t changes and finally k_t are the time-varying parameters - the total fertility indices. ($\varepsilon_{x,t}$ is the error term with characteristics of white noise). The estimation of b_x and k_t is based on Singular

Value Decomposition (SVD) of matrix of age-specific fertility rates, presented e.g. by Bell, Monsell [1], Lee, Carter [7] or Hyndman, Booth [4]. The age-specific fertility rates $f_{x,t}$ at age x and time t according to Lee, Tuljapurkar [8] create $35 \times T$ dimensional matrix

$$\mathbf{F} = \mathbf{A} + \mathbf{B}\mathbf{K}^{\top} + \mathbf{E} \tag{3}$$

and the identification of Lee-Carter model is ensured by

$$\sum_{x=15}^{49} b_x = 1 \qquad \text{and} \qquad \sum_{t=1}^{T} k_t = 0 \tag{4}$$

Finally,

$$a_x = \frac{\sum_{t=1}^{T} f_{x,t}}{T} \tag{5}$$

is the simple arithmetic average of age-specific fertility rates. For predicting the future age-specific fertility rates it is necessary to forecast the values of parameter k_t only. This forecast is mostly calculated by *ARIMA(p,d,q)* models with or without drift, (by Box, Jenkins [2] methodology). The values of the parameters a_x and b_x are independent of time and the prediction using the Lee-Carter model is according to Lee, Tuljapurkar [8] purely extrapolative. Our decomposition is applied only to 1998–2012 dataset, because according to the development shown in the Figure 1 (right) is evident, that this shape of distribution is stable reproduction period of Czech females which can continue to the future.

3 Results

In RStudio software (R Development Core Team [12]) we estimate the parameters a_x (age-specific fertility profiles independent in time), b_x (additional age-specific components determine how much each age group changes when k_t changes) and k_t (time-varying parameters - the total fertility indices) for Lee-Carter's model based on data matrix of 1998–2012 only, using the SVD method implemented in the package "demography" (Hyndman [5]). We can see these parameters in the Figure 2.

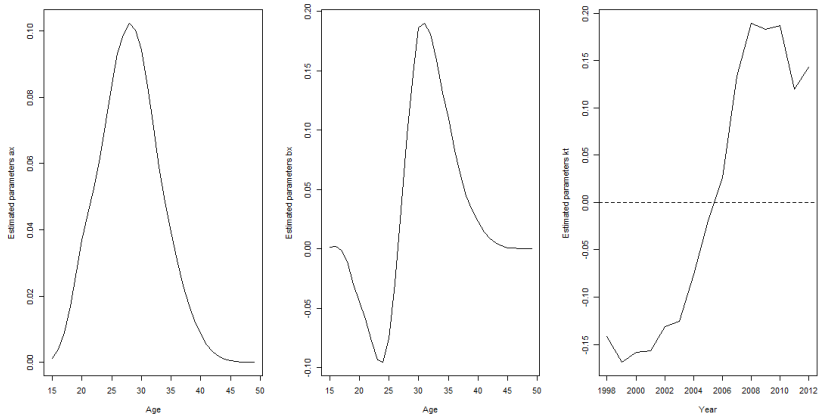


Figure 2 Estimates of age-specific fertility profiles independent in time (parameter a_x, left), the additional age-specific components determine how much each age group changes when k_t changes (parameter b_x, middle) and estimates of the time-varying parameters k_t - the fertility indices (right). *Source: author's calculations and illustration.*

After that we calculated predictions of the parameters k_t up to the year 2050 based on the methodological approach of individual time series modelling ARIMA, (Box, Jenkins [2], Melard, Pasteels [9]) and ran by "forecast" package in RStudio software (Hyndman [5]). Parameters of ARIMA models are displayed in Table 1. It is important to note that the dataset of 1998–2012 is too short for providing forecasts up to the year 2050. Therefore, we use dataset from 1925 to 2012 as a support for our analysis. Given that the resulting predictions of k_t parameters are the same as using both dataset of 1925–2012 and 1998–2012, our results are correct. Realization of predictions of k_t parameters is shown in Figure 3. 95% both-side confidence intervals are slightly closer at the case of model based on 1998–2012 dataset.

Database 1925–2012:	ARIMA (1,1,0)	Database 1998–2012:	ARIMA (1,1,0)
AR(1)	0.3494	AR(1)	0.4782
s.e.	0.1000	s.e.	0.2244
t-stat.	3.4940	t-stat.	2.1310
AIC	-142.98	AIC	-45.690
BIC	-138.05	BIC	-44.410

Table 1 Estimated parameters of two ARIMA models for parameter k_t. We used our knowledge of the shape of the ARIMA model for database 1925–2012 and applied it to the shortened model. Future predictions are comparable despite the fact that the database 1998–2012 is very short for quality prediction. *Source: author's calculations and illustration.*

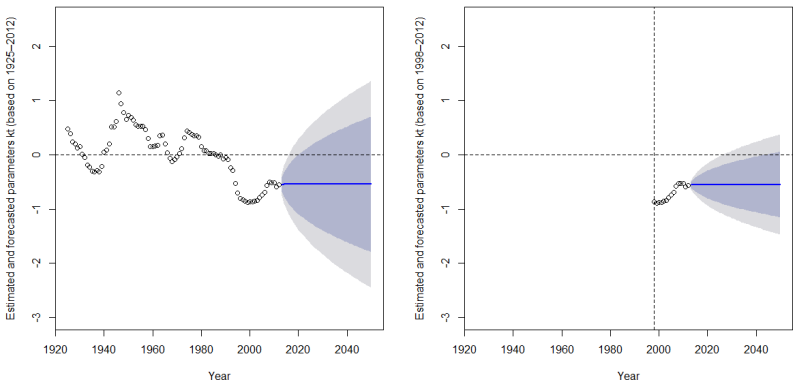


Figure 3 Prediction of the time-varying parameters k_t - fertility indices. On the left side is the model based on data of 1925-2012, on the right side is the model based on data of 1998–2012. There is no significant difference when using different lengths of time series. *Source: author's calculations and illustration.*

Now we evaluate Lee-Carter model on the basis of approach, which is presented by Charpentier, Dutang [6] or Simpach, Dotlacilova, Langhamrova [14]. Using RStudio software we display the Pearson's residues. Lee-Carter's model is evaluated on the basis of the residues by age x and of the residues at time t. The most residues are concentrated around 0, a lot of variability is explained by the estimated model. The Pearson's residues for estimated Lee-Carter model are shown in the Figure 4, where residues by age x dependent on time t are on the left and the residues at time t dependent on age x are on the right.

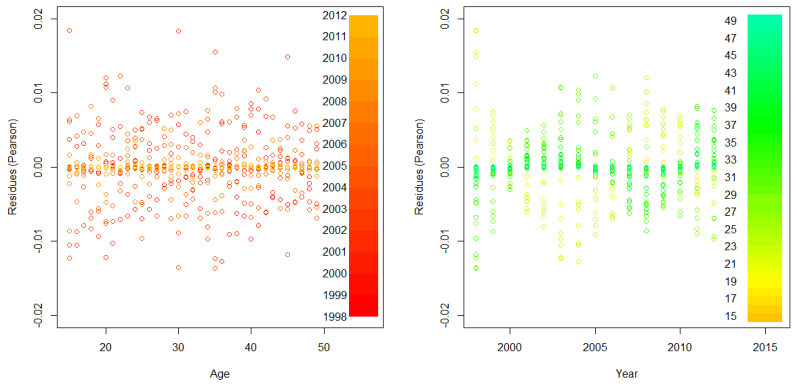


Figure 4 Diagnostic control of the Lee-Carter's model - Pearson's residues for model based on data of 1998–2012. *Source: author's calculations and illustration.*

Based on the estimated parameters $\hat{a}_x$, $\hat{b}_x$ and $\hat{k}_t$ of Lee-Carters model we can estimate the future values of $f_{x,t}$ as

$$f_{x,t} = \hat{a}_x + \hat{b}_x \times \hat{k}_t. \tag{6}$$

Estimated values (left) and the empirical values with the attached estimates (right) of age-specific fertility rates $f_{x,t}$ based on Lee-Carter's model are displayed in 3D perspective chart in the Figure 5 (top). The

estimated values based on the expert expectations by Eurostat (left) and then the empirical values with these attached estimates (right) are displayed below.

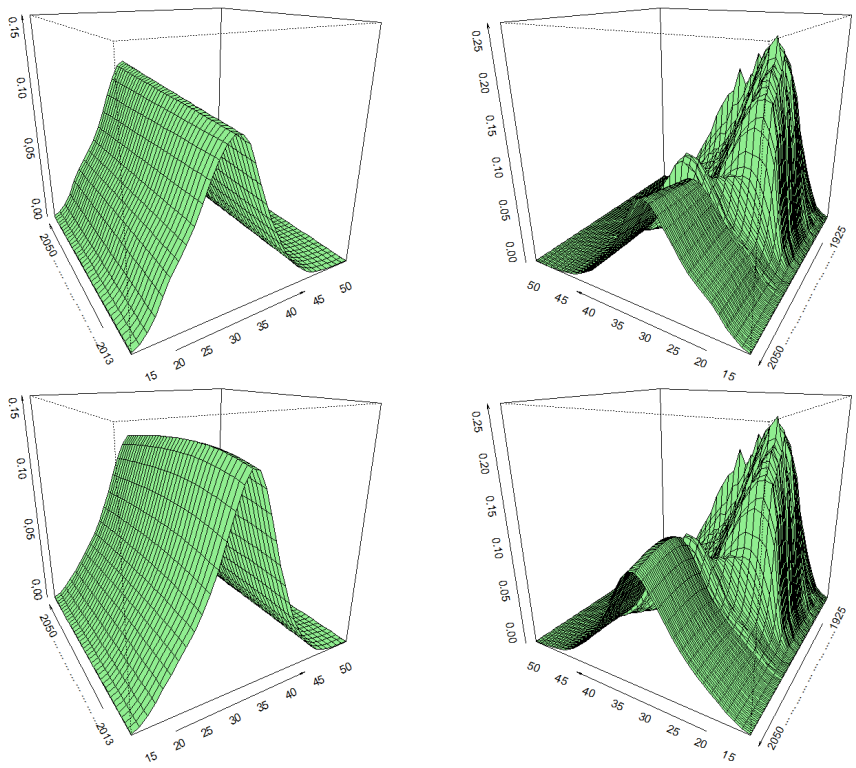


Figure 5 Forecasted values of age-specific fertility rates $f_{x,t}$ of Czech females for the period 2013 to 2050 by Lee-Carter's model based on actual data matrix for the period 1998 to 2012 (top left) and the empirical values of these rates for the period 1925 to 2012 with an forecast attachment (top right). Bottom left are $f_{x,t}$ of Czech females for the period 2013 to 2050 forecasted by Eurostat and bottom right are empirical values with these forecasted values. *Source: author's calculations and illustration, Eurostat database.*

Estimates calculated by the Lee-Carter model are initially underestimated bellow the empirical values of 2012. This is probably due to the deflection of the average fertility profile independent of time (parameter a_x), because shortly before 2000 there was the level of fertility in the Czech Republic significantly lower than today. The forecast assumes a gradual (linear) increase up to the year 2050. On the other hand, the estimates published by Eurostat in its low variant are overestimated over the empirical values of 2012. At a time when the prediction was processed, Eurostat has already calculated with a higher fertility than really occurred. According to Eurostat expectations there is evident an increase of fertility in approximately next 20 years, and later the fertility should start to fall again.

4 Conclusion

The aim of this paper was to estimate the main components explaining the level of fertility in the Czech Republic by the Lee-Carter model and calculate predictions of age-specific fertility rates up to the year 2050. Results were compared to expert estimates by Eurostat in the low variant and there were shown that they are not quite the same. However, if we use estimates of the Lee-Carter model and prediction of Eurostat in the low variant together, we can get a reliable picture of the future level of age-specific fertility rates. Each of these approaches has a different methodology for the estimation, and because the results are not dramatically different, we may take them as a reliable upper and lower limits of the future development of fertility in the Czech Republic. According to the results of prediction by Lee-Carter model and expert expectations by Eurostat is impossible that the level of total fertility rate in the Czech Republic in 2050 would almost reached the simple reproduction of the population.

Acknowledgements

This paper was supported by grant No. P402/12/G097 DYME "Dynamic Models in Economics" of the Czech Science Foundation and also by long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Bell, W. R., Monsell, B.: Using principal components in time series modelling and forecasting of age-specific mortality rates. *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 154–159, (1991).

[2] Box, G. E. P., Jenkins, G.: *Time series analysis: Forecasting and control*. San Francisco, Holden-Day, 537 p., (1970).

[3] Caputo, M., Nicotra, M., Gloria-Bottini, E.: Fertility Transition: Forecast for Demography. *Human biology*, Vol. 80, Iss. 4, pp. 359–376, (2008).

[4] Hyndman, R. J., Booth, H.: Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, Vol. 24, Iss. 3, pp. 323-342, (2008).

[5] Hyndman, R. J.: *demography: Forecasting mortality, fertility, migration and population data*. R package v. 1.16, URL: http://robjhyndman.com/software/demography/, (2012).

[6] Charpentier, A., Dutang, Ch.: *L'Actuariat avec R*. [working paper]. Decembre 2012. Paternite-Partage a lindentique 3.0, France de Creative Commons, 215 p., (2012).

[7] Lee, R. D., Carter, L. R.: Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, Vol. 87, pp. 659–675, (1992).

[8] Lee R. D., Tuljapurkar, S.: Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association*, Vol. 89, pp. 1175–1189, (1994).

[9] Melard, G., Pasteels, J. M.: Automatic ARIMA Modeling Including Intervention, Using Time Series Expert Software. *International Journal of Forecasting*, Vol. 16, pp. 497–508, (2000).

[10] Myrskylae, M., Goldstein, J. R., Cheng Yen-Hsin., A.: New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals. *Population and Development Review*, Vol. 39, Iss. 1, pp. 31-56. (2013).

[11] Peristera, P., Kostaki, A.: Modeling fertility in modern populations. *Demographic Research*, Vol. 16, No. 6, pp. 141-194, (2007).

[12] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/, (2008)

[13] Rueda, C., Rodriguez, P.: State space models for estimating and forecasting fertility. *International Journal of Forecasting*, Vol. 26, Iss. 4, pp. 712-724, (2010).

[14] Simpach, O., Dotlacilova, P., Langhamrova, J.: Effect of the Length and Stability of the Time Series on the Results of Stochastic Mortality Projection: An application of the Lee-Carter model. *Proceedings ITISE 2014*, Granada: University of Granada, pp. 1375-1386, (2014).

[15] Simpach, O., Langhamrova, J.: Stochastic Modelling of Age-specific Mortality Rates for Demographic Projections: Two Different Approaches. *Mathematical Methods in Economics 2014*, Olomouc: Palacky University in Olomouc, pp. 890-895, (2014).

[16] Simpach, O., Pechrova, M.: Assessing the impact of standard of living on the life expectancy at birth using Vector Autoregressive Model. *Mathematical Methods in Economics 2013, PTS I and II*, Jihlava: College of Polytechnics Jihlava, pp. 921–926, (2013).

Use of linear programming method to constructing a model for reduction of emission in a selected company

Eva Šlaichová[1], Eva Štichhauerová[2], Magdalena Zbránková[3]

Abstract. Corporate social responsibility and environmental policy have become the key criteria in considering of competitiveness of the companies. In the frame of these issues it is possible to use also methods of the operation research, while linear programming seems to be one of the most efficient techniques. It can be used to solve problems dealing also with environmental and ecological topics. Application of the linear programming generally enables the managers to maximize the profit and minimize the costs on the other hand and of course, it is necessary to involve the ecological costs in their calculations. This contribution presents the results of the pilot research focused on the use of operations research methods in this field. The authors of the contribution primarily took into consideration the linear programming, particularly application and results achieved by this method. The analysis was carried out and the environmental aspects of activities of the company were evaluated – particularly production of CO_2 emissions. The ecologic-economical model is tested on the real data of the selected enterprise. The contribution also presents general applicability of the model

Keywords: emission reduction, mathematical modeling, linear programming, operation analysis.

JEL Classification: C02, Q52
AMS Classification: 90C05

1 Introduction

One of the currently mostly studied topics are greenhouse gas emissions and their significant component CO_2, so called carbon dioxide emissions, which have a significant negative effect on human health and the economy. In the Czech Republic, it deals particularly with a loss of drinking water sources, reduction of agricultural yields and increased morbidity of the population [2]. The question is whether between the amount of produced emissions and fuel consumption is a direct proportion. Then it would have been possible to prove that the environmental and economic aspects are not mutually exclusive during the provision of transport services.

In the frame of these issues it is possible to use also methods of the operation research, while linear programming seems to be one of the most efficient techniques because it can be used to solve problems dealing also with environmental and ecological topics. The authors of the contribution decided to use two simple assignment models with different objective functions. However the multi-objective (multi-criteria) linear programming with the weighted sum of these objectives as a compromise function is also possible to use to solve this case study.

This contribution presents the results of the case study from the real logistic company (with modified data) focused on comparison of two models, model optimizing the transportation plan while minimizing emissions of CO_2 and the same model minimizing fuel costs.

2 Literature Overview

Linear programming can be used in many business areas. The first is the production planning [7]. This area is focused on the production plan, where it is necessary to combine the production of several types of products in order to achieve the maximum possible profit [9]. Other applications are so called traffic problems. The criterion in this case, is, on the contrary, the minimal costs [8]. Linear programming also enables solutions of assignment problems [10]. Linear assignment problem is essentially a special case of the traffic problems, where there is the same number of resources as well as customers and requirements of each customer and also the capacity of each

[1] Faculty of Economics, Technical University of Liberec, Department of Business Administration and Management, Studentská 1402/2, Liberec 1, Czech Republic, eva.slaichova@tul.cz.
[2] Faculty of Economics, Technical University of Liberec, Department of Business Administration and Management, Studentská 1402/2, Liberec 1, Czech Republic, eva.stichhauerova@tul.cz.
[3] Faculty of Economics, Technical University of Liberec, Department of Business Administration and Management, Studentská 1402/2, Liberec 1, Czech Republic, magdalena.zbrankova@tul.cz.

source are equal to one. At the same time, the assignment follows that it is not possible to divide these customer requirements, i.e. such as to satisfy a particular customer, so that it will meet the requirement of half of one source and the remaining half from another source.

Methodological approach in the application of a linear programming in the decision-making process of the company´s management follows:

1. In the first step it is always necessary to identify and define the problem.
2. The second stage is to build an economic model, which states:
   a) an aspect of the evaluation, in most situations it is the maximal profit or minimum cost,
   b) characteristics of the processes in the system, these are activities that affect the objective of the analysis,
   c) identification of the factors that enter into the process, those can be energy, working capital or material,
   d) description of the interrelationships of all the above-mentioned elements of the business model.
3. The third stage is the construction of a mathematical model. The mathematical model converts data from an economic model to determine the number of variables and the mathematical relationship between the variables. The part of this phase is also testing of a mathematical model and verification of the model [3].

Corporate social responsibility and environmental policy have become the key criteria in considering of competitiveness of the companies. According to this statement, linear programming can be applied also in constructing the ecologic economical model for the industrial company profit optimization, see e.g. [11], [1] and [4]. The professional literature also deals with the successful use of this method for optimizing of the model considering vehicle assignment and segment emission constraints for energy saving [5]. Another advantage of linear programming is that it can be used in planning and scheduling of the production in a combination with an appropriate software [6].

3 Optimization of the transportation plan while minimizing emissions of CO2

3.1 Data and methodology

This contribution presents the results of the case study on the example of the logistic company. The aim is to compare two models, model optimizing the transportation plan while minimizing emissions of CO_2 and the same model minimizing fuel costs. The data were modified at the request of the company.

Linear programming was applied to set both models because all the constraints can be described by the linear equations and inequations. Calculations were processed by the software tool Solver provided by Microsoft Excel 2007.

The first model assumes that company wants to optimize the transportation plan while minimizing emissions of CO_2.

3.2 The structure of the model

The aim of the logistical company is to minimize production of a total volume of CO_2 emissions via assigning trucks to routes from the main depot to the target depot every day. The premise is that transportation is carried out at all routes every day. The aim is to minimize the total volume of CO_2 emissions produced per day when assigning trucks to routes from the main depot to the target depots. The premise is that every day the transportation is carried out on all routes just once. The objective function z is described by expression (1).

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} e_{ij} * x_{ij} \rightarrow \min \qquad (1)$$

$$i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$$

Where the coefficient e_{ij} is the volume of CO_2 emissions produced by the i-th vehicle on the road from the main depot to depot j-th and back (ie. on the j-th line). If the number of cars in the fleet is m and n is the number of routes, the maximum number of possible combinations of vehicle and the route is $m * n$.

Assigning cars to routes has to be followed in order to:

- Each route was secured by just one car;
- Each car was assigned to no more than to one route.

An obligatory condition (2) indicates that the variable x_{ij} can have only two values, 0 and 1:

$$x_{ij} \in \{0;1\} \tag{2}$$

From the fleet logistics company at maximum of one car can be assigned to a route between the capital and the target depot. Capacity of the fleet cannot be exceeded. Thus, not more than one of the variables x_{ij} must be for the i-th car equal to 1, and this must be applied to all vehicles. This is captured by so called *line restrictions* given by inequation array (3):

$$\begin{aligned} &x_{11} + x_{12} + \ldots + x_{1n} \leq 1 \\ &x_{21} + x_{22} + \ldots + x_{21} \leq 1 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &x_{m1} + x_{m2} + \ldots + x_{mn} \leq 1 \end{aligned} \tag{3}$$

Each route between the depot and the depot target must be implemented by one car, ie. at least one vehicle must be assigned to the route. Thus, just one of the variables x_{ij} must be for the j-th route equal to 1, and this must be applied to all routes. This is captured by so called *columnar restrictions* given by equation array (4):

$$\begin{aligned} &x_{11} + x_{21} + \ldots + x_{m1} = 1 \\ &x_{12} + x_{22} + \ldots + x_{m2} = 1 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &x_{1n} + x_{2n} + \ldots + x_{mn} = 1 \end{aligned} \tag{4}$$

3.3 The application of the model on the real data

Necessary information on the composition of the fleet of the logistics company can be seen in the Table 1 (data were modified at the request of the company). This is a truck with a maximum load volume between 15-19 m^3. Limitations in the case of a vehicle B13 is the maximum length of the route without refueling (CNG), suggesting that it may be assigned only to routes within a distance of 400 km.

Type of the vehicle*	Indicati-on of the vehicle B_i	Year of origin	Emission of CO_2 [g/km]	Fuel Consumption [l/km]*
Renault Master 2.8 DTI	B1	1999	287	0.108
Fiat Ducato Maxi 2.8 JTD	B2	2002	235	0.072
Mercedes-Benz Sprinter 308 CDI	B3	2004	282	0.107
Fiat Ducato 33 Multijet 120 L3H2	B4	2007	218	0.082
Mercedes-Benz Sprinter 315 CDI	B5	2009	252	0.104
Fiat Ducato 2.3 JTD 17H L4H3 Maxi	B6	2012	225	0.085
Renault Master 2.8 DTi	B7	2012	215	0.081
Iveco Daily 35S15V Euro 5	B8	2013	187	0.094
Mercedes-Benz Sprinter 319 CDI	B9	2014	205	0.078
Mercedes-Benz Sprinter 313 D Blue Efficiency +	B10	2014	195	0.075
Mercedes-Benz 319 CDI	B11	2014	207	0.080
VW LT35 - MAXI JUMBO	B12	2003	254	0.103
Fiat Ducato 3.0 Natural Power CNG L2H2	B13	2011	239	0.088**

* guidance combined fuel consumption in litres per 1 km

** consumption of gas CNG v kilogrammes per 1 km

Table 1 CO_2 emissions and fuel consumption by type of vehicle

Target Destination Depot	Route Indication A_j	Route Lenght [km]
Ústí nad Labem	A1	214
Olomouc	A2	526
Hradec Králové	A3	250
Ostrava	A4	708
Brno	A5	376
Plzeň	A6	224
České Budějovice	A7	268
Liberec	A8	242
Chodov (KV)	A9	324
Jihlava	A10	226
Velké Přílepy + Praha 10*	A11	126 (tj. 92+34)

* Two routes secured by one car.

Table 2 The length of the route between the main depot in Modletice and target depots, including the return journey

Structural factors e_{ij} of the objective function (1) were obtained by multiplying the volume of CO_2 emissions *i*-th vehicle g/km (see Table 1) and the route length between the depot and the *j*-th depot in km (see Table 2). Exception are coefficients $e_{13,2}$ $e_{13,4}$ and so it was necessary to lay them equal to the value of 999,999, to avoid assigning vehicles B13 to the routes longer than 400 km (depot Olomouc and Ostrava). The value 999,999 is used as so called prohibitive rate.

Linear programming was solved in MS Excel Solver tool. While respecting the constraints of the domain (2), the inequation array (3) and the equation array (4) was found optimal combination for assignment trucks from the fleet on individual routes between the main depot in Modletice and target depots (see Table 3 in chapter 6). The minimum possible amount of CO_2 emissions produced by vehicles on routes between the depots is **738,262 g** daily.

4 Optimization of the transportation plan while minimizing fuel costs

The second model assumes that company wants to optimize the transportation plan while minimizing fuel costs.

4.1 The structure of the model

The aim is to minimize logistics costs on a daily fuel consumption when assigning trucks to routes from the main depot to the target depots. The premise is that every day the transportation on all routes is carried out just once.

The objective function *z* is described by expression (5).

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} * x_{ij} \rightarrow \min$$
$$i = 1, 2, \ldots, m; j = 1, 2, \ldots, n \tag{5}$$

Where the coefficient c_{ij} is the consumption cost of the *i*-th vehicle at the *j*-th path. The maximum number of possible combinations of vehicle and the route is $m * n$.

As in the previous case, again, it must be ensured that each line was secured by just one car (see equation array (4)), and each car has to be assigned to just one route (see inequation array (3)).

4.2 The application of the model on the real data

Due to the fact that the company's fleet consists of both diesel trucks, as well as one CNG gas, consumption was converted into monetary terms. For this purpose, the price of one liter of diesel was set at CZK 32 and the price of 1 kg of CNG gas for CZK 26.

Structural coefficients c_{ij} objective function (5) were obtained by multiplying the monetary value of consumed fuel of *i*-th vehicle in CZK/km (see Table 1) and the length of the *j*-th km route (see Table 2). Coefficients $c_{13,2}$ and $c_{13,4}$ were laid equal to the value of 999,999, to avoid assigning vehicles B13 routes to relevant depots Olomouc and Ostrava, distant from the main depot Modletice more than 400 km.

While respecting the constraints of the domain (2), the inequation array (3) and the equation array (4), an optimal combination of assignment of trucks to individual routes was found in MS Excel (see Table 3 in chapter 6). Minimum overall fuel consumption amounts to about **CZK 8,945** per day.

5 Comparison of the models

The total cost for fuel consumption per day in a situation that the vehicles are assigned to routes with respect to the requirement for a minimum total volume of produced CO_2 is CZK 9,329, it means an increase about 4.29 % (CZK 384) more than within the cost-optimal combination.

Total daily volume of CO_2 emissions in a situation where the company respects the requirement for minimum consumption of fuel, is 766,062 g, which is about 3.77% (27,800 g) more than the combination of vehicles minimizing CO_2 emissions.

Destination Depot	**Route**	**Cars assigned to minimize:**	
		total CO2 emissions per day	**total cost of the consumption per day**
Ústí nad Labem	A1	B5	B12
Olomouc	A2	B10	B10
Hradec Králové	A3	B4	B7
Ostrava	A4	B8	B2
Brno	A5	B9	B13
Plzeň	A6	B13	B8
České Budějovice	A7	B7	B11
Liberec	A8	B6	B4
Chodov (KV)	A9	B11	B9
Jihlava	A10	B2	B6
Velké přílepy + Praha 10	A11	B12	B5
(unused car)		B1	B1
(unused car)		B3	B3
The total volume of CO2 emissions per day		**738,262 g**	766,062 g
The total cost for fuel consumption per day		CZK 9,329	**CZK 8,945**

Table 3 Comparison of the models

6 Conclusion

From the model results it cannot be implied that the amount of emissions produced was directly proportional to fuel consumption, rather the contrary. In the case that the company decides to give priority to optimize CO_2 emissions, the total cost of the fuel consumption will be about CZK 384 per day higher than by the model, when the company would have preferred aspect of the optimal costs. In the case of saving of one kilogram of CO_2, the cost of fuel will increase by 13.81 CZK. Although according to the results, between low emissions and costs can be find a positive relationship. If the company prefer vehicles that have lower emissions for longer transport routes, it is possible to reduce overall emissions, thereby reducing the carbon. As indicated by the results, a vehicle that was not deployed to transportation were not meeting the requirements of EURO II and III Those vehicles were produced before 2004 and their emissions are among the highest (more than 280 g per kilometer). At the same time, these vehicles have the highest consumption (Tab.1), so that even when selecting the vehicles to minimize the total cost of the fuel consumption, these vehicles would not be put into operation as well. From this perspective, it is worth considering a further economic evaluation of whether it would be economically and environmentally efficient to replace the "outdated" new vehicles that fulfill stricter emission standards for pollutants EURO VI.

Regarding the amount of fuel costs, there is scope for reducing overall costs due to the economic and defensive driving. Thanks to economical driving it is possible to achieve fuel savings (eg. by saving 1 l/100 km when running 3,484 km per day and cost 32 CZK/l, savings could be around 1,116 CZK). Economical driving in

addition to fuel economy carries even further: it is possible to calculate the savings of brakes, tires, clutch, etc. Safe and defensive driving style, which is linked to the driver's ability to anticipate and deal with crisis situations, eventually to avoid them, can translate eg. even in lower accident rates and other related benefits (e.g. insurance, liability, etc.). It follows that drivers should learn the rules of economic and defensive driving and should be encouraged to respect them. A financial award can be offered to them in the case of achieving fuel savings in a certain amount for a certain periods.

Monitoring of CO_2 emissions can be realized by analyzing the driving book, or through conversion calculator Surplus consumption. These data can be a starting point for further optimization and reduction of emissions. The company can decide to track the carbon footprint of transportation services, learn to recognize the true costs and generate opportunities to reduce them. This allows to present its environmental responsibilities to existing and potential customers and gain a certain competitive advantage. Since currently increasing the number of customers who are interested in the impact of the product on the climate and the environment, green marketing can be a result of a growing number of new customers.

References

[1] Ahmed, S. I.; et al. A Linear Programing Approach for Landfill Gas Utilization for Renewable Energy Production. In: *Applied Mechanics and Materials,* Trans Tech Publications, Trnava, 2014, 619-624.
[2] EurActiv. *Emise CO2 z automobilů: v EU i ČR se střetávají dva přístupy*. [online] 2013 [cit. 2015 - 5 - 8] Available from: http://www.euractiv.cz/obchod-a-export0/clanek/emise-co2-z-automobilu-v-eu-i-cr-se-stretavaji-dva-pristupy-automobilovy-prumysl-emisni-cile-hodac-acea-010618
[3] Gros, I.: *Matematické modely pro manažerské rozhodování*. VŠCHT, Praha, 2009.
[4] Hao, D.; et al: Prediction of Potential for Carbon Emission Reduction based on LEAP Model. *International Journal of Advancements in Computing Technology,* **5** (2013), 715-722.
[5] Hui, H.; et al: A multi-objective train-scheduling optimization model considering locomotive assignment and segment emission constraints for energy saving. *Journal of Modern Transportation.* **21**(2013), 9-16.
[6] Koblasa, F. and Manlig, F.: Application of Adaptive Evolution Algorithm on Real-World Flexible Job Shop Scheduling Problem. In: *Mathematical Methods in Economics*. Olomouc, 2014, 425 - 430.
[7] Lord, M. S.; et al: Linear Programming & Optimizing the Resources. *Interdisciplinary Journal Of Contemporary Research In Business*, **11** (2013), 701 – 705.
[8] Paramet, L.: Global optimization method for mixed transportation network design problem: A mixed-integer linear programming approach. *Transportation Research,* **5** (2011), 808–827.
[9] Slack, N.: *Operations Management.* Pearson Education Limited, Edinburgh, 2013.
[10] Turčok, L.; et al: Application of Linear Programming Method on Selected Examples from the Enterprises. In: *Applied Mechanics and Materials*. Trans Tech Publications, Trnava, 2014, 201-206.
[11] Zapletal, F., and Němec, R.: The Usage of Linear Programming to Construction of Ecologic-economical Model for Industrial Company Profit Optimization. In: *Mathematical Methods in Economics*. Karviná, 2012, 1010-1015.

Testing dependencies in the progress of day surgery intervention among adults by adjusted residual analysis

Vincent Šoltés[1], Beáta Gavurová[2]

Abstract. Slovak hospitals represent the weakest segment of our health care system. The problem is not only the outdated material-technical base, but also the lack of patient satisfaction, demotivation of doctors, nonfunctional continuing education of physicians and other health care professionals, etc. The introduction of day surgery is one of the potential opportunities in savings funds of health insurance providers. Day surgery is beneficial not only for patients but also for health care providers, as well as health insurance providers. According to the Ministry of Health of the Slovak Republic, one of the solutions to the lack of funds is optimization the network of hospital beds. Those aspects justify the difficulty of addressing this issue, to which we contributed the declared partial outputs under our collaboration with the Ministry of Health and National Health Information Centre. In this paper, we focused on the analysis on day surgery of seniors in Slovakia, with respect to test rate of hospitalization and age we used the analysis of cross tables implemented in the statistical software SPSS, version 19. To determine the cause of the significance of relationship between two categorical characters we used so-called "Adjusted Residual". The results are part of the research report for the Ministry of Health.

Keywords: Day Surgery, Healthcare System, Cross tables, Adjusted Residual

JEL Classification: C55, I12
AMS Classification: 91C99

1 Reduction of beds in the economy of Slovak health care system

In the last two decades, the issue of funding health care has become a high priority. The economic crisis accompanied by the need of reducing public budget deficit, creates constant pressure to reduce costs arising in different areas [2], [10]. One of the areas in which we see a clear possibilities of saving not only in the health care system, but also in national economic context, is the provision of health care through the use of day surgery (DS). DS has become a form of saving funds in the health care system, not only in the world, where has long been successfully established and in constant progression, but also in Slovakia. DS can be defined as an operation or procedure during which the patient is admitted or discharged from a surgical emergency at the same day. Economic meaning of cutting cost actions in each kind of human activity is important factor increasing empirical observed performance [8], [9]. From an economic point of view, this area provides opportunities to reduce operating costs of hospitals and also to reduce requested reimbursement from health insurance providers, but can also have a positive effect on the region, for example health insurance, which should be reflected in the form of faster re-engaging of patient in the labour force [11], [12]. Currently is DS in Slovakia and abroad increasingly regarded as a standard procedure scheduled, suitable not only for the patient himself and his family, but also for the health care providers (HCP) and the entire health care system. DS is suitable for doctors due to the high efficiency and rationalization of work, while intervention of day surgery may be performed only by top professionals with the capability to estimate the postoperative process. From global point of view, significant boom of DS intervention occurred in the period 1970 – 1980, especially in U.S., Canada, UK and Australia, when there was growth of number of DS units [7]. DS in Slovakia has started to be used until 1998, but significant progress in its development has not yet been recorded. Among the countries there are still significant differences in the use of DS [3], [4], [5]. As the main determinants of these differences can be noted the existence of various regulations and incentives in different countries, different financial compensations of DS, resistance to change during the introduction of new procedures of doctors and anaesthesiologists [13]. Development of DS is effective only if it is accompanied by complementary reduction of the number of beds [6]. The category of the number of beds is also linked to the process of hospitalization. This indicator in Slovakia has not been changed significantly in the long term period. Between years 2003 - 2013 for women accounted for approximately 600,000 hospitalizations per year, for men around the 450,000. Further development of DS in Slovakia associated with the planned develop-

[1] *Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia, vincent.soltes@tuke.sk*
[2] *Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia, beata.gavurova@tuke.sk*

ment of integrated health care centers under the auspices of Ministry of Health of the Slovak Republic (MoHSR), Health Policy Institute, as well as with the construction of new hospital complexes in Slovakia should lead to a greater progress in the development of DS. Trenčín region was approved by the MoHSR as a pilot region in the definition, preparation and planning for integrated reform of health care system. The main reasons for the selection of Trenčín region are the existing regional implementation capacities (at regional level), its position as a Slovak representative in the International Health Equity Project 2020, ownership of regional acute central hospitals by regional government and consolidated network of outpatient specialists in these hospitals. For this reason we focused in the analytical section on the analysis of DS intervention in Trenčín region (number of HCP performancing DS intervention in this region is 42, the number of daily beds is 252) and Košice region with the highest number of HPCs aimed at DS (78), and for this reason also the highest number of daily beds for patients of DS in Slovakia on 31th of December, 2014 (381 daily beds).

2. Analysis of DS

DS intervention is possible to sort by five characters on the basis of the annual statement J (MoHSR) 1-01 and the Bulletin of the MoHSR from the first of March 2006, section 9-16, part 23 - Vocational guidance MoHSR about DS intervention. It is the number of operated patients and from that the number the hospitalized patients after intervention (we can get by difference from it the number of outpatients after intervention), then the age of operated patients to the next day, divided into two groups: paediatric patients, "juniors" (age from 0 to maximum 18 years and 364 days) and adult patients, "seniors" (minimum age 19 years). Another feature is region, in which territory was DS intervention reported by the relevant unit realized. DS intervention is reported in every region of Slovakia. The fourth feature is the year in which was DS intervention realized. We had data for the four years, from 2009 to 2012. The fifth feature is a group of seven specialized fields under the MoHSR Bulletin. Each realized DS intervention belongs to one of them. Fields are: Surgery, Orthopaedics, Emergency surgery and Plastic surgery (next "Surgery"), Gynaecology and Obstetrics (next "Gynaecology"), Ophthalmology, Otorhinolaryngology (next "ORL"), Urology, Dentistry and Gastroenterology surgery and gastroenterology. DS intervention for the last two fields is reported at the minimum rate and therefore is not part of our further analyses.

2.1 Research methods

To test dependency of the proportion of hospitalized adults on the year we used cross tables, where we test dependency between pairs of characters. During testing we compare real absolute frequencies to expected absolute frequencies, which should be in the case of independence issues. The null hypothesis of the test is: two characters are unrelated, the alternative hypothesis is: two characters are related.

If the p-value of the test is greater than 0.05, we accept the null hypothesis (the two characters are unrelated). If the p-value is less than 0.05, then the null hypothesis is rejected (the two characters are related). This is standard procedure for small number of tests and small absolute frequencies in boxes of cross tables. To test relation between the rate of hospitalization and year we used the analysis of cross tables implemented in the statistical software SPSS version 19. During the test are compared real absolute frequencies in each cell of table with expected absolute frequencies, which should be in the case of independence characters. In the case of larger number of tests and larger absolute frequencies, it is appropriate to reinforce the critical p-value. We have chosen p-value of 0.001for the critical value of chi-square test. To determine the cause of the significance of relationship between two categorical characters we use, so called "Adjusted Residual" [1], which is defined by the formula:

$$AR_{ij} = \frac{f_{ij} - E_{ij}}{\sqrt{E_{ij}\left(1-\frac{r_i}{N}\right)\left(1-\frac{c_j}{N}\right)}}, \tag{1}$$

where:

f_{ij} is real frequency in the i-th row and the j-th column of table, (2)

$c_j = \sum_{i=1}^{R} f_{ij}$ is sum of the j-th column, $r_i = \sum_{j=1}^{C} f_{ij}$ is sum of the i-th row, (3)

$N = \sum_{i=1}^{R} r_i = \sum_{j=1}^{C} c_j$ is total sum of table, $E_{ij} = \frac{r_i c_j}{N}$ is expected frequency. (4)

AR indicates the amount of deviation between the actual and the expected frequency, as well as the direction of deviation, what was also our objective of analysis.

The standard condition for critical value of residue is | AR |> 2.0 (p <0.05) or | AR |> 2.6 (p <0.01). However, in our case, the number of operated patients in tables is large (several thousand), and also we carry many tests, which results in very high incidence of formal significant residues. Therefore, we consider as a significant difference between the real and expected frequency the value | AR |> 4.9 (two-side quantile of standard normal distribution for p-value 10^{-6}). In results we indicate the value of chi-square test, while the main indicator of significance is the value of residue. The research problem is to determine whether proportion of hospitalized adults depends on year for individual regions and fields. We test the hypothesis that characters: hospitalization of adult after surgery intervention (first character) and year in which intervention was realized (second character) are independent compared with the alternative hypothesis that the characters are dependent. We also assume that the presence of different types of DS intervention in each field is approximately the same in each year.

2.2 Results and evaluation

In the text we present due to the content and range limitations of contribution only two cross tables in Trenčín region (Table 1) and Košice region (Table 2), while in overall assessment of testing dependencies of proportion development of hospitalized adults on year, are included values of all 8 regions. Residues with value |AR|> 4.9 are in bold.

Table 1 Operated adults vs. years for specialized fields in Trenčín region

Field			Year				Total
			2009	2010	2011	2012	
Surgery	Outpatients	Count	2685	4443	5727	4726	17581
		AR	4,5	-5,5	21,2	-19,5	
	Inpatients	Count	302	**772**	**305**	**1162**	2541
		AR	-4,5	**5,5**	**-21,2**	**19,5**	
	Total	Count	2987	5215	6032	5888	20122
Gynaecology and Obstetrics	Outpatients	Count	603	1106	2195	2434	6338
		AR	2,5	-13,4	5,0	4,2	
	Inpatients	Count	0	**50**	**2**	7	59
		AR	-2,5	**13,4**	**-5,0**	-4,2	
	Total	Count	603	1156	2197	2441	6397
Ophthalmology	Outpatients	Count	888	1264	1806	2315	6273
		AR	0,6	0,7	-0,7	-0,4	
	Inpatients	Count	0	0	1	1	2
		AR	-0,6	-0,7	0,7	0,4	
	Total	Count	888	1264	1807	2316	6275
Otorhinolaryngology	Outpatients	Count	89	293	563	937	1882
		AR	-18,4	-2,5	1,2	10,8	
	Inpatients	Count	**76**	39	43	**11**	169
		AR	**18,4**	2,5	-1,2	**-10,8**	
	Total	Count	165	332	606	948	2051
Urology	Outpatients	Count	87	125	170	134	516
		AR	1,6	2,0	2,5	-5,9	
	Inpatients	Count	0	0	0	**13**	13
		AR	-1,6	-2,0	-2,5	**5,9**	
	Total	Count	87	125	170	147	529

Interpretation of test results for the Trenčín region:

About all fields except the field of Ophthalmology we can reliably say that the proportion of hospitalized elderly depends on the year in which the patient was operated. The proportions of seniors hospitalized at least in one of the four explored years (|AR|> 4.9) contributed to this fact.

Fields: Surgery (Chi-Sq = 622.504, P-Value < 0.001), **Gynaecology and Obstetrics** (Chi-Sq = 179.510, p-Value < 0.001), **Urology** (Chi-Sq = 34.633, P-Value < 0.001) and **Otolaryngology** (Chi-Sq = 381.445, p-Value < 0.001):

We can reliably say that the proportion of hospitalized adult patients is dependent on the year in which the patient was operated. Shares of hospitalized patients have significantly contributed to the argument of dependency, in the three years on surgery, in two years on gynaecology and in otorhinolaryngology. In the field of surgery in 2011 was the most significantly lower proportion of hospitalized adult patients (305 from 6032 operated), as we might expect in the case of independence. On the other hand, in 2010 (field of Gynaecology) and 2009 (field of Otorhinolaryngology) were recorded the most significant higher proportions of hospitalized adult patients, as we might expect in the case of the independence of explored signs. In the field of urology was significantly higher proportion of hospitalized adult patients in 2012. We have looked at the development and discrepancies in day surgery intervention in each year in the Košice region (Table 2).

Table 2 Operated adult patients versus years for specified fields in the Kosice region

Field				Rok				Total
				2009	2010	2011	2012	
Surgery		Outpatients	Count	2886	4442	5263	4525	17116
			AR	19,7	24,4	13,4	-49,4	
		Inpatients	Count	**107**	**197**	**606**	**2272**	3182
			AR	**-19,7**	**-24,4**	**-13,4**	**49,4**	
	Total		Count	2993	4639	5869	6797	20298
Gynaecology and Obstetrics		Outpatients	Count	4112	4515	6043	5984	20654
			AR	-14,3	31,6	10,8	-23,5	
		Inpatients	Count	**1679**	**232**	**1297**	**2655**	5863
			AR	**14,3**	**-31,6**	**-10,8**	**23,5**	
	Total		Count	5791	4747	7340	8639	26517
Ophthalmology		Outpatients	Count	4713	5348	6877	8114	25052
			AR	-23,8	7,0	4,9	9,3	
		Inpatients	Count	**161**	**0**	**20**	**0**	181
			AR	**23,8**	**-7,0**	**-4,9**	**-9,3**	
	Total		Count	4874	5348	6897	8114	25233
Otorhinolaryngology		Outpatients	Count	325	309	1184	1253	3071
			AR	8,8	-39,8	11,2	20,5	
		Inpatients	Count	**16**	**723**	**187**	**56**	982
			AR	**-8,8**	**39,8**	**-11,2**	**-20,5**	
	Total		Count	341	1032	1371	1309	4053
Urology		Outpatients	Count	131	617	643	989	2380
			AR	4,1	9,1	9,0	-17,2	
		Inpatients	Count	0	**5**	**8**	**271**	284
			AR	-4,1	**-9,1**	**-9,0**	**17,2**	
	Total		Count	131	622	651	1260	2664

Interpretation of test results for the Košice region:
With a high degree of reliability for each field we can state that the proportion of hospitalized adult patients is dependent on the year in which the operation was carried out. In all five specialized fields recorded shares of hospitalized adult patients in all four years contributed to this fact.
Fields: Surgery (Chi-Sq = 2537.196, P-Value < 0.001) and **Urology** (Chi-Sq = 295.551, P-Value < 0.001):
The most significant contribution to the argument of dependency was the year 2012, when 2,272 adult patients were hospitalized from 6797 operated on the surgery, and 271 from a total number of 1260 adult patients operated in the field of urology.
Fields: Gynaecology and Obstetrics (Chi-Sq = 1433.912, P-Value < 0.001) and **Otorhinolaryngology** (Chi-Sq = 1619.108, P-Value < 0.001):
The most significant contribution to the argument of dependency was the year 2010 when 232 adult patients were hospitalized from 4747 operated on the gynaecology, and 723 from 1032 adult patients operated in the field of Otorhinolaryngology.

Field: Ophthalmology (Chi-Sq = 572.599, P-Value < 0.001)
In 2009, there was the most significantly different (higher) share of inpatients (161 hospitalized adult patients from 4874 operated patients) as we might expect in the case of independency.
The analogous way was dealt in other regions. Because of the content and scope limitation of this paper, in its next part we state summary test results of all regions following defined aim of paper.

3 Assessment testing dependency of the proportion of hospitalized seniors on the year

Based on data of all regions and all specified fields (as listed in Table 3) can be reliably maintained that the proportion of hospitalized adult patients is dependent on year.

Table 3 Operated adult patients in all regions of SR and in all specified fields versus years

Patients \ Year	2009	2010	2011	2012	Total
Outpatients	58 911	73 515	107 862	120 656	360 944
Inpatients	5 908	5 580	10 461	24 893	46 842
Total	64 819	79 095	118 323	145 549	407 786
Share of inpatients	9,1%	7,1%	8,8%	17,1%	11,5%

Table 3 has showed that the average proportion of hospitalized adult patients in Slovakia for years 2009-2012 and fields is 11.5%. The share was lower in the years 2009-2011 and higher in 2012. The results of testing dependency of the proportion of hospitalized adult patients on year in the region and in the specialized field were compiled in Table 4. In this table are listed fields for each regions, about which we can reliably say that the proportion of hospitalized adult patients in the field depends on the year, and also individual years (in bold), when were recorded significantly higher (Year +) or significantly lower (Year -) contributions hospitalized adult patients, as we would expect in the case of independence of examined signs using confidence limits |AR|> 4.9 ($p < 10^{-6}$). Of all regions, only in Banská Bystrica region there are fields (Surgery and Urology) with the same types of significant deviation of shares: in 2012 an increased proportion of hospitalized adult patients, while between 2009 and 2011 reduced share. If we release the limit of significance to the condition |AR|> 3.3 ($p < 0.001$), we will have three other pairs of fields in three different regions (the default font). Košice region: Surgery and Urology (increased share in 2012, reduced share in 2009 -2011). Nitra region: Otorhinolaryngology and Urology (increased in 2012, decreased in 2010 and 2011). Žilina region: Ophthalmology and Urology (increased in 2012, decreased in 2009-2011). In the Table 5 is given a frequency of years according to the direction of the share deviation of the expected value.

Table 4 Mapping of the years on significantly increased, as well as on decreased shares of hospitalized adults in regions and specialized fields.

Region	Field	Year +	Year -	Region	Field	Year +	Year -
BC	Surg a	**2012**	**2009,** 2010**, 2011**	PV	Surg	**2010**	**2009**
	Gyn	**2012**	**2009, 2010, 2011**		Gyn	**2010**	**2009, 2011, 2012**
	ORL	**2009,** 2011**, 2012**	**2010**		ORL	**2012**	**2009, 2010, 2011**
	Urol a	**2012**	**2009,** 2010**, 2011**	TA	Surg	**2012**	**2009, 2010,** 2011
BL	Surg	**2011, 2012**	**2010**		Gyn	**2011, 2012**	**2009, 2010**
	Gyn	**2009, 2011**	**2010**		ORL	**2012**	**2011**
	Oph	**2009**	**2010, 2011, 2012**		Urol	**2012**	2009**, 2010**
	ORL	**2010**	**2009, 2012**	TC	Surg	**2010, 2012**	2009**, 2011**
	Urol	**2011**	2009, 2010, **2012**		Gyn	**2010**	**2011,** 2012
KI	Surg b	**2012**	**2009, 2010, 2011**		ORL	**2009**	**2012**
	Gyn	**2009, 2012**	**2010, 2011**		Urol	**2012**	-
	Oph	**2009**	**2010, 2011, 2012**	ZI	Surg b	**2012**	**2009, 2010, 2011**
	ORL	**2010**	**2009, 2011, 2012**		Oph	**2012**	2009, 2010, 2011
	Urol b	**2012**	2009, **2010, 2011**		ORL	**2012**	2009, **2010, 2011**
NI	Surg	**2012**	**2009, 2010, 2011**		Urol b	**2012**	-
	Gyn	**2010**	**2011, 2012**				
	ORL b	**2012**	**2010, 2011**				
	Oph b	**2012**	2010, **2011**				

a – fields with the same type of significant deviation of share of hospitalized adult patients in a given region and in years |AR| > 4,9 ($p < 10^{-6}$). b – fields with the same type of significant deviation of share of hospitalized adult patients in a given region and in years |AR| > 3,3 ($p < 0,001$).
Banská Bystrica (BC), Bratislava (BL), Košice (KI), Nitra (NI), Prešov (PV), Trnava (TA), Trenčín (TC), Žilina (ZI).

Table 5 Frequency of years according to the direction of the share deviation of hospitalized adult patients

Year	+	-	Share +
2012	22	8	73%
2009	6	13	32%
2010	7	22	17%
2011	4	20	17%

The highest share of significantly higher proportion of hospitalized day surgery adult patients (73%) for the five analyzed fields of day surgery (Surgery, Gynaecology and Obstetrics, Ophthalmology, Otolaryngology and Urology) and for the period 2009-2012 was in 2012. For the following year 2009, share was 32%. For the years 2010 and 2011, the proportion was the same 17%.

Conclusion

Day surgery in Slovakia has been insufficient regulated for a long time. It was implemented in the process of increasing effectiveness of the health system in Slovakia, but some important existential links have still been missing. Despite of several-years legislative support by the MoHSR, the proportion of day surgery compared to other countries is very low. Day surgery development process is needed to be analyzed in view of the three pro-

cesses: medical, economic and social. Their perfect knowledge will enable us to understand the importance of interaction development process with parameters of access to health care in Slovakia, quality, economic cost-effectiveness, efficiency and equality. Setting up an active health policy leading to the positive progress in the development DS in our country requires searching compromises between the declared parameters, what is very problematic. We have revealed by our analysis unequal development between the number of completed day surgery intervention in different regions in a group of adults and revealed differences were specified. It is very difficult to propose compelling reasons for this situation. One of them is meaningful usage own reporting methodologies in the health facilities induced by current pricing strategies of HIP. This creates the need for implementation of multidimensional analysis focusing on the support of DS intervention and development their regional disparities, as well as on the detection risk groups of patients for the correct setting concrete type of intervention. This also includes setting up clearly defined requirements for personnel and technical equipment of the workplaces, while not only in surgery for adults but also in children's surgery can be performed high proportion of common, and typical for children, surgical procedures under DS. These facts have been highly relevant just in the current process of integrated health care centres in the context of increasing effectiveness of Slovak health system.

Acknowledgment
This work was supported by the VEGA Project No. 1/0929/14 "*Multidimensional economic and financial evaluation of the implementation process and the use of one day health care and quantification of the financial impact on the health care system in the Slovak Republic".*

References

[1] Agresti, A. *Categorical Data Analysis* (2nd Ed.). New York: John Wiley & Sons, Inc., 2002.
[2] Bem, A. and Michalski, G. The financial health of hospitals. V4 countries case in: *Sociálna ekonomika a vzdelávanie*. Ľapinová, Gubalová (ed.), 2014, Univerzity Mateja Bela v Banskej Bystrici, 2014, 1-9.
[3] Buzink, S., Šoltés, M., Radoňak, J., Fingerhut, A., Hanna, G. and Jakimowicz, J. Laparoscopic Surgical Skills Programme: Preliminary Evaluation of Grade I Level 1 Courses by Trainees. Videosurgery and other Miniinvasive Techniques, 2012, **7**(3)188 – 192.
[4] Castells, X., Alonso, J., Castilla, M. et al. Outcomes and Costs of Outpatient and Inpatient Cataract Surgery: A Randomised Clinical Trial. J Clin Epidemiol, **54(**1), 23 – 29.
[5] Castoro, C., Bertinato, L., Baccaglini, U., Drace, C.A. and McKee, M. Policy Brief. Day Surgery: Making it Happen. Brussels: WHO European Centre for Health Policy, 2007.
[6] Gavurová, B., Šoltés, V., Kafková, K., Černý, Ľ. *Vybrané aspekty efektívnosti slovenského zdravotníctva. Jednodňová zdravotná starostlivosť a jej rozvoj v podmienkach Slovenskej republiky*. Košice: EkF TUKE, 2013, 275 p.
[7] Jarrett, P., Staniszewski, A. The Development of Ambulatory Surgery and Future Challenges. In: Lemos, P. Jarrett, P. a Philip, B. (eds): Day Surgery Development and Practice. London, UK: International Association for Ambulatory Surgery (IAAS), 2006, 21 – 34.
[8] Michalski, G. Full operating cycle influence on food and beverages processing firms characteristics. In: *Agricultural Economics - Zemědělská ekonomika*, 2015, **61**.
[9] Michalski, G. Value maximizing corporate current assets and cash management in relation to risk sensitivity: Polish firms case, *Economic Computation and Economic Cybernetics Studies and Research*. 2014, **48**(1), 259-276.
[10] Michalski, G. The Efficiency of Liquid Assets Investment: Taxed and Not Taxed Healthcare Organizations Case. Wroclaw University of Economics Scientific Mongraphs: Research Papers of Wroclaw University of Economics, 2012, No. 226, Wroclaw. http://dx.doi.org/10.2139/ssrn.2143698
[11] Raisova, M., Buleca, J. and Michalski, G. Food processing firms inventory levels in hard times. 2004-2012 Slovak, Czech and Polish enterprises case. In: *Procedia Economics and Finance*, Elsevier, 2014, **12**, 557-564, DOI:10.1016/S2212-5671(14)00379-7, .
[12] Szczygiel, N., Rutkowska-Podolowska, M. and Michalski, G. Information and Communication Technologies in Healthcare: Still Innovation or Reality? Innovative and Entrepreneurial Value - creating Approach in Healthcare Management, In: *5th Central European Conference in Regional Science Conference Proceedings* / Nijkamp Pater (ed.), Technical University of Košice, 2015, 1020-1029.
[13] Šoltés, M. and Radoňak, J.: Hodnota demografických a anamnestických údajov pre predpoveď náročnosti elektívnej laparoskopickej cholecystektómie. In: *Slovenská chirurgia:* časopis Slovenskej chirurgickej spoločnosti, 2012, **9**(1), 23-26.

Linguistic approximation using fuzzy 2-tuples in investment decision making

Tomáš Talášek[1], Jan Stoklasa [2], Jana Talašová[3]

Abstract. The paper focuses on linguistic approximation methods for the outputs of mathematical decision support models in the context of investment decision making. Since the decision makers in this context are frequently laymen in mathematics, the ability of the models to provide understandable and easily interpretable results is of great importance. We explore the possibilities of using the fuzzy 2-tuples concept introduced by Herrera et al. [2] for linguistic approximation and propose a method of linguistic approximation of fuzzy-number-type outputs suitable for the use in investment decision support. The performance of the proposed method is discussed on a practical example of mutual fund selection.

Keywords: Linguistic approximation, linguistic modelling, multiple criteria decision making, decision support, fuzzy 2-tuples, mutual funds.

JEL classification: C44
AMS classification: 90B50, 91B06, 91B74

1 Introduction

In linguistic modelling it is necessary to be able to assign appropriate labels (known values of a linguistic variable) to general fuzzy sets. In practical applications it is often required to obtain fuzzy numbers as outputs from mathematical models, subnormal or multimodal fuzzy sets on $\mathbb{R}$ (normal and unimodal fuzzy sets) are difficult to interpret and use by practitioners. This paper therefore concentrates on assigning linguistic labels to fuzzy numbers. This process is called *linguistic approximation*. From the mathematical point of view linguistic approximation is a mapping from a given class of fuzzy numbers on $\mathbb{R}$ to a set of linguistic values (labels) of a linguistic variable $\mathcal{V}$ whose mathematical meanings are modelled by fuzzy numbers on $\mathbb{R}$. In this paper we propose a new method for the linguistic approximation of fuzzy-number-evaluations that is based on the idea of fuzzy 2-tuples [2]. We showcase the proposed method on the outputs of a multi-stage decision support model for investment decision making proposed in [6] and further elaborated in [7].

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0,1]$. For each $x \in U$ the value $A(x)$ is called a *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an *α-cut* of A for any $\alpha \in [0,1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0,1]$ (this implies A is *unimodal*); (3) $\mathrm{Supp}(A)$ is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if $\mathrm{Supp}(A)$ is a subset of an interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$. Each fuzzy number A is determined by

[1]Palacký University, Olomouc, Faculty of Science, Dept. of Math. Analysis and Applications of Mathematics, 17. listopadu 1192/12, 77146, Olomouc, Czech Republic, tomas.talasek@upol.cz
[2]dtto, jan.stoklasa@upol.cz
[3]dtto, jana.talasova@upol.cz

$A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall\alpha \in (0,1]$, and the closure of the support of A $\mathrm{Cl}(\mathrm{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. An *intersection* of two fuzzy sets A and B is a fuzzy set $(A\cap B)$ defined as follows: $(A\cap B)(x) = \min\{A(x), B(x)\}$, $\forall x \in U$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. If $A \in \mathcal{F}_N(U)$ is a linear fuzzy number and c is a real number, then $A + c = (a_1 + c, a_2 + c, a_3 + c, a_4 + c)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [1].

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [4]) of the interval $[a, b]$, i.e. for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([8]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X.

We are frequently required to be able to represent fuzzy sets by real numbers, this procedure is called *defuzzification*. In applications an approximation of a fuzzy number A by its center of gravity (COG) t_A is frequently used. The *center of gravity* of a fuzzy number A defined on $[a, b]$, except for fuzzy numbers where $a_1 = a_2 = a_3 = a_4$ (so called *fuzzy singletons*), is defined by the formula $t_A = \int_a^b A(x)x\,dx / \int_a^b A(x)dx$. The center of gravity of a fuzzy singleton is defined as $COG_A = a_1$. Other possible defuzzification methods are discussed in [3]. A *distance* of fuzzy numbers $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}, B = \left\{[\underline{b}(\alpha), \overline{b}(\alpha)]\right\}_{\alpha\in[0,1]}$ can be defined by the formula $d(A, B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\,d\alpha$ (see e.g. [5] for other possible approaches).

3 Proposed method for linguistic approximation

Let us consider an evaluation scale $[a, b]$ and a mathematical model, the outputs of which are triangular fuzzy numbers $E_j \in \mathcal{F}_N([a, b])$, $j = 1, \ldots, m$ (more general types of fuzzy numbers can also be considered, but within this paper we restrict our investigation to triangular ones). These fuzzy numbers represent evaluations of some objects (alternatives). Such outputs can be obtained e.g. in situations where each expert provides a fuzzy evaluation in the form a triangular fuzzy number (all the experts are using the same evaluation scale) and an overall evaluation is computed by a fuzzy weighted average of these expert evaluations. Our aim is to propose a linguistic approximation method for these evaluations that would be understandable to the users of the outputs (decision makers/evaluators) and at the same time would provide more specific information than commonly used linguistic approximation methods (see e.g. [5]).

First we define a linguistic variable $(Eval, \mathcal{T}(Eval), [a, b], G, M)$. The set $\mathcal{T}_B(Eval) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\} \subset \mathcal{T}(Eval)$ constitutes a *basic term set* of $Eval$, consisting of all the terms the decision maker wants to use for evaluation purposes. $M(\mathcal{T}_i)$ is a triangular fuzzy number on $[a, b]$, for all $i = 1, \ldots, n$. The elements of the basic term set must be well understood by the decision maker and their fuzzy number meanings $M(\mathcal{T}_i) = T_i \in \mathcal{F}_N([a, b])$ for all $i = 1, \ldots, n$ specified in cooperation with him/her. For the purposes of this paper we will suppose that the fuzzy evaluations we need to linguistically approximate are not significantly more uncertain then the meanings of the elements of the basic terms set. Let us also consider that $T_1, \ldots, T_n$ form a uniform Ruspini fuzzy partition of $[a, b]$:

$$\begin{aligned} M(\mathcal{T}_1) &= (a, a, a + \Delta), \\ M(\mathcal{T}_i) &= (a + (i-2)\cdot\Delta, a + (i-1)\cdot\Delta, a + i\cdot\Delta),\ i = 2, \ldots, n-1, \\ M(\mathcal{T}_n) &= (b - \Delta, b, b), \end{aligned} \tag{1}$$

where $\Delta = (b - a)/(n - 1)$.

A fuzzy set representing feasible evaluations on $\mathbb{R}$ is defined as

$$FE(x) = \begin{cases} 1 & \text{if } x \in [a, b], \\ 0 & \text{if } x \notin [a, b]. \end{cases} \tag{2}$$

Non-uniform partitions and partitions by non-triangular fuzzy numbers will be the subject of further study and hence are left out of the scope of this paper.

Since the basic term set of *Eval* contains only n elements, we might not be able to find a linguistic approximation of some of the possible fuzzy evaluations that would fit well enough (using just these n linguistic terms). We are however restricted to the use of the elements of the basic term set, as the linguistic terms it contains are the only ones that the decision maker clearly understands. To resolve this issue, we suggest to use the concept of fuzzy 2-tuples (which in this context translates into shifting the meanings of the basic terms to either side within the specified universe) and thus introduce the following syntactic rule to describe the results of such shifts and thus add *derived terms* to $\mathcal{T}_B(Eval)$:

$$\mathcal{T}(Eval) = \mathcal{T}_B(Eval) \cup \{\mathcal{T}_i^{\beta};\, \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta),\, i = 1, \ldots, n\},$$

where $\mathcal{T}_i^{\beta}$ is "$\mathcal{T}_i$ *shifted by* β" (see Figure 1; the size of the shift can be described linguistically as well; it will be shown in the numerical example). $\mathcal{T}_i^{\beta}$ can be expressed as a 2-tuple $(\mathcal{T}_i, \beta)$, where $\mathcal{T}_i$, $i = 1, \ldots, n$ is an elementary linguistic term and $\beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta)$ expresses the shift of $\mathcal{T}_i$, the sign of β indicates the direction of the shift. For $\beta = 0$, $\mathcal{T}_i^{\beta}$ coincides with $\mathcal{T}_i$, $i = 1, \ldots n$.

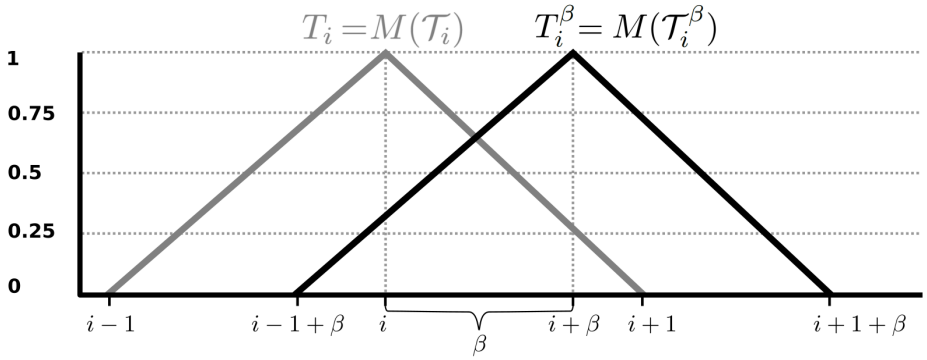


Figure 1 Shifting the meaning of a basic term set element $\mathcal{T}_i$ by β

To define a syntactic rule, we extend the universe $[a, b]$ to $[a - \Delta, b + \Delta]$ (this is a technicality to simplify the notation) and extend the meaning of $\mathcal{T}_1$ and $\mathcal{T}_n$ from T_1 and T_n to T_1' and T_n' respectively in the following way (note that the meanings of $\mathcal{T}_1$ and $\mathcal{T}_n$ remain unchanged on $[a, b]$).

$$\begin{aligned} T_1' &= (a - 0.5 \cdot \Delta, a - 0.5 \cdot \Delta, a, a + \Delta), \\ T_n' &= (b - \Delta, b, b + 0.5 \cdot \Delta, b + 0.5 \cdot \Delta), \end{aligned} \tag{3}$$

and for $\mathcal{T}_2, \ldots, \mathcal{T}_{n-1}$ we leave the meaning unchanged, that is

$$T_i' = T_i,\ i = 2, \ldots, n - 1. \tag{4}$$

The meanings of the elements of the derived term set are now computed using the formula

$$M(\mathcal{T}_i^{\beta}) = (T_i' + \beta) \cap FE,\ i = 1, \ldots, n,\ \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta). \tag{5}$$

It is however not reasonable to move T_1 to the left and T_n to the right, hence for $\mathcal{T}_1$ we will consider $\beta \in [0, 0.5 \cdot \Delta)$ and for $\mathcal{T}_n$ we will consider $\beta \in [-0.5 \cdot \Delta, 0]$.

The linguistic approximation $\mathcal{T}^* \in \mathcal{T}(Eval)$ of a fuzzy evaluation $E_j \in \mathcal{F}_N([a, b])$ is computed by

$$M(\mathcal{T}_{i^*}^{\beta^*}) = \underset{\substack{i=1,\ldots,n \\ \beta \in [-0.5 \cdot \Delta, 0.5 \cdot \Delta)}}{\arg\min} d(E_j, T_i^{\beta}). \tag{6}$$

The result of this novel linguistic approximation method is a fuzzy 2-tuple $(\mathcal{T}_{i^*}, \beta^*)$. In case there are more solutions $(\mathcal{T}_{i^*}, \beta^*)$ to (6), all such fuzzy 2-tuples are presented to the decision maker. To obtain a fully linguistic description of the evaluation, β can be interpreted (described linguistically) using e.g. Table 1. For example $(Good, +0.1)$ translates into *slightly better than good*.

Negative β value	Linguistic description	Positive β value	Linguistic description
$[-0.05 \cdot \Delta, 0 \cdot \Delta)$	About	$(0 \cdot \Delta, 0.05 \cdot \Delta]$	About
$[-0.2 \cdot \Delta, -0.05 \cdot \Delta)$	Slightly worse than	$(0.05 \cdot \Delta, 0.2 \cdot \Delta]$	Slightly better than
$[-0.35 \cdot \Delta, -0.2 \cdot \Delta)$	Worse than	$(0.2 \cdot \Delta, 0.35 \cdot \Delta]$	Better than
$[-0.5 \cdot \Delta, -0.35 \cdot \Delta)$	Noticeably worse than	$(0.35 \cdot \Delta, 0.5 \cdot \Delta)$	Noticeably better than

Table 1 Linguistic labels for the interpretation of values of β.

4 Short example of the proposed linguistic approximation method

Let us consider a multiple-criteria decision support system proposed in [6]. The multi-stage model for mutual fund selection first assesses investor's investment aim and his/her investment horizon. In the next stage the risk profile of the investor is taken into account. Based on the information obtained in the first two stages an evaluation of each mutual fund under consideration is computed. This evaluation is obtained in form of a fuzzy number (triangular fuzzy numbers are used to approximate more complex outputs if necessary; see Figure 2). The best fit linguistic approximation of the fuzzy evaluations was performed. As a result the decision maker (investor) was provided with a linguistic description of the overall evaluation of each mutual fund and with a numerical value representing the center of gravity of the respective fuzzy evaluation. These results are summarized in Table 2. The linguistic approximation used in [6] assigns one of the five elements of the linguistic scale presented in Figure 3 (or a combination of these terms) to each overall fuzzy evaluation of a mutual fund. Much information is therefore lost in the process and the linguistic label may not fit well.

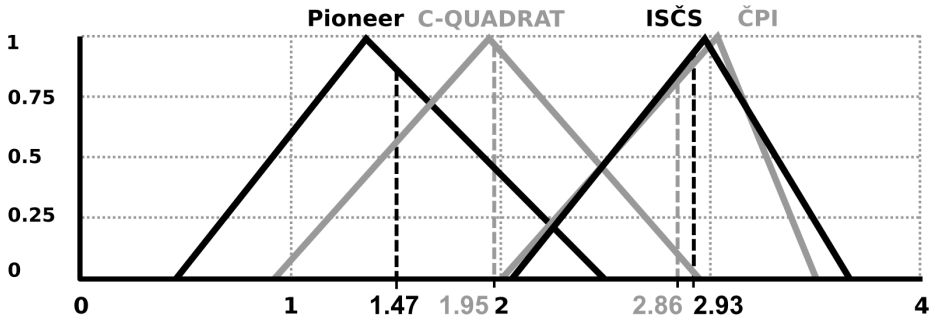


Figure 2 Overall evaluations of four mutual funds according to [6]

Alternatives	Overall evaluation	Output of mathematical model [6]		Proposed linguistic approximation	
		Linguistic approximation	COG	2-tuple	Linguistic approximation
Pioneer	(0.504, 1.404, 2.516)	Between Bad and Average	1.468	(Bad, +0.457)	Noticeably better than bad
ISČS	(2.116, 2.992, 3.656)	Good	2.932	(Good, -0.008)	About good
ČPI	(2.012, 3.036, 3.524)	Good	2.864	(Good, +0.014)	About good
C-QUADRAT	(0.928, 1.920, 2.960)	Average	1.948	(Average, -0.073)	Slightly worse than average

Table 2 Evaluations of four mutual funds. Results provided by [6] compared with the results obtained by the linguistic approximation method proposed in this paper.

Let us now approach the fuzzy evaluations obtained in [6] presented in Table 2 with the linguistic approximation method proposed in this paper. We can depart from the original evaluation scale used in [6], that is $(Eval_{ex}, \{Very\,bad, Bad, Average, Good, Excellent\}, [0, 4], G, M)$, where the meanings of the elements of the basic term set are summarized in Figure 3. Using the distance defined in preliminaries we obtain the linguistic approximations summarized in Table 2.

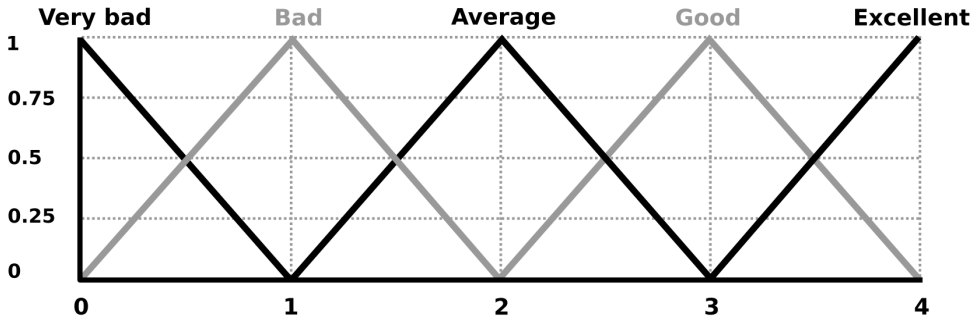


Figure 3 Meanings of the basic linguistic terms of $Eval_{ex}$ for $n = 5$ on the evaluation universe $[0, 4]$. These meanings correspond with the meanings of the values of the linguistic scale used in [6].

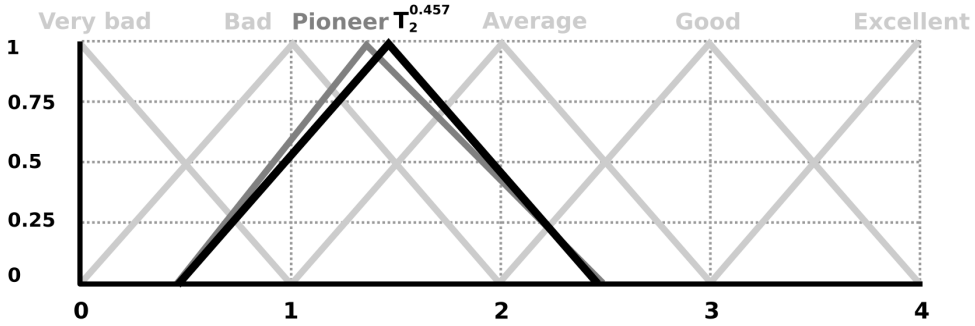


Figure 4 Illustration of the Pioneer mutual fund evaluation – its linguistic approximation by the fuzzy 2-tuple based method.

5 Discussion and conclusion

A novel linguistic approximation method for fuzzy numbers is presented in this paper. The method draws from the fuzzy 2-tuple concept, which allows it to provide a combination of linguistic term and its necessary shift (expressed numerically or linguistically). Using an initial linguistic terms set specified by the decision maker (the fuzzy number meanings of the linguistic terms are defined in cooperation with decision maker) it constructs a linguistic variable with substantially extended term set. The grammar and syntactic rule necessary to construct this linguistic variable are introduced in the paper. The linguistic approximation itself then uses a distance of fuzzy numbers to find the linguistic term that is closest to the approximated fuzzy number. The method performs particularly well when the approximated fuzzy numbers are of a similar shape and uncertainty as the meanings of the basic terms set. When the basic linguistic terms set is linearly ordered, the fuzzy 2-tuple linguistic approximation also provides means for ordering of the approximated fuzzy evaluations. This feature can prove useful both in decision making and evaluation (rankings of alternatives etc.). A short numerical example from the area of financial decision making showcasing the proposed method is also presented.

The example compares a classic "best-fit" approach to the linguistic approximation fuzzy outputs of mathematical models with the proposed fuzzy 2-tuple based method. We can clearly see, that although the same basic term set (with the same meanings) and the same distance was used for linguistic approximation in [6] and here, the linguistic approximations obtained by the methodology proposed in this paper provide more insights in the evaluations. The fuzzy 2-tuple based approximation operates with the basic terms well understood by the decision maker and uses small shifts of their meanings (numerically quantified; linguistic labelling of these shifts is also possible) to reflect differences between the fuzzy evaluations and the predefined meanings of the basic terms.

If we focus on the *Pioneer* fund from the example, we can see, that Talášek et al. [6] suggest a linguistic approximation involving two neighbouring basic linguistic terms – in fact it is suggested that the appropriate linguistic approximation lies somewhere in between these two basic terms. No indication is suggested as to which of the original basic terms is closer (consider that the basic terms are those that are well understood by the decision maker). The approach proposed in this paper provides a similar result. It however uses a single basic linguistic term and describes its necessary modification (shift of meaning). The decision maker is thus provided with a single linguistic label (which is similar to classic approaches to linguistic approximation) and provides an additional piece of information regarding the difference of the approximated fuzzy set and the meaning of the basic term which can be used if needed. The output of the proposed fuzzy 2-tuple based approximation is presented in Figure 4. The overall evaluation of the mutual fund Pioneer is depicted together with its linguistic approximation expressed by the fuzzy number $T_2^{0.457}$. The most promising feature of the proposed method is the use of a single well understood linguistic label along with a simple linguistic or numerical modifier of its meaning.

The effect of different distances of fuzzy numbers, different shapes of membership functions and different types of scales on the usefulness of the proposed method in practical applications, as well as the use of the proposed methof for the ranking of alternatives will be the subject of further study. Also empirical research concerning the understandability and intuitiveness of fuzzy 2-tuple based outputs for practitioners is planned.

Acknowledgements

Supported by the grant IGA PrF 2015 013 Mathematical models of the Internal Grant Agency of Palacký University, Olomouc and by the grant GA14-02424S Methods of operations research for decision support under uncertainty of the Grant Agency of the Czech Republic.

References

[1] Dubois, D., Prade, H.: *Fuzzy sets and systems: theory and applications*, Academic Press, 1980.

[2] Herrera, F., Martinez, L.: An approach for combining linguistic and numerical information based on the 2-tuple fuzzy linguistic representation model in decision-making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **8**(5), (2000), 539–562.

[3] Leekwijck, W. V., Kerre, E. E.: Defuzzification: criteria and classification. *Fuzzy sets and systems*, **108**(2) (1999), 159–178.

[4] Ruspini, E.: A New Approach to Clustering. *Inform. Control*, **15**, (1969), 22–32.

[5] Stoklasa, J.: *Linguistic models for decision support.* Lappeenranta: Lappeenranta University of Technology, 2014.

[6] Talášek, T., Bohanesová, E., Stoklasa, J., Talašová, J. Investment decision making using fuzzy scorecards–mutual funds selection, *Proceedings of the 32nd International Conference on Mathematical Methods in Economics 2014, September 2014*, Palacký University, Olomouc, 1021–1026.

[7] Talášek, T., Bohanesová, E., Talašová, J.: Two–step method for multiple criteria investment decision making: mutual funds selection using FuzzME software. *International Journal of Mathematics in Operational Research*, (accepted, to appear).

[8] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning I, II, III. *Inf. Sci.*, **8**, (1975), 199–257, 301–357, **9**, (1975), 43–80.

The Pros and Cons of Extending a DSGE Model with More Elaborated Foreign Blocks

Jaromír Tonner[1], Osvald Vašíček[2]

Abstract. This contribution aims at discussing various ways how foreign block can be modelled in a small open economy DSGE model (ranging from a reduced form VAR model to a QPM-style model of the foreign economy). These extensions are assessed mainly on ex-post forecasting performance. We find out that a QPM-style model can bring an improvement of predictive ability up to 20 percent. The necessary assumption for achieving it (for the Czech case) is, however, the autoregressive parameter of foreign interest rates very close to one what indicates model misspecification. Another and a very simple way how the predictive power of a DSGE model can be improved is a correct setting of parameter describing 'forward-lookiness' of the UIP condition.

Keywords: dsge, foreign block, UIP condition, predictive ability

JEL classification: C44
AMS classification: 90C15

1 Motivation

The DSGE model[1] is nowadays a powerful tool for a monetary policy analysis and forecasting at many central banks. Its strength lies not only in achieving these two purposes, but above all it is an apparatus for generating sensitivity and alternative scenarios. This brings a better intuition about risks around a baseline scenario. The main risks for small open economies are nowadays from abroad, as modellers say 'from the rest of the world'. From this point of view it is crucial how sensitivity scenarios concerning development in the rest of the world are generated, and in turn how consistent they are inside. Generally, there are two ways in which the consistent development of foreign variables can be incorporated in the forecast for a small open economy. The first is to generate an outlook for foreign variables outside the model. The second is to implement an internally consistent foreign block within a DSGE model. This work should examine the pros and cons of particular approaches to the extension of the g3 model with more elaborated foreign blocks (ranging from a reduced form VAR model to a QPM-style model of the foreign economy[2]).

The current g3 model (see [1] for details) contains the foreign block as a diagonal VAR model (only the diagonal elements are not zero). Foreign variables are fixed on a forecasting horizon as fully anticipated. In such a setting there is no guarantee of consistency in foreign variable outlooks, so consistent foreign scenarios must be generated outside the g3 model; usually they are generated by a NiGEM model (http://nimodel.niesr.ac.uk).

The following figure shows the above-mentioned inconsistency between the outlook for foreign GDP and foreign prices in the EU on the one side and the market outlook for foreign interest rates on the other side (outlook from the 3rd inflation report of 2015). This discrepancy is especially notable in the outlook for the year 2017. Foreign GDP will have returned to a steady state and foreign inflation is expected to be at steady state levels at the beginning of 2016. However, the market outlook for foreign interest rates is expected to stay at very low levels till the end of 2017. It seems improbable that the ECB will not increase interest rates further in 2017 when GDP and inflation have reached targeted values.

[1]Czech National Bank, Monetary Department, Na Příkopě 28, Prague, Jaromir.Tonner@cnb.cz
[2]Masaryk University, Deparment of Economics, Lipová 41a, Brno, osvald@econ.muni.cz

[1]All experiments in this article are carried out within the current Czech National Bank model, which is called g3 model. See [1] for details about the g3 model.

[2]More complex structures than QPM would not bring any improvement with regard to the fact that models are then linearized and eventually result in a VAR model

If we fixed such an outlook in a small open economy model, it would result in dramatic changes in domestic interest rates at the beginning of the forecast, i.e. in 2015, because foreign variables are assumed to be fully anticipated. However such a recommendation would be rather implausible for policy makers in a small open economy, because uncertainty about foreign variables in 2017 is very high.

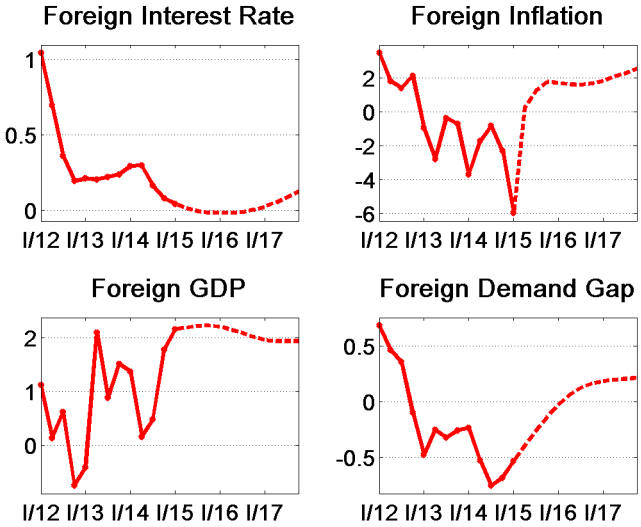


Figure 1 Inconsistency of Foreign Variables Outlook

2 Foreign Block Models

There are basically two ways how the foreign economy can be represented in a small DSGE model: simple VAR model or QPM-style model. In the rest of the text the following holds: n_t^* is a foreign demand, $\tilde{\pi}_t^*$ is a foreign inflation, i_t^* is a foreign nominal interest rate, r_t^* is a foreign real interest rate, $\dot{n}_t^*$ is the foreign demand growth, $\dot{Z}_t$ is an implied aggregate technology, $a\dot{X}_t$ is an export specific technology, $a\dot{O}_t$ is an openness technology, $a\dot{Q}_t$ is a change in quality, $\dot{s}_t$ is a change in nominal exchange rate, $prem_t$ is a debt elastic premium and b_t is a scaled net foreign assets. Absence of index t means the steady state value of a given variable. ϵ denotes a corresponding shock.

2.1 Diagonal VAR Model

The current g3 model contains the foreign block as a diagonal VAR model (only diagonal elements are not zero). That is why consistent foreign scenarios must be generated outside the g3 model.

$$\begin{bmatrix} \log(\frac{n_t^*}{n^*}) \\ \log(\frac{\tilde{\pi}_t^*}{\tilde{\pi}^*}) \\ \log(\frac{i_t^*}{i^*}) \end{bmatrix} = \begin{bmatrix} 0.75 & 0 & 0 \\ 0 & 0.30 & 0 \\ 0 & 0 & 0.80 \end{bmatrix} \begin{bmatrix} \log(\frac{n_{t-1}^*}{n^*}) \\ \log(\frac{\tilde{\pi}_{t-1}^*}{\tilde{\pi}^*}) \\ \log(\frac{i_{t-1}^*}{i^*}) \end{bmatrix} + \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \begin{bmatrix} \epsilon^{n^*} \\ \epsilon^{\tilde{\pi}^*} \\ \epsilon^{i^*} \end{bmatrix}$$

$$\dot{n}_t^* = \frac{n_t^*}{n_{t-1}^*} \frac{\dot{Z}_t a\dot{X}_t a\dot{O}_t}{a\dot{Q}_t}$$

Note here that foreign demand (which is observed at levels and not at growths) must be stationarized concretely by four technologies (aggregate, export-specific, openness and quality). The aggregate technology plays a prominent role because it stationarized the whole model.

2.2 VAR Model

The general VAR model is suggested in [2] and is denoted as *CTW* in the text. This modification is a generalization of the diagonal VAR model. It permits the co-movements of foreign variables and thus the correct setting of the foreign block's parameters (mainly standard deviations of shocks) is required.

$$\begin{bmatrix} \log(\frac{n_t^*}{n^*}) \\ \log(\frac{\tilde{\pi}_t^*}{\tilde{\pi}^*}) \\ \log(\frac{r_t^*}{r^*}) \end{bmatrix} = \begin{bmatrix} 0.75 & -0.00 & -0.00 & 0.00 \\ 0.10 & 0.30 & -0.10 & 0.00 \\ 0.00 & -0.00 & 0.60 & 0.00 \end{bmatrix} \begin{bmatrix} \log(\frac{n_{t-1}^*}{n^*}) \\ \log(\frac{\tilde{\pi}_{t-1}^*}{\tilde{\pi}^*}) \\ \log(\frac{r_{t-1}^*}{r^*}) \\ \log(\frac{\dot{Z}_t}{\dot{Z}}) \end{bmatrix} + \begin{bmatrix} 0.01 & 0 & 0 \\ 0.00 & 0.01 & 0 \\ 0.00 & 0.00 & 0.01 \end{bmatrix} \begin{bmatrix} \epsilon^{n^*} \\ \epsilon^{\tilde{\pi}^*} \\ \epsilon^{r^*} \end{bmatrix}$$

$$r_t^* = \frac{i_t^*}{\tilde{\pi}_{t+1}^*},\ \dot{n}_t^* = \frac{n_t^*}{n_{t-1}^*} \frac{\dot{Z}_t a \dot{X}_t a \dot{O}_t}{a\dot{Q}_t}$$

2.3 QPM Model

A simple QPM-style gap model represents the external economy[3]. The foreign demand and foreign inflation equations contain forward-looking components and the interest rate equation is a simple Taylor rule. We have already carried out an initial estimation in DYNARE, explored the model properties and set a calibration.

$$\log(\hat{n}_t^*) = (1-\rho_{\hat{n}^*}) \log(\hat{n}_{t+1}^*) + \rho_{\hat{n}^*} \log(\hat{n}_{t-1}^*) - \rho_{\hat{r}^*} \log(\hat{r}_{t-1}^*) + \epsilon_t^{\hat{n}^*}$$

$$\log(\tilde{\pi}_t^*) = (1-\rho_{\tilde{\pi}^*}) \log(\tilde{\pi}_{t+1}^*) + \rho_{\tilde{\pi}^*} \log(\tilde{\pi}_{t-1}^*) + \rho_{2,\hat{n}^*} \log(\hat{n}_t^*) + \epsilon_t^{\tilde{\pi}^*}$$

$$\log(i_t^*) = \rho_{i^*} \log(i_{t-1}^*) + (1-\rho_{i^*}) \log\left(\tilde{\pi}^* r^* \left(\frac{\tilde{\pi}_{4,t+4}^*}{(\tilde{\pi}^*)^4} \right)^{mp^{\text{infl}}} (\hat{n}_t^*)^{mp^{\text{o-gap}}} \right) + \epsilon_t^{i^*}$$

$$\hat{n}_t^* = \frac{n_t^*}{n^*},\ \tilde{\pi}_t^{4*} = \tilde{\pi}_t^* \tilde{\pi}_{t-1}^* \tilde{\pi}_{t-2}^* \tilde{\pi}_{t-3}^*,\ \hat{r}_t^* = \frac{r_t^*}{r^*},\ r_t^* = \frac{i_t^*}{\tilde{\pi}_{t+1}^*},\ \dot{n}_t^* = \frac{n_t^*}{n_{t-1}^*} \frac{\dot{Z}_t a \dot{X}_t a \dot{O}_t}{a\dot{Q}_t}$$

What we find interesting and will discuss is the way in which foreign demand is linked to the domestic economy. Foreign demand is observed at levels (not at gaps) and the outlook for foreign demand is also fixed at levels by a foreign demand shock $\epsilon_t^{\hat{n}^*}$. However this choice is not exclusive, there are other approaches, e.g. foreign demand can be observed at gaps, and outlook can be fixed at gaps. In this case, however, there is one more degree of freedom. It relies on the construction of the foreign demand gap (e.g. a choice of lambda for a Hodrick-Prescott filter). The outlook for foreign demand would no longer be fixed at its Consensus Forecast levels, but at computed gaps. Results are discussed in the section 3.

ρ_{i^*}	0.992	$\sigma_{\epsilon^{i^*}}$	0.0019
$\rho_{\hat{n}^*}$	0.499	$\sigma_{\epsilon^{\hat{n}^*}}$	0.0177
$\rho_{\tilde{\pi}^*}$	0.262	$\sigma_{\epsilon^{\tilde{\pi}^*}}$	0.0056
$\rho_{\hat{r}^*}$	0.112	$mp^{\text{o-gap}}$	2.597
$\rho_{2,\hat{n}^*}$	0.019	mp^{infl}	0.151

Table 1 Model's Parameters

2.4 Forward-lookiness of the UIP Condition

One of the key equation in a small open DSGE model is the uncovered interest rate parity (UIP) condition. In the g3 model it takes a form of

$$i_t = i_t^* \dot{s}_{t+1}^{\rho_s} \dot{s}_{t-1}^{-(1-\rho_s)} \dot{s}^{2(1-\rho_s)} e^{\epsilon_t^{uip}} prem_t$$

$$\log(prem_t) = \rho_{prem} \log(prem_{t-1}) + (1-\rho_{prem}) \log(1) + \epsilon_t^{prem} - \zeta b_t$$

[3] See [3] for details.

where parameter ρ_s plays a role as 'forward-lookiness' of the UIP condition. How section 3 shows, it has a crucial impact on a predictive ability of the g3 model (model change is denoted as *rho*).

3 Predictive Power

In this subsection we evaluate the prediction errors of in-sample simulations while at the same time considering the extent of model change. The table reports the RMSE of eight-step prediction of main variables, such as policy rate (IR), CPI inflation (CPI), nominal exchange rate (ER), real GDP (GDP), nominal wage growth (W), real consumption (C), real investments (I), real exports (X), real imports (N), and their deflators (PC, PJ, PX and PN). The numbers in the table correspond to the increase (-), or decrease (+) in prediction ability in percentage points of the prediction ability of the original model.

From the point of view of a predictive ability power, the concept *QPM*, *QPM gap* (foreign demand gap observed, but fixed at levels) and a setting of $\rho_s = 0.6$ seems to be the most effective.

model/var	*CTW*	*QPM*	*QPM gap*	*QPM gap obs*	$\rho_s = 0.4$	$\rho_s = 0.6$	$\rho_s = 0.7$	$\rho_s = 0.8$
IR	6	-6	-10	-20	13	-24	-0	28
CPI	-0	-0	0	-0	-0	0	0	-0
ER	6	-33	-31	-21	-25	-15	-0	17
GDP	-0	2	3	20	-3	-2	-0	2
W	5	22	13	30	-12	-9	-0	5
C	-2	15	11	20	9	-3	-0	4
I	0	11	10	20	2	-0	-0	0
X	-1	-1	-1	-6	-3	-1	-0	1
N	-0	-0	0	8	-1	-0	-0	0
PC	1	1	1	-0	-2	-0	-0	0
PI	1	-5	-5	-3	-2	-1	0	1
PX	1	-10	-8	-8	-13	-7	-0	9
PN	2	-12	-12	-10	-8	-4	0	3
SUM	18.9	-16.0	-29.8	27.0	-44.7	-67.2	-0.0	71.9

Table 2 In-Sample Simulation - Predictive Power

4 Impulse Responses

The foreign demand shock in the current g3 model leads to higher wages which overweight appreciated exchange rate and thus to higher inflation. Interest rates react and increase. Real GDP increases as consumption and exports are higher. There is no reaction of foreign interest rates and inflation. When *ctw fb* or *qpm fb* models are employed, foreign interest rates and inflation increase as a reaction to higher foreign demand. It implies higher domestic inflation (inflation is imported despite lower wages and appreciated exchange rate) and higher domestic interest rates. Relatively high interest rates lead to lower consumption, but higher foreign demand to higher export growth.

Higher foreign prices in the g3 model lead to appreciation in order to mitigate improved competitiveness. Inflation is imported even exchange rate is appreciated. It leads to higher domestic interest rates. Real consumption decreases due to higher domestic interest rates, but exports are higher as our goods become more competitive. In the case of *ctw fb* or *qpm fb* models higher foreign inflation leads to higher foreign interest rate and lower foreign demand what strengthens effects to domestic variables but preserves directions.

Higher foreign interest rates in the g3 lead to depreciation, inflation and to higher domestic interest rates (UIP condition). This leads to lower wages and real consumption, but exports are supported by depreciated exchange rate. In the case of *ctw fb* or *qpm fb* models higher foreign interest rates lead to lower foreign inflation and lower foreign demand what strengthens effects to domestic variables but

preserves directions. The only exception is real export in *qpm fb* where effects of lower foreign demand are stronger than effects of depreciated exchange rate.

There are not big differences between impulse responses of the current g3 model and model *rho* where forward-lookiness of the UIP condition is changed to $\rho_s = 0.6$.

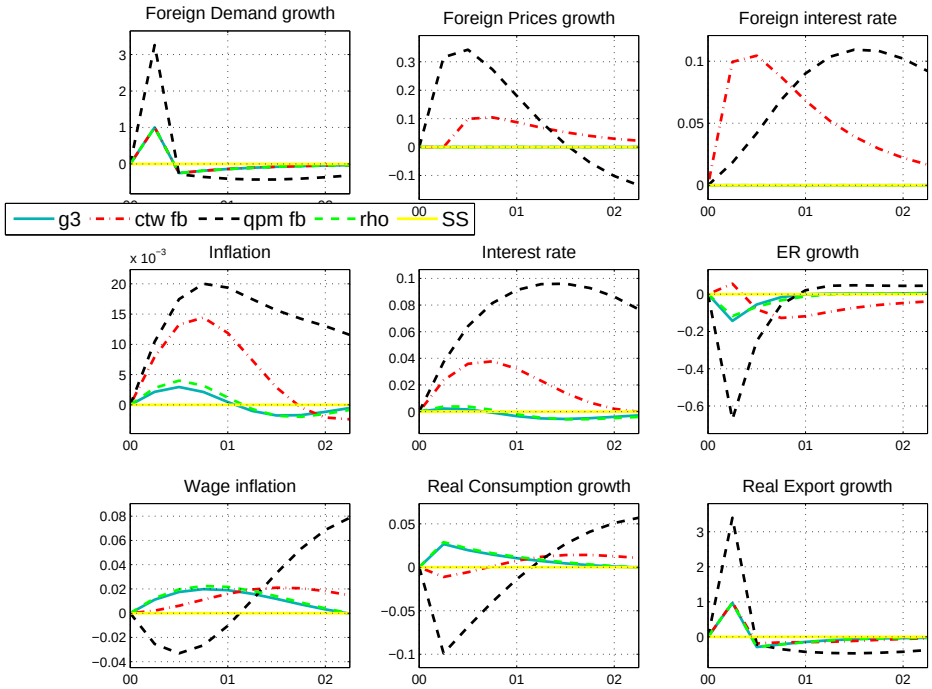


Figure 2 Foreign Demand Shock

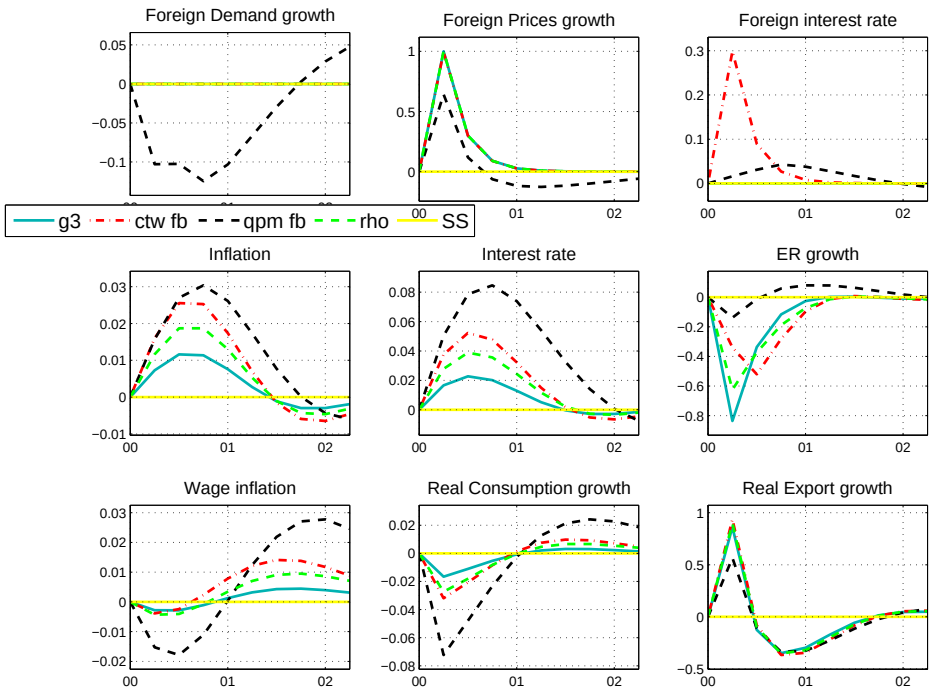


Figure 3 Foreign Prices Shock

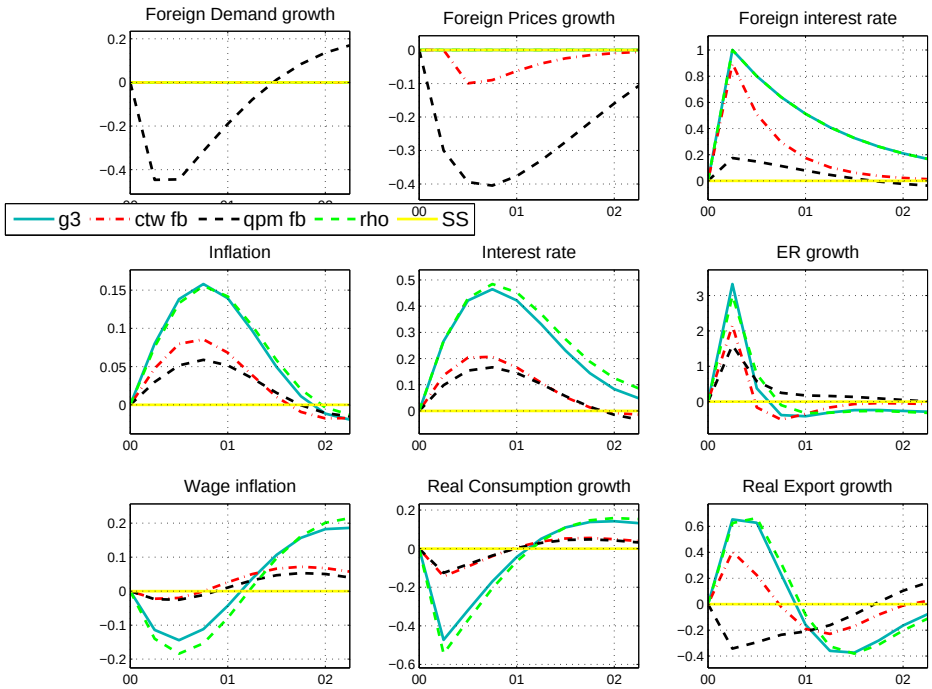


Figure 4 Foreign Interest Rate Shock

5 Summary

This text discusses the pros and cons of extending the g3 model with more elaborated foreign blocks. We consider the simplest VAR block to QPM style model of a foreign economy. These extensions are assessed on ex-post forecasting performance and on the intuitiveness of the implied model properties (impulse responses, decompositions to shocks). The main result is that the predictive power of a model can be improved by 16 percent (for the Czech case) when we employ a QPM style model of a foreign economy compared to a simple VAR block. However, the necessary condition is setting of autoregressive parameter of foreign interest rates near to one. This result indicates model misspecification and further research into this model is needed. We also discuss a setting of the parameter describing forward lookiness of the UIP condition. This parameter seems to be crucial for the overall prediction ability of a model. The value 0.6 seems to be the best (the Czech case) as lower or higher value brings bigger prediction errors.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1235/2014. This support is gratefully acknowledged. The views expressed here are those of the authors and do not necessarily reflect the position of the Czech National Bank.

References

[1] Andrle, M., Hlédik, T., Kameník, O., and Vlček, J.: *Implementing the New Structural Model of the Czech National Bank*. Czech National Bank, Working Paper Series 2, 2009.

[2] Christiano, L. J., Trabandt, M., and Walentin, K.: *Introducing Financial Frictions and Unemployment into a Small Open Economy Model*. Journal of Economic Dynamics and Control, vol. 35, 2011, 1999–2041.

[3] *The Czech National Bank's Forecasting and Policy Analysis System* CNB, Na Příkopě 28, Prague, (2003).

A Mathematical Model of Institutional Choice

Peter Toumanoff[1]

Abstract. This paper demonstrates that the choice of institutions can be examined using a neoclassical analytical framework. The choice of an institution is described by the equilibrium of a mathematical model that minimizes transactions costs subject to a transacting constraint, in a manner similar to the cost-minimization model of the firm. The model introduces the notion that institutions are characterized by vertical and horizontal relationships, which affect both the production of transactions through a transactions function, and the cost of transactions, through a transactions cost function. The equilibrium illustrates that particular exogenously determined parameters affect the mix of vertical and horizontal relationships that make up the optimum governing institution. It also determines the level of transactions costs as a function of the rate of transacting. The model is used to predict how the mix of vertical and horizontal relations and, therefore, the nature of institutions, respond to changes in parameters such as the number of transactors, asymmetry of knowledge, measurability of the object of exchange, negotiating and enforcement costs, etc.

Keywords: Institutions, transactions costs, governance mechanisms.

JEL Classification: D23, I22, P12, P50
AMS Classification: 91B32, 91B38

1 Introduction

The role that social institutions play in determining economic behavior has long been understood to be both vitally important and very difficult to incorporate into theory, especially neoclassical theory. Theoretical paradigms such as the Austrian school, the evolutionary economists, the public choice school, the new economic historians, and the new institutional economists have dealt explicitly and directly with the effects of and reasons for social institutions. Their efforts represent significant progress toward the understanding of economic behavior which is not well explained by neoclassical theory, such as the long-run dynamics of social change, collective choice, so-called 'market failures', the internal organization of the firm, and so on. This paper is motivated by the belief that it is worthwhile to provide an integrating analytical framework for investigating the choice of institutions that uses neoclassical methods and techniques. It is intended to provide a bridge between these other paradigms and neoclassical theory, allowing their insights to become a more integral part of the teaching of economics and the formation of public policy.

This paper presents a model of institutional choice that uses the techniques of neoclassical mathematical optimization. Many of the propositions implicit in the existing literature can be derived from the model, and comparative statics analysis can predict and explain how institutions change in response to changes in the external environment. In the next section a neoclassical model of institutional choice is presented in its most general and abstract form. Section 3 suggests a more specific functional form for transactions costs, and Section 4 uses the specific transactions cost function to derive comparative statics results. Finally, the conclusion suggests how the model of institutional choice may be applied to some important contemporary economic issues.

2 An Optimal Choice Model of Institutions

The common-sense postulates of the literature on markets and hierarchies can be described in a formal choice-theoretic framework. The model offered here resembles the neoclassical cost-minimization model of the firm. In the latter, combinations of inputs are chosen to minimize production cost subject to an output constraint. In the model of institutional choice, combinations of vertical and horizontal relationships are chosen to minimize transactions costs subject to a transactions constraint.

2.1 Definitions and Propositions

Definition 1. ***Transaction:*** *The performance of some task involving two or more agents. Economic transactions involve the rendering of commodities, services, or productive inputs from agent*(*s*) *to agent*(*s*)*, directly or indirectly, singly or in combination.*

[1] Marquette University, Department of Economics, Box 1881, Milwaukee, WI 53201-1881, peter.toumanoff@mu.edu.

Proposition 1. *For a transaction to take place, participants must be identified and their roles must be implemented. These requirements cause transactions to be costly.*

Definition 2. ***Vertical relationships:*** *Agents have a superior-subordinate relationship.*

Proposition 2. *Vertical relationships are relatively more effective at implementing transactions than at identifying them.*

Definition 3. ***Horizontal relationships:*** *Agents participate voluntarily and have equal status.*

Proposition 3. *Horizontal relationships are relatively more effective at identifying transactions than at implementing them.*

Definition 4. Institution: *Social mechanism that governs aspects of social behavior. Institutions provide the rules and norms that determine people's relationships to one another while engaging in transactions.*

Proposition 4. *The most effective institutions represent a mix of vertical and horizontal relationships that take advantage of what each does best.*

Proposition 5. *External factors related to the nature of the transactions being undertaken determine the optimal institutional choice relationships,*

2.2 The General Model

Transactions are accomplished by means of relationships among agents, or

$$T = f(V, H), \; f_V, f_H > 0 \tag{1}$$

in which T represents the rate of transacting (or task performance), V represents the number of vertical relationships, and H represents the number of horizontal relationships. The rate of transacting changes as the rate of task performance changes, and requires corresponding changes in the number of agents and relationships.

Consider a production process as an activity that requires transactions that are governed by some institutional arrangement. As described by Williamson (1975, 1981, and 1985) and Dietrich (1994), governance can take a wide variety of forms. These range from the simple hierarchy, representing all vertical relationships, to progressively more decentralized forms such as the U-form and the M-form, which incorporate more horizontal relationships, to more extreme forms of horizontal organization such as producer cooperatives or 'putting-out' systems. In the latter forms, individual input suppliers act as relatively independent contractors, choosing how much to supply and at what price by mutual contract. The possibility of multiple organizational forms that are capable of carrying out the same activity suggests that there is a technology of organization that is embodied in the transactions function. According to this technology, substitution between vertical and horizontal relationships is possible in the production of transactions. Such substitution can be described graphically by the level curves of the transactions function, which I call isotran curves. An isotran represents combinations of vertical and horizontal relationships that are capable of generating a given rate of transacting. Organizational innovations like the development of the M-form by Dupont and General Motors in the 1930s represent qualitative changes in the transactions function, and have the effect of shifting the isotran curves.

Transactions costs are defined as the opportunity cost of resources devoted to the identification and implementation of transactions, where implementation includes both negotiation and enforcement. Transactions costs are a function of V, H, and parameters α_i, $i = 1,2,...,n$, or

$$TC = C(V, H; \alpha_i), \; C_V, C_H > 0 \tag{2}$$

The parameters represent external conditions, related to the nature of the tasks being performed, which affect the cost of utilizing vertical and horizontal relationships. These are factors that affect the identification and implementation of transactions. A relevant parameter may be the measurability of the object of a transaction. For example, the quality of a consumer product is sometimes difficult to measure at the time of a commodity transaction. Transactions costs are affected because, in such a case, safeguards assuring product quality must be negotiated ex ante or enforcement of a certain level of quality must be implemented ex post.

The objective of the optimization model is to choose the combination of vertical and horizontal relationships that minimizes transactions costs, subject to maintaining the rate of transacting at some constant level, $T = T_0$. The Lagrangian objective function is

$$\min \mathcal{L} = C(V, H; \alpha_i) + \lambda(T_0 - f(V, H)) \tag{3}$$

The necessary and sufficient conditions require that

$$\mathcal{L}_V = C_V - \lambda f_V = 0 \tag{4}$$

$$\mathcal{L}_H = C_H - \lambda f_H = 0 \tag{5}$$

$$\mathcal{L}_\lambda = T_0 - f(V, H) = 0 \tag{6}$$

$$\begin{vmatrix} C_{VV} - \lambda f_{VV} & C_{VH} - \lambda f_{VH} & -f_V \\ C_{HV} - \lambda f_{HV} & C_{HH} - \lambda f_{HH} & -f_H \\ -f_V & -f_H & 0 \end{vmatrix} = C_V \; f_V^{\,2} \left(\frac{d\left(\frac{C_H}{C_V}\right)}{dH} - \frac{d\left(\frac{f_H}{f_V}\right)}{dH} \right) < 0 \tag{7}$$

Equations (4) through (7) describe an equilibrium at which the isotran is tangent to a level curve for the transactions cost function (which can be called 'isotranscost') and that the isotran is more convex (less concave) than the isotranscost. The equilibrium is pictured in Figure 1, which describes an optimum combination of horizontal and vertical relationships (H^*, V^*). The optimum combination minimizes the costs of transacting at a rate of T_0. At point (H^*,V^*), the T_0 isotrans is tangent to the optimum isotranscost. Transactions costs are minimized at a level of TC^*. The optimum values of the choice variables, $V^*(\alpha_i, T_0)$, $H^*(\alpha_i, T_0)$, and $\lambda^*(\alpha_i, T_0)$, are functions of the parameters, α_i and T_0, as is the indirect objective function, $TC^* = C^*(V^*,H^*;\alpha_i)$. The optimum value of the Lagrange multiplier, λ^*, is interpreted as $\partial TC^*/\partial T_0$, or the rate at which transactions costs change as the rate of transacting changes. Therefore, λ^* represents marginal transactions costs. If institutions can be interpreted as aggregations of vertical and horizontal relationships, the equilibrium of the transactions cost-minimizing model can be interpreted as describing the optimal institutional arrangement for the governance of T_0 transactions. Note that, by describing the optimum values of the choice variables, V^* and H^*, as functions of the parameters, the model provides a mathematical statement of Propositions 4 and 5, that the most effective institutions represent a mix of vertical and horizontal relationships that take advantage of what each does best, and that external factors related to the nature of the transactions being undertaken determine the optimal institutional choice relationships. The external factors are the parameters of the optimization model.

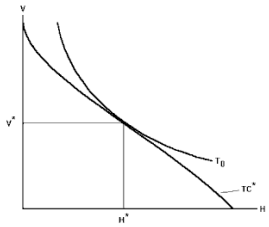


Figure 1 Optimal Combinations of Vertical and Horizontal Relationships

The next step in the development of a usable model of institutional choice is to examine how the optimal mix of vertical and horizontal relationships in institutions changes in response to changes in T_0 and the other parameters. At the level of generality embodied in this model, such comparative statics results are not forthcoming. However, it is possible to conclude that changes in the parameters α_i will affect the equilibrium through their effect on the isotranscost. Changes which make the curve more negatively sloped will favor more vertically oriented institutions. Changes that make the curve less negatively sloped favor more horizontally oriented institutions. To investigate such changes in more detail, a more specific form for the transactions cost function is developed in the next section.

3 A Transactions Cost Function

The transactions cost function offered in this section is a first attempt to construct a working model of institutional choice. Both identification costs and implementation costs for both vertical and horizontal relationships must be accounted for. Identification costs are considered first.

3.1 Identification Costs

Identification costs are incurred in the course of transacting when resources are devoted to discover who the appropriate participants to a transaction are. Sometimes an individual has specialized knowledge; information pertinent to a transaction that is known by that individual, but not by anyone else. For example, only an individual consumer knows his or her tastes and preferences, and only an input supplier knows his or her opportunity costs. An individual's specialized knowledge of tastes or opportunity costs is difficult or impossible to communicate to another. Hayek's (1945) notion of the "particular circumstances of time and place" is a good example of specialized knowledge that is difficult to communicate.

In a vertical relationship, potential transactors have a superior-subordinate relationship in which the superior must identify the appropriate subordinate, whose participation is non-voluntary. This task is made more difficult the more specialized and unobservable the knowledge of the individual subordinate. Let the function $\Theta(\sigma, \delta)$ represent the cost of identifying a non-voluntary agent. The arguments σ and δ measure the degree to which the knowledge is specialized and unobservable, respectively. I assume that Θ is increased by increases in σ and δ; that is, the partial derivatives Θ_σ and Θ_δ are positive.

Another problem for a superior attempting to identify a vertical relationship is posed by the number of agents that are appropriate for a particular transaction relative to an entire population of potential transactors. Identifying appropriate non-voluntary subordinates to perform some task is quite difficult if the number of such appropriate subordinates is small relative to the entire population of potential transactors. Consider the example of identifying the appropriate consumers of running shoes. To find a single individual out of an entire population of n potential agents whose size, athleticism and tastes are such that the benefits of using the shoes exceeds the opportunity costs of providing them is the same as finding one distinct element out of a set of n. To find two consumers of running shoes is the same as finding two distinct elements out of the set, and so on. A hierarchy seeking to identify the V agents out of a population of P who are appropriate for running shoes has $\frac{P!}{(PV)!V!} = \binom{P}{V}$ possible combinations to choose from. If the shoe is appropriate for the entire population, then identification of the appropriate transactors becomes somewhat easier for the hierarchy because there is only one combination of n distinct elements out of a set of n.

These problems for agents in a vertical relationship suggest that identification costs can be characterized by a function like

$$ID^V = \Theta(\sigma,\delta)V\binom{P}{V} \tag{8}$$

In this function, P represents the population of potential transactors. This function increases at an increasing and then decreasing rate until V equals approximately 50% of the population. After that point, it decreases until, as the number of vertical relationships approaches the entire population of potential transactors, the cost to the superior of identifying transactors approaches $\Theta(\sigma, \delta)V$.

Identification of appropriate transactors can also be carried out by horizontal relationships. Because agents participate voluntarily in horizontal relationships, the identification costs are, simply, the costs to the individual of voluntarily identifying herself as a transactor. Examples of this cost may be the time devoted to searching for transactions, or resources diverted for advertising. Every agent in a horizontal relationship faces such costs, therefore identification costs can be treated as proportional to horizontal relationships, or

$$ID^H = \eta H, \quad \eta > 0 \tag{9}$$

The parameter η represents the cost per agent of voluntarily identifying herself as a transactor. Identification costs for transactions involving both vertical and horizontal relationships can be treated as the sum of ID^V and ID^H, or

$$ID(V,\ H) = \Theta(\sigma,\delta)V\binom{P}{V} + \eta H \tag{10}$$

3.2 Implementation Costs

Implementation costs are incurred as transactors negotiate and enforce the terms of a transaction. When the object of a transaction is easily measured, and when exclusive use of it is easily enforced, implementation of the transaction is simplified. Conversely, when the object of a transaction cannot be measured, or exclusive use cannot be enforced, implementation of the transaction is greatly complicated. This is because performance of the transaction is difficult to monitor, and so safeguards must be negotiated and guarantees enforced. As suggested earlier, determining the quality of a commodity presents measurability problems. Examples of excludability problems are prominent in the public goods and externalities literatures, such as common fisheries or preventative measures for infectious diseases.

In a hierarchical structure of vertical relationships a superior has to communicate with and monitor each subordinate involved in a transaction. The cost of doing so depends on the measurability and excludability of the activities that are undertaken in the realization of the transaction. Therefore, IMP^V characterize the implementation costs associated with vertical relationships as

$$IMP^V = \Phi(\mu,\chi)V, \quad \Phi_\mu, \Phi_\chi < 0 \tag{11}$$

In equation (11), the function $\Phi(\mu,\chi)$ represents the cost of communicating a message regarding an activity from one transactor to another and verifying that the task is accomplished. The parameters μ and χ represent the measurability and excludability of the object of the transaction. The partial derivatives $\Phi\mu$ and $\Phi\chi$ are both assumed to be negative, as the more measurable and/or excludable the object of a transaction, the less costly is communicating and verifying the transaction.

Horizontal relationships involved in transacting also communicate and verify messages, and therefore also suffer implementation costs. However, in horizontal relationships bilateral negotiations among the participants must take place. Potentially, the number of separate relationships that must be made equals $H(H–1)/2$. Furthermore, the nature of the negotiations among transactors in horizontal relationships differs from those in vertical relationships because participation in the transaction is voluntary rather than compulsory. For these reasons, IMP^H characterize the implementation costs associated with horizontal relationships as

$$IMP^{H} = \Phi(\mu,\chi)\frac{\gamma}{2}H(H-1), \quad \gamma > 0 \tag{12}$$

The function $\Phi(\mu,\chi)$ is as above, the multiplier associated with measurability and excludability. The second multiplier in equation (12), γ, allows for the possibility of differences in voluntary vs. compulsory participation in transactions. Although positive, γ can be greater or less than one. A value of γ greater than one indicates that the voluntary nature of horizontal relationships increases the implementation cost relative to the compulsory nature of vertical relationships. This might be the case if transactors must be convinced to participate in a transaction. Conversely, a value of γ less than one indicates that the voluntary nature of horizontal relationships decreases the implementation cost relative to the compulsory nature of vertical relationships. This might be the case if enforcement of a transaction is simplified when transactors are acting in their self-interest, which must be the case if participation is voluntary.

Total implementation costs is the sum of equations (11) and (12), or

$$IMP = \Phi(\mu,\chi)\left(V + \frac{\gamma}{2}H(H-1)\right) \tag{13}$$

Transactions costs in total are the sum of identification costs (eq. 10) and implementation costs (eq.14), or

$$C(V,H) = \Theta(\sigma,\delta)V\binom{P}{V} + \eta H + \Phi(\mu,\chi)\left(V + \frac{\gamma}{2}H(H-1)\right) \tag{14}$$

The function increases initially with increases in V and H and the isotranscosts are concave. For large values of V, however, and depending on the values of $\Theta(\sigma, \delta)$ and $\Phi(\mu, \chi)$, transactions costs may decrease with increases V, and the isotranscosts may become convex or even positively sloped. The function behaves this way because the identification costs first increase and then decrease with increases in vertical relationships.

4 The Comparative Statics of Institutional Choice

The model of institutional choice, presented in general form in Section 2, is re-explored in the following section using the transactions cost function represented by equation (14). Using a more specific form of a transactions cost function permits comparative statics analysis of the model. If we replace $C(V, H; \alpha_i)$ with equation (14) in equations (4), (5) and (6), we obtain first- and second-order conditions that describe an equilibrium for which the transactions cost-minimizing values of V and H depend on the values of the parameters T_0, σ, δ, μ, χ, P, η, and γ, or $V^* = V^*(T_0,\sigma,\delta,\mu,\chi,\eta,\gamma,P)$ and $H^* = H^*(T_0,\sigma,\delta,\mu,\chi,\eta,\gamma,P)$. Although explicit forms cannot be found for these functions the effects of changes in the parameters on the values of V^* and H^* can be found by implicit differentiation. Changes in several of the parameters are found to have unambiguous effects. These are the implementation parameters, μ, χ and γ, and the voluntary identification parameter, η.[2] We find that $\frac{\partial V^*}{\partial \mu} < 0$, $\frac{\partial H^*}{\partial \mu} > 0$, $\frac{\partial V^*}{\partial \chi} < 0$, $\frac{\partial H^*}{\partial \chi} > 0$, $\frac{\partial V^*}{\partial \gamma} > 0$, $\frac{\partial H^*}{\partial \gamma} < 0$, $\frac{\partial V^*}{\partial \eta} > 0$, and $\frac{\partial H^*}{\partial \eta} < 0$.

These results tell us that improvements in the measurability of an object of a transaction, μ, favor more horizontally-oriented institutional arrangements, whereas decreases in measurability favor more vertically-oriented institutional arrangements. This is because vertical relationships have a comparative advantage in implementing transactions, so changes that make implementation more difficult, like a decrease in the measurability of the transacted item, favor vertical organization. Changes that make implementation easier, like an improvement in

[2] Mathematical expressions for these partial derivatives available from the author on request.

measurability, favor horizontal organization. Recognize that a change in excludability of the object of the transaction, χ, has the same effects for the same reasons.

The parameter γ represents relative differences in voluntary vs. compulsory participation in transactions. The higher is γ, the greater the cost of implementing a transaction via horizontal relationships relative to vertical relationships. The partial derivatives with respect to γ suggest that, if voluntary implementation of transactions becomes more costly relative to compulsory implementation, the optimum adjustment is toward more vertically oriented institutional arrangements.

The parameter η represents the cost of an agent's voluntarily identifying herself through a horizontal relationship as a transactor. Depending on the transaction, it may be as simple as taking a grocery item to a checkout counter, or it may involve much more, like hiring a broker to find a property. If this cost should increase, the optimal institutional structure changes to a combination including more vertical relationships and fewer horizontal.

5 Conclusion

This paper demonstrates that the choice of institutions can be examined using a neoclassical analytical framework. Institutions are treated as collections of vertical and horizontal relationships among agents that govern transactions. The paper postulates the existence of a transactions function and a transactions cost function, both of which depend on combinations of vertical and horizontal relationships. Doing so permits the choice of an institution to be described by the equilibrium of a model that minimizes transactions costs subject to a transacting constraint, in a manner similar to the cost-minimization model of the firm. The equilibrium illustrates that particular attributes such as measurability of an object of exchange affect the mix of vertical and horizontal relationships that make up the optimum governing institution. It also determines the level of transactions costs as a function of the rate of transacting.

The comparative statics analysis of the model suggests that transactions that involve relatively small fractions of the population, and for which the objects of exchange are easily measured and excludability is not a problem, favor relatively more horizontally oriented institutions. This finding is consistent with the observation that markets almost universally govern the distribution of privately consumed commodities. If the object of an exchange is not easily measured, and when the costs of voluntary implementation are relatively high, more vertically oriented institutions are called for. Transactions that involve the majority of the population and for which excludability is difficult tend to favor more vertically oriented institutions. Thus we see that governments usually undertake the provision of public goods. The more specialized and unobservable the knowledge pertinent to transacting, the more horizontally oriented institutions are favored, if the number of transactors is a small fraction of the population

There are many potential uses for a model of institutional choice like the one described in this paper. An obvious application is the reformation of the economic systems of post-communist and communist states. Communist economic and political systems are relatively hierarchical, authoritarian, and vertical in nature. They are relatively effective at implementing easily identified tasks that require population-wide mobilization of resources, like undertaking a military operation or extensive development. They are much less effective at identifying very complex tasks that involve diverse participants, like satisfying consumer tastes or adapting to the constantly changing requirements of global trade. Therefore, the theory of institutional choice can be applied to explain the successes and failures of communist economies, and the direction of change in the post-communist world. Demographic, geographic, and economic differences among the post-communist nations affect each country's optimal mix of economic institutions, and so it can be expected that each will follow a different path in the future. For example, the Czech Republic, which has a relatively sophisticated economy and is well positioned for global trade has introduced more market institutions in its economic system more rapidly than a country like Romania, which has a much more primitive economy and is not as well positioned to engage in global trade is more hierarchical and less market-oriented.

References

[1] Dietrich, Michael: *Transaction Cost Economics and Beyond: Towards a New Economics of the Firm*. Routledge, London, 1994.
[2] Hayek, F. A.: The Use of Knowledge in Society, *American Economic Review*, **35** (1945), 519-530.
[3] Williamson, Oliver E.: *Markets and Hierarchies: Analysis and Antitrust Implications*. The Free Press, New York, 1975.
[4] Williamson, Oliver E.:. The Modern Corporation: Origins, Evolution, Attributes. *Journal of Economic Literature*, **19**, No. 4. (1981), 1537-1568.
[5] Williamson, Oliver E.: *The Economic Institutions of Capitalism*, The Free Press, New York, 1985.

Method of Non-weighted Average Absolute Deviation in Context of Measuring Income Inequality by Gini Coefficient: case study of Visegrad group countries

Kamila Turečková[1], Eva Kotlánová[2]

Abstract: There are many possibilities how to measure income inequality in economies. This article presents use of the method of non-weighted average absolute deviation for expression and comparison of income inequality in four selected countries in Central Europe known as Visegrad Group. The aim of this paper is to use this new method to measure how the income of individuals differs from a situation when everybody would have self-same income (the situation of perfect income equality). Then the results of this new method are compared with results of measuring income inequality by standard method using Gini coefficient. The analysis of income distribution of four selected European economies is made in period 2005–2013 and its order is be made in quintiles based on empirical data from the Statistics on Living Conditions and Welfare published by Eurostat. According to results of comparison of measuring via standard method using Gini coefficient and the new method of non-weighted average absolute deviation, authors recommend to use the new one especially for its easy calculation based on commonly available data in the required format provided by statistical offices.

Keywords: Gini Coefficient, Income Inequality, Lorenz Curve, Method of Non-weighted Average Absolute Deviation.

JEL Classification: C13, D31, I32,
AMS Classification: 62P20

1 Introduction

There are many possibilities how to look on or measure standards of living in selected countries and societies. One of the best known is GDP per capita. Despite the fact that this indicator could reach relatively large value, it does not predicate about differences of incomes in society. Another indicator we could hear about very often is average wage. Not even its amount is a guarantee of economic well-being. It is usual that over 50% of working population of the country could not reach this amount. That is why is also necessary to focus on income dispersion and to express income inequality in the economy because income inequality creates a perilous space for social, interpersonal and economic discrimination. One of the best known and used measures of income inequality is Gini coefficient and its graphical representation – Lorenz curve [11].

The aim of this paper is to apply other method to measure and express income inequality in case of Visegrad group inhabitants between years 2005 and 2013. This analysis of income inequality is based on method of non-weighted average absolute deviation and its result of measuring income inequality is here compared to results calculated by Gini coefficient. There will be also mapped the changes in income inequality of Czech, Slovak, Polish and Hungarian residents in the selected period of time. The measurement of income inequality is so important because, for example, some studies found the relationship between income inequality and economic growth [6] or income inequality and society's level of mortality [5].

The text of the article below this Section 1, Introduction, will be organized as follow: Section 2 provides same theoretical introduction to measurement of income inequality, in detail the method of non-weighted average absolute deviation and Gini coefficient are characterized here. Section 3 is oriented to methodology and characterization of data which was used. Section 4 contains the empirical analysis of the income inequality in four selected European countries using the method of non-weighted average absolute deviation and method of determining the Gini coefficient. There is also indicated the development of income inequality during ana-

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Economics and Public Administration, Univerzitní nám. 1934/3, Karviná 733 40, tureckova@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Economics and Public Administration, Univerzitní nám. 1934/3, Karviná 733 40, kotlanova@opf.slu.cz.

lyzed period with evaluation of countries and their ranking. Finally the conclusion, Section 5, highlights some major conclusions of detailed analysis of income inequality and also mentions results of measurement of income inequality by standard method using Gini coefficient. Then the results are compared to the result of income inequality obtained by method of non-weighted average absolute deviation.

2 Measurement of Income Inequality

Income inequality is a natural part of every human society because allocation of income is mostly considerably unequal through people and households in each society and economy. There is a lot of methods how to measure or express income inequality. Among well-known of them belong: Lorenz curve, Gini coefficient, Coefficient of income inequality S80/S20 (or Quintile share ratio or S80/S20 Ratio), Atkinson index, Theil index, Robin Hood index (also known as Hoover or Schutz index) and Variation coefficient [1], [7], [9]. Income inequality is analysed here using Gini coefficient and new method of non-weighted average absolute deviation. That is why these methods described below in more details. For more information about other indexes see for example [1], [2], [7], [8], [9], [10], [11], [12], [14] or [15].

2.1 Gini coefficient

Gini coefficient measures the extent to which the distribution of income within a country deviates from a perfectly equal distribution. A coefficient of 0 expresses perfect equality where everyone has the same income, while a coefficient of 1 expresses full inequality where only one person has all the income [3].

$$G = \frac{A}{A + B} \tag{1}$$

where G = $\langle 0;1 \rangle$ (see Figure 1a).

Gini coefficient is the most widely used method of measuring of concentration in income inequality and may be based on the Lorenz curve and it is a numeric representation of divergence of Lorenz curve from Straight line of equality when the Gini ratio of concentration is calculated by dividing A area and sum of A area and B area, then:

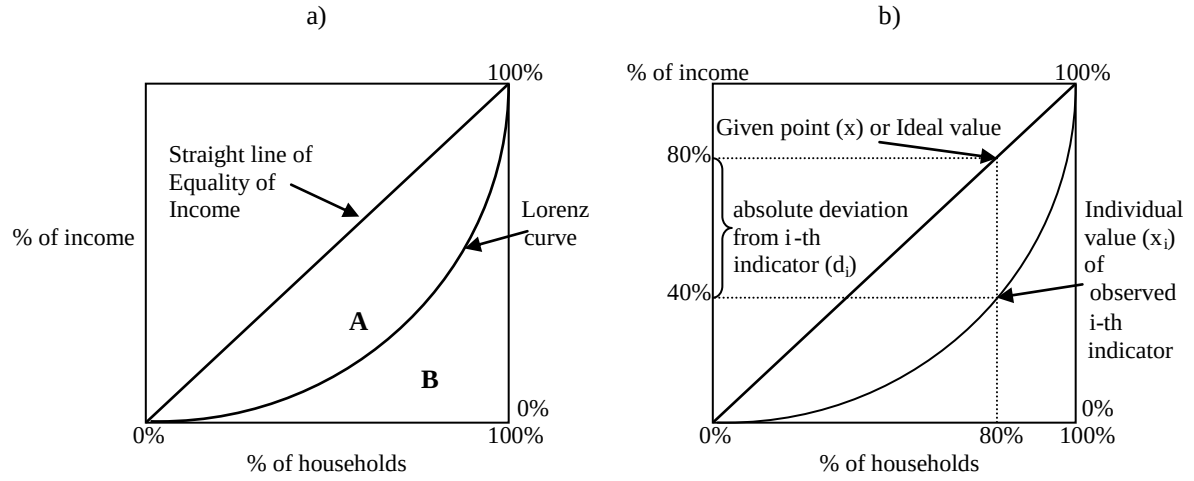


Figure1 Lorenz curve as a source for determine Gini coefficient and value of income inequality measured by method of non-weight average absolute deviation

The main advantage of using this coefficient is its easy interpretation and the fact that it measures income inequality regardless of a population in the country and regardless of a size of economy. It could be used to compare of income distribution among different groups of population as well as single regions in one country [12] .

Objections to the crudity and ambiguity of this measure arise from the fact that every Lorenz curve has just the only one Gini coefficient but two societies with the same Gini coefficient could have different Lorenz curve. In view of this fact it is recommended to supplement results of Gini coefficient with conclusions of another analyses or indexes [11].

2.2 Method of non-weighted average absolute deviation

Method of average deviation reflects the degree of variability, defined as arithmetic average of absolute deviations of individual values of observed indicators from the selected value (given point). Generally the deviation is reckoned from the ideal value, recommended value, central value that is constructed as some type of average, median, mean of the data set and other. These values, chosen here, represent values for ideal distribution of income in society, which means the value of expressing absolute equality in income for each inhabitant. In general, absolute deviation is constructed on the basis of this formula 1:

$$d_i = \left| x_i - (x) \right| \tag{2}$$

where: d_i presents the absolute deviation from i-th indicator,

x_i presents the i-th indicator (data element, variable),

(x) is the chosen given point.

Own value of non-weighted average absolute deviation we obtained from the formula 2:

$$\bar{d}_i = \frac{\sum_{i=1}^{P} \left| x_i - \left(\bar{x}\right) \right|}{n_i} \tag{3}$$

where: $\bar{d}_i$ presents the average absolute deviation from i-th indicator,

n_i presents the number of values of i-th indicator that we have available,

$(\bar{x})$ is the arithmetic mean of i-th indicator.

Through graphical representation of Lorenz curve it is possible to illustrate also this method of measurement of income inequality (see Figure 1b). Using this method is preferable due to data which are provided by statistics in the context of income distribution among different population groups (households). There is also much easier to use and apply the method of non-weighted average absolute deviation to express income inequality than count Gini coefficient because the results of both methods are essentially identical (see conclusion). Gini coefficient is based on the calculation of the Lorenz curve, the calculation of which is not so easy. It is possible to formulate Lorenz curve as a distribution function F(x) as an equation:

$$L(F) = \frac{\int_{-\infty}^{x(F)} x \cdot f(x) \cdot dx}{\int_{-\infty}^{\infty} x \cdot f(x) \cdot dx} = \frac{\int_{0}^{F} x(F') \cdot dF'}{\int_{0}^{1} x(F') \cdot dF'}, \tag{4}$$

where: x(F) is inverse function to the distribution function F(x).

Using of methods of non-weighted average absolute deviation in context of income inequality measurement is specified in these areas: (I) indicator (x) is the ideal percentage value of income which is set in concrete percentage of households in society, for example, 40% of households get precisely 40 % of total income ((x) = 40%). (II) variable x_i presents real money of income of household cumulated into relevant quintiles (quartiles, deciles, percentiles aso.). Value of non-weighted average absolute deviation for measurement of income inequality (for our purposes we denote D_{ii}) can reach values from 0 to 100 and if value of non-weighted average absolute deviation is lower (the more close to 0) than less income inequality is between the richest and poorest households in society. Perfect income equality in the society would occur in a situation when the value would come out zero (D_{ii}=0). For more information see [13].

3 Methodology and data

From a methodological perspective, the work is based on data from Table 1 gained by Eurostat, concretely from the Statistics on Income and Living Conditions (SILC) – Distribution of income by quintiles as a share of national equivalised income for Czech Republic, Slovak Republic (Slovakia), Poland and Hungary. The covered period includes years 2005-2013 because of missing credible data which is not available for a longer period. Income understood as a total disposable income of a household that is calculated by counting personal income received by all members of the household plus income received at household level.

Disposable household income includes all income from work (employee wages and self-employment earnings), private income from investment and property, transfers between households and all social transfers received in cash including old-age pensions [3].

Calculations of Gini coefficient and value of non-weighted average absolute deviation (D_{ii}) are based on calculations using formulas (1), (2) and (3). All these measures of income inequality were described in the text above. The software used was MS Excel. All calculations and graphical analysis are authors own.

Country	Czech Republic					Poland				
Year/Households	20%	40%	60%	80%	100%	20%	40%	60%	80%	100%
2005	9.8	14.4	17.5	22.2	36.0	6.4	12.1	16.6	22.6	42.4
2006	10.1	14.5	17.7	22.0	35.7	7.2	12.4	16.8	22.8	40.7
2007	10.1	14.5	17.7	22.1	35.6	7.6	12.8	17.0	22.5	40.1
2008	10.3	14.7	17.8	21.9	35.3	7.8	12.8	16.9	22.4	40.1
2009	10.3	14.7	17.7	21.7	35.6	8.0	12.9	17.0	22.6	39.5
2010	10.2	14.7	17.7	21.9	35.5	7.9	13.0	17.2	22.7	39.2
2011	10.1	14.6	17.7	22.0	35.6	7.9	13.0	17.3	22.6	39.1
2012	10.1	14.7	17.8	22.0	35.3	7.9	13.0	17.3	22.8	39.0
2013	10.4	14.7	17.7	22.0	35.2	8.0	13.1	17.4	22.7	38.9
Country	Hungary					Slovak Republic				
2005	9.1	14.1	17.7	22.1	37.0	9.1	14.7	18.2	22.5	35.5
2006	7.5	13.0	16.8	21.5	41.2	9.3	14.2	17.4	21.3	37.7
2007	9.6	14.6	18.0	22.5	35.3	10.0	14.9	18.2	22.3	34.6
2008	9.8	14.6	18.2	22.3	35.2	10.0	15.1	18.5	22.6	33.8
2009	9.9	14.7	18.2	22.4	34.8	9.8	14.8	18.1	22.5	34.9
2010	10.0	14.8	18.4	22.7	34.2	9.3	14.6	18.1	22.7	35.3
2011	9.3	14.0	17.8	22.7	36.3	9.2	14.8	18.2	22.7	35.1
2012	9.2	14.2	17.9	22.5	36.3	9.3	14.6	18.4	23.2	34.6
2013	8.9	14.0	17.7	22.3	37.1	9.4	15.1	18.6	23.2	33.7

Table 1 Households net money income (%)

Table 1 shows distribution of income for each country´s households classified into 5 groups of 20% of the total households - quintiles - for analyzed years.

4 Empirical Analysis

Empirical analysis was made on the basis of the disposable household income of selected countries. Figure 2 shows the development of values of average absolute deviation (D_{ii}) for Visegrad group countries.

Figure 2 shows the development of values of an average absolute deviation for Visegrad group for the years 2005-2013 where the differences in level and development of income distribution are visible. It shows that the best income equality from analysed countries has Czech Republic followed by Slovak Republic. The worst income equality was in Poland. How we can also see from the Figure 2 the situation in income distribution between households is getting better in Poland while development of income distribution is highly volatile in Hungary.

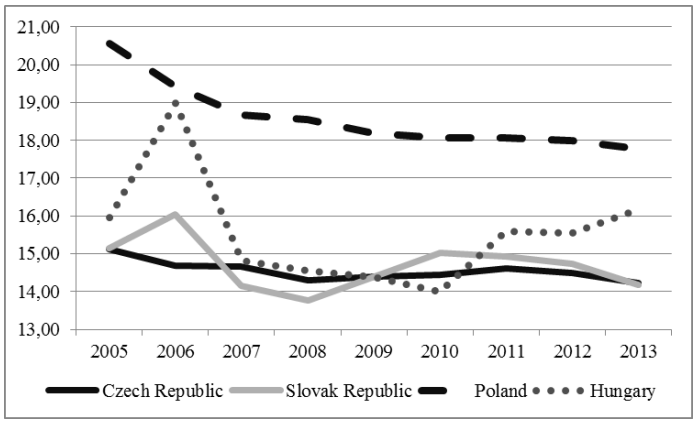


Figure 2 Development of values of income inequality express by D_{ii} in selected central European countries (4), 2005-2013

Table 2 shows the calculated values that express income inequality by using a "new" method of non-weight average absolute deviation and standard method of Gini coefficient. In both case of the calculation of income inequality, the lower value of the indicator corresponds to a more even distribution of income in selected countries.

Country	Czech Republic		Hungary		Poland		Slovak Republic	
Year/Variable	D_{ii}	Gini	D_{ii}	Gini	D_{ii}	Gini	D_{ii}	Gini
2005	15,125	0,525	15,950	0,537	20,550	0,601	15,150	0,529
2006	14,675	0,518	18,975	0,577	19,425	0,585	16,050	0,536
2007	14,650	0,518	14,825	0,523	18,675	0,575	14,150	0,513
2008	14,300	0,513	14,550	0,519	18,550	0,572	13,775	0,510
2009	14,400	0,514	14,375	0,517	18,175	0,568	14,400	0,517
2010	14,450	0,515	14,000	0,513	18,075	0,567	15,025	0,527
2011	14,600	0,517	15,600	0,534	18,075	0,567	14,925	0,526
2012	14,500	0,516	15,550	0,533	18,000	0,567	14,725	0,525
2013	14,225	0,512	16,175	0,541	17,775	0,564	14,175	0,518
Correlation coefficient	0,997921		0,999743		0,998633		0,977852	
F-test	1 678,537		13 602,366		2 554,470		152,797	

Table 2 Values of average absolute deviation (D_{ii}), Gini coefficients (Gini) and correlation coefficients for selected central European countries (4), 2005-2013

Table 2 also displays that the correlation coefficients for values gets from method of average absolute deviation and values of Gini coefficient are extremely positively high which means that there is an enormous significant dependence between these two variables. Significance of correlation coefficient was verified by F-test. Correlation coefficient = 0 was set as a null hypothesis (H_0). Based on results of F-test we reject the null hypothesis thus correlation coefficient is significant on significance level lower than 0.01.These conclusions were also confirmed by further analysis and calculations of income inequality in other economies and societies. Since the correlation between the results obtained with the method of non-weighted average absolute deviation and Gini coefficient is significant, it is advisable to use the method of non-weighted average absolute deviation to express the deviation in income inequality instead of Gini coefficient which calculation is considerably more difficult. Also using methods of average absolute deviation to express income inequality is recommended offers due to the format of provided data on income distribution in society by statistical offices. Non-weighted average absolute deviation method can expand the existing portfolio of methods to measure and express income inequality between households in society.

5 Conclusions

There exist many traditional and standard ways how to measure or describe income inequality in our society and economy. In this paper we have paid attention to the alternative method of measuring and expressing income inequality through method of non-weighted average absolute deviation. It was used to map changes in income distribution of four central European countries, especially the countries collectively labeled as Visegrad group (Czech Republic, Poland, Hungary and Slovak Republic) during the period of years 2005-2013. This method was supplemented by using method of measuring Gini coefficient because of confirmation the relevance of calculations and conclusions of income inequality. The highest income equality reached Czech Republic from the analyzed group of countries while the worst income equality was in Poland.

Authors recommend the use of measurement and comparison of income inequality method of non-weighted average absolute deviation for its easy calculation based on commonly available data in the required format provided by statistical offices and the same interpretation results in the context of the Gini coefficient that calculation is more difficult.

Acknowledgements

This paper was supported by the project SGS/13/2015 "Influence of Selected Macroeconomic and Microeconomic Determinants on the Competitiveness of Regions and Firms in Countries of the Visegrad Group Plus".

References

[1] Atkinson, A. B.: On the Measurement of Inequality. *Journal of Economic Theory* **2** (1970), 244 – 263.
[2] Dalton, H.: *The measurement of the inequality of incomes*. [online]. [2014-08-19]. Available from: http://www.ophi.org.uk/wp-content/uploads/Dalton-1920.pdf.
[3] EUROSTAT. Category: Living condition glossary. [online]. [2014-12-12]. Available from: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Category: Living _conditions_glossary.
[4] EUROSTAT. Statistics: Income, Social Inclusion and Living conditions [online]. [2015-01-10]. Available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_ inclusion_living_conditions/data/database.
[5] Kawachi, I., Kennedy, P. B., Lochner, K. and D. Prothrow-Stith: Social Cupital, Income Inequality, and Mortality. *American Journal of Public Health* **87** (9), 1997, 1491 – 1498.
[6] Kuznets, I.: Economic Growth and Income Inequality. *American Economic Review* **65** (1), 1955, 1 – 28.
[7] Lapáček, M.: *Příjmová nerovnost a ekvivalenční stupnice*. 2007.[online]. [2014-08-10]. Available from: http://nf.vse.cz/download/veda/workshops/inequality.pdf.
[8] Litchfield, J. A.: Inequality: Methods and Tools. Text for World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance. [online]. [2014-08-22]. Available from: http://www.worldbank.org/poverty/inequal/index.htm.
[9] Schutz, R.:On the Measurement of Income Inequality. *The American Economic Review* **41** (1), 1951, 107 – 122.
[10] Turečková, K. a E. Kotlánová: Income inequality in the Czech Republic and Slovak Republic. In: *Proceedings of the International Workshop Ten years of Membership of the Czech and Slovak Republic in the European Union*. Karviná: OPF SU Karviná, 2014a, 240-247.
[11] Turečková, K. a E. Kotlánová: Poverty analysis and measuring income inequality in Czech Republic. In: *Conference Proceedings from 32nd International Conference Mathematical Methods in Economics MME 2014*. Olomouc: PU Olomouc, 2014b, 1063-1067.
[12] Turečková, K.: Příjmové nerovnosti a jejich měření. *Acta Academica Karviniensia*, vol. 1, 2007, 191 -198. Turečková, K.: Measurement of Income Inequality by Method of Non-weighted Average Absolute Deviation: Case of South European Countries. In: *Crafting Global Competitive Economies: 2020 Vision Strategic Planning& Smart Implementation*. Norristown: IBIMA, 2014, 1620-1628.
[13] Večerník, J.:Earnings disparities and income inequality in CEE countries: an analysis of development and relationships. *Luxembourg Income Study Working Paper Series*, Working Paper, No. 540, 2010. [online]. [2014-08-20]. Available from: <http://ssrn.com/abstract =1806991>.
[14] Wolff, E. N.: *Poverty and income distribution*. 2nd Ed. Chichester, UK: Wiley-Blackwell, 2009.

Sensitivity analysis of a DSGE model with time-varying parameters identified by a nonlinear particle filter

Stanislav Tvrz[1], Osvald Vašíček[2]

Abstract. In this paper, we assess the robustness of the time-varying parameter estimates of a dynamic stochastic model of a small open economy general equilibrium with financial frictions that were obtained with the use of a nonlinear particle filter. First, we assess the sensitivity to the selection of the subset of model parameters that are assumed to be time-varying. For that purpose, the model is estimated in three different configurations of the time-varying and constant parameters. Second, we examine the sensitivity of the time-varying estimates with respect to the calibration of the initial value of the adhesion parameter. The value of a time-varying parameter is given by the model law of motion as a weighted average of the last known value of the parameter and its initial value plus stochastic innovation. The weights are determined by the adhesion parameter that is common for all the time-varying parameters. Thus, the adhesion parameter determines the general tendency of the time-varying parameters to return to their respective initial values. Three different calibrations of the adhesion parameter are compared in the paper. The model is estimated using the data of the Czech economy.

Keywords: DSGE model, time-varying parameters, particle filter, sensitivity analysis.

JEL classification: E32, E44, E58
AMS classification: 91B64

1 Introduction

Due to the unprecedented development of the Czech economy during the period of the Great Recession, we find it reasonable to investigate the question of possible structural changes of the Czech economy. We follow the direction of research of structural stability of the Czech economy of Vašíček *et al.* [7], who found evidence of parameter drifting in the Czech data. Also, we continue and elaborate on the research agenda introduced in Tvrz and Vašíček [4] and [5], where we estimated a small open economy DSGE model of the Czech economy with time-varying parameters with help of a nonlinear particle filter. The aim of this paper is to assess the robustness of the nonlinear particle filter algorithm for estimation of the time-varying parameters of a DSGE model with financial frictions.

First, we assess the sensitivity to the selection of the subset of model parameters that are assumed to be time-varying. For that purpose, the model is estimated in three different configurations of the time-varying and constant parameters. Second, we examine the sensitivity of the time-varying estimates with respect to the calibration of the initial value of the adhesion parameter. The value of a time-varying parameter is given by the model law of motion as a weighted average of the last known value of the parameter and its initial value plus stochastic innovation. The weights are determined by the adhesion parameter that is common for all the time-varying parameters. Thus, the adhesion parameter determines the general tendency of the time-varying parameters to return to their respective initial values. Three different calibrations of the adhesion parameter are compared in the paper.

[1]National Bank of Slovakia, Imricha Karvaša 1, 813 25 Bratislava, Slovak Republic, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic, tvrz@mail.muni.cz, The views expressed here are those of the authors and do not necessarily reflect the position of the National Bank of Slovakia.
[2]Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic, osvald@econ.muni.cz

2 The model

Since we focus on the period of financial and subsequent economic crisis, a DSGE model with financial frictions is used for the analysis. In our paper we use the model framework developed by Shaari [2] that includes financial accelerator mechanism proposed by Bernanke *et al.* [1].This appropriately complex medium-sized model of a small open economy incorporates important real as well as nominal rigidities and allows us to describe the Czech economy in a reasonable detail. We chose Bernanke *et al.* [1] specification of financial frictions because it explicitly incorporates the interest rate spread and it was the uncoupling of the policy interest rate and the commercial interest rates that was the main manifestation of the financial frictions in the Czech economy during the Great Recession. Structure of the model is quite standard, therefore, we will describe only the most important features of the model.

2.1 Households

The households receive wages for supplied labour, government transfers, profits made by retailers and domestic and foreign bonds returns. Domestic bonds pay fixed nominal return in domestic currency while foreign non-contingent bonds give a risk adjusted nominal return denominated in foreign currency. The debt-elastic risk premium contains exogenous AR(1) component of risk-premium or uncovered interest parity shock. The households then spend their earnings on consumption and domestic and foreign bonds acquisition.

2.2 Entrepreneurs

The entrepreneurs play two important roles in the model. They run wholesale goods producing firms and they produce and own the capital. Market of intermediate goods as well as capital goods market is assumed to be competitive. The wholesale goods production is affected by domestic productivity AR(1) shock and the capital goods production is subject to capital adjustment costs. Entrepreneurs finance the production and ownership of capital K_t by their net-worth N_t and borrowed funds. Cost of borrowed funds is influenced by borrower's leverage ratio via external finance premium $EFP_t = [N_t/(Q_{t-1}K_t)]^{-\chi}$, where Q_t is a real price of capital or Tobin's Q and χ is financial accelerator parameter. To maximize profit, the entrepreneurs choose the optimal level of capital and borrowed funds. Each turn a proportion $(1-A_t^{NW})\varsigma$ of entrepreneurs leaves the market and their equity $(1-A_t^{NW})\varsigma V_t$ is transferred to households in a form of transfers. A_t^{NW} is a shock in entrepreneurial net worth. It influences the development of net worth by changing the bankruptcy rate of entrepreneurs and its positive innovations increase the survival rate of entrepreneurs. Its logarithmic deviation from steady state is assumed to evolve according to AR(1) process. ς is the steady-state bankruptcy rate.

2.3 Retailers

Next, there are two types of retailers in the model. Domestic goods retailers and foreign goods retailers. Both types of retailers are assumed to operate in the conditions of monopolistic competition. Domestic good retailers buy domestic intermediate goods at wholesale price and sell the final domestic goods to the consumers. Foreign good retailers buy goods from foreign producers at the wholesale price and resell the foreign goods to the domestic consumers. The difference between foreign wholesale price expressed in domestic currency and final foreign goods price, i.e. deviation from law of one price is determined by exogenous AR(1) shock. By Calvo-type price setting and inflation indexation of the retailers the nominal rigidities are introduced into the model.

2.4 Central bank

The central bank determines the nominal interest rate in accordance with forward/backward-looking Henderson McKibbin Taylor interest rate rule. Deviations of interest rate from the interest rate rule are explained as monetary policy i.i.d. shocks.

2.5 Foreign sector

The foreign economy variables - real output, CPI inflation and nominal interest rate, are modelled using a structural VAR(1) model.

2.6 Time-varying parameters

The time-varying parameters are defined as unobserved endogenous variables with following law of motion,

$$\theta_t = (1 - \alpha_t^\theta) \cdot \theta_{t-1} + \alpha_t^\theta \cdot \overline{\theta} + \nu_t^\theta \tag{1}$$

where θ_t is a general time-varying parameter, $\overline{\theta}$ is initial value of this parameter, α_t^θ is a time-varying adhesion parameter common for all the remaining time-varying parameters and $\nu_t^\theta \sim N(0, \sigma_\nu^\theta)$ is exogenous innovation in the value of parameter θ_t. Setting of the adhesion parameter α_t^θ influences the tendency of the time-varying parameter θ_t to return to its initial value $\overline{\theta}$. With $\alpha^\theta = 0$, the time-varying parameter would be defined as random walk, while with $\alpha^\theta = 1$, the parameter would be white noise centred around the initial value $\overline{\theta}$. The adhesion parameter α^θ is common for all the remaining time-varying parameters and is itself considered time-varying with adhesion set to 0.01.

3 Estimation technique

Nonlinear particle filter (NPF) is used to identify the unobserved states of the DSGE model, including the time-varying parameters, in this paper. In this section, we briefly describe the main principles of this nonlinear particle filter.

3.1 Nonlinear particle filter

Unlike basic Kalman filter that is optimal only for linear systems with Gaussian noise, the nonlinear particle filter is a more sophisticated tool that can be used even for nonlinear state-space systems with non-Gaussian noise. In this section, we provide only the basic principles of the algorithm. A detailed description can be found for example in Van Der Merwe *et al.* [6] or Yano [8]. In a condensed form, the NPF algorithm can be described as follows:

1. Initialization: $t = 0$, set the prior mean $\overline{x}_0$ and covariance matrix P_0 for the state vector x_t.
2. Generating particles: Draw a total of N particles $x_t^{(i)}$, $i = 1, \ldots, N$ from distribution $p(x_t)$ with mean $\overline{x}_t$ and covariance matrix P_t.
3. Time Update: $t = t + 1$, for each particle $(i = 1, \ldots, N)$ propagate the particle into future with the use of nonlinear transition and measurement equation and calculate means $\overline{x}_{t|t-1}$, $\overline{y}_{t|t-1}$ and covariance matrices $P_{t|t-1}$, $P_{y,y}$ and $P_{x,y}$.
4. Kalman filter: Calculate Kalman gain $K_t = P_{x,y} \left(P_{y,y}\right)^{-1}$, $\overline{x}_t = \overline{x}_{t|t-1} + K_t(y_t - \overline{y}_{t|t-1})$ and $P_t = P_{t|t-1} - K_t P_{y,y} \left(K_t\right)^T$
5. Continue by step 2 until $t = t_{max}$.

In our application we performed 20 runs of the NPF with 30.000 particles each for the second order approximation of the nonlinear DSGE model.

3.2 Initial values

Before the application of the NPF algorithm we estimated the model with constant parameters to obtain the estimates of model parameters. The posterior means of the model parameters were used as initial values of the time-varying parameters ($\overline{\theta}$). The standard deviations of time-varying parameters innovations (σ_ν^θ) were set equal to the 10 % of the posterior means of the structural parameters. Constant

model parameters were estimated using Random Walk Metropolis-Hastings algorithm as implemented in Dynare toolbox for Matlab. Two parallel chains of 1.000.000 draws each were generated during the estimation. First 50 % of draws were discarded as a burn-in sample. The scale parameter was set to achieve acceptance rate around 30 %. Estimation of the model with constant parameters is described in detail in Tvrz *et al.* [3].

4 Data

Quarterly time series of eight observables were used for the purposes of estimation. These time series cover the period between 1999Q1 and 2014Q4 and contain 64 observations each. Time series of real gross domestic product (GDP), harmonised consumer price index (CPI), 3-month policy interest rate and real investment are used for the domestic economy. The foreign economy is represented by the 17 euro area countries. Seasonally adjusted time series of real GDP, harmonised CPI and 3-month policy interest rate are used. Time series of CZK/EUR real exchange rate are also used. These seasonally adjusted time series were obtained from the Eurostat, Czech National Bank, European Central Bank databases. The original time series were transformed prior to estimation so as to express the logarithmic deviations from their respective steady states. Logarithmic deviations of most observables from their trends were calculated with the use of Hodrick-Prescott filter ($\lambda = 1600$). Time series of the domestic and foreign CPI inflation were demeaned.

5 Empirical results

Table 1 presents the priors and posteriors of the structural parameters of the model with constant parameters. Obtained results were used to calibrate the particle filter.

			Prior		Posterior	
Parameter		**Distribution**	**Mean**	**Std**	**Mean**	**Std**
Υ	Consumption habit	Beta	0.60	0.05	0.5937	0.0515
Ψ	Inv. elast. of lab. supply	Gamma	2.00	0.50	1.2431	0.3232
ψ^B	Debt-elastic risk premium	Gamma	0.05	0.02	0.0223	0.0073
η	Domestic/foreign elast. subst.	Gamma	0.65	0.10	0.5989	0.0725
κ	Price indexation	Beta	0.50	0.10	0.2665	0.0678
γ	foreign goods preference bias	Beta	0.40	0.10	0.2931	0.0583
θ_H	Domestic goods Calvo	Beta	0.70	0.10	0.8060	0.0245
θ_F	Foreign goods Calvo	Beta	0.70	0.10	0.7932	0.0324
ψ^I	Capital adjustment costs	Gamma	20.00	5.00	29.800	5.2436
Financial frictions						
Γ	Steady-state leverage ratio	Gamma*	1.50	0.50	1.4654	0.0470
ς	Steady-state bankruptcy rate	Beta	0.025	0.005	0.0313	0.0058
χ	Financial accelerator	Gamma	0.05	0.01	0.0379	0.0070
Taylor rule						
ρ	Interest rate smoothing	Beta	0.50	0.20	0.8612	0.0217
β_π	Inflation weight	Gamma*	1.50	0.20	1.9400	0.2419
Θ_y	Output gap weight	Gamma	0.50	0.20	0.1159	0.0422

*Shifted gamma distribution - parameters are assumed to take values from $\langle 1, \infty)$.

Table 1 Estimated structural parameters

Comparison of filtered trajectories of the time-varying structural parameters conditional on the choice of time-varying and constant model parameters is presented in Figure 1. At first, only the structural parameters are considered time-variyng, next the shock AR(1) parameters and finally the foreign SVAR(1) block parameters are added. The common adhesion parameter is set to $\alpha^\theta = 0.25$. Comparison of the filtered trajetories depending on the calibration of the adhesion parameter α_0^θ is depicted in Figure 2. Only structural parameters are considered time-varying in this case.

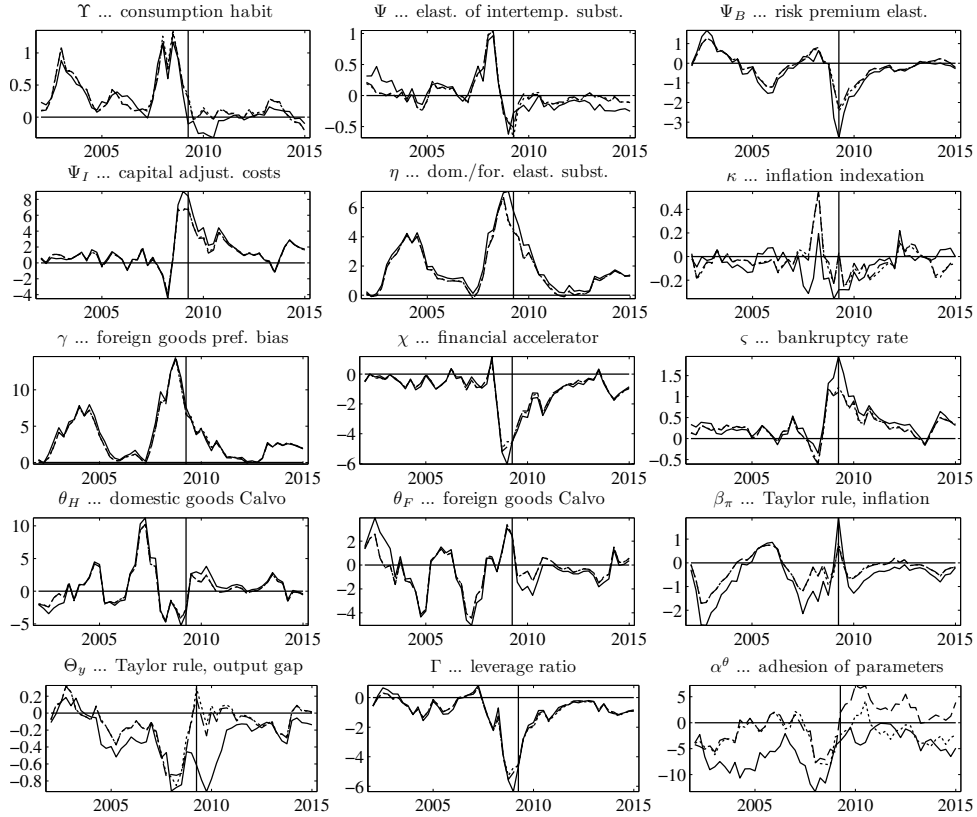


Figure 1 Parameter choice comparison (deviations in per cent of initial values)

Note: solid line - SP, dashed line - SP + AR(1), dotted line - SP + AR(1) + SVAR(1), vertical line - 2009Q1

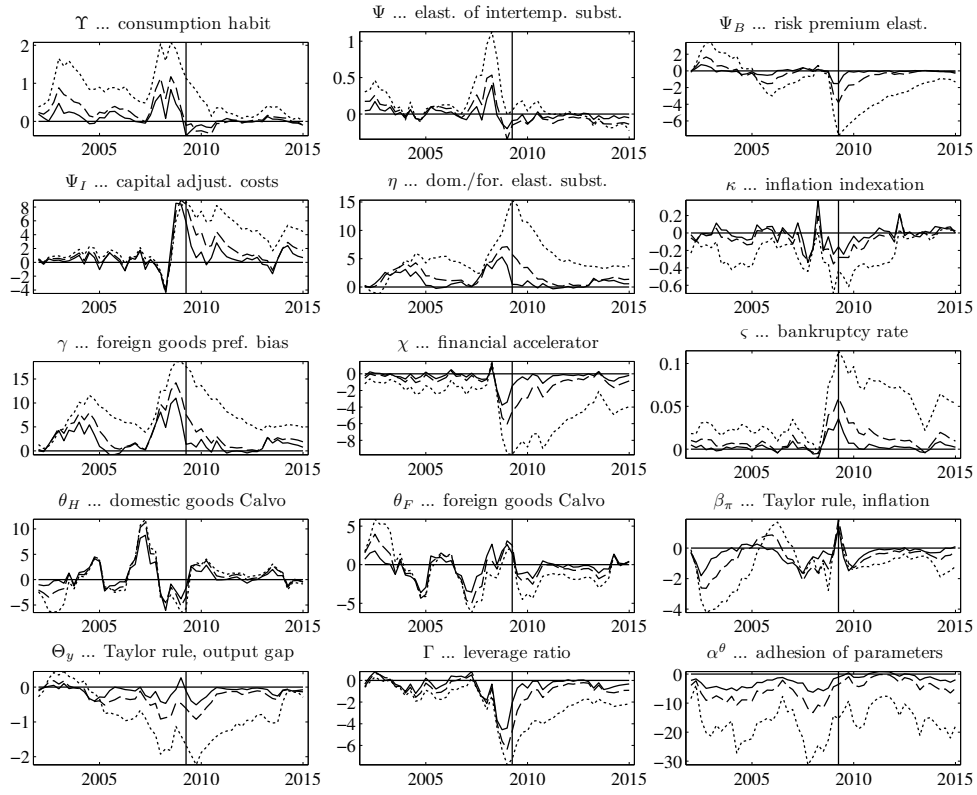


Figure 2 Adhesion calibration comparison (deviations in per cent of initial values)

Note: solid line - $\alpha_0^\theta = 0.5$, dashed line - $\alpha_0^\theta = 0.25$, dotted line - $\alpha_0^\theta = 0.1$, vertical line - 2009Q1

6 Conclusion

According to the presented results, nonlinear particle filter is a relatively robust tool for identification of unobserved trajectories of time-varying parameters in DSGE models. The filtered trajectories of structural parameters with different initial values of common adhesion parameter are very similar. The periods where given parameters increased and decreased are identified almost identically. The differences in maximum deviations from initial values are only minor between the two alternatives with $\alpha_0^\theta = 0.25$ and $\alpha_0^\theta = 0.5$. With $\alpha_0^\theta = 0.1$ the parameters deviate further from their initial values and especially during the Great Recession they stay away for a longer period of time. This is given by the fact that with $\alpha_0^\theta = 0.1$ the model definition of the time-varying parameters approaches a random walk. Most of the trajectories of the time-varying structural parameters were nearly unaffected by the enlargement of the subset of time-varying model parameters. Only the trajectory of the output gap weight in the Taylor rule changed distinctively after the enlargement. This was caused by the fact that the smoothing parameter was considered constant in the baseline specification and was included in the subset of time-varying parameters with the shock AR(1) parameters. Since the smoothing parameter appears only in the interest rate rule it influenced the filtration of remaining time-varying parameters in that equation to some extent.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1235/2014. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated.

References

[1] Bernanke, B. S., Gerlter, M., and Gilchrist, S.: *The Financial Accelerator in a Quantitative Business Cycle Framework.* In: Taylor, J.B., and Woodford, M (eds.) Handbook of Macroeconomics, vol. 1, Elsevier, Amsterdam, 1999, 1341–1393.

[2] Shaari, M. H.: *Analysing Bank Negara Malaysia's Behaviour in Formulating Monetary Policy: An Empirical Approach.* Doctoral thesis, College of Business and Economics, The Australian National University, 2008.

[3] Tvrz, S., Tonner, J., and Vašíček, O.: *Financial Accelerator Mechanism in a Small Open Economy: DSGE Model of the Czech Economy.* In: Ramík, J., and Stavárek, D. (eds.). Proceedings of 30th International Conference Mathematical Methods in Economics, Karviná: Silesian University, School of Business, 2012, 944–949.

[4] Tvrz, S., and Vašíček, O.: *Nonlinear DSGE model of the Czech economy with time-varying parameters.* In: Vojáčková, H. (ed.) Proceedings of the 31st International Conference Mathematical Methods in Economics. Jihlava: College of Polytechnics Jihlava, 2013, 968–973.

[5] Tvrz, S., and Vašíček, O.: *Nonlinear DSGE model of a small open economy with time-varying parameters: Czech economy in a period of recession.* In: Talašová, J., Stoklasa, J., and Talášek, T. (eds.) Proceedings of the 32nd International Conference Mathematical Methods in Economics. Olomouc: Palacký university, 2014, 1068–1073.

[6] Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E.: *The Unscented Particle Filter.* CUED Technical Report 380, Cambridge University Engineering Department, 2000.

[7] Vašíček, O., Tonner, J., and Polanský, J.: *Parameter Drifting in a DSGE Model Estimated on Czech Data.* Finance a úvěr – Czech Journal of Economics and Finance, vol. 61, no. 5, UK FSV, Praha, 2011, 510–524.

[8] Yano, K.: *Time-varying Analysis of Dynamic Stochastic General Equilibrium Models Based on Sequential Monte Carlo Methods.* ESRI Discussion Paper No. 231. Economic and Social Research Institute, Tokyo, Japan, 2010.

Could Prediction Error Help in Fractal Analysis of Time Series?

Lucie Tylova[1], Martin Dlask[1], Jaromir Kukal[1], Quang Van Tran[1]

Abstract. Fractal analysis provides important characterization of any stock market. Time series can be considered as a sample of fractal Brownian motion (fBm). In the case of European stock market indices, logarithmic transform of daily values is necessary as preprocessing. Original fBm is traditionally converted to fractional Gaussian noise (fGn) by differentiation and then analysed in frequency and time domains. Resulting Hurst exponent is a measure of stock fractal behaviour. Novel approach to Hurst exponent estimation is based on fGn generation using forward or backward linear predictors or symmetric smoother, respectively. Prediction error is supposed to be fGn and analysed using various techniques. The paper evaluates efficiency of proposed methods and compares them with traditional one. The second aim is to compare individual stock market indices and interpret their fractal behaviour.

Keywords: Stock market, time series, stock index, fBm, fGn, linear predictor, Hurst exponent.

JEL classification: C53
AMS classification: 60925

1 Introduction

Technical analysis of stock market plays an important role in investment decision making. Dynamics of stock market is traditionally investigated via time series analysis focusing on market indices values, their changes and predictions, fractal behaviour and similarities. This paper compares various approaches to given problem.

There are many subjective and objective factors, which change the stock price in general. From the point of one individual stock market index, the change of values of other indices is only a random factor that can cause their chaotic behaviour. Therefore, we considered fractal data analysis as a suitable technique for both time series and predictor examination.

2 Traditional approach to Hurst exponent estimation

Let m be the length of time series $\{a_k\}_{k=1}^{m}$ where $a_k > 0$ is individual daily value of stock market index which will be subject of statistical analysis. Standard market investigation is based on stock value differences: absolute, relative, or logarithmic ones. Due to symmetry property we prefer logarithmic differences in the form

$$x_k = \ln \frac{a_{k+1}}{a_k} \tag{1}$$

for $k = 1, \ldots, m-1$.

There is a number of methods that can be used for fractal dimension estimation. Under the fBm assumption of investigated signal, every approach is based on different property such as number of zero-crossing, rescaled range, power spectrum or Holder continuity. Since the fBm can be converted to fGn with the same value of parameter H by means of differencing or cumulative sum vice versa, it is sufficient to analyse one type of the fractal stochastic process in one method.

[1]CTU in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, Trojanova 13, Praha 2, Czech Republic, tylovluc@fjfi.cvut.cz

2.1 R/S method

The R/S method, a classical technique of Hurst exponent [4] estimation, repeatedly separates the original sampling of a signal into disjoint segments and calculates mean, standard deviation and range for each of them. Considering the division into r segments with m elements, the R/S statistic is given by

$$(R/S)_m = \frac{1}{r}\sum_{j=1}^{r}\frac{R_j}{S_j}, \tag{2}$$

where S_j and R_j indicates standard deviation and range at interval j, respectively. For different interval lengths m, the R/S statistics $(R/S)_m$ is subsequently plot against m in log-log plot to obtain Hurst parameter using the power law

$$\mathrm{E}\left[(R/S)_m\right] \propto m^H. \tag{3}$$

2.2 Zero-crossing method

Zero-crossing method, independently investigated by Azais [1], Feuerverger [3] and Couerjolly [2] provides point estimate of Hurst parameter H based on the number of zero-crossing of fGn sample. Couerjolly suggests to calculate the relative number of zero crossing S_N with respect to the sample length N, which converges almost surely to function $\frac{\theta(H)}{\pi}$, where $\theta(H)$ is defined as

$$\theta(H) = \begin{cases} \arctan\left(\frac{\sqrt{1-r(H)^2}}{r(H)}\right) & \text{for } r(H) > 0 \\ \frac{\pi}{2} + \arctan\left(\frac{-r(H)}{\sqrt{1-r(H)^2}}\right) & \text{otherwise} \end{cases}$$

and $r(H) = 2^{2H-1} - 1$. For finite N large enough, the Hurst parameter approximately equals

$$H \approx \frac{1}{2}\left(1 + \log_2(1 + \epsilon|\cos(\pi \cdot S_N)|)\right), \tag{4}$$

where $\epsilon = \operatorname{sgn}\left(H - \frac{1}{2}\right)$ is assumed to be known.

2.3 Whittle estimator

Whittle estimator [10] is a method for estimation of fractional parameter d of ARFIMA(p, d, q) process. The covariance function of fGn with $H = d + \frac{1}{2}$ obeys the same power law as ARFIMA(0,d,0), however their spectral densities differ. The aim is to estimate the parameter d_{opt} by solving the minimization problem

$$d_{\text{opt}} \in \underset{d\in\left(-\frac{1}{2};\frac{1}{2}\right)}{\operatorname{argmin}} Q(d) \quad \text{for} \quad Q(d) = \int_{-\pi}^{\pi}\frac{I(\lambda)}{f(\lambda, d)}\mathrm{d}\lambda, \tag{5}$$

where $I(\lambda)$ is the experimental spectral density of investigated sample and $f(\lambda, d)$ is theoretical density of ARFIMA(0,d,0) in the form

$$f(\lambda, d) = \frac{1}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d}. \tag{6}$$

2.4 Quadratic variations

Istas and Lang [5] designed a method for Holder parameter estimation of random process. Since fBm is Holder continuous with parameter $H - \epsilon$ for every $\epsilon > 0$, the upper estimate of it can lead to a good estimation of Hurst exponent H. For a sample of random process $X(i), i = 0, ..., n$ and arbitrary vector $\vec{a} \in \mathbf{R}^{p+1}$ satisfying the condition, that sum of its elements equals zero, they defined quadratic variation with sampling step Δ

$$U(\vec{a}, n, \Delta) = \frac{1}{n}\sum_{j=0}^{n-p}\left(\Delta_{\vec{a}}X_j\right)^2 \quad \text{where} \quad \Delta_{\vec{a}}X_j = \sum_{i=0}^{p} a_i X((i+j)\cdot\Delta). \tag{7}$$

Istas and Lang proved, that the quadratic variation $U(\vec{a}, n, \Delta)$ converges to variance $\sigma^2_{\vec{a},\Delta}$ of finite difference $\Delta_{\vec{a}} X_j$ as n tends to infinity. The variance can be expressed by following formula

$$\sigma^2_{\vec{a},\Delta} \approx -C \cdot (-1)^D \sum_{k=0}^{p} \sum_{l=0}^{p} a_k a_l |k-l|^{2H} \cdot \Delta^{2H}, \tag{8}$$

for constants $C \in \mathbf{R}$, $D \in \mathbf{N}$ and Hurst exponent H. It is possible to estimate the parameter H by utilizing different choices of $\vec{a}$ for nonlinear regression model (8).

3 Prediction error as fGn?

Traditional analysis of chaotic time series is frequently based on assumption about fGn property of logarithmic differences. The main hypothesis beyond the scope of this paper is that prediction error has also fGn property in the case of linear model. But the process of log–difference observation differs from evaluation of prediction error history. Therefore, adequate Hurst exponents of data differences and prediction errors may differ. Novel methodology of prediction fGn analysis brings an alternative view to prediction quality.

Let m be the length of time series $\{x_k\}_{k=1}^m$. Let n be the model size as a number of parameters. The *Forward Linear Predictive Model* (FLPM) has the form

$$x_k = \sum_{i=1}^{n} a_i x_{k-i} + e_k, \tag{9}$$

for $k = n+1, \ldots, m$ where e_k is the model error in the k-th measurement and a_i is the model parameter for $i = 1, \ldots, n$. Formula (9) represents the traditional *Autoregressive Model* (AR) [6].

The predictive AR model (9) can also be used in the opposite time direction. The resulting *Backward Linear Predictive Model* (BLPM) is

$$x_k = \sum_{i=1}^{n} a_i x_{k+i} + e_k, \tag{10}$$

where e_k is again the model error, but for $k = 1, \ldots, m-n$.

The third one is *Symmetric Linear Smoothing Model* (SLSM), and thus with lower prediction error for smooth signals. Supposing n is even, the adequate model is

$$x_k = \sum_{i=1}^{n/2} a_i x_{k-i} + \sum_{i=1}^{n/2} a_{n/2+i} x_{k+i} + e_k, \tag{11}$$

where e_k is the model error for $k = n/2+1, \ldots, m-n/2$.

FLPM, BLPM, and SLSM are also suitable for Alzheimer's disease diagnosis [9] from EEG and can be easily extended to non-linear ANN predictors [8].

4 Application to stock market indices

Daily data of stock market indices were subject of individual investigation during period from 12 December 1991 to 12 March 2015. There are four European indices (CAC40, DAX, FTSE, SMI), two Asian (HSI, NIKKEI), and three North American (SP500, NASDAQ, TSX). At first, the original time series was transformed to logarithmic differences in order to obtain logarithmic return series. Optional predictors and smoother convert logarithmic returns to model residuals. Therefore, the chaotic behaviour of the differences may differ from the chaotic behaviour of residuals in the case of prediction and smoothing.

Traditional methods are used for Hurst parameter estimation of original logarithmic differences and model errors of FLPM, BLMP, and SLSM, considering them as fGn. For each stock market and model residua we will estimate the parameter H by means of using R/S method (R/S), Zero-crossing method (ZC), Whittle estimator (WE) and Quadratic variations (QV). Resulting Hurst exponents can be treated as input values for multi-criteria decision problem. The stock market with overall highest Hurst parameter

index	R/S	ZC	WE	QV	RANK	AIA
CAC40	0.5394	0.4745	0.5223	0.5300	22	2.46
DAX	0.5725	0.5007	0.5388	0.5258	16	1.82
FTSE	0.5138	0.5502	0.4625	0.5396	24	2.68
HSI	0.5307	0.5664	0.4958	0.4981	23	2.44
NASDAQ	0.5717	0.5631	0.5790	0.5202	14*	1.12**
NIKKEI	0.5054	0.4583	0.4957	0.4978	34	3.42
SMI	0.5972	0.5303	0.5358	0.5853	11***	0.90***
SP500	0.5745	0.4722	0.5147	0.4618	25	2.71
TSX	0.5496	0.6107	0.5376	0.5457	11***	1.20*

Table 1 Hurst exponents and predictability via log. differences

is also the most predictable and therefore suitable for investments. To determine this crucial value, we use AIA and RANK multi-criterion decision methods [7].

Concerning the analysis of raw logarithmic differences, Table 1 presents Hurst parameter estimation together with multi-criteria view. The SMI and TSX stock market obtained the highest Hurst exponent according to both multi-criterion decision methods. The least predictable is NIKKEI with scores 34 and 3.42 from RANK and AIA, respectively.

index	R/S	ZC	WE	QV	RANK	AIA
CAC40	0.5237	0.4856	0.5332	0.4990	18*	1.71
DAX	0.5242	0.4768	0.5140	0.4993	22	2.01
FTSE	0.5217	0.5031	0.5016	0.4997	20	1.88
HSI	0.4819	0.5174	0.4696	0.5001	23	2.64
NASDAQ	0.4706	0.5547	0.5237	0.4996	19	1.65*
NIKKEI	0.5056	0.4707	0.4767	0.4978	31	3.25
SMI	0.5510	0.4784	0.5197	0.5001	15**	1.40**
SP500	0.5298	0.5217	0.5266	0.5013	8***	0.84***
TSX	0.5006	0.5673	0.4876	0.4969	24	2.34

Table 2 Hurst exponents and predictability via forward prediction

Next, we investigated fractal behaviour of forward prediction model with H estimates in Table 2. The SP500 is the stock market with the least chaotic residua since the resulting value from both AIA and RANK method is by far the smallest of all coefficients. Whereas the characteristics of this stock market are a little bit inconspicuous in table 1, in the case of BLMP errors, the method can lead to more accurate results.

index	R/S	ZC	WE	QV	RANK	AIA
CAC40	0.5522	0.4957	0.5392	0.4988	17**	1.39*
DAX	0.5451	0.4837	0.5306	0.4991	17**	1.66
FTSE	0.5278	0.5087	0.5009	0.4990	23	2.01
HSI	0.4976	0.5127	0.4655	0.5001	23	2.52
NASDAQ	0.4860	0.5461	0.5209	0.4989	22	1.82
NIKKEI	0.5355	0.4722	0.4796	0.5006	23	2.21
SMI	0.5686	0.4810	0.5117	0.4999	17**	1.46**
SP500	0.5447	0.5170	0.5247	0.4995	14***	1.23***
TSX	0.5134	0.5568	0.4854	0.4968	24	2.40

Table 3 Hurst exponents and predictability via backward prediction

Hurst exponents obtained from backward prediction in Table 3 show similar dependence as in the case of forward prediction. The SP500 has again the highest value of Hurst exponent in the multi-criteria point of view, however, the values are not as much convincing as in the previous case.

index	R/S	ZC	WE	QV	RANK	AIA
CAC40	0.1868	0	0	0	25.5	3.56
DAX	0.1679	0	0	0	26.5	3.88
FTSE	0.1607	0	0	0	27.5	4.00
HSI	0.2015	0.0062	0	0	18.5	3.19
NASDAQ	0.1944	0.0358	0.2187	0	13.5***	1.68**
NIKKEI	0.2012	0	0	0	22.5	3.32
SMI	0.2206	0	0.0878	0	15.5*	2.60*
SP500	0.1930	0.0475	0	0.0177	14**	1.46***
TSX	0.2093	0.0252	0	0	16.5	2.66

Table 4 Hurst exponents and predictability via symmetric smoothing

The case of symmetric smoother is also a kind of ex post data analysis of low practical importance. Results of smoothing error analysis are collected in Table 4. Adequate Hurst exponent estimates are close to zero which is a typical phenomenon for a signal with discontinuities.

5 Conclusion

Various techniques (R/S, ZC, WE, QV) produced slightly different point estimates of Hurst exponent in particular cases. Utilizing RANK and AIA methods for multiple-criteria decision-making, we analysed both pure logarithmic differences and errors of prediction with following general conclusions:

- the RANK method provided similar decisions as AIA approach, therefore the methods can substitute each other,
- Swiss financial index SMI together with Canadian industrial index TSX exhibit slow changes, therefore higher values of Hurst exponents were obtained in the case of logarithmic differences,
- concerning forward prediction error, US mixed index SP500 and Swiss financial index SMI exhibited high values of Hurst exponents, because financial markets are domain of more sophisticated predictive analysis,
- backward prediction error of low practical importance was used for comparison and exhibited similar properties as forward prediction,
- symmetric smoothing only generates a noise with discontinuities, therefore it is not suitable for fractal analysis,
- point estimates of H are very close to $1/2$ (white noise model) in the case of log. differences and corresponding confidence intervals include $H = 1/2$: therefore, our model is comparable with white noise model,
- both R/S and QV methods have highest value of H for log. differences in the case of the best compromise solution,
- QV method has highest value of H for forward prediction in the case of the best compromise solution,
- therefore, Quadratic variations (QV) method is recommended as representative instead of RANK or AIA compromise.

Eventually, we conclude that applying QV method or combination of four dimension estimation methods together with RANK or AIA decision can help to analyse both logarithmic differences and forward predictors of stock market indices and can produce results that might be useful for investment recommendations.

Acknowledgements

The paper was created with the support of CTU in Prague, Grant SGS14/208/OHK4/3T/14.

References

[1] Azais, J. M.: Conditions for convergence of number of crossings to the local time: Application to stable processes with independent increments and to Gaussian processes. *Probability and mathematical statistics* **11.**, 1 (1990), 19–36.

[2] Couerjolly, J. F.: Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *Journal of statistical software* **5.**, 7 (2000), 1-53.

[3] Feuerverger, A., Hall, P., Wood, A. T. A.: Estimation of Fractal Index and Fractal Dimension of Gaussian Process by Counting the Number of Level Crossings. *Journal of Time Series Analysis* **15.**, 6 (1994), 587-606.

[4] Hurst, H. E.: Methods of Using Long-Term Storage in Reservoirs. *ICE Proceedings* **5.**, 5 (1956), 519–543.

[5] Istas, J., Lang P.: Quadratic variations and estimation of the local Holder index of a Gaussian process. *Annales de l'Institut Henri Poincare Probability and Statistics* **33.**, 4 (1997), 407–436.

[6] Priestley, M. B.: *Non-linear and Non-stationary Time Series Analysis*, Academic Press, 1988.

[7] Triantaphyllou, E.: *Multi-Criteria Decision Making: A Comparative Study*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[8] Tylova, L., Kukal, J., Vysata, O.: Spectral Analysis of Predictive Error in Alzheimer's Disease Diagnostics, *Neural Network World* **23** (2013), 427–434.

[9] Tylova, L., Kukal, J., Vysata, O.: Predictive Models in Diagnosis of Alzheimer's Disease from EEG, *Acta Polytechnica* **53** (2013), 94–97.

[10] Whittle, P.: Estimation and information in stationary time series. *Arkiv for Matematik* **3.**, 5 (1953), 423–434.

The Marketization of the Education Industry in the Czech Republic

Gabriela Vaceková[1], Jana Soukopová[2]

Abstract. The marketization of nonprofit organizations is of both theoretical relevance and practical importance. In the presented paper, we examined the scope of nonprofit marketization in the post-transitional context of the Czech Republic in order to determine whether the nonprofit commercial revenues are replacements for or additions to other funding sources. We focused on the Czech education industry. The research sample covered 89 public secondary schools from the South Moravian Region. The methodology was based on econometric and statistical methods using linked open data; the time sequence covered three years (2012 – 2014). We based our approach on the work of McKay et al. [20], adapted for Czech conditions. The authors used the Herfindahl-Hirschman Index and Modern Portfolio Theory to measure revenue diversification. This enabled the exploration of whether and to what extent the education industry in the Czech Republic has succumbed to market forces.

Keywords: marketization, commercialization, nonprofit organizations, education industry, Czech Republic.

1 Introduction

All over the world, nonprofit organizations (hereinafter referred to as NPOs) are increasingly turning to commercial activities. The commercialization involves a growing market orientation and entrepreneurial activism (cf. [23], [28], [34]). The diversity of the manifestations of the commercialization process is reflected in a host of definitional approaches [19]. Salamon [26] referred to commercialization as the "marketization of welfare"; Weisbrod [36] understood it in terms of the increasing tendency of nonprofits to develop new enterprises, charge fees, and produce goods for sale. According to Tuckman [31], commercialization occurs when NPOs "decide to produce goods and services with the explicit intent of earning a profit". "Marketization" is a related term used to refer to NPOs becoming "more market driven, client driven, self-sufficient, commercial, or business-like" [20].

Following a theory and research review that indicated the specificity of particular issue areas in relation to nonprofit commercialization, we focused on the education industry. This field could be considered to be an ideal mixed-mode marketplace where the processes of overlapping and interaction of non-profit and for-profit activities of various organizations might be easily observed. This is encapsulated in the concept of hybrid organizations, which emphasizes the merging of principles from different fields [21].

The attention must be placed on the need to achieve financial stability and sustainability for these organizations [30]. The theoretical part of the paper was thus based on relevant economic theories of the nonprofit sector (e.g., [12-13], [35-36], [10] [27], [16-17], [25]), allowing a comprehensive classification of financial resources in line with actual trends in international research.

According to the resource dependency theory, NPOs use commercial income as a replacement for lost government grants and private revenues [7-8], [32], [20]. However, nonprofit scholars have provided little empirical evidence for or against this theory. Kerlin and Pollak [23], favoring institutional theory [22], [29], showed that there is little evidence that the increase in commercial revenues between 1982 and 2002 was associated with declines in government grants and private contributions. This perspective suggests that the survival of an NPO requires that it conform to the institutional environment in which it exists [18]. In this case, the nonprofit commercial revenues would serve as a complement to other funding sources.

Hence, the aim of the paper was to examine the scope of nonprofit marketization in the Czech Republic based on the case of the education industry in order to determine whether the nonprofit commercial revenues of

[1] Masaryk University, Faculty of Economics and Administration, Department of Public Economics, Lipová 41a, 602 00 Brno, gabriela.vacekova@econ.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Department of Public Economics, Lipová 41a, 602 00 Brno, soukopova@econ.muni.cz

secondary schools in South Moravian Region are replacements for or additions to other funding sources, while focusing on their portfolio structure. We asked the following research questions:

RQ1: *Are nonprofit commercial revenues of educational NPOs an addition to or a replacement for other financial sources?*

RQ2: *What is the scope of the profit-oriented earned-income activities of educational NPOs and their portfolio structure?*

2 Data and methodology

The methodology is based on econometric and statistical methods using linked open data from free Czech Ministry of Finance databases (ÚFIS and MONITOR) and from secondary school annual reports. The analysis period covered three years, from 2012 to 2014, and the sample consisted of all 89 public secondary schools in the South Moravian Region.

We used statistical analysis (mean, median, etc.) and correlation analysis to show the correlations between commercial revenue, total income, and other explanatory variables. This enabled us to show the bivariate relationships between all the variables, and to indicate whether any were positive and statistically significant.

We next used the Herfindahl-Hirschman Index (HHI) and Modern Portfolio Theory (MPT) to measure the diversification of the NPO commercial revenues. Secondary schools derive commercial revenues in different ways, e.g., sales of services, goods, or own products, and/or rental of assets. We questioned what determined the commercial revenue mix in secondary schools.

The methodology used was developed by [6], [11], [5], [37] and adapted for Czech conditions with the HHI. The baseline equation of the commercialization diversification index *DI* is:

$$DI = \sum_{i=1}^{N} \left(\frac{r_i}{R} \right)^2 \tag{1}$$

Where N is the number of revenue commercialization sources, r_i is the i-th commercialization revenue, and R is the total commercialization revenue. $DI \rightarrow \min$, $0 < DI \leq 1$, $DI = 1$ if there is only one commercial revenue source.

The specific context within which we examined this issue was the education industry in the Czech Republic, specifically different types of secondary schools in South Moravian Region. Table 1 shows the structure of the research sample.

Type of Secondary School	Number of Schools
College preparatory school	27
Brno	9
other municipalities	18
Secondary vocational school	62
Brno	15
other municipalities	37

Table 1 Number of different types of secondary schools

In the analysis, we examined the research sample as a whole, and afterwards we used cluster analysis while focusing on college preparatory schools and secondary vocational schools separately. We also examined the secondary schools from the city of Brno in comparison to other municipalities in order to determine whether the results connected to marketization differed.

3 Results and discussion

Table 2 shows the descriptive statistics of average values (2012-14) for public secondary schools from the South Moravian Region.

A great statistical variance in the values was evident. This could influence the results of other analyses. Therefore, we split the sample into two sets according to different school types (college preparatory schools and

secondary vocational schools). We also focused on secondary schools in the city of Brno as compared to other municipalities. Tables 3 and 4 show the results of the splits.

	Total	Mean	SD	Median	Range
Total income	1 152 674.96	38 529.56	24 573.73	32 008.69	9 010.74-163 310.54
Institutional income	1 119 285.23	39 974.47	23 075.44	31 909.61	9 007.67-162 627.73
Commercial revenues	33 389.73	1 192.49	2 577.23	596.38	0.00-17 417.95
Commercial profit	7 724.07	275.86	714.56	129.89	-9.59-6 152.56

Table 2 Descriptive statistics of research sample (*n*=89)

	Total	Mean	SD	Median	Range
Total income	839 273.99	31 084.22	12 035.86	28 676.19	11 505.65-66 237.58
(Brno)	(486 951.97)	(34 782.28)	(12 885.24)	(32 768.51)	
Institutional income	813 471.12	30 128.56	10 326.18	28 367.55	11 505.65-50 905.57
(Brno)	(464 980.96)	(33 212.93)	(10 213.81)	(32 589.24)	
Commercial revenue	25 802.87	955.66	3 243.34	273.25	0.00-17 417.95
(Brno)	(21 971.01)	(1 569.36)	(4 404.16)	(323.44)	
Commercial profit	3 337.99	123.63	186.33	73.86	0.00-987.72
(Brno)	(1 555.65)	(111.12)	(65.89)	(107.10)	

Table 3 Descriptive statistics of college preparatory schools (*n*=27)

	Total	Mean	SD	Median	Range
Total income	1 152 674.96	38 529.56	27 691.02	34 284.93	9 010.74-163 310.54
(Brno)	(662 939.61)	(41 786.77)	(33 629.76)	(41 454.13)	
Institutional income	1 119 285.23	39 974.47	25 878.33	34 202.29	9 007.67-162 627.73
(Brno)	(648 616.20)	(43 241.08)	(31 991.94)	(47 906.78)	
Commercial revenue	33 389.73	1 192.49	2 184.79	891.17	0.00-17 417.95
(Brno)	(14 323.41)	(954.89)	(2 447.95)	(710.07)	
Commercial profit	7 724.07	275.86	835.94	168.41	-9.59-6 152.56
(Brno)	(2 349.26)	(156.62)	(1 162.05)	(158.76)	

Table 4 Descriptive statistics of secondary vocational schools (*n*=62)

From the results provided in Tables 2-4, it is clear that nonprofit commercial revenues were a complement to other financial sources (given the ratio to total income), especially for the institutional incomes of secondary schools. No difference was found between college preparatory schools and secondary vocational schools.

The correlation coefficients in Table 5 show correlations between commercial revenues and total income for the whole sample and the two sets of secondary schools.

	2012	2013	2014	2012-14
Whole sample	0.4668	0.4037	0.4105	0.4282
College preparatory schools (Brno)	0.7230	0.8298	0.5884	0.6244
Secondary vocational schools (Brno)	0.4128	0.4240	0.3349	0.4253

Table 5 Correlation coefficients

This result was not surprising. It was connected to the fact that the share of commercial profit did not exceed 3 percent at college preparatory schools, while it exceeds 10 percent in some of the secondary vocational schools, and there are schools with more than 30 percent.

In the following text we present the results of an analysis that addressed the research question: *What is the scope of the profit-oriented earned-income activities of education NPOs and their portfolio structure?*

The profit and loss statements of secondary schools contained the following types of commercial revenues:

- sales of services;
- sales of goods;
- sales of own products;
- rentals;
- other income.

Table 6 contains descriptive statistics of different types of commercial revenues.

	Total	**Mean**	**SD**	**Median**	**Range**
Services	68 903.51	931.13	1752.467	274.44	1-9 994.19
Goods	3 537.24	294.77	400.404	52.26	2.55-1 199.5
Own products	1 256.53	157.07	148.213	112.60	1-360.91
Rentals	24 603.75	410.06	588.068	146.30	1-2 897.90
Other	33 993.22	596.37	3200.645	68.16	0.05-2 4484.38

Table 6 Descriptive statistics of research sample (n=89)

From the results presented in Table 6, it is evident that the sale of services provided the largest share of commercial revenue (almost 50 percent). Other income (25 percent) and rentals (18 percent) formed an important part of the commercial revenues as well. Other types of commercial activities (sales of goods or own product) were only minor.

When we examined the diversification of commercial income according to the HHI and MPT (1), the result for the whole sample ($DI_{n=89}$) was 0.372686, which means that there was a considerable diversification of commercial revenues. Why did secondary schools approach revenue diversification? One possible answer could be that the diversification of commercial revenues can increase community buy-in and organizational legitimacy [9], [2].

Results for college preparatory schools varied, DIn=27 = 0.636798. College preparatory schools in the city of Brno had a higher diversification index: DIn=9 = 0.829049. This may mean that, thanks to a steady supply of institutional support and a stable number of students, college preparatory schools in Brno did not need to offer services beyond the sale of services and rentals. For other municipalities, the HHI was similar to the whole sample $DI_{n=18}$ = 0.445107.

The diversification of commercial revenues was larger for secondary vocational schools: $DI_{n=89}$ = 0.520137, which was expected and understandable. Secondary vocational schools have a better chance of selling their own products and services (horticultural services, electrical engineering, health, etc.). In Brno, the diversification of commercial revenues was even higher $DI_{n=15}$ = 0.404939 than other municipalities, $DI_{n=37}$ = 0.665846, which could be linked to the urban/rural division (cf. [15], [3]).

This research is still in progress; more analyses need to be done in order to better understand the underlying processes affecting the nonprofit marketization in the Czech Republic. Nevertheless, these pilot results provided important new insights into the marketization phenomenon in the Czech Republic.

4 Conclusion

The aim of the presented paper was to examine the scope of nonprofit marketization in the Czech Republic. Based on the case of the education industry, we sought to determine whether the nonprofit commercial revenues of secondary schools in the South Moravian Region were replacements for or additions to other funding sources, while focusing on their portfolio structure.

Our results were both surprising and expected. In the transitional context of the Czech Republic, we could not prove the assumptions of the resource dependency theory. The case of secondary schools in the South Moravian Region favored the institutional theory, as the commercial revenues could be seen only as an addition to other funding sources. In relation to the Hirschman-Herfindal Index (HHI), it was obvious that HHI is a

suitable and useful tool for measuring commercialization diversification. Although the HHI was developed primarily for measuring the concentration of competition, it may serve other purposes, as was confirmed in the research by [6], [11], [37].

NPOs can become business-like in any dimension in more or less comprehensive ways, maintaining alternative orientations to varying degrees [21]. It has been argued that hybridity has long been, and will remain, a characteristic of NPOs (e.g. [4]; [21]). A significant amount of scientific literature has been devoted to this issue. However, this is not the case in the Czech Republic, even though the trend towards emerging hybridization in the transitional economies of CEE has increased significantly in recent years as more NPOs explore revenue generation opportunities [33]. A complex approach that would enable reflection on the specific nature of hybridity in a transitional context as well as on current public debates and policy making discourses on the subject is lacking. Future research should fill this gap.

References

[1] Current, J., and Weber, C.: Application of facility location modelling constructs to vendor selection problems. *European journal of operational research*, **76**(1994), 387-392.

[2] Bielefeld, W.: What affects nonprofit survival? *Nonprofit Management and Leadership*, **5**(1994), 19–36.

[3] Bielefeld, W., and Murdoch, J. C.: The locations of nonprofit organizations and their for-profit counterparts: An exploratory analysis. *Nonprofit and Voluntary Sector Quarterly*, **33**(2004), 221-246.

[4] Brandsen, T., Van de Donk, W., and Putters, K.: Griffins or chameleons? Hybridity as a permanent and inevitable characteristic of the third sector. *Intl Journal of Public Administration*, **28**(2005), 749-765.

[5] Carroll, D. A., and Stater, K. J.: Revenue diversification in nonprofit organizations: Does it lead to financial stability? *Journal of Public Administration Research and Theory*, **19**(2009), 947–966.

[6] Chang, C. F., and Tuckman, H. P.: Revenue diversification among non-profits. *Voluntas*, **5**(1994), 273–290.

[7] Crimmins, J. C., and Keil, M.: *Enterprise in the nonprofit sector*. Americans for the Arts, 1983.

[8] Eikenberry, A. M., and Kluver, J. D.: The marketization of the nonprofit sector: civil society at risk? Public administration review, **64**(2004), 132-140.

[9] Galaskiewicz, J.: Estimating point centrality using different network sampling techniques. *Social Networks*, **13**(1991), 347–386.

[10] Grand, J.: Quasi-markets and social policy. *The Economic Journal*, (1191), 1256-1267.

[11] Greenlee, J. S., and Trussel, J. M.: Predicting the financial vulnerability of charitable organizations. *Nonprofit Management and Leadership*, **11**(2000), 199–210.

[12] Hansmann, H. The role of nonprofit enterprise. *Yale law journal*, (1980), 835-901.

[13] Hansmann, H. Economic theories of nonprofit organization. In *The nonprofit sector: A research handbook*. 1987, 27-42.

[14] Hansmann, H.: When does worker ownership work? ESOPs, law firms, codetermination, and economic democracy. *Yale Law Journal*, (1990), 1749-1816.

[15] Henisz, W. J., and Delios, A.: Uncertainty, imitation, and plant location: Japanese multinational corporations, 1990-1996. *Administrative science quarterly*, **46**(2001), 443-475.

[16] James, E.: The public/private division of responsibility for education: an international comparison. *Economics of Education Review*, **6**(1987), 1-14.

[17] James, E. Economic theories of the nonprofit sector: A comparative perspective. In *The third sector, comparative studies of non-profit organizations.* De Gruyter, Berlin, New-York, 1990, 21-30.

[18] Kerlin, J. A., and Pollak, T. H.: Nonprofit Commercial Revenue A Replacement for Declining Government Grants and Private Contributions? *The American Review of Public Administration*, **41**(2011), 686-704.

[19] LeRoux, K. M.: What drives nonprofit entrepreneurship? A look at budget trends of metro Detroit social service agencies. *The American Review of Public Administration*, **35**(2005), 350-362.

[20] McKay, S., Moro, D., Teadsdale, S., and Clifford, D.: The marketization of charities in England and Wales. *Voluntas*, **26** (2011),336–354.

[21] Maier, F., Meyer, M., and Steinbereithner, M.: Nonprofit Organizations Becoming Business-Like A Systematic Review. *Nonprofit and Voluntary Sector Quarterly*, (2014), 0899764014561796.

[22] Meyer, J. W., and Rowan, B.: Institutionalized organizations: Formal structure as myth and ceremony. *American journal of sociology*, (1977), 340-363.

[23] Nicholls, A., and Cho, A. H.: Social entrepreneurship: The structuration of a field. In *Social entrepreneurship: New models of sustainable social change.* Oxford: Oxford University Press, 2006.

[24] Powell, W. W., and Owen-Smith, J.: Universities as creators and retailers of intellectual property: Life-sciences research and commercial development. In *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector.* Cambridge: Cambridge University Press, 2000.

[25] Salamon, L. M.: Of market failure, voluntary failure, and third-party government: Toward a theory of government-nonprofit relations in the modern welfare state. Nonprofit and Voluntary Sector Quarterly, **16**(1987), 29-49.
[26] Salamon, L. M.: The Marketization of Welfare: Changing Nonprofit and For-Profit Roles in the American Welfare State. *Social Service Review*, **67**(1993), 16–39.
[27] Salamon, L. M., and Anheier, H. K.: Social origins of civil society: Explaining the nonprofit sector cross-nationally. *Voluntas: International journal of voluntary and nonprofit organizations*, **9**(1998), 213-248.
[28] Sharir, M., and Lerner, M. Gauging the success of social ventures initiated by individual social entrepreneurs. *Journal of world business*, **41**(2006), 6-20.
[29] Scott, W. R. *Institutions and organizations*. Thousand Oaks: SAGE, 1995.
[30] Svidroňová, M. Sustainability strategy of non-government organisations in Slovakia. *E+M Ekonomie a management*, (2013), 3-30.
[31] Tuckman, H. P. Competition, commercialization, and the evolution of nonprofit organizational structures. *Journal of Policy Analysis and Management*, **17**(1988), 175-194.
[32] Young, D.: New trends in the US non-profit sector: Towards market integration. *The Non-Profit Sector in a Changing Economy*. OECD: Paris, 2003.
[33] Vaceková, G., and Prouzová, Z.: *Non-profit Commercial Revenues. Evidence from the Czech Republic*. WP No. 16, 2014.
[34] Weerawardena, J., McDonald, R. E., and Mort, G. S.: Sustainability of nonprofit organizations: An empirical investigation. *Journal of World Business*, **45**(2010), 346-356.
[35] Weisbrod, B. A.: *The voluntary nonprofit sector: An economic analysis*. Lexington Books, 1977.
[36] Weisbrod, B. A. The Nonprofit Economy. Cambridge: Harvard University Press, 1988
[37] Wicker, P., Longley, N., and Breuer, C.: Revenue volatility in German nonprofit sports clubs. *Nonprofit and Voluntary Sector Quarterly*, **44** (2013), 5–24.

Alterations of the Causal Model to Make the Causal Effect Identifiable

Lucie Váchová[1]

Abstract. In a number of disciplines, economy included, the causal relations among variables exist. Based on economic theories the causal model can be designed to represent the relations among variables (here the causal model means a diagram in which a link between variables appears whenever the causal connection exists). Under certain conditions the causal effect of intervention (i.e., the impact of setting of the fixed value by an external force on the specific set of variables in the diagram) is identifiable as discussed by Galles and Pearl in [1]. The aim of this paper is to point out different types of alterations executed in a model where specific causal effect is not identifiable in order to make its identifiability possible. The model that includes unmeasurable variable is used. It is shown that even minor alterations of original model can cause that the effect can be expressed without necessity to employ unobservable variable and thus becomes computable.

Keywords: causal effect, identifiability, intervention, unmeasurable variable.

JEL classification:
AMS classification:

1 Introduction

The contribution is focused on identification of the causal effect in causal models and it considers Pearl's approach to causality. The example of using this approach in economic application is shown for instance in [5]. Economic situations (variables and relations among them) used in the paper are significantly simplified. The basic model (see Figure 1) pictures the relations among only four variables represented by nodes A, B, C and D. For the purpose of the application we can assume that these nodes are representatives of for example unmeasurable variable crisis A, and measurable variables selling price of the product B, the quantity of the product sold C and profit D. All variables are considered to be discrete and have finite number of values.

First of all let us introduce some basic notions regarding the causal effect identification. A *causal structure* [3] of a set of variables V is a directed acyclic graph (DAG) in which each node corresponds to a distinct element of V, and each link represents a direct functional relationship among the corresponding variables.

A *causal model* [3] is a pair $M = \langle S, \Theta_S \rangle$ (S is a causal structure, Θ_S is a set of parameters that are compatible with S). The parameters Θ_S assign a function $x_i = f_i(pa_i, u_i)$ to each $X_i \in V$ and a probability measure $P(u_i)$ to each U_i, where pa_i are particular values of variables from PA_i, i.e., of the parents of X_i in S and where each U_i is a random disturbance (*unmeasured variable*) distributed according to $P(u_i)$, independently of all other U (u_i represents an instantiation of U_i).

An *intervention* (denoted $do(X = x)$ or $do(x)$ or $\hat{x}$) is setting the fixed value by an external force on the specific set of variables in the diagram [3].

Given two disjoint sets of variables X and Y, a *causal effect* of X on Y is a function from X to the space of probability distributions on Y and is denoted by $P(y \mid \hat{x})$. For each realization x of X, $P(y \mid \hat{x})$ gives the probability of $Y = y$ induced by deleting from the causal model all equations corresponding to variables in the set X and substituting $X = x$ in the remaining equations [3]. That means the value of x that is selected by the control (the intervention) is used in the new equations [4].

[1]University of Economics, Prague/Faculty of Management, Department of Exact Methods, Jarosovska 1117/II, 37701, Jindrichuv Hradec, Czech Republic, vachova@fm.vse.cz

The causal effect of X on Y is *identifiable* from a graph G if the quantity $P(y \mid \hat{x})$ can be computed uniquely from any positive probability of the observed variables (see, e.g., [2]). Provided that all variables in model are measurable, the causal effect for atomic intervention (i.e., intervention on a single variable, $do(X_i = x_i^{'})$) can be computed using *truncated factorization* [3] formula as:

$$P(x_1, \ldots, x_n|\hat{x}_i^{'}) = \begin{cases} \prod_{j \neq i} P(x_j|pa_j) & \text{if} \quad x_i = x_i^{'} \\ 0 & \text{if} \quad x_i \neq x_i^{'} \end{cases} \tag{1}$$

A *back-door path* [1] from X to Y in a graph G is a path from X to Y that contains an arrow pointing into X.

In case that unmeasured variables are involved, the following four graphical conditions of Galles and Pearl (see [1]) can be used to identify a causal effect.

Theorem 1 (Galles and Pearl 1995)**.** *Let X and Y denote two singleton variables in a causal model characterized by a directed acyclic graph (DAG) G. A necessary and sufficient condition for identifiability of $P(y \mid \hat{x})$ in a graph G is that G satisfies one of the following four conditions:*

1. *There is no directed path from X to Y in G.*
2. *There is no back-door path from X to Y in G, that is $(X \perp Y)_{G_{\underline{X}}}$.*
3. *There exists a set of nodes S that blocks all back-door paths from X to Y such that $P(s \mid \hat{x})$ is identifiable.*
4. *There exist sets of nodes Z_1 and Z_2 such that:*
   - *(i) No element of Z_2 is a descendant of X;*
   - *(ii) Z_1 blocks every directed path from X to Y (i.e., $(Y \perp X \mid Z_1)_{G_{\overline{Z_1}\,\overline{X}}}$);*
   - *(iii) Z_2 blocks all back-door paths between Z_1 and Y in $G_{\overline{X}}$ (i.e., $(Y \perp Z_1 \mid Z_2)_{G_{\overline{X}\,\underline{Z_1}}}$);*
   - *(iv) Z_2 blocks all back-door paths between X and Z_1 (i.e., $(X \perp Z_1 \mid Z_2)_{G_{\overline{X}}}$);*

Where $G_{\overline{X}}$ is a graph derived from G by deleting all arrows heading to nodes in X and $G_{\underline{Z_1}}$ is a graph constructed from G by deleting all arrows emerging from nodes of Z_1.

2 Identification of causal effect

2.1 Basic model

In all presented examples it is our aim to compute the causal effect $P(d|\hat{b})$, i.e. the causal effect of the intervention on variable B (selling price) on variable D (profit). First, let us consider the basic model as pictured in Figure 1. The joint probability distribution of this model is expressed as

$$P(a, b, c, d) \quad = P(a) \cdot P(b|a) \cdot P(c|a, b) \cdot P(d|b, c).$$

The intervention on B (represented by $\hat{b}$ in the following formulas) is stated by removing the B factor

$$P(a, c, d|\hat{b}) \quad = P(a) \cdot P(c|a, b) \cdot P(d|b, c).$$

And the causal effect of the intervention is finally expressed through summation over A and C

$$P(d|\hat{b}) \quad = \textstyle\sum_{a,c} P(a) \cdot P(c|a, b) \cdot P(d|b, c). \tag{2}$$

It can be easily proven that the desired effect is not identifiable since none of the four conditions in the Theorem 1 is satisfied. Condition 1 cannot be satisfied based on the existence of directed paths between B and D. Condition 2 is not satisfied because the back-door path exists from B to D, the path

$$B \leftarrow A \rightarrow C \rightarrow D.$$

For any possible set S the Condition 3 cannot hold since the unmeasured variable A cannot be removed from computations and thus the causal effect $P(c \mid \hat{b})$ is not identifiable. Since the directed path from B to D is the direct edge, there is no such set that would block this path. It means the Condition 4 does not hold.

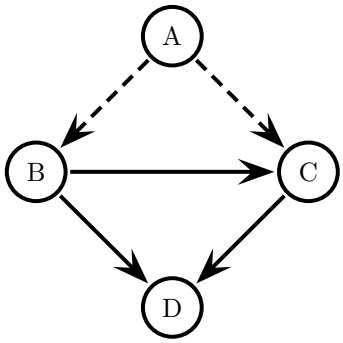


Figure 1 A graph shows the relations of variables under consideration.

2.2 Possible alterations

Let us consider the impact of some possible changes in the basic model. Since the impossibility to compute the causal effect is the result of the unmeasurable variables being the part of the model, firstly the attention will be focused on variable A. In case this variable was incorrectly labeled as unmeasurable even though in fact it is only unmeasured, it brings the immeadiate solution to the problem of non-identifiability of the desired causal effect and it can be computed as follows. From the original joint probability distribution

$$P(a,b,c,d) \quad = P(a) \cdot P(b|a) \cdot P(c|a,b,) \cdot P(d|b,c).$$

using the truncated factorization formula (1) we get

$$P(a,c,d|\hat{b}\,) \quad = P(a) \cdot P(c|a,b) \cdot P(d|b,c).$$

And since all the variables are now measurable we can compute the causal effect as

$$P(d|\hat{b}\,) \quad = \textstyle\sum_{a,c} P(a) \cdot P(c|a,b) \cdot P(d|b,c).$$

Let us return to the situation assuming that variable A is indeed unmeasurable. Following the definiton of identifiable causal effec as mentioned in Section 1, we must look for such alterations in the basic model that would enable us to compute the causal effect without necessity to employ this variable. We will discuss several possibilities, namely deleting the edges, changing directions of edges, and introduction of new variable into the basic model.

First group of changes introduced here consists of deleting certain edges from the basic model. By deleting the links between nodes we in fact deny the relations between variables represented by these nodes. Thus such an alteration in a model should be allways supported by the theory that shows the original form of model was not accurate. For the purpose of the paper, let us assume that the edges of interest can be deleted from the model. Since our aim is to eliminate the influence of the unmeasurable variable A from the computations, we will first focus on the edges between A and B, and A and C respectively (the altered models are shown in Figure 2).

The removal of the edge between A and B leads to the following joint probability distribution

$$P(a,b,c,d) \quad = P(a) \cdot P(b) \cdot P(c|a,b) \cdot P(d|b,c).$$

We infer the causal effect of intervention on B on variable D as

$$P(d|\hat{b}\,) \quad = \textstyle\sum_{a,c} P(a) \cdot P(c|a,b) \cdot P(d|b,c). \tag{3}$$

We can see that the resulting representation of the causal effect is exactly the same expression we obtained when computing the causal effect in the basic model (compare Formulas 2 and 3). So the desired causal effect cannot be computed and this alteration is not helpful.

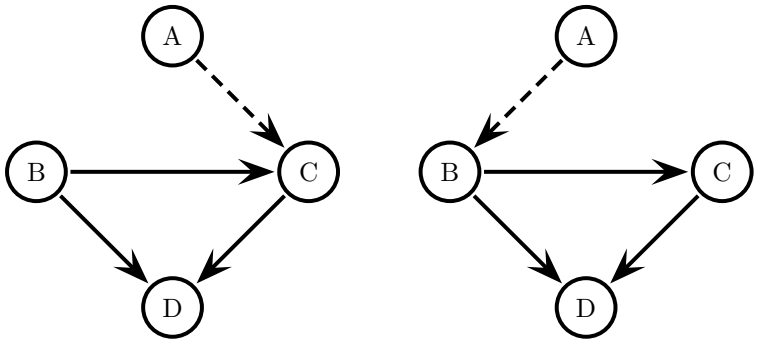


Figure 2 The left model presents the removal of the edge between A and B. The right part shows the model after removing the edge between A and C.

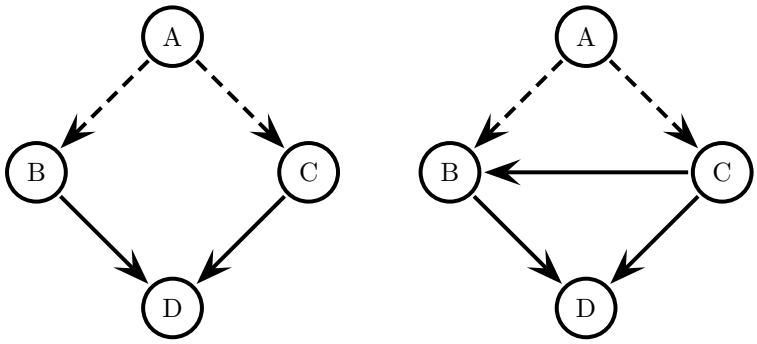


Figure 3 The left model presents the removal of the edge between B and C. The right model shows the situation where the direction of the edge between B and C was changed.

The situation changes in case of removal of the edge between A and C. Employing the intervention and using the summation over A and C we get

$$P(d|\hat{b}) \quad = \sum_c P(c|b) \cdot P(d|b,c) \sum_a P(a). \tag{4}$$

Now we can see that the given causal effect can be expressed without the unmeasurable variable A (since it can be easily removed from Formula 4) and it makes it computable.

Let us study the relation between variables B and C. We will introduce two alterations - the removal of the edge (and thus situation when causal relation between them does not exist) and wrong direction of the edge in the basic model (the relation under consideration exists but it was incorrectly identified what the cause and the effect are). See Figure 3 to understand the structure of these altered models.

Similarly to the previous examples we infer the causal effect. In case of removal the edge between$B$ and C we obtain

$$P(d|\hat{b}) \quad = \sum_c P(d|b,c) \sum_a P(a) \cdot P(c|a). \tag{5}$$

And this causal effect again is computable.

For the situation of changed orientation of the edge (from $B \to C$ to $B \leftarrow C$) we obtain the joint probability distribution

$$P(a,b,c,d) \quad = P(a) \cdot P(b|a,c) \cdot P(c|a) \cdot P(d|b,c).$$

and the causal effect

$$P(d|\hat{b}) \quad = \sum_c P(d|b,c) \sum_a P(a) \cdot P(c|a). \tag{6}$$

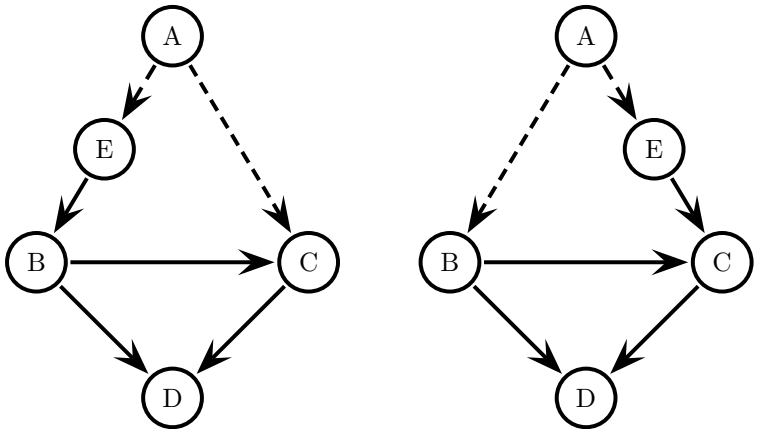


Figure 4 The left model presents the introduction of the new node between A and B. The right model shows the model after adding the new node between A and C.

We can see that the formula for computing this causal effect is the same as in the previous case (compare Formulas 5 and 6).

Finally we will focus on the introduction of new node into the basic model. Again we have to consider suitable location for such an alteration. Let us study four different situations - adding the node between A and B, between A and C (see Figure 4) and between B and C with different orientation in each example (see Figure 5) and discuss its results. Here the added node always represents a measurable variable.

Adding the node E between nodes A and B (in application it can be for example the variable production costs) will result in the causal effect

$$P(d|\hat{b}) \quad = \sum_{c,e} P(d|b,c) \sum_a P(a) \cdot P(c|a,b) \cdot P(e|a).$$

It can be easily proven that this causal effect is not computable since none of the conditions from the Theorem 1 is satisfied.

When it comes to situation of adding the node E between the nodes A and C (in case we are able to measure for instance the purchasing power of the customer), the joint probability distribution is in the form of

$$P(a,b,c,d,e) \quad = P(a) \cdot P(b|a) \cdot P(c|b,e) \cdot P(d|b,c) \cdot P(e|a).$$

After intervention on B and summation over A, C, E the result is computable causal effect

$$P(d|\hat{b}) \quad = \sum_{c,e} P(c|b,e) \cdot P(d|b,c) \sum_a P(a) \cdot P(e|a).$$

When adding the node E between the nodes B and C (as shown in left part of Figure 5) and given the intervention on B and the summation over A, C, E we obtain

$$P(d|\hat{b}) \quad = \sum_{c,e} P(d|b,c) \cdot P(e|b) \sum_a P(a) \cdot P(c|a,e).$$

Independencies in the model allows us to compute the causal effect as

$$P(d|\hat{b}) \quad = \sum_{c,e} P(d|b,c) \cdot P(e|b) \sum_{b'} P(b') \cdot P(c|b',e).$$

Since the variable A is not used in this formula the effect can be computed.

Similar situation appears in the last presented example. In case we manage to replace the relation between B and C with their observable common cause (such as the quality of the product), we will get the causal effect

$$P(d|\hat{b}) \quad = \sum_{c,e} P(d|b,c) \cdot P(e) \sum_a P(a) \cdot P(c|a,e).$$

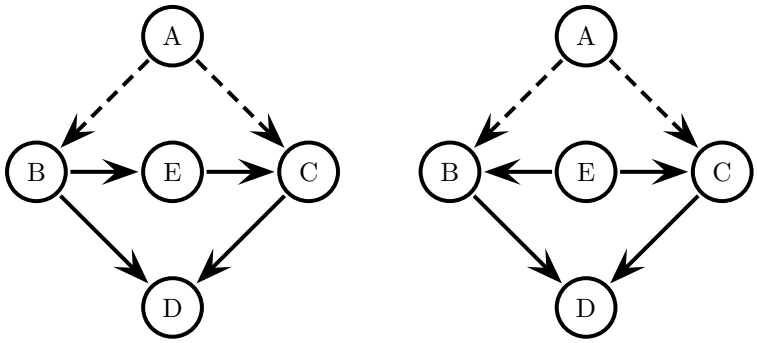


Figure 5 The model presents the introduction of the new node between B and C.

Again, using the independencies in the causal diagram it can be modified to the form of

$$P(d|\hat{b}) \quad = \sum_{c,e} P(d|b,c) \cdot P(e) \sum_{b'} P(b') \cdot P(c|b',e).$$

We managed to express the causal effect without necessity to use unmeasurable variable A, thus the causal effect is computable.

3 Conclusion

The contribution focuses on some possible alterations of simple causal model in order to enable to compute the causal effect of intervention on the specific variable. It is shown that number of alterations is available to be used, but only some of them are helpful. The computations used in the paper are based on the Galles and Pearl therorem (see Theorem 1). One of the aims of the paper is to show that different alterations can be used that would result in the model with computable causal effect. It provides the possibility to choose such an alteration that corresponds to the resources. We can apply an alteration that responds to the situation of interest and for which we have available data.

Acknowledgements

The research was supported by GA ČR under grant no. 15-00215S.

References

[1] Galles, D., and Pearl, J.: Testing identifiability of causal effects,*Uncertainty in Artificial Intelligence* **11** (1995), 185–195.

[2] Pearl, J.: Causal diagrams for empirical research, *Biometrika* **82.4** (1995), 669–688.

[3] Pearl, J.: *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2009.

[4] Tian, J., Pearl, J.: A General Identification Condition for Causal Effects. In: *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, AAAI Press/The MIT Press, Menlo Park, CA, 2002, 567–573.

[5] Vachova, L., Bina, V.: A Note on Computation of Causal Effect in Economic Models. In: *MME 2013: Proceedings of 31th International Conference Mathematical Methods in Economics* (Vojáčková, Hana), Jihlava, 2013, 974–979.

Modelling extremal claim severity using generalized Pareto regression model

Jiri Valecky[1]

Abstract. The paper is focused on modelling claim severity in insurance in which it is crucial to cope with the occurrence of large claims to obtain a reliable and accurate model. We present an approach to modelling extremal claim severity using regression model based on the individual characteristics (rating factors) of policyholders because the model based on the risk theory for all claims is not as convenient when we need to obtain estimates of individual claims. Thus, we assume that the individual claim severity follows mixed distribution using the gamma distribution and generalized Pareto (GP) distribution with various parameters for all of the policyholders. However, while the claim frequency is known, the frequency of regular and extremal claims is not. So, we also present a method how to identify the threshold value (location parameter of GP distribution) which defines the extremal claims and we estimate the generalized Pareto regression model for extremal events using the real block of motor hull insurance policies.

Keywords: claim severity, extremes, extreme value theory, generalized linear models, insurance, generalized Pareto regression.

JEL classification: C18, C58, G22
AMS classification: 60E05, 62J12, 62P05

1 Introduction

We present an approach to modelling individual large claim severity although the large claim frequency is not observable. To model large claim severity, we use generalized linear models (GLMs) under the assumption of generalized Pareto (GP) distribution and the frequency of large claims is estimated using threshold value which indicates the occurrence of extremal claim. The application of this model is important mainly for premium pricing, estimating individual claim severity, as well as estimating solvency capital requirement (SCR) for underwriting risk according to the Solvency II.

In general, the crucial issue of modelling claim severity is the occurrence of large (extremal) claims. As inference, the empirical distribution of claim severity is heavy-tailed and common parametrized distributions do not fit the data. Thus, the application of extreme value theory (EVT) appear to be valuable, mainly the peak over the threshold approach. However, it is necessary to know the threshold value which can be determined ad hoc or using the excess function. In addition, it is also possible to estimate this parameter using method of moments, method of probability weighting moments or maximum likelihood method. Recent reviews of this problem can be found in [3] or [10].

By contrast to collective risk model, the individual risk model focuses on the random sum of claims in which the individual claim severity follows different probability distribution and to estimate distribution for each policy is necessary. However, the number of observations for each policy is not sufficient, thereby to determine the probability distribution and threshold value is not plausible. By contrast, we may assume that individual claim severity follows the same family of distributions, but with different parameters which depend on the individual characteristics (rating factors). In addition, the threshold value also differs because the same level for all of the policies yield insufficient data fit, as well as the small probability of small claims is not respected.

We use the general concept of GLMs which were introduced by [11]. The first regression analysis using individual rating factors and also one of the first separate analysis of claim frequency-severity appeared

[1]VSB-TU/Faculty of Economics, Department of Finance, Sokolska tr. 33, 701 21 Ostrava, jiri.valecky@vsb.cz

in [8], while the first application of GLMs was used to model claim frequency for marine insurance and claim size for motor insurance in [11]. More applications of GLMs occurred mostly after 1990s when the insurance market was being deregulated in many countries and the GLMs were used to undertake a tariff analysis, for example [1], [5] or [13], or to set premiums, e.g. [4], [12] or [14].

Thus, the goal of the paper is to present an approach suitable for modelling individual large claims. The remainder of the paper is organized as follows. A generalized Pareto (GP) regression is described in the Section 2. The empirical GP regression model, which was estimated on the real block of motor hull insurance policies, is presented in Section 4. Section 5 provides the conclusions. All of the computations were performed in MATLAB.

2 Generalized Pareto regression

Let's consider a random variable Y which follows the GP distribution $GP(\sigma,\kappa)$ with the cumulative distribution function (cdf)

$$F(y;\sigma,\kappa,\theta)=\begin{cases}1-\left(1+\kappa\frac{y-\theta}{\sigma}\right)^{-\frac{1}{\kappa}}, & 1+\kappa\frac{y-\theta}{\sigma}>0,\kappa\neq 0,\\ 1-\exp\left(-\frac{\kappa}{\sigma}\right), & y-\theta>0,\kappa=0,\end{cases} \tag{1}$$

where κ is the shape parameter, θ is the location parameter and σ represents the scale parameter.

Now, we extend the GP distribution to a regression model by taking κ,σ as a functions of individual rating factors, see, e.g. [7]. Thus, we assume random independent claim severity $Y_1,\ldots,Y_n$ and let $\mathbf{x_i}$ be a vector of rating factors related to the Y_i such that

$$Y_i|\mathbf{x}_i\sim GP\left(\sigma\left(\mathbf{x}_i\right),\kappa\left(\mathbf{x}_i\right)\right),i=1,...,n, \tag{2}$$

where $\sigma\left(\mathbf{x}_i\right),\kappa\left(\mathbf{x}_i\right)$ are functions of rating factors, that is, $\kappa=h_1\left(\mathbf{x}_i,\beta_1\right),\sigma=h_2\left(\mathbf{x}_i,\beta_2\right)$ with vectors of regression coefficients $\beta=(\beta_1,\beta_2)$. The log-likelihood is in the form of

$$\log L(\beta)=\sum\log f\left(Y_i;\sigma\left(\mathbf{x}_i\right),\kappa\left(\mathbf{x}_i\right)\right), \tag{3}$$

where f is the GP density function

$$f(y;\sigma,\kappa,\theta)=\begin{cases}\frac{1}{\sigma}\left(1+\kappa\frac{y-\theta}{\sigma}\right)^{-\frac{1}{\kappa}-1}, & 1+\kappa\frac{y-\theta}{\sigma}>0,\kappa\neq 0,\\ \frac{1}{\sigma}\exp\left(-\frac{\kappa}{\sigma}\right), & y-\theta>0,\kappa=0.\end{cases} \tag{4}$$

The log-likelihood function may be rewritten in the form of

$$\log L(\beta)=-\log\sigma\left(\mathbf{x}_i\right)-\left(\frac{1}{\kappa\left(\mathbf{x}_i\right)}+1\right)\log\left(1+\kappa\left(\mathbf{x}_i\right)\frac{Y_i-\theta}{\sigma\left(\mathbf{x}_i\right)}\right). \tag{5}$$

However, also the threshold value θ varies across the policies. Therefore, we assume that the threshold value depends on the rating factors, thus $\theta\left(\mathbf{x_i}\right)=h_3\left(\phi,\mathbf{x_i}\right)$. In addition, the level of $\theta\left(\mathbf{x_i}\right)$ must ensure sufficient number of observations to obtain reliable estimates, see [6] and [7]. It is possible to formulate the mixed-integer programming problem which allows to estimate $\beta=(\beta_1,\beta_2,\phi)$ for the GP regression model in which k observations fall above $\theta\left(\mathbf{x_i}\right)$, see [2],

$$\max_{\beta_1,\beta_2,\beta_3,\delta}\sum_{i=1}^{n}\left\{-\log\sigma\left(\mathbf{x}_i\right)-\left(\frac{1}{\kappa\left(\mathbf{x}_i\right)}+1\right)\log\left(1+\kappa\left(\mathbf{x}_i\right)\frac{Y_i-\theta\left(\mathbf{x}_i\right)}{\sigma\left(\mathbf{x}_i\right)}\right)\right\}\left(1-\delta_i\right) \tag{6}$$

subject to

$$\begin{aligned}&\beta_1,\beta_2,\phi\in R,\\ &Y_i+M\delta_i\geq\theta\left(\mathbf{x}_i\right),\quad i=1,\ldots,n,\\ &\sum_{i=1}^{n}\left(1-\delta_i\right)=k,\\ &\delta_i\in\{0,1\},\qquad\qquad i=1,\ldots,n,\end{aligned} \tag{7}$$

with M a big number. Finding the solution to this problem is quite complex and the quantile regression, [9], can be used to facilitate the estimation process. We find the p-th $(0 < p < 1)$ quantile regression coefficient estimates as a solution to the optimization problem

$$\min_{\phi_p} \sum_{i=1}^{n} \left(p(Y_i - \theta(\mathbf{x_i}))^+ + (1-p)(Y_i - \theta(\mathbf{x_i}))^- \right) \tag{8}$$

with $x^+ = \max(0, x)$ and $x^- = \max(0, -x)$. The $\theta(\mathbf{x_i})$ estimates are inserted into (5) to obtain β.

To assess the goodness of fit of the GP regression, we may use the exponential quantile plot. First, the variable R_i are calculated using the formula

$$R_i = \begin{cases} \frac{1}{\kappa(\mathbf{x}_i)} \log\left(1 + \kappa(\mathbf{x}_i) \frac{Y_i}{\sigma(\mathbf{x}_i)}\right), & \kappa(\mathbf{x}_i) \neq 0, \\ \frac{Y_i}{\sigma(\mathbf{x}_i)}, & \kappa(\mathbf{x}_i) = 0, \end{cases} \tag{9}$$

which yields a standard exponential random variable and the coordinates in the exponential quantile plot are, see [6],

$$\left(-\log\left(1 - \frac{i}{n+1}\right), R_i \right), i = 1, \ldots, n, \tag{10}$$

where R_i is ordered variable $R_1 \leq R_2 \leq \ldots \leq R_n$. When the GP regression fit the data well, we expect the points close to the unit diagonal.

3 Empirical GP regression

In this section, we present the GP regression model estimated using a real block of insurance policies. The data sample encompassed the characteristics of claims occurred during the years 2004–2010 (22,274 observations). The following rating factors were included: vehicle age (`agecar`), engine power divided by engine volume (`kwvol`·1000), owner's age (`ageman`), indicator of company/private car (`company`), gender of policyholder (`gender`), type of fuel (`fuel`) and vehicle price (`price`). Next table summarizes the basic characteristics of the claim severity and continuous rating factors. It is obvious that all continuous variables are non-normal, as well as claim severity.

Rating factor	Mean	Standard deviation	Skewness	Kurtosis
`severity`	52,210.12	72,625.59	17.50	1,026.25
`agecar`	4.28	2.93	0.78	4.31
`ageman`	46.02	13.94	0.33	2.413
`kwvol`	44.25	8.95	0.70	5.44
`price`	474,959.10	407,200.20	4.99	53.92

Table 1: Basic characteristics of data

We assumed that the individual thresholds were determined as a percentage of the price, i.e. $\theta = \phi \cdot$ `price`. In addition, we assumed positive shape and scale parameters, therefore we used exponential functions for h_1, h_2.

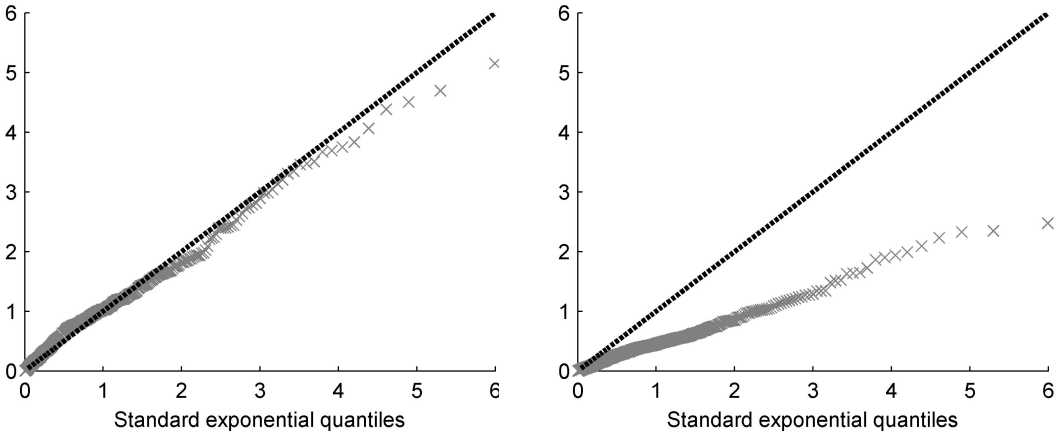
The percentage of the price was estimated by quantile regression at the level of 0.6424 and the regression parameters were obtained using the maximum likelihood method. However, to avoid the problem of local minimum, it is necessary to run the estimation process starting from different initial points or use some algorithm for global optimization. The estimated parameters are summarized in Table 2.

We also estimated parameters of GP regression model, in which the constant threshold value was estimated at the level 9,624. We evaluated the fit of the both GP regression models using the exponential quantile plot.

As shown in Figure 1, the ordered variable R_i of GP regression in which the covariate dependent threshold was assumed scatter quite well around the first diagonal, while the quantile plot for the GP regression with constant threshold indicated insufficient data fit.

Rating factor	Shape parameter κ	Scale parameter σ
intercept	-10.9775	9.2034
`agecar`	-0.0181	-0.1127
`ageman`	-0.1575	0.0002
`kwvol`	-0.0183	0.0271
`fuel`	-0.1191	0.2285
`company`	3.4577	-0.0674
`gender`	0.0108	-0.1066

Table 2: Parameter estimates



(a) Covariate dependent threshold (b) Constant threshold

Figure 1: Ordered random variable R of generalized Pareto regression models

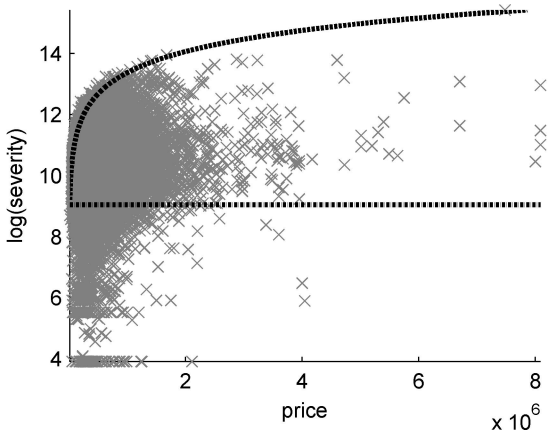


Figure 2: Covariate dependent (dotted line) and constant threshold (dashed line)

To make the picture complete, Figure 2 shows the covariate dependent and constant thresholds for various price level and the points above the dotted and dashed line represent the large claims which occurred during the given period. For higher clarity we used the log scale on the y axis. It is obvious that the model with constant threshold parameter considers significantly more claims to be a large claim which yielded the insufficient fit of the data as shown in Figure 1.

4 Conclusion

Although the large claim frequency is not observable, we presented an approach to modelling individual large claim severity using the real block of insurance policies. To model large claim severity, we used generalized linear models (GLMs) under the assumption of generalized Pareto distribution. The frequency of large claims was estimated using threshold value which indicated the occurrence of extremal claim.

The application of the model is suitable for modelling individual large claims. It can be also used to analyse the influence of rating factors on the distribution of large claims. In addition, another application consists in undertaking tariff and the model is also crucial for modelling claims in general. However, the distribution of large claim must be combined with the distribution of small claims and also the probability of large claim occurrence is necessary to estimate.

Thus, the GP regression may be considered as a flexible tool for modelling large claims in general. In addition, it is necessary to respect covariate dependent threshold, otherwise the GP regression fits the data insufficiently. It is also possible to test the statistical significance of estimated coefficients and re-estimate the model after omitting non-significant rating factors.

Acknowledgements

This paper was solved within the project SP2015/75 and within Operational Programme Education for Competitiveness Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Andrade-Silva, J. M.: An application of Generalized Linear Models to Portuguese Motor Insurance,. In: *Proceedings XXI ASTIN Colloquium*, New York, 1989, 633.

[2] Beirlant, J., De Waal, D. and Ferro, C. A.: *Statistics of extremes*. John Wiley & Sons Inc., Chichester, 2004.

[3] Bermudez, P. Z. and Kotz, S.: Parameter estimation of the generalized Pareto distribution - Parts I and II, *Journal of Statistical Planning and Inference* **140** (2010), 1353-1388.

[4] Branda, M.: Optimization Approaches to Multiplicative Tariff of Rates Estimation in Non-Life Insurance, *Asia-Pacific Journal of Operational Research* **31** (2014), 145–162.

[5] Brockman, M. J. and Wright, T. S.: Statistical motor rating: making efficient use of your data, *Journal of the Institute of Actuaries* **119** (1992), 457–543.

[6] Coles, S.: *An Introduction to Statistical Modelling of Extreme Values*. Springer, London, 2001.

[7] Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds, *Journal of the Royal Statistical Society B* **52** (1990), 393–442.

[8] Kahane, Y. and Levy, H.: Regulation in the insurance industry: Determination of premiums in automobile insurance, *Journal of Risk and Insurance* **42** (1975), 117–132.

[9] Koenker, R. and Bassett, G.: Regression quantiles, *Econometrica* **46** (1978), 33–50.

[10] Mackay, E. B. L. and Challenor, P. G.: A comparison of estimators for the generalized Pareto distribution, *Ocean Engineering* **38** (2011), 1338–1346.

[11] McCullagh, P. and Nelder, J. A.: *Generalized linear models*. Chapman & Hall, London, 1989.

[12] Murphy, K. P., Brockman, M. J. and Lee, P. K. W.: Using Generalized Linear Models to Build Dynamic Pricing Systems for Personal Lines Insurance, In: *Casualty Actuarial Society Forum*, 2000, 107–140.

[13] Renshaw, A. E.: Modelling the claims process in the presence of covariates, *ASTIN Bulletin* **24** (1994), 265–285.

[14] Zaks, Y., Frostig, E. and Levikson, B.: Optimal pricing of a heterogeneous portfolio for a given risk level, *ASTIN Bulletin* **36** (2006), 161–185.

Financing of renewal of assets by depreciation

František Vávra[1], Tomáš Ťoupal[2]

Abstract. Depreciation has its cost and financial aspect. Depreciation can be taken as a source of funds in company in case of profit and is used to cover operating expenses. This leads to their appreciation by rate of return costs and generates fund. There are two forms of renewal of assets. The first form is financing from accumulated and appreciated depreciation charge and the second one is acquisition of an asset by some type of credit. Depreciation charges can´t cover the financial need of renewal of assets. The value of an asset increases by inflation and technical improvement, credit financing is the value of an asset increased by interest. It is necessary to take into account appreciation of depreciation's resources. This paper is interested in conditions where the appreciated resources reach the level of "time price" of analyzed asset or there is created sufficient amount of money to cover both principal and accrued interest. There is solved the problem of its creation and time to reach a parity between the created resources and the need of financing, the relations between the rate of return costs, the index of price changes and corresponding interest rate and finally a tax effect.

Keywords: Depreciation, Accumulative Financing, Credit Financing.

JEL Classification: G31, G32
AMS Classification: 62P05

1 Basic notation

Let us first introduce some basic notation.

C_k is a purchase price of acquired assets at the end of the k −th period. In the case of a clear period will be used C without an index.

i represents a mean annual index of price growth in sector where acquired intangible assets are used. Concrete realizations are obtained from $i_k = \frac{C_k - C_{k-1}}{C_{k-1}}$ and i means a long-term representative value from all obtained values.

r $r > 0$, means a representative long-term achieved rate of return costs, i.e. $r = \frac{costs + net\ income}{costs} - 1 = \frac{net\ income}{costs}$.

o_k indicates an accounting depreciation charge of intangible assets for the k −th period.

n is the economic life of an asset.

s_k denotes an instalment for k −th period at the end of the k −th period in the case of acquisition of fixed assets on some type of credit, $s_k = q_k + u_k$.

q_k is a repayment of principal.

u_k is the interest portion of instalment (accrued interest).

p means a representative interest rate including all costs of credit in the case of acquisition of fixed assets on some type of credit, thus $p = RPSN$.

[1] University of West Bohemia, Faculty of Applied Sciences, Department of Mathematics, Technická 8, 306 14 Pilsen, Czech Republic, vavra@kma.zcu.cz
[2] University of West Bohemia, Faculty of Applied Sciences, Department of Mathematics, Technická 8, 306 14 Pilsen, Czech Republic, ttoupal@kma.zcu.cz

f_k represents a value of accounting fund at the end of the k −th period in the case of fund or amortisation financing (see below).

z_k is an income tax base for the k-the period.

t_k is a tax payment for the k −th period.

d is a tax rate of „corporate income tax".

The philosophy of financing by depreciation is based on the fact that the depreciation charges are used (or there is possible to say „invested") in production (to cover running costs) which means their appreciation by $(r * 100)\%$ in each period. In the case of predefined conditions (the company is in "black numbers"), it is possible to fully financing the renewal of assets by depreciation charges. The basic procedures of the renewal of assets is financing by "its" accumulated and appreciated own means (creation of the fund of renewal of fixed assets) and acquisition of assets by some type of credit.

2 Financing by accumulated and appreciated own funds

An increase in the price of renewed assets at time of acquisition is

$$\begin{gathered} C_{k+1} = (1+i)C_k; C_0 = C \\ C_k = (1+i)^k C. \end{gathered} \tag{1}$$

The creation of the fund of renewal of fixed assets from depreciation charges can be described as

$$\begin{gathered} f_{k+1} = (1+r)f_k + o_{k+1};\ f_0 = 0 \\ f_k = \sum_{j=1}^{k} o_j (1+r)^{k-j}. \end{gathered} \tag{2}$$

The economic life is the shortest period (such smallest n) where the following inequality is hold

$$\sum_{j=1}^{n} o_j (1+r)^{n-j} \geq (1+i)^n C, \tag{3}$$

if this condition is satisfied (when the value of fund reaches or exceeds first a new acquisition price).

The inequality (3) can be easily modified by dividing both sides by the initial acquisition price

$$\sum_{j=1}^{n} \frac{o_j}{C} (1+r)^{n-j} \geq (1+i)^n \equiv \sum_{j=1}^{n} \vartheta_j (1+r)^{n-j} \geq (1+i)^n, \tag{4}$$

where $\vartheta_j = \frac{o_j}{C}$ is a proportional part of the acquisition price depreciated in the j −th period. It means the ratio of uniquely specifying depreciation method. For the interpretation, there is consider the following rearrangement:

$$\sum_{j=1}^{n} \frac{\vartheta_j}{(1+r)^j} \geq \left(\frac{1+i}{1+r}\right)^n. \tag{5}$$

Thus, the present value of depreciation charges must reach at least the present value of the future value of the acquisition price considering the „inflation" change, all with using the discount rate of the rate of return costs. From the equality $\sum_{j=1}^{n} \vartheta_j = 1$, it is also

$$\sum_{j=1}^{n} \frac{\vartheta_j}{(1+r)^j} < 1. \tag{6}$$

In the case of straight line depreciation is obtained $\vartheta_j = \frac{1}{n}$ for $j = 1, \dots, n$, then

$$\sum_{j=1}^{n} \frac{\vartheta_j}{(1+r)^j} = \frac{1}{nr}(1-(1+r)^{-n}). \tag{7}$$

And the inequality (5) is rearranged in the following form

$$\begin{gathered} \frac{1}{nr}(1-(1+r)^{-n}) \geq \left(\frac{1+i}{1+r}\right)^n, \\ ((1+r)^n - 1) \geq nr(1+i)^n. \end{gathered} \tag{8}$$

Then the basic task is to find out whether there is such $n > 0$ that satisfies the inequality above. Of course, if it exists then it is questionable whether this number is meaningful. It is also useful to know how such n looks like. This is possible to solve it by some type of numerical method e.g. bisection method. The described methodology is shown in the following figure.

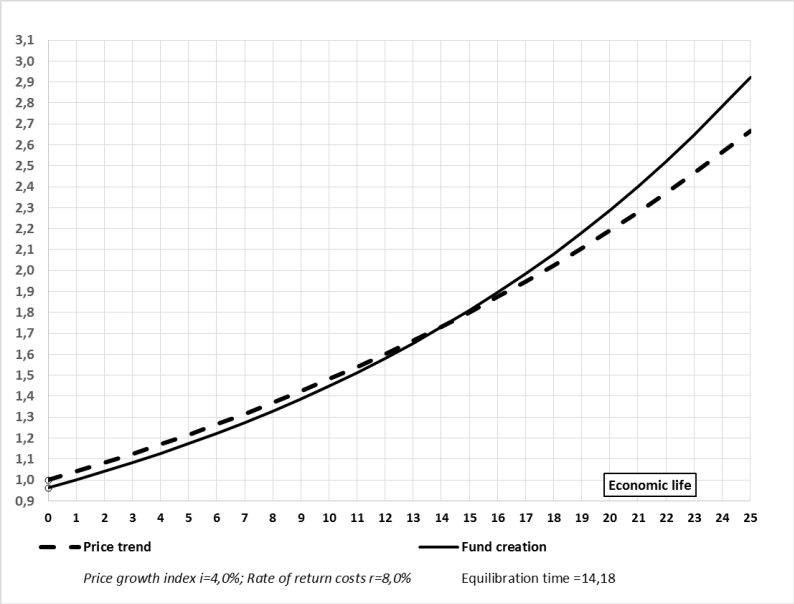


Figure 1 Linear depreciation and creation of the fund of renewal of fixed assets

3 Credit financing and amortization financing

The creation of the amortization fund from depreciation charges can be described as

$$\begin{gathered} f_{k+1} = (1+r)f_k + (o_{k+1} - s_{k+1});\ f_0 = 0, \\ f_k = \sum_{j=1}^{k} (o_j - s_j)(1+r)^{k-j}. \end{gathered} \tag{9}$$

For the instalments is hold

$$C = \sum_{j=1}^{m} \frac{s_j}{(1+p)^j}, \tag{10}$$

where m is the repayment period of the credit, C is the acquisition price of an asset and $s_j = 0$ for $j > m$.

The economic life is the shortest time (such smallest $n \geq m$), for which is hold the inequality

$$f_n = \sum_{j=1}^{n} (o_j - s_j)(1+r)^{n-j} \geq 0. \tag{11}$$

If this inequality is satisfied then (after dividing by $(1+r)^n$) there is obtained the following inequality

$$\sum_{j=1}^{n} \frac{o_j}{(1+r)^j} \geq \sum_{j=1}^{n} \frac{s_j}{(1+r)^j}. \tag{12}$$

In short, if there exists some solution of above described problem, then it is necessary to create inside the amortization fund containing enough resources to cover all instalments. Furthermore, analogous to the previous case, it is true that the present value of depreciation charges must reach at least the present value of the credit instalments, all with using discount rate of the rate of return costs.

If there are taken again the relative values $\vartheta_j = \frac{o_j}{C}$ and $\sigma_j = \frac{s_j}{C}$, then it is possible to obtain the inequality

$$\sum_{j=1}^{n} \frac{\vartheta_j}{(1+r)^j} \geq \sum_{j=1}^{n} \frac{\sigma_j}{(1+r)^j} \tag{13}$$

and this condition is independent on the acquisition price. However, this condition links a depreciation schedule with instalment regime. Furthermore, this condition can be easily rearranged as

$$\sum_{j=1}^{n} \frac{\vartheta_j - \sigma_j}{(1+r)^j} \geq 0. \tag{14}$$

From here it is possible to make some conclusions. For instance, this inequality is not satisfied by the assumption of „classical" instalments (uniform portion of principal plus accrued interests), i.e. $\sigma_j > \frac{1}{n}$, straight-line depreciation $\vartheta_j = \frac{1}{n}$ and the same period of depreciation and repayment of credit. An example, where it is possible to finance the credit by depreciation is shown in Figure 2 (classic instalment, accelerated accounting depreciation $\vartheta_j = \frac{2(n-j+1)}{n(n+1)}$ and n is depreciation period).

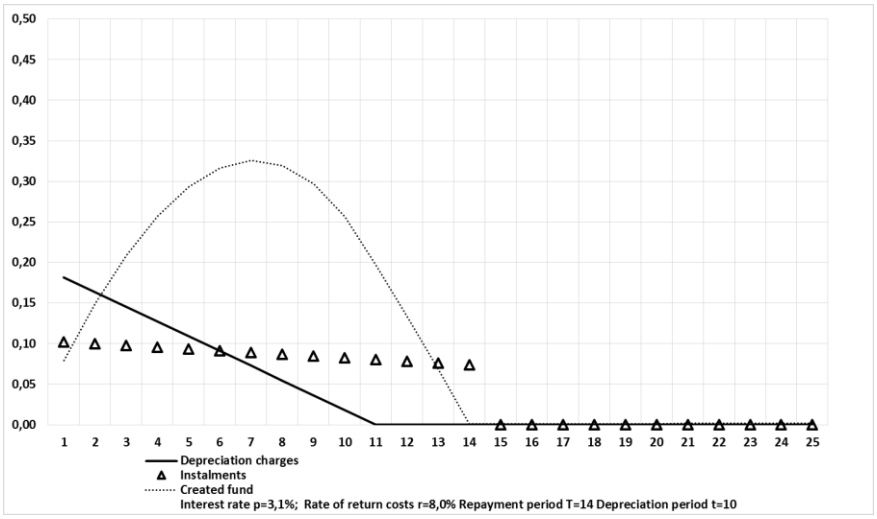


Figure 2 Accelerated depreciation and creation of the amortization fund of renewal of fixed assets

There is necessary to use the „additional financing" from the share in the proceeds of assets in the case of straight-line depreciation. The depreciation charges (appreciated) do not cover „interest need". The essence of this situation is in the fact that the depreciation charges are appreciated until the acquisition of new asset by financing of the accumulated funds. In contrary, there is a continuous drawdown for each instalment in the case of amortization financing.

4 Tax view

4.1 Financing by own accumulated funds

The creation of innovation fund is subjected to „income tax“, just the appreciated part of this fund. For this reason is possible to use the following relations.

The income tax base is obtained as

$$z_{k+1} = rf_k = r\sum_{j=1}^{k} o_j(1+r)^{k-j}. \tag{15}$$

The income tax is obtained as

$$t_{k+1} = dr(1+r)^k \sum_{j=1}^{k} \frac{o_j}{(1+r)^j}. \tag{16}$$

Here it is useful, that the present value of depreciation charges discounting by the rate of return costs has a significant impact on paid tax.

4.2 Credit financing, amortization financing

The creation of the amortization fund is subjected to „income tax“, just appreciated part of this fund and repayment of credit principal (more precisely, they are going out of already taxed money). Conversely, interests are deductible part of the base of „income tax“.

The income tax base is obtained as

$$z_{k+1} = rf_k + s_{k+1} - u_{k+1} \tag{17}$$

The income tax obtained as

$$\begin{gathered} t_{k+1} = d(rf_k + s_{k+1} - u_{k+1}), \\ t_{k+1} = d\left(r\sum_{j=1}^{k}(o_j - s_j)(1+r)^{k-j} + q_{k+1}\right), \\ t_{k+1} = d\left(r(1+r)^k \sum_{j=1}^{k} \frac{(o_j - s_j)}{(1+r)^j} + q_{k+1}\right). \end{gathered} \tag{18}$$

Again the present value of depreciation charges has a significant impact on paid tax - here is depreciated by the present value of applied instalments - discounting by the rate of return costs. There is evident from the expression for the amount of tax that with this type of financing will (in most practical case, majority $(o_j - s_j) < 0$) be lowering tax base.

5 Conclusion and discussion

The presented models can be used as an analytical apparatus for comparing the different types of innovation financing of fixed assets. They are also important for theoretical microeconomic, for example see a significant influence of „the present value of applied depreciation charges“ on the creation of the innovation funds. As already indicated here, for his perfection it is required to include the financing (additional) from the proceeds of analyzed asset (division of proceeds between the financier and operator). It is also inconsiderable to use it for tax optimization (creation of funds with significantly lower tax ratio than nominal or effective).

References

[1] *Czech Accounting Standard for Businesses No. 013*, Long-lived tangible and intangible assets.
[2] *IAS 16 Property*, Plant and Equipment.
[3] Vávra F., Ťoupal T., Wagnerová E., Marek P.: *Economical and Financial View of Ageing*, Financial Management of Firms and Financial Institutions, Ostrava, 2009.
[4] Edgerton D.: *Valuation and Depreciation*, CPA Australia Ltd, Australia, 2013.

Variant forms of the Raiffa solution to the bargaining problem

Milan Vlach [1]

Abstract. Like the Nash solution to the cooperative bargaining problem, several alternative solutions suggested by Raiffa in the early 1950's were originally proposed for the two-person case. All of these solutions have a number of appealing properties, and all of them have been extended for the case of more than two persons. We discuss several modifications and generalizations of the discrete Raiffa solution for a number of domains of the n-person case. Moreover, we present some open problems concerning axiomatic characterization.

Keywords: cooperative bargaining, discrete Raiffa solution, stepwise solution, axiomatization.

JEL classification: C78, C71
AMS classification: 91A12

1 Introduction

Since Nash's papers [4] and [5] on bargaining, and Raiffa's studies [7] and [8][1] on arbitration schemes in the beginning of 1950's, it has become customary to formulate an *n-player bargaining problem* as a nonempty collection $\mathcal{B}$ of pairs (S, d) where S is a nonempty subset of n-dimensional real linear space R^n and d is a point in R^n. The elements of S are interpreted as utility n-tuples that the players can obtain (each ith player obtains value of the ith component of the tuple) by reaching a unanimous agreement on an element of S, and d as the outcome when the players do not agree unanimously on any point in S. The set S is called the *bargaining set* and the point d is called the *disagreement point* or *threat point* or *status quo point*, depending on the underlying interpretation.

A *solution* or a *solution function* on $\mathcal{B}$ is a function f from $\mathcal{B}$ to R^n such that, for each instance (S, d) of $\mathcal{B}$, the value $f(S, d)$ of f belongs to S. A *solution* or a *solution outcome for an instance* (S, d) of problem $\mathcal{B}$ is the value $f(S, d)$ of solution function f at (S, d). Thus, to prevent the possibility of confusion, we should be mindful that, in the bargaining literature, the word "solution" is often used in two different meaning; it may mean a function or a value of a function.

These definitions are too general for having a meaningful theory of bargaining. Therefore, to have a meaningful theory, the instances (S, d) forming a bargaining problem and solution functions should have some reasonable properties. For example, in the canonical (Nash) problem, it is assumed that each instance (S, d) satisfies the following conditions: S is compact and convex, d belongs to S, and there is at least one point in S that is better than d for each player. There is an extensive game-theoretic literature on bargaining and arbitration, in which a multitude of solution concepts are proposed and analyzed, see [9] and [16] for surveys.

Throughout the paper we use the following notation. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ from the n-dimensional space R^n, we write $x < y$ and $x \leq y$ if, respectively, $x_i < y_i$ and $x_i \leq y_i$ for each i. The converse relations $>$ and $\geq$ between elements of R^n are defined analogously. The scalar product of x and y from R^n is denoted by xy. The sets $\{x \in R^n : x \geq 0\}$ and $\{x \in R^n : x > 0\}$ are denoted by R^n_+ and R^n_{++}, respectively. If A is a subset of R^n and x is a point in R^n, then we denote the sets $\{a + x : a \in A\}$ and $\{a - x : a \in A\}$ by $A + x$ and $A - x$, respectively. Similarly, if λ is a real number, we define λA as the set $\{\lambda a : a \in A\}$. Moreover, we define the sets A^+_x and A^{++}_x by $A^+_x = \{y \in A : y \geq x\}$ and $A^{++}_x = \{y \in A : y > x\}$, respectively. If (S, d) is an instance of a bargaining problem, then we say that

[1]Institute of Information Theory and Automation, Pod Vodárenskou věží 4, Prague, Czech Republic, and Charles University, Faculty of Mathematics and Physics, Department of Theoretical Computer Science and Mathematical Logic, vlach@utia.cas.cz or milan.vlach@mff.cuni.cz

[1]See also discussion in Chapter 6 of [3].

S_d^+ is the *individually rational part* of (S,d), and S_d^{++} is the *strict individually rational part* of (S,d). An instance (S,d) is said to be *d-comprehensive* if, for each $x \in S$, the set $\{y : d \leq y \leq x\}$ is included in S.

2 Nash's solution

A particular solution function of a bargaining problem can be defined in a number of different ways. For example, the Nash solution function to the Nash two-player problem[2] can be specified either explicitly by assigning to each instance (S,d) the maximizer of the product $(x_1 - d_1)(x_2 - d_2)$ over the individually rational part of (S,d), or implicitly by requiring f to satisfy the following four conditions:

1. Pareto optimality: $(f(S,d) + R_+^2) \cap S = f(S,d)$ for each (S,d).
2. Symmetry: $f_1(S,d) = f_2(S,d)$ for each (S,d) such that $(x_2,x_1) \in S$ whenever $(x_1,x_2) \in S$.
3. Scale invariance: For each (S,d), if A is a positive affine transformation of R^2 to itself, then $f(A(S),A(d)) = A(f(S,d))$.
4. Independence of irrelevant alternatives: For every pair $(S,d),(T,d)$ of instances such that $S \subseteq T$, if $f(T,d)$ belongs to S, then $f(T,d) = f(S,d)$.

Note that neither of these two definitions specifies some explicit bargaining procedure. It is therefore desirable to support every proposed solution function by specifying rules of some non-cooperative bargaining game the solution of which would give the same solution. This line of research aimed at bridging the gap between the non-cooperative and cooperative models was initiated by Nash who proved that his solution can be obtained as the limit of a sequence of the Nash equilibria of certain non-cooperative games. This approach is known under the name of the Nash Program, see [12]. However, in this paper we do not deal with such non-cooperative procedures. For a detailed discussion of such strategic approach to bargaining, we refer to [6] and [12].

3 Stepwise solutions

Probably most of cooperative bargaining problems in real life are resolved gradually by some step-by-step procedures during which the players reach intermediate agreements, and where these intermediate agreements are used as new disagreement points.

Recently, Diskin et al. [2] have formalized this idea by proposing to consider a solution concept which is composed of two solution functions. One function specifies an interim agreement and the other specifies the terminal agreement. More precisely, a *stepwise solution* of a bargaining problem $\mathcal{B}$ is a pair (f,g) of functions from $\mathcal{B}$ into R^n such that, for each instance (S,d) of $\mathcal{B}$,

- both $f(S,d)$ and $g(S,d)$ belong to S,
- and $f(S,d) = f(S,g(S,d))$.

3.1 The discrete Raiffa solution

One of the solution functions proposed by Raiffa in the early 1950's, called now the *discrete Raiffa solution*, is defined as the limit of a sequence $\{x^k\}$ of points from S generated as follows: Let (S,d) be an instance of $\mathcal{B}$. Define $u(S,d) = (u_1(S,d),u_2(S,d))$ and $m(S,d) = (m_1(S,d),m_2(S,d))$ by

$$u_1(S,d) = \max\{x_1 : (x_1,x_2) \in S_d^+\},\ u_2(S,d) = \max\{x_2 : (x_1,x_2) \in S_d^+\},\ m(S,d) = \frac{1}{2}u(S,d) + \frac{1}{2}d.$$

Set $x^0 = m(S,d)$, and continue inductively by defining $x^k = m(S,x^{k-1})$.

The convexity of S and the definition of function m guarantee that, for each k, the point x^k belongs to S and $x^{k+1} \geq x^k$. Because S is compact, we know that the sequence $\{x^k\}$ is convergent and its limit

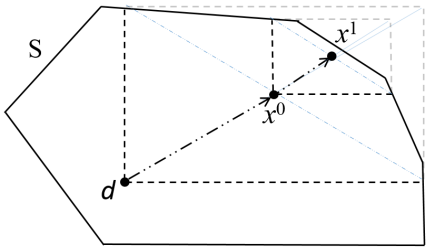


Figure 1 The discrete Raiffa solution

belongs to the Pareto frontier of S. Moreover, if the Pareto frontier of S is piecewise linear, then the convergence may be finite, see Fig. 1.

It can easily be seen that the discrete Raiffa's solution is a stepwise solution (f, g) where f and g are defined as follows:

$$f(S,d) = d + \frac{1}{2}(u(S,d) - d), \tag{1}$$
$$g(S,d) = d^{\infty}(S,d), \tag{2}$$

where $u(S,d) = (u_1(S,d), u_2(S,d))$, and $d^{\infty}(S,d)$ is the limit of the sequence $\{d^k(S,d)\}$ of points constructed inductively by $d^0(S,d) = d$ and $d^{k+1}(S,d) = f(S,d^k)$.

3.2 Generalized discrete Raiffa's solution

The definitions of functions f and g in the previous subsection immediately suggest an extension of discrete Raiffa's solution to problems with more than two players. Namely, for n players, we define $u(S,x) = (u_1(S,x), u_2(S,x), \ldots, u_n(S))$ where $u_i(S,x) = \max\{x_i : x \in S_d^+\}$ for each i, and then we set

$$f(S,d) = d + \frac{1}{n}(u(S,d) - d), \tag{3}$$
$$g(S,d) = d^{\infty}(S,d). \tag{4}$$

In fact, Diskin et al. generalized the two-player Raiffa solution and its extensions to more than two players by proposing the following family of stepwise solutions.

The *set of generalized Raiffa solutions* is a family of stepwise bargaining solutions $\{(f^p, g^p)\}_{0<p\leq 1}$ where, for each p, $0 < p \leq 1$, f^p and g^p are defined by

$$f^p(S,d) = d + \frac{p}{2}(u(S,d) - d), \tag{5}$$
$$g^p(S,d) = d^{\infty}(S,d), \tag{6}$$

where $d^{\infty}(S,d)$ is the limit of the sequence $\{d^k(S,d)\}$ of points constructed inductively by

$$d^0(S,d) = d \text{ and } d^{k+1}(S,d) = f^p(S,d^k).$$

However, when considering this approach for problems with more than two players, we have to take into consideration that some properties of higher dimensional spaces may essentially differ from the two dimensional spaces. This requires, at least for a number of problems, to adjust the domain of feasible instances. In this respect, Diskin et al. assume that each of the bargaining sets is also d-comprehensive.

It turns out (see Diskin et al. [2]) that a stepwise bargaining solution (f, g) satisfies the following seven axioms if and only if it is a generalized Raiffa solution.

Axiom 1. $g(S,d) = g(S, f(S,d))$.

[2]Recall that $\mathcal{B}$ is composed of instances (S,d) in which S is a convex compact subset of R^2 containing d and at least one point (x_1, x_2) such that $x_1 > d_1$ and $x_2 > d_2$.

Axiom 2. $g(S,d)$ is individually rational.

Axiom 3. If $f(S,d)$ is individually rational, and if d is not Pareto optimal in S, then $f(S,d) \neq d$.

Axiom 4. If all players are symmetric in (S,d), then they are also symmetric in $f(S,d)$.

Axiom 5. $f(A(S), A(d)) = A(f(S,d))$.

Axiom 6. If $S \subseteq T$, then $f(S,d) \leq f(T,d)$.

Axiom 7. If $S_d = T_d$, then $f(S,d) = f(T,d)$.

3.3 The Shapley-Shubik solution

Because much of modern economic theory is based on ordinal preferences, it is of interest to strengthen the condition of scale invariance by requiring invariance not only with respect to positive affine transformations but with respect to a larger set of utility transformations; namely, with respect to all strictly increasing transformations of individual utilities. However, it is known (see [13] or [15]) that, in the case of two-player bargaining problem considered by Nash, all solution functions satisfying this stronger requirement are quite uninteresting. This impossibility result, however, does not hold for the Nash problem with more than two players, and also not for two player problems whose bargaining sets are finite.

Such an ordinal solution for the three-player problem has been constructed by Shapley and Shubik already in [14], see also [15]. Recently Safra and Samet [10], [11] have generalized the Shapley-Shubik solution for problems with more than three players and demonstrated that there is even a continuum of ordinal solutions that are efficient and symmetric.

The construction of ordinal solutions for the three-player problem is based on the observation that if Q is a Pareto surface in R^3 and (a_1, a_2, a_3) is a point in $R^3 \setminus Q$, then there is a unique point (b_1, b_2, b_3) on the other side of Q such that the points $(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)$ belong to S. Using this fact, one can define the Shapley-Shubik solution outcome for each instance (S,d) as the limit of the sequence $\{x^k\}$ of points defined by setting $x^0 = d$ and defining x^k to be the unique point $x^{k+1} = (x_1^{k+1}, x_2^{k+1}, x_3^{k+1})$ determined by the property that the points $(x_1^k, x_2^{k+1}, x_3^{k+1}), (x_1^{k+1}, x_2^k, x_3^{k+1}), (x_1^{k+1}, x_2^{k+1}, x_3^k)$ belong to the Pareto surface of S. It can easily be seen that such solutions can be formalized as stepwise solutions.

3.4 The modified Raiffa solution

Recently Budinský, Valenčík et al. [1] studied a class of problems that can be viewed as three-player bargaining problems $\mathcal{B}$ consisting of instances (S,d) in which S is the intersection of R^3_+ and a three-dimensional convex cone containing the origin and having its vertex in R^3_{++}.

The following example can serve as illustration of problems studied in [1]. Let u, v, w be positive real numbers and η be a real number from the unit interval $[0,1]$. Consider the inequality

$$(x_1 - u) + (x_2 - v) + (x_3 - w) + \eta R(x - u, y - v, z - w) \leq 0 \tag{7}$$

with three real unknowns x_1, x_2, x_3, where h is the following function of three real variables ξ_1, ξ_2, ξ_3:

$$h(\xi_1, \xi_2, \xi_3) \quad = \quad \max(|\xi_1|, |\xi_2|, |\xi_3|). \tag{8}$$

Now define S as the intersection of the solution set of inequality (7) and R^3_+, and choose as d any point of R^3_+ which satisfies (7) strictly. It can be verified that the solution set of (7) is a three-dimensional convex cone with vertex at (u, v, w). Fig. 2 depicts the case given by $(d_1, d_2) = (0,0)$, $\eta = 0.5, u = 6, v = 4, w = 2$. To solve the problem, Budinský, Valenčík et al. [1] proposed a procedure that can be viewed as a modification of Raiffa's procedure by incorporating the basic observation used in the construction of the Shapley-Shubik ordinal solution.

In general, the solution outcome to an instance (S,d) is constructed as the limit of the sequence $\{y^k\}$ of points from S defined as follows. Set $y^0 = d$, and let (x_1, x_2, x_3) be the point obtained from y^0 by one step of the Shapley-Shubik procedure. Construct the next point y^1 by the same averaging that is used in the discrete Raiffa procedure, but now using the points

$$(y_1^0, x_2, x_3), \ (x_1, y_2^0, x_3), \ (x_1, x_2, y_3^0)$$

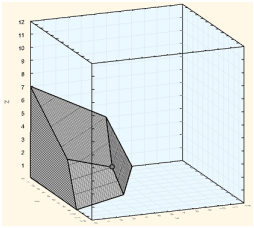


Figure 2 A cone in the Chebyshev metric

instead of using the points

$$(u_1(S, y^0), y_2^0, y_3^0),\ (y_1^0, u_2(S, y^0), y_3^0),\ (y_1^0, y_2^0, u_3(S, y^0)).$$

Then continue in the same way, that is, construct y^{k+1} from y^k as follows. First use the fact that there is a unique point (x_1, x_2, x_3) such that the points $(y_1^k, x_2, x_3), (x_1, y_2^k, x_3), (x_1, x_2, y_3^3)$ belong to S and then set

$$y^{k+1} = \frac{1}{3}((y_1^k, x_2, x_3) + (x_1, y_2^k, x_3) + (x_1, x_2, y_3^k)).$$

Again the convexity of S guarantees that, for each k, the point y^k belongs to S and $y^{k+1} \geq y^k$. Because the set S is compact, we know that the sequence $\{y^k\}$ converges to a point in S; in fact to a point on the Pareto surface of S.

4 Notes and Comments

The procedure proposed in the previous section seems to be new. It can be used for solving the three-player problems to which it is possible to apply the Shapley-Shubik procedure, provided that the results of the averaging taken from the discrete Raiffa procedure belong to the bargaining set. Therefore some non-convexity in bargaining sets can be permitted.

It would be interesting to compare the procedure with other available procedures on some standard classes of problems. For example, there is some evidence that the proposed modification is faster than the Shapley-Shubik and discrete Raiffa procedures when applied to standard instances. Extensions to problems with more than three players[3] would also be of interest. Also it is worth noting that, for many classes of bargaining sets, the triples of points used in the procedure form a unique finite von Neumann-Morgenstern stable set.

However, the main open question consists in establishing systems of axioms that define the proposed stepwise solution uniquely on reasonable classes of bargaining problems.

5 Acknowledgment

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 DYME – Dynamic Models in Economics.

References

[1] Budinský, P., Valenčík, R. et al.: *Theory of Redistribution Systems (in Czech)*. Vysoká škola finanční a správní, EUPRESS, Prague, 2010.

[2] Diskin, A., Koppel, M., and Samet, D.: Generalized Raiffa solutions, *Games and Economic Behavior* **73** (2011), 452–458.

[3] The procedure can be modified to be applicable to the two player problems, but then it is identical to the discrete Raiffa procedure.

[3] Luce, R. D., and Raiffa, H.: *Games and Decisions: Introduction and Critical Survey*. Wiley, New York, 1957.

[4] Nash, J.: The Bargaining Problem, *Econometrica* **18** (1950), 155–162.

[5] Nash, J.: Two Person Cooperative Games, *Econometrica* **21** (1953), 128–140.

[6] Osborn, M. J., and Rubinstein, A.: *Bargaining and Markets*. Academic Press, San Diego, California, 1990.

[7] Raiffa, H.: Arbitration schemes for generalized two-person games. Report No. M720-1, R30, Engineering Research Institute, University of Michigan, Ann Arbor, June, 1950.

[8] Raiffa, H.: Arbitration Schemes for Generalized Two Person Games. In: *Contributions to the Theory of Games II* (H. W. Kuhn and A. W. Tucker eds.), Princeton University Press, Princeton, New Jersey, 1953, 361–387.

[9] Roth, A. E.: textitAxiomatic models of bargaining. Springer Verlag, Berlin, Heidelberg, New York, 1979.

[10] Safra, Z., and Samet, D.: An ordinal solution to bargaining problems with many players, *Games and Economic Behavior* **46** (2004), 129–142.

[11] Samet, D., and Safra, Z.: A family of ordinal solutions to bargaining problems with many players, *Games and Economic Behavior* **50** (2005), 89–106.

[12] Serrano, R.: Fifty years of the Nash program, *Investigaciones Economicas* **29** (2005), 219–258.

[13] Shapley, L.: Utility comparison and the theory of games. In: *La Décision, edited by Guildband*, G. Th., Editions du CNRS, Paris, 1969, 251–263. Reprinted in: *The Shapley value* (A.E. Roth ed.), Cambridge University Press, 1988, 307–319.

[14] Shapley, L. S., and Shubik, M.: *Game Theory in Economics*, Chapter 4: Preferences and Utility. R-904/4-NSF, Published by The Rand Corporation, 1974.

[15] Shubik, M.: *Game Theory in the Social Sciences*. The MIT Press, Cambridge, Mass., 1982, 1984.

[16] Thomson, W.: Bargaining and the theory of cooperative games: John Nash and beyond. Working Paper No. 554, University of Rochester, 2009.

Statistical analysis of competing risks in an unemployment study

Petr Volf[1]

Abstract. This study continues in the theme of contribution Volf (2010) and extends it considerably. While the previous paper was devoted mainly to the analysis of real incidence of competing events, the present one is much more concerned with the analysis of dependence of these events (more precisely, of random variables - latent times to events). To do it, we discuss first the problem of identifiability of marginal and joint distributions of competing random variables. Then, the copula models are utilized in order to express the dependence. Finally, the Gauss copula is used to solution of a real example with unemployment data.

Keywords: statistics, survival analysis, competing risks, copula, unemployment data.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

The problem of competing risks, except in the field of reliability, biostatistics and medical studies, is also often studied in demography, labour statistics, and in econometrics generally. In the insurance mathematics the setting of competing risks is sometimes called the multiple decrement model (c.f. Arnold and Brockett, 1983). The interest in the problem dates back to 70-ties of the last century. From the beginning it was revealed that in the competing risks setting the background model may not be identifiable. A proof and an example of this phenomenon is given in Tsiatis (1975), some instances of identifiable (or not) models are presented in Basu and Ghosh (1978). In these classical studies the notion of copula has not been used yet. Just later it was recognized that the use of copula for multi-dimensional continuous distribution can lead to a 'nice' closed form of the model. Therefore we are facing the problem of reasonable copula selection. Fortunately, it is known (cf. Zheng and Klein, 1995), that the selection of copula type is not crucial to a good fit of the model, that the finding proper value of its parameter (connected with correlation) is much more important.

The outline of the paper is the following: The next section introduces the scheme of competing risks, presents the method of analysis of competing events incidence, and points to the problem of possible non-identifiability of their marginal distributions. We shall mention also certain identifiability results in the framework of regression models. Then the notion of copula is recalled and used in competing risks model formulation. As a particular example, in Section 3 the Gauss copula is introduced and the procedure of simultaneous maximum likelihood estimation of marginal distributions and correlation is shown. This approach is applied in Section 4 containing a real example. We use the data on unemployed people from Han and Hausman (1990). There are two competing chances to re-gain an employment, we analyze their marginal and joint distribution (i.e. also their dependence), with the aid of Gauss copula model.

2 Competing risks and incidence

Let us recall the competing risks situation: Certain event (e. g. a failure of a device) can be caused by K reasons. It means that there are K (possibly dependent) random variables $T_j, j = 1, ..., K$, some-

[1]Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, Praha 8, Czech Republic, volf@utia.cas.cz

times accompanied by a variable C of random right censoring (C is then independent of all T_j). Let $\overline{F}_K(t_1,...,t_K) = P(T_1 > t_1,...,T_K > t_K)$ be the joint survival function of $\{T_j\}$. However, instead the 'net' survivals T_j we standardly observe just 'crude' data (sometimes called also 'the identified minimum') $Z = \min(T_1,...,T_K,C)$ and the indicator $\delta = j$ if $Z = T_j$, $\delta = 0$ if $Z = C$. Such data lead us to direct estimation of the distribution of $Z = \min(T_1,...,T_K)$, for instance its survival function $S(t) = P(Z > t) = \overline{F}_K(t,...,t)$. Further, we can estimate so called **incidence densities**

$$f_j^*(t) = dP(Z = t, \delta = j) = -\frac{\partial \overline{F}_K(t_1,...,t_K)}{\partial t_j}|(t_1 = ... = t_K = t),$$

and also their integrals, **cumulative incidence functions**

$$F_j^*(t) = \int_0^t f_j^*(s)\,ds = P(Z \le t, \delta = j).$$

Notice that $\lim F_j^*(t) = P(\delta = j) < 1$ if $t \to \infty$, $S(t) = 1 - \sum_{j=1}^K F_j^*(t)$.

A more practical form of the cumulative incidence function (more convenient for statistical estimation) uses so called **cause–specific hazard functions** for events $j = 1, 2, \ldots, K$:

$$h_j^*(t) = \lim_{d\to 0} \frac{P(t \le Z < t + d,\, \delta = j \,|\, Z \ge t)}{d}.$$

Overall hazard rate for $Z = \min(T_1,...,T_K)$ is then:

$$h^*(t) = \lim_{d\to 0} \frac{P(t \le Z < t + d \,|\, Z \ge t)}{d} = \sum_{j=1}^K h_j^*(t),$$

by integration the cumulated hazard rates $H_j^*(t)$, $H^*(t)$ are obtained. Consequently, $S(t) = P(Z > t) = \exp(-H^*(t))$. Then $f_j^*(t) = h_j^*(t) \cdot S(t)$ and the cumulative incidence functions can be written as

$$F_j^*(t) = P(Z \le t, \delta = j) = \int_0^t S(s) \cdot h_j^*(s)\,\mathrm{d}s.$$

As both components, i.e. S and h_j^*, are estimable consistently by standard survival analysis methods, it follows that there also exist consistent estimates of F_j^*, see for instance Lin (1997), Scheike and Zheng (2008) in a regression context, also Volf (2010).

2.1 Problem of non-identifiability

However, in general, from data (Z_i, δ_i), $i = 1, \ldots, N$ it is not possible to identify neither marginal nor joint distribution of $\{T_j\}$. A. Tsiatis (1975) has shown that for arbitrary joint model we can find a model with independent components having the same incidences, i.e. we cannot distinguish the models. Namely, this 'independent' model is given by cause-specific hazard functions $h_j^*(t)$. In other words, even if the model is parametrized and the MLE yields consistent estimates, in general we do not know parameters of which model are estimated.

The situation can be better in the case of a regression model, because the covariates provide an additional information, especially when their structure is rich enough. There are numerous results showing conditions for full model identifiability, let us mention here Heckman and Honoré (1989) and their proof of identifiability in the Cox or the AFT model cases. Lee (2006) investigated more general transformation models of regression. Berg et al. (2007) have studied two competing transition rates from unemployment state. They have used a discrete-time multiplicative regression model with latent heterogeneities and their main identifying assumption is basically the same as that of Heckman and Honoré, namely that exit rates should not vary with the observed covariates in exactly the same way. However, all these studies rely on an assumption that the dependence structure (in the next section given by a copula parameter) does not change with covariates. If it is not the case, the problem of identifiability arises anew.

Further, as a consequence of the Tsiatis (1975) result, in competing risks models without regressors it is necessary to make certain functional assumptions about the form of both marginal and joint distribution in order to identify them. Several such cases are studied in Basu and Ghosh (1978) and in some other papers.

2.2 Competing risks and copula

In the sequel we shall consider just 2 competing events, i.e. random variables S, T, censoring variable C, and observed data – realizations of N i.i.d. random variables $Z_i = min(S_i, T_i, C_i), \delta_i = 1, 2, 0$, $i = 1, 2, ..., N$. The notion of copula offers a way how to model multivariate distributions, namely the joint distribution function $F_2(s, t)$ of S, T:

$$F_2(s, t) = C(F_S(s),\, F_T(t),\, \theta), \tag{1}$$

F_S, F_T are marginal distribution functions of S, T, $C(u, v, \theta)$ is a copula, i.e. a two-dimensional distribution function on $[0, 1]^2$, with uniformly on $[0, 1]$ distributed marginals U, V, θ is a copula parameter. The parameter is, as a rule, uniquely connected with correlation of U, V, hence also with correlation of S, T. It is seen that the use of copula allows to model the dependence structure separately from the analysis of marginal distributions. Hence, the identifiability of the copula (and its parameter) and marginals can be considered as two separate steps.

Zheng and Klein (1995) proved that when the copula is known, the marginal distributions are estimable consistently (and then the joint distribution, too, from (1)), even in non-parametric (so that quite general) setting. However, in general, also value of θ is needed, because (again due to Tsiatis, 1975) without fully determined copula we are not able to distinguish between the 'true' model and corresponding independent one. On the other hand, Zheng and Klein also discussed importance of proper selection of copula form. As it has already been said, the knowledge (or a good estimate) of parameter θ is much more crucial for correct model of joint distribution. As a consequence, because the knowledge of copula type is still an unrealistic supposition, we can try to use certain sufficiently flexible class of copulas, as approximation, and concentrate to reliable estimation of its parameter.

3 Gauss copula

There exist a large number of different copula functions, among them for instance a set of Archimedean copulas. However, we concentrate here to one rather universal and flexible copula type, namely to Gauss copula (in our setting used for connecting just two random variables). Let X, Y be standard normal random variables $\sim N(0, 1)$ tied with (Pearson) correlation $\rho = \rho(X, Y)$. We denote ϕ, φ univariate standard normal distribution function and density and by $\phi_2(x, y)$, $\varphi_2(x, y)$ corresponding 2-dimensional functions. Then

$$\varphi_2(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}\Big\}$$

with $\boldsymbol{x} = (x, y)'$ and Σ the covariance matrix $[1, \rho; \rho, 1]$. If we define $U = \phi(X)$, $V = \phi(Y)$, we obtain a 2-dimensional distribution on $(0, 1)^2$ with the copula

$$C(u, v) = \phi_2\big(\phi^{-1}(u),\, \phi^{-1}(v)\big). \tag{2}$$

Naturally, $\rho(U, V) \neq \rho(X, Y)$ (though they are rather close, as a rule), while Spearman's correlations coincide, namely $\rho_{\mathrm{SP}}(X, Y) = \rho_{\mathrm{SP}}(U, V) = \rho(U, V)$. We can connect also density functions. Let $c(u, v)$ be the joint density of (U, V), then

$$c(u, v) = \frac{\varphi_2(x, y)}{\varphi(x) \cdot \varphi(y)},$$

again with $u = \phi(x)$, $v = \phi(y)$. As we are primary interested in the model for dependence of competing variables S, T, let us assume that their joint distribution function is given by Gauss copula (2),

$$F_2(s, t) = \phi_2\big(\phi^{-1}(F_S(s)),\, \phi^{-1}(F_T(t))\big), \tag{3}$$

and $S = F_S^{-1}(\phi(X))$, $T = F_T^{-1}(\phi(Y))$. Again $\rho_{\mathrm{SP}}(S, T) = \rho_{\mathrm{SP}}(U, V)$, and "initial" $\rho = \rho(X, Y)$ is the only parameter describing the dependence of S and T. It, naturally, differs from $\rho(S, T)$, however, all values $\rho(S, T)$ can be achieved by convenient choice of $\rho(X, Y)$. Let us remark here that the real dependence among S, T can be much more complicated, nevertheless the use of Gauss copula offers here certain rather simple and sufficiently flexible (as regards the correlation) set of distributions.

3.1 Estimation in Gauss copula model

When parameter ρ is known, copula (2) is fully defined and from Zheng, Klein (1995) it follows that the distribution of (S,T) can be estimated, in parametric and even non-parametric setting. However, without knowledge of ρ nonparametric model is not identifiable and in the parametric setting explicit proofs of identifiability are available for just certain types of marginal distributions. That is why in the following example we shall assume log-normal marginal distributions. Their identifiability in a framework of Gauss copula follows from the result of Basu and Ghosh (1978, Sect. 7), as log-normal variables are the same monotone transformation of normal variables. It also means that after log transformation of data we can work with Gauss marginal distributions. Naturally, the fit of chosen model to the data has to be tested.

The estimation procedure will be based on the maximum likelihood method. The data are (Z_i, δ_i), $i = 1, \ldots, N$, the likelihood function then has the form

$$L = \prod_{i=1}^{N} \left\{ -\frac{\partial}{\partial s} \overline{F}_2(s,t) \right\}^{I[\delta_i=1]} \cdot \left\{ -\frac{\partial}{\partial t} \overline{F}_2(s,t) \right\}^{I[\delta_i=2]} \cdot \overline{F}_2(s,t)^{I[\delta_i=0]},$$

evaluated at $s = t = Z_i$, with $\overline{F}_2(s,t) = P(S > s, T > t) = 1 - F_S(s) - F_T(t) + F_2(s,t)$. From transformation (3) it follows that $F_2(s,t) = \phi_2(x,y)$ with $x = \phi^{-1}(F_S(s))$, $y = \phi^{-1}(F_T(t))$. Hence, when we put $X_i = \phi^{-1}(F_S(Z_i))$, $Y_i = \phi^{-1}(F_T(Z_i))$, we obtain after some computation – integration of 2-dimensional Gauss density $\varphi_2(x,y)$, that

$$L = \prod_{i=1}^{N} \left\{ f_S(Z_i) \left[1 - \phi_1(Y_i; \rho X_i, 1-\rho^2) \right] \right\}^{I[\delta_i=1]} \cdot$$

$$\cdot \left\{ f_T(Z_i) \left[1 - \phi_1(X_i; \rho Y_i, 1-\rho^2) \right] \right\}^{I[\delta_i=2]} \cdot \left\{ 1 - F_S(Z_i) - F_T(Z_i) + \phi_2(X_i, Y_i) \right\}^{I[\delta_i=0]}, \qquad (4)$$

where $\phi_1(x; \mu, \sigma^2)$ denotes the distribution function of normal distribution ($N(\mu, \sigma^2)$, evaluated at x. Parameter ρ is hidden in ϕ_1 and in ϕ_2. Distributions of S and T are present both explicitly and also implicitly, in transformed X_i, Y_i. It is seen that the problem of maximization is not an easy task and has to be solved by a convenient search procedure.

4 Application

Han and Hausman (1990) have analyzed the data on unemployment duration, with two competing chances to leave the unemployment state, either by obtaining a new job (our variable S) or by a recall to the former employer (variable T). Some histories of unemployment were censored (by a variable C), independently, i.e. terminated from other reasons. The data on together 1051 people are collected in Table III of Han and Hausman (with several insignificant misprints which we have corrected), the time is in fact discrete, data are aggregated to weeks. The data show some (rare, however) non-regularities, visible also in the graph of cause-specific hazard rates in Figure 1, for instance significantly larger numbers of events in 26-th week which may be related to a change of support after the first half-year of unemployment. We analyzed first the incidence of both competing variables separately. It was assumed that cause-specific hazards were constant during each week and that each person could experience just one (potential) event in a week (i.e. that for each person i the values S_i, T_i, C_i – though two of them 'not-realized' – must be different). Figure 1 shows estimated cause-specific hazard rates, their cumulated sums, and finally estimated cumulated incidence functions, together with estimated distribution function of $\min(S,T)$.

Han and Hausman had also an information on several covariates which was not available to us. They therefore used a discrete version of Cox regression model. It has to be said that they also used certain not fully correct approximations, for instance substituting Gumbel distribution by the Gauss one. In fact their estimate of correlation was not significant, i.e. they actually have shown that in the framework of their model (with their approximations) the risks are conditionally independent, given the covariates.

We concentrated to the analysis of competing risks in the framework of the Gauss copula model. Further, we assumed log-normal marginal distributions of S and T in order to assure the identifiability. Reasonability of such an assumption was checked in the following graphical way: $M = 500$ samples of new competing-risks data, each of extent $N = 100$, were randomly generated from estimated model. From each sample, the cumulated incidence function was estimated. They are plotted on Figure 2 and

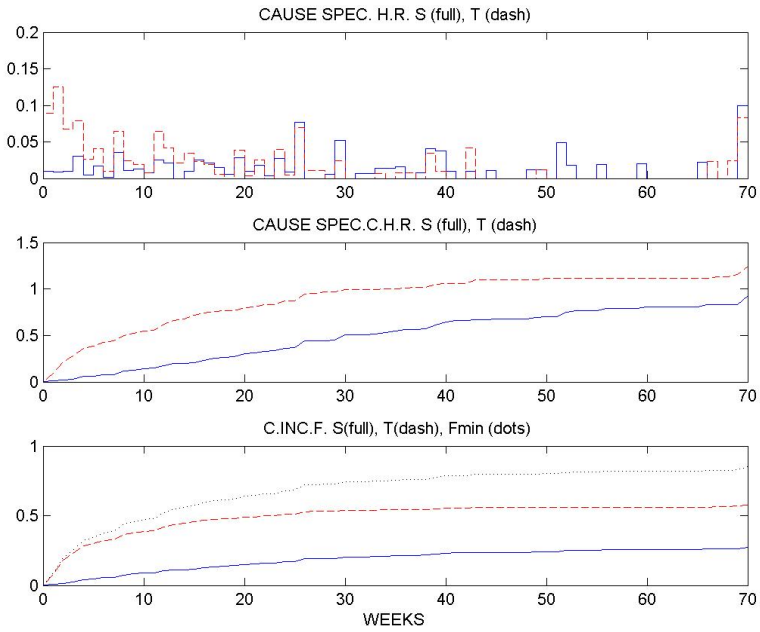


Figure 1 Estimated cause-specific hazard functions (above), cumulated cause-specific hazards (middle), cumulated incidence functions (below), for S (full) and for T (dashed curves). Dotted is estimate of F_{min}.

compared with cumulated incidence functions obtained from real data (thick curves). It is seen that the 'clouds' of generated curves are around real ones, in both cases. It could be taken as a graphical goodness-of-fit test supporting our idea of log-normal distributions.

A random search procedure for maximum of likelihood (4) brought us to the following final results:

$$\mu_S = 2.7966,\ \sigma_S = 1.2299,\ \mu_T = 2.4749,\ \sigma_T = 1.5852,\ \rho = 0.8620.$$

Here $\rho = \rho(X, Y)$ of corresponding standard Gauss variables (see definition of Gauss copula in preceding section), while numerically computed $\rho(U, V) = 0.849$ and, finally, $\rho(S, T) = 0.519$. It is large positive, indicating strong dependence between both competing variables. As the solution of the MLE was based on a numerical optimization procedure, we were not able to assess confidence intervals of involved parameters. In fact, graphs on Figure 2 provide at least partial information on the whole model reliability. As an alternative we have considered also Weibull marginal distributions, also connected by Gauss copula. While estimated correlation was comparable, achieved maximum of likelihood was smaller and the graphical comparison as in Figure 2 indicated significantly worse fit of this model.

5 Conclusion

We have studied the problem of competing risks with the focus on assessing the dependence of competing random variables and identifying their marginal as well as joint distributions. The joint distribution was expressed with the aid of a copula, the case of Gauss copula was investigated in more details. Proposed model was then utilized in an example with real unemployment data. Statistical analysis revealed positive correlation between times to both competing events. On the other hand, even the experience with artificial data indicates that in the framework of chosen Gauss copula competing risks model the log-likelihood is flat and the convergence of computations to its (hardly detectable) maximum is rather slow.

Acknowledgements

The research was supported by the grant No. 13-14445S of the Grant Agency of the Czech Republic.

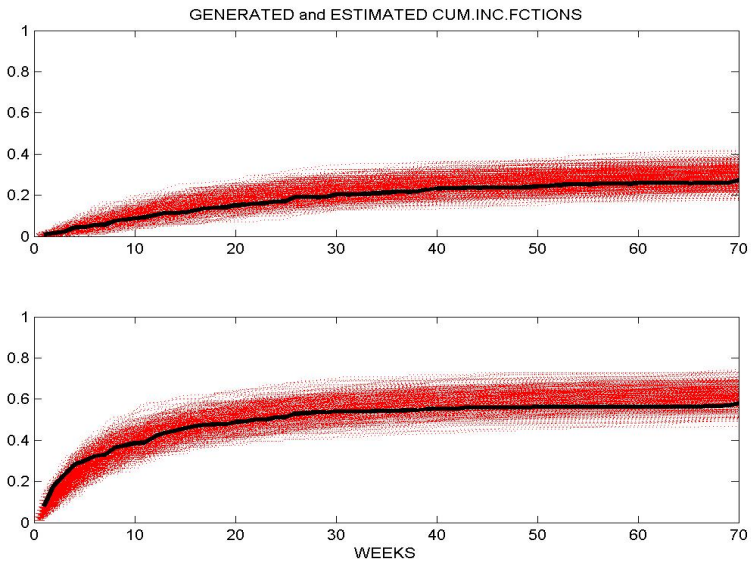


Figure 2 Set of cumulated incidence functions estimated from generated data, above for S, below for T, thick curves – cumulated incidence functions from real data (the same as in Figure 1 below)

References

[1] Arnold, B.C. and Brockett, P.L.: Identifiability for dependent multiple decrement/competing risk model, *Scand. Actuarial J.* 1983, 117–127.

[2] Basu, A.P. and Ghosh, J.K.: Identifiability of the Multinormal and Other Distributions under Competing Risks Model, *Journal of Multivariate Analysis* **8** (1978), 413–429.

[3] Van den Berg, G.J., van Lomwel, A.G.C., and van Ours, J.C.: Nonparametric estimation of a dependent competing risks model for unemployment durations, *Empirical Economics* **34** (2008), 477-491

[4] Han, A. and Hausman J.A.: Flexible parametric estimation of duration and competing risk models, *J. of Applied Econometrics* **5** (1990), 1–28.

[5] Heckman, J.J. and Honoré, B.E.: The identifiability of the competing risks model, *Biometrika* **76** (1989), 325–330.

[6] Lee, S.: Identifcation of a competing risks model with unknown transformations of latent failure times, *Biometrika* **93** (2006), 996–1002.

[7] Lin, D.Y.: Non-parametric inference for cumulative incidence functions in competing risks studies, *Statistics in Medicine* **16** (1997), 901–910.

[8] Scheike, T.H. and Zhang, M.: Flexible competing risks regression modelling and goodness-of-fit, *Lifetime Data Analysis* **14** (2008), 464–483.

[9] Tsiatis, A.: A nonidentifiability aspects of the problem of competing risks, *Proc. Nat. Acad. Sci. USA* **72** (1975), 20–22.

[10] Volf, P.: On statistical modeling of incidence of competing events, with application to labor mobility analysis. In: *Proceedings of the MME 2010*, JCU Ceske Budejovice, 2010, 670–675.

[11] Zheng, M. and Klein, J.P.: Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika* **82** (1995), 127–138.

Bimodality testing of the stochastic cusp model

Jan Voříšek[1], [2]

Abstract. Multimodal distributions are popular in many areas: biology (fish and shark population), engineering (material collapse under pressure, stability of ships), psychology (attitude transitions), physics (freezing of water) etc. There were a few attempts to utilize multimodal distributions in financial mathematics as well (e.g. [2], [6], [5]).

Cobb et al. [4] described a class of multimodal distributions belonging to the exponential family, which has unique maximum likelihood estimators and showed a connection to the stationary distribution of the stochastic cusp catastrophe model. Moreover was shown, how to identify bimodality for given parameters of the stochastic cusp model using the sign of Cardans discriminant.

A statistical test for bimodality of the stochastic cusp model using maximum likelihood estimates is proposed in the paper as well as the necessary condition for bimodality which can be used for simplified testing to reject bimodality. By proposed methods is tested the bimodality of exchange rate between USD and GBP in the periods within the years 1975 - 2014.

Keywords: multimodal distributions, stochastic cusp model, statistical bimodality test.

JEL classification: C46
AMS classification: 62F03

1 Introduction

Stochastic cusp model has its name thanks to classification of singularities by Vladimir Arnold [1] within deterministic models of catastrophe theory proposed by Rene Thom [9], where cusp model is the most used one. Its deterministic version was popularized in the 1970s and in the 1980s was developed the theory of the stochastic version by Lauren Cobb [3]. He also proposed the numerical maximum likelihood method and methods of moments for estimation of parameters. There are other methods for estimation of catastrophe models, however the paper will utilize the properties of maximum likelihood estimators for the statistical testing.

The specification of the stochastic cusp model is such that the probability density function belongs to the class of generalized exponential distributions. It is convenient that probability density function of the stochastic cusp model accommodates variable skewness, kurtosis, and even bimodality. Bimodality of the model is accessed by Cardan's discriminant, which is negative, when the probability density function has two modes. These properties as well as the existence of estimating methodology encouraged empirical research especially in behavioural science and psychology. The distinct perception of agents in the market was the motivation for the usage of stochastic cusp model in the financial mathematics (eg. by Zeeman [10], Creedy and Martin [5] or Fernandes [6]).

Despite the frequent usage of the stochastic cusp model the statistical test for negativeness of Cardan's discriminant, which implies bimodality, calculated from estimated parameters, was not proposed. This paper tries to fill this gap by approximating the distribution of estimated Cardan's discriminant using delta method. The approximate distribution is than used to propose a statistical test, where the null hypotheses is rejected, if there is enough statistical evidence against bimodality of the distribution.

[1]Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation, Department of Econometrics, Pod Vodárenskou věží 4, CZ-182 08 Prague 8, Czech Republic, vorisek@karlin.mff.cuni.cz

[2]Faculty of Mathematics and Physics, Charles University Prague, Department of Probability and Mathematical Statistics, Sokolovská 83, CZ-186 75 Prague 8, Czech Republic, vorisek@karlin.mff.cuni.cz

The rest of paper is organized as follows. Section 2.1 describes a cusp model and its connection to the multimodal distributions, section 2.2 shows different approaches to estimate parameters and the statistical test of bimodality is proposed in the Section 2.3. Section 3 utilizes the methodology for the real data example of USD, GBP exchange rate and demonstrates different situations, where the bimodality is rejected. Section 4 summarizes the results and concludes.

2 Cusp model

2.1 Connection to multimodal distributions

This section summarizes the connection of the stochastic cusp model to the exponential class of multimodal distributions. The standard parametrization of both approaches is shown as well as the transformation from one to another following Cobb et al. [4].

The generalized exponential family of distributions is characterized by probability density function

$$f_k(x) = \xi(\beta) \exp\left[\int_a^x \frac{g(s)}{v(s)} ds\right], \tag{1}$$

where $g(x) = \sum_{i=0}^{k} \beta_i x^k$ is polynomial function of order $k > 0$ and function $v(x)$ has one of the form:

$$\begin{array}{rcl} \text{typeN}: & v(x) = 1 & -\infty < x < \infty \\ \text{typeG}: & v(x) = x & 0 < x < \infty \\ \text{typeI}: & v(x) = x^2 & 0 < x < \infty \\ \text{typeB}: & v(x) = x(1-x) & 0 < x < 1. \end{array} \tag{2}$$

Let (a, b) be an open interval, where $v(x)$ is positive and $\xi(\beta)$ is normalizing constant for $\int_a^b f_x(x)dx$ to be unity.

Cusp model in polynomial parametrization is given by the polynomial:

$$g(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3, \tag{3}$$

where $b_3 < 0$ and $v(x) = 1$. Then the probability density function will be

$$f_p(x) = \xi_p(\mathbf{b}) \exp\left[b_0 x + \frac{b_1}{2} x^2 + \frac{b_2}{3} x^3 + \frac{b_3}{4} x^4\right], \tag{4}$$

where $\xi_p(\mathbf{b})$ is normalizing constant depending on parameters b_0, b_1, b_2 and b_3.

To get the probability density function of standard cusp model parametrization:

$$f_c(z) = \xi_c(\theta) \exp\left[\alpha z + \frac{\beta}{2} z^2 - \frac{1}{4} z^4\right], \tag{5}$$

where $z = \frac{x-\lambda}{\sigma}$, $\xi_c(\theta)$ is the normalizing constant and θ stands for parameters $\alpha, \beta, \lambda, \sigma$, one needs to utilize the following substitutions:

$$\begin{aligned} \sigma &= (-b_3)^{-1/4}, \\ \lambda &= -b_2/(3b_3), \\ \beta &= (b_1 + b_2\lambda)\sigma^2, \\ \alpha &= \sigma g(\lambda). \end{aligned} \tag{6}$$

The cusp distribution in both types of parametrization may also be characterized by nonlinear diffusion processes. Let $\sigma^2(x) := v(x)$ and $\mu(x) := \frac{1}{2}\left(g(x) + \sigma^2(x)'\right)$. Then $f_p(x)$ is the stationary density function of a process x_t that is driven by the stochastic differential equation:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \tag{7}$$

where W_t is a standard Wiener process.

For identifying the bimodality (or unimodality) of the cusp probability density function (5), one needs to calculate the Cardan's discriminant

$$\delta_C = \left(\frac{\alpha}{2}\right)^2 - \left(\frac{\beta}{3}\right)^3, \tag{8}$$

which is negative, when the probability density function is bimodal and positive in unimodality case. The parameters α (*asymetry*) and β (*bifurcation*) are invariant with respect to changes in λ (*location*) and σ (*scale*), as is δ_C, and they have following approximate interpretations. If $\delta_C \geq 0$ then α measures skewness and β kurtosis and when $\delta_C < 0$ then α indicates the relative height of the two modes and β their relative separations.

2.2 Estimation of parameters

For the estimation of the polynomial parametrization (4) Cobb et al. [4] proposed a moment recursion relations which connects $k+1$ parameters to the first $2k$ moments of the probability density function. However, these estimates can not be utilized for the testing of bimodality, but could serve after the transformation (6) as a starting values for the numerical search for the maximum likelihood estimators $\hat{\theta}$ of the cusp probability density function (5). By following the theory of exponential families (e.g. Lehman [8]) we know that maximum likelihood estimators of the polynomial form exist, are unique can be found for example by a Newton-Raphson search. Other possible way is to use an R-package called "cusp" [7], which uses R build in function *optim* for maximizing the log likelihood of observed values. The variance matrix of the MLE is estimated using the Hessian matrix of the log likelihood function. Then the asymptotic distribution of the MLE is:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, I^{-1}\right), \tag{9}$$

where I is the Fisher information matrix.

2.3 Formulation of the test

The necessary condition for Cardan's discriminant to be negative is the positivity of parameter β. This condition could be statistically tested at first using the MLE result $\hat{\beta} \xrightarrow{d} N(\beta_0, \hat{\sigma}^2_\beta)$, where $\hat{\sigma}^2_\beta$ denotes the estimate of $\hat{\beta}$ variance from the Hessian matrix of the log likelihood function. The null hypothesis would be $H0 : \beta = 0$ with the one sided alternative $H1 : \beta < 0$. Rejection of the null hypothesis implies rejection of the negativity of Cardan's discriminant, which means rejection of bimodality.

For the direct testing of bimodality serves the approximate distribution of Cardan's discriminant δ_C (10) derived using delta method from (9). The delta method yields

$$\sqrt{n}\left(\hat{\delta}_C - \delta_0\right) \xrightarrow{d} N\left(0, \nabla h(\theta)^T I^{-1} \nabla h(\theta)\right), \tag{10}$$

where the function $h(\theta) = \left(\alpha^2/4 - \beta^3/27\right)$ transforms parameters of cusp probability density function (5) into Cardan's discriminant. This result allows to propose a statistical test for bimodality: we would like to test $H0 : \delta_C = 0$ against $H1 : \delta_C > 0$. Similarly to testing only parameter β, the test statistics has asymptotically normal distribution and rejecting of null hypothesis means the rejection of bimodality.

3 Empirical testing of bimodality

	1975-2014	1993	2003	2011
mean	1.699	1.502	1.635	1.603
median	1.639	1.497	1.624	1.607
std. dev.	0.235	0.036	0.052	0.030
skewness	0.859	-0.057	0.759	-0.379
kurtosis	3.914	2.680	2.983	2.207
minimum	1.042	1.418	1.550	1.534
maximum	2.455	1.593	1.668	1.627

Table 1 Descriptive statistics.

The performance of the statistical test will be illustrated using the exchange rate of USD and GBP, which is inspired by article [5], which suggested using multiple equilibria model for exchange rate. The data represents indicative middle market (mean of spot buying and selling) rates as observed by the Bank's Foreign Exchange Desk in the London interbank market around 4pm from the beginning of the year 1974 till the end of 2014. The basic statistical characteristic are presented in Table 1.

	1975-2014	1993	2003	2011
$\hat{\alpha}$	-1.760	0.353	-0.857	0.402
$\hat{\beta}$	0.108	-2.766	0.706	0.682
$\hat{\delta}_C$	0.775	0.815	0.171	0.029
p-val δ_C	0.000	0.314	0.021	0.193
sd $\hat{\delta}_C$	0.064	1.686	0.084	0.033
p-val β	0.916	0.087	0.988	0.988
sd $\hat{\beta}$	0.078	2.036	0.311	0.302

Table 2 Descriptive statistics.

The years 1993, 2003 and 2011 (Figure 2, 3 and 4, histogram and estimated probability density function) were chosen for demonstrative purposes of statistical testing of bimodality proposed in section 2.3. The results of the testing are summarized in Table 2, where *p-val* denotes p-value of corresponding test (i.e. test of Cardan's discriminant or parameter β alone). From the results we see, that we reject the null hypothesis (bimodality) for the year 1993 due to test of parameter β (at the significance level 0.1) and for the year 2003 due to the test of δ_C (at the significance level 0.05) as well as for the whole sample period. The rejecting of bimodality in the year 1993 is caused by negative estimate of beta parameter. On the other hand, in the year 2003 is Cardan's discriminant made statistically higher than zero by the estimate $\hat{\alpha}$ more distant from zero compared to the β. By the same reason we reject bimodality of the whole sample period. Year 2011 demonstrates the lack of statistical evidence against bimodality, hence we can not reject bimodality by any test.

4 Conclusion

In this paper is proposed the statistical test for bimodality of the stochastic cusp model using Cardan's discriminant and a simpler test builded on parameter beta and a necessary condition for bimodality, which can be used for a rejection of the bimodality as well. In more formal words, tests allow to reject null hypothesis of bimodality of the probability density function of the stochastic cusp model. The tests were performed on the USD, GBP exchange rate in chosen years to demonstrate the abilities of the tests to reject bimodality and different reasons for rejection of bimodality were pointed out. In general, these test could be used for testing of bimodality, whenever the estimated parameters has multivariate normal distribution as is the case of maximum likelihood estimation.

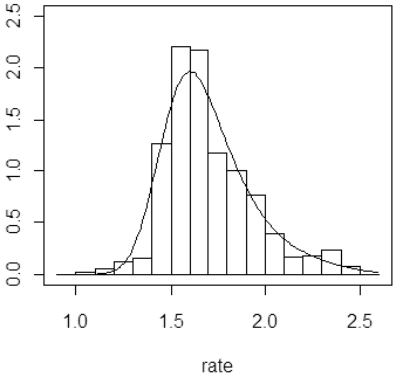


Figure 1 Data 1975-2014

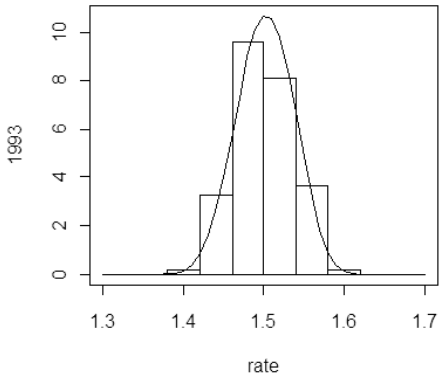


Figure 2 Data 1993

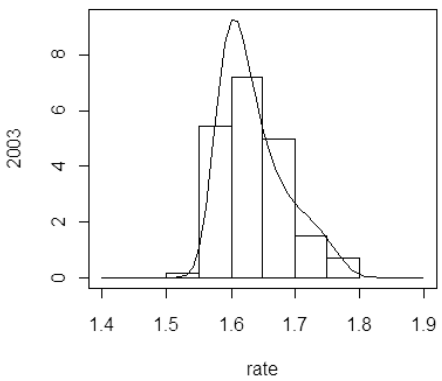


Figure 3 Data 2003

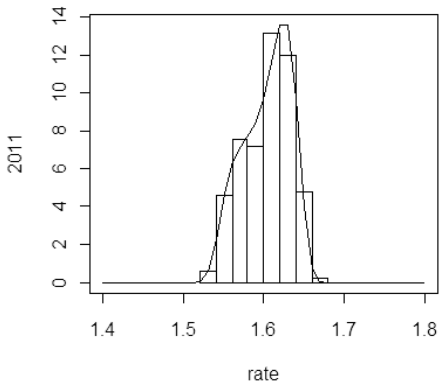


Figure 4 Data 2011

Acknowledgements

Support from the Czech Science Foundation by the grant No. 402/09/0965 is gratefully acknowledged.

References

[1] Arnold, V. I.: *Catastrophe Theory* Springer-Verlag, Berlin, 2004.

[2] Barunik J., Kukacka J.: Realizing stock market crashes: stochastic cusp catastrophe model of returns under time-varying volatility. *Quantitative Finance* (2014), 1–15. http://dx.doi.org/10.1080/14697688.2014.950319

[3] Cobb, L.: Parameter estimation for the cusp catastrophe model. *Behavioral Science*, **vol. 26**, Issue 1 (1981), 75-78.

[4] Cobb, L., Koppstein, P., and Chen, N. H. : Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family. *Journal of the American Statistical Association* **vol. 78**, No. 381 (1983), 124–130.

[5] Creedy, J., and Martin, V. : Multiple Equilibria and Hysteresis in Simple Exchange Models. *Economic Modelling* **vol. 10**, Issue 4 (1993), 339–347.

[6] Fernandes, M.: Financial crashes as endogenous jumps: estimation, testing and forecasting. *Journal of Economic Dynamics and Control* **vol. 30**, Issue 1 (2006), 111–141.

[7] Grasman, R. P. P. P., van der Maas, H. L. J., and Wagenmakers, E-J.: Fitting the Cusp Catastrophe in R: A cusp Package Primer. *Journal of Statistical Software* **vol. 32**, Issue 8 (2009), 1–28.

[8] Lehman, E. L.: *Testing Statistical Hypotheses*. Wiley, New York, 1959.

[9] Thom, R.: *Structural Stability and Morpohogenesis*. Benjamin, New York, 1975.

[10] Zeeman, E. C.: On the unstable behaviour of stock exchanges. *Journal of Mathematical Economics* **vol. 1**, Issue 1 (1974), 39–49.

Wavelet Coefficients Energy Redistribution and Heisenberg Principle of Uncertainty

Miloslav Vošvrda[1], Jaroslav Schürrer[2]

Abstract. This paper introduces the relation of Heisenberg Principle of Uncertainty to maximum Wavelet level of decomposition where the Wavelet coefficients typically provide sparse representation. In Quantum physics, Heisenberg Principle of Uncertainty states that we cannot know the exact position and momentum of a particle simultaneously. In time frequency domain, Heisenberg Principle states that we cannot exactly know information about time and frequency simultaneously. Time delta multiplied by frequency delta is greater than some arbitrary constant. This restriction has different consequences in Fourier transformation and Wavelet transformation, which is more suitable for non-stationary times series analysis. Wavelets fits into this principle because a basic Wavelet is characterized by short time and high frequency. However, when the Wavelet is stretched then it has longer time and lower frequency. This principle is inherent in the nature of things and has nothing to do with numerical precision of the Wavelet analysis.

First part of the paper summarizes Heisenberg Principle of Uncertainty, Wavelet transformation and signal energy. Second part presents Wavelet analysis of Apple Inc. stock daily closing price, showing energy redistribution depending on the Wavelet decomposition level based on the Wavelet choosen for the decomposition and the level of decomposition.

Keywords: Heisenberg Principle of Uncertainty, Wavelet Transformation, signal energy, signal entropy.

JEL classification: C44
AMS classification: 90C15

1 Heisenberg Principle of Uncertainty definition in time frequency domain

Theoretical part of this paper summarizes various definitions of Heisenberg Principle of Uncertainty from the reference literature and defines basic signal properties and also Discrete Wavelet Transformation which is used later in numerical part. We start with definition by Gabor atoms and their relation to Heisenberg Uncertainty and continue with Windowed Fourier transformation and Continuous Wavelet transformation.

1.1 Gabor atoms

Then Heisenberg Uncertainty as a analogy was first used in frequency domain by Gabor during time frequency dictionary definition constructed from waveforms of unit energy $||\phi_\gamma|| = 1$. The following paragraph is based on pioneer work of Mallat in Wavelet theory [1]. Lets us start with definition of time localization u of ϕ_γ and spread of ϕ_γ around u.

$$u = \int t.|\phi_\gamma(t)|^2 dt \text{ and } \sigma^2_{t,\gamma} = \int |t-u|^2.|\phi_\gamma(t)|^2 dt \tag{1}$$

[1]Institute of Information Theory and Automation CAS, Pod Vodrenskou v 4, Prague , Czech Republic, vosvrda@utia.cas.cz
[2]Masaryk Institute of Advanced Studies, Czech Technical University, Kolejn 2637/2a, Prague, Czech Republic , jaroslav.schurrer@muvs.cvut.cz

For frequency localization and spread we have similar equations

$$\xi = (2\pi)^{-1} \int \omega . |\hat{\phi}_\gamma(\omega)|^2 d\omega \text{ and } \sigma^2_{\omega,\gamma} = (2\pi)^{-1} \int |\omega - \xi|^2 . |\hat{\phi}_\gamma(\omega)|^2 d\omega \tag{2}$$

The Fourier Parseval formula shows dependency of $\langle f, \phi_\gamma \rangle$ on the values of $f(t)$ and $\hat{f}(\omega)$. Rectangle of size $\sigma_{t,\gamma} \times \sigma_{\omega,\gamma}$ is centered at (u, ξ) for (t, ω).

$$\langle f, \phi_\gamma \rangle = \int_{-\infty}^{+\infty} f(t).\phi^\star_\gamma(t)dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega).\hat{\phi^\star}_\gamma(\omega)d\omega \tag{3}$$

Heisenberg box representation of Gabor atom ϕ_γ as is depicted in following figure 1. This box may be seen as "quantum of information over an elementary resolution cell."[1]. This rectangle has minimum surface defined

$$\sigma_{t,\gamma}.\sigma_{\omega,\gamma} \geq \frac{1}{2} \tag{4}$$

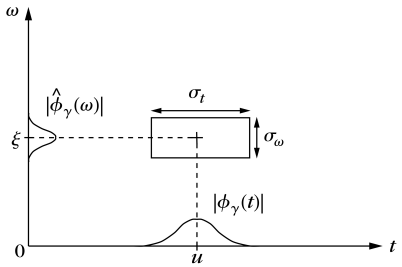


Figure 1 Heisenberg box of Gabor atom [1]

1.2 Windowed Fourier Transformation

Windowed Fourier Transformation uses time window $g(t)$ which is translated in time and frequency and has unit norm $||g|| = 1$. Each dictionary atom $g_{u,\xi}$ is projected by Windowed Fourier Transformation see Figure 2 (left side).

$$Sf(u, \xi) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t).g(t - u).e^{-i\xi t} \tag{5}$$

1.3 Continuous Wavelet Transformation

Wavelet dictionary is based on the mother Wavelet ψ which is scaled by parameter s and translated by parameter u

$$\psi_{u,s} = \frac{1}{\sqrt{s}}.\psi\Big(\frac{t - u}{s}\Big)dt \tag{6}$$

and satisfying condition of zero mean

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{7}$$

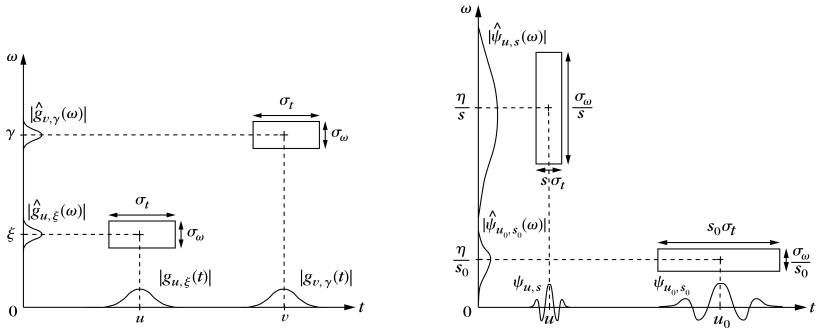


Figure 2 Energy spread of two windowed Fourier Transformation (left) and two Wavelets (right) [1]

The term $\frac{1}{\sqrt{s}}$ ensures energy conservation during scale shift. Continuous Wavelet Transformation projects function f at any scale and positon of mother Wavelet

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t).\frac{1}{\sqrt{s}}.\psi^*\Big(\frac{t-u}{s}\Big)dt \tag{8}$$

Wavelet atom has Heisenberg box with fixed area where parameter s varies as is depicted in Figure 2 (right side). Wavelets have time and frequency localization in comparation with Fourier atoms. When the s parameter varies , the time and frequency changes, but the area of box is still constant. For width window we have good resolution in frequency an low in time and vice versa for narrow window good resolution in time and low in frequency.

2 Signals and their properties

Let $f(x)$ be a function defined on the interval $(-\pi, \pi)$ such that $f^2(x)$ has a finite integral on that interval. If a_n, b_n are the Fourier coefficients of the function $f(x)$ than Bessel's Inequality states

$$\pi\Big(2a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\Big) \leq \int_{-\pi}^{\pi} f^2(x)dx \tag{9}$$

The energy E of 2π periodic function $f(x)$ is defined

$$E = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)dx \tag{10}$$

and Bessel's inequality can be than written in following form:

$$\Big(2a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\Big) \leq E \tag{11}$$

where term $(a_n^2 + b_n^2)$ represents energy. We define discrete signal as a sequence of numbers $\{x(n)\}, n \in \mathbb{Z}$ satisfying following equation

$$\sum_{n\in\mathbb{Z}} |x(n)| < \infty \tag{12}$$

which states that signal has to be bounded. This condition is necessary prerequisite for Discrete Wavelet Transform. For discrete signal $x(n)$ we also define his energy $E = x_1^2 + x_2^2 + ... + x_n^2, n \in \mathbb{Z}$. If equation (12) is satisfied we speak about signal with finite energy. This definition is based on Bessel's Inequality and energy definition for periodic function.

2.1 Signal comparison

Root Means Square Error (RMS Error) between two signals $x(n)$ and $y(n)$ is defined as

$$RMSError = \sqrt{\frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}{n}} \tag{13}$$

3 Discrete Wavelet Transformation - DWT

The Wavelet function has two important parameters: scaling s and translation u and must satisfy admissibility and regularity condition. Admissibility means that Wavelet has zero average in time domain (must be oscillatory). Regularity requires smoothness and concentration in both time and frequency domains. Discrete Wavelet Transformation uses scaling parameter as a power of two ([5])

$$s = 2^{-j} \tag{14}$$

and the time shift becomes

$$u = k.2^{-j} = k.s \tag{15}$$

Substituting equations (14) and (15) into base function in (6) we get

$$\psi_{u,s} = \psi\Big(\frac{t-u}{s}\Big) = \psi\Big(\frac{t-ku}{s}\Big) = \psi(s^{-1}t - k) = \psi(2^j.t - k) \tag{16}$$

The mother Wavelet at scale j and translation k is then defined as

$$\psi_{j,k} = 2^{j/2}\psi(2^j.t - k) \tag{17}$$

and similarly the scaling function at scale j and translation k is defined

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j.t - k) \tag{18}$$

The scaling and mother functions can be constructed as a linear combination of translations with the doubled frequency of a the base scaling function $\phi(2t)$ and the base mother function $\psi(2t)$ with equations

$$\phi(t) = \sum_{k=-\infty}^{+\infty} \sqrt{2}h_o(k)\phi(2t-k) \qquad \psi(t) = \sum_{k=-\infty}^{+\infty} \sqrt{2}h_1(k)\phi(2t-k) \tag{19}$$

where $h_0(k)$ and $h_1(k)$ are Wavelet filter coefficients. For scaling Haar Wavelet we have following coefficients $h_0(0) = h_0(1) = 1/\sqrt{2}$ and for Haar mother Wavelet we have $h_1(0) = 1/\sqrt{2}$ and $h_1(1) = -1/\sqrt{2}$. Daubechies 4 filter coefficients are $h_0(0) = 0.4830, h_0(1) = 0.8365, h_0(2) = 0.2241, h_0(3) = -0.1294$ and Daubechies 4 Wavelet coefficients are $h_1(0) = -0.1294, h_1(1) = -0.2241, h_1(2) = 0.8365, h_1(3) = -0.4830$.

We also notice that signal has to be bounded to perform DWT and that DWT ensures energy conservation on the first level of decomposition. The first level of DWT is performed by computing the first trend and the first fluctuation coefficients from the original signal. The second level DWT is performed by computing the second trend and the second fluctuation for the first trend only and so on for the other decomposition levels.

4 Numerical results

Numerical part demonstrates computation corollaries of Heisenberg's Uncertainty Principle. As a data source we use daily price for Apple Inc. (AAPL) for the last 15 years. This sample has 3813 data points where each value represents value in USD and the following graph shows stock price progress during the evaluated period. It clearly shows that this signal has finite energy thus we can use DWT where coefficients length depends on the input signal length and the selected Wavelet filter.

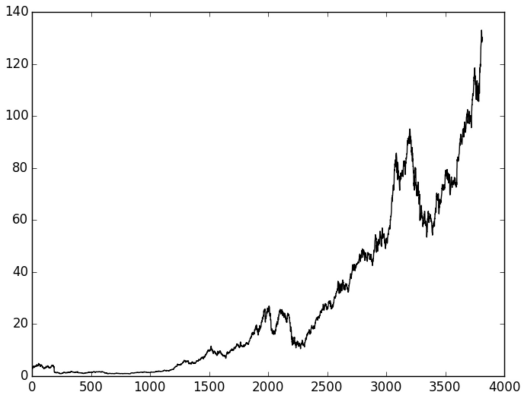


Figure 3 AAPL price in USD from 3.1.2000 to 27.2.2015.

For the Wavelet analysis we use a single prototype function (wavelet), which can be thought of as a band pass filter. In practise the filter coefficients for Haar are well known from theory and Daubechies 4 orthogonal wavelet with two vanishing moments with coefficients "in [1], which include values for different Wavelet filters". We use our own code written in Python, because implemented Wavelet Transformations in various software packages use built in mechanism for energy preservation to get around consequences of Heisenberg's Uncertainty Principle. Verification can be done very easily. We select testing signal with length sufficient at minimum to level 4 of decomposition and compute trend coefficients with selected Wavelet package and with own program using filter described in section 3. Than we compare energy of coefficients on particular levels of decomposition. If energy is not approximately same than Wavelet package uses energy preservation algorithm.

Figure 4. depicts development of original AAPL price (top), Haar 1 and Haar 5 (bottom) trend coefficients where we can see how trend looks like with increasing level of Wavelet level decomposition thus coefficients reduction from 3813 to 119 (number of coefficients is on vertical axis).

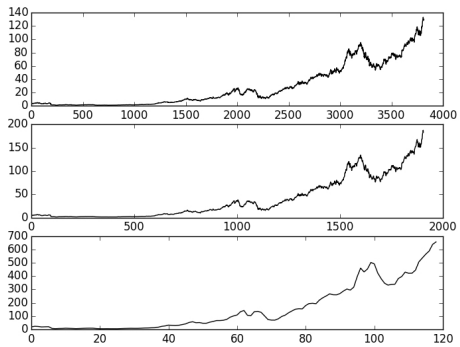


Figure 4 AAPL price in USD, Haar 1 and Haar 5 trend coefficients

The following table summarises energy preservation depending on the level of decomposition. We note that maximum useful level of decomposition for the given input signal length and wavelet filter length is 9. Theoretically there is a maximum level of decomposition equal to 11 with Haar Wavelet, but in this case energy of Wavelet trend coefficient drops to 20% of the original signal.

Level of decomposition	Haar	Daubechies 4
1	99.74	98.90
2	99.73	98.81
3	98.66	98.59
4	98.61	92.20
5	98.52	82.48
6	91.68	70.85
7	79.80	69.65
8	63.08	46.53
9	62.27	43.08

Table 1 Energy distribution on DWT decomposition level (trend coefficients)

Different row values between Haar and Daubechies 4 Wavelets in table 1 have relation to various Wavelets families used during DWT where each Wavelet family has different length of compact support. With increasing level of decomposition we lose signal energy associated with trend coefficients at certain levels of decomposition.

5 Conclusion

In this paper we presented an overview of Heisenbergs Uncertainty Principle and its relation to decomposition level of Discrete Wavelet Transformation. The Heisenbergs Uncertainty Principle can also be perceived as a limitation of fixed amount energy localization in small time interval. We cannot compact energy into decrescent time intervals. Numerical part analyzed real stock data for Apple Inc. showing how energy percentage drops down for trend coefficients with increasing level of decomposition and different Wavelets used for the analysis. We also note the fact that available Wavelet packages include built in mechanism for energy preservation and we mention a simple algorithm how to check this fact.

Acknowledgements

We would like to express our gratitude for supporting from the Czech Science Foundations through grant P402/12/G097. Thanks for supporting go to Masaryk Institute of Czech Technical University in Prague as well.

References

[1] Getreuer, P.: *Filter Coefficients to Popular Wavelets.*
http://www.mathworks.com/matlabcentral/fileexchange/5502-filter-coefficients-to-popular-wavelets, Updated 01 Dec 2009.

[2] Fugal, D., and Jang, J.: *Conceptual Wavelets in Digital Signal Processing.* Space & Signals Technical Publishing, P.O. Box 1771 San Diego CA, 2009.

[3] Mallat, S.: *A Wavelet Tour of Signal Processing.*

[4] Tan, L, L.: *Conceptual Wavelets in Digital Signal Processing.* Space & Signals Technical Publishing, P.O. Box 1771 San Diego CA, 2009.

[5] Tang, Y., Y.: *Digital Signal Processing.* Academic Press, 225 Wyman Street, Waltham, MA, 2013.

[6] Vidakovic, B.: *Statistical Modeling by Wavelets.* Wiley-Interscience, 111 River Street Hoboken, 1999.

Analysis of the index of economic freedom using structural models

Michal Vrabec[1], Luboš Marek [2]

Abstract. We use exploratory factor analysis to explore the potential latent variable structure of the Index of Economic Freedom. We want to test the hypotheses and parameter estimation of factors that are discovered. We also want to compare this model with the results of path models with observed variables and structural models with latent variables. Another goal of our work is to compare the classification of individual countries through cluster analysis.

Keywords: Index of Economic Freedom, factor analysis, structural models, cluster analysis.

JEL Classification: C, (G – H).
AMS Classification: 62H25

1 Introduction: structure of EFW

An index of economic freedom defines and attempts to measure the economic freedom of a country or a region. There are several indices of economic freedom; they are published annually by different organizations. The best-known among them are Economic Freedom of the World (EFW) by the Heritage Foundation in cooperation with the Wall Street Journal (EFW) [1], and the Index of Economic Freedom by the Canadian Fraser Institute (IEF) [2].

The index values result from step-by-step aggregation of a large number of partial indices. For example, EFW consists of 40 partial indices, which are aggregated into the resulting index at four levels. In the next to top level, five aggregated indices are considered [1]. The aggregation of the Economic Freedom of the World index from the 40 partial indices is shown in Figure 1 below.

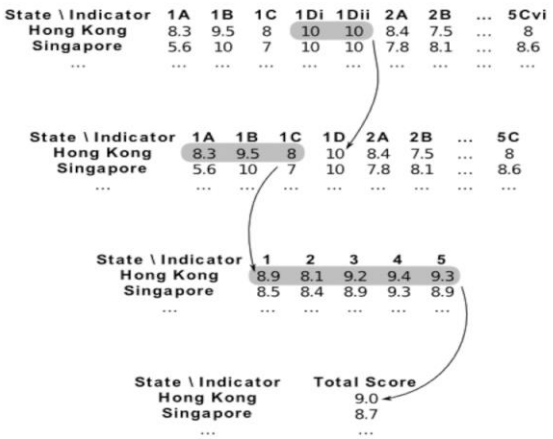


Figure 1 Aggregation of the EFW index (Source: The Heritage Foundation)

The following list describes the hierarchical structure of the partial indices involved in the design of the final index.

[1] University of Economics, Department of Statistics and Probability, W. Churchill Sq. 4, Prague 3, 130 67 Czech Republic, vrabec@vse.cz.
[2] University of Economics, Department of Statistics and Probability, W. Churchill Sq. 4, Prague 3, 130 67 Czech Republic, marek@vse.cz.

1) **Size of Government**
   a) Government consumption
   b) Transfers and subsidies
   c) Government enterprises and investment
   d) Top marginal tax rate
      i) Top marginal income tax rate
      ii) Top marginal income and payroll tax rate
2) **Legal System and Property Rights**
   a) Judicial independence
   b) Impartial courts
   c) Protection of property rights
   d) Military interference in rule of law and politics
   e) Integrity of the legal system
   f) Legal enforcement of contracts
   g) Regulatory restrictions on the sale of real property
   h) Reliability of police
   i) Business costs of crime
3) **Sound Money**
   a) Money growth
   b) Standard deviation of inflation
   c) Inflation: most recent year
   d) Freedom to own foreign currency bank accounts
4) **Freedom to Trade Internationally**
   a) Tariffs
      i) Revenue from trade taxes (% of trade sector)
      ii) Mean tariff rate
      iii) Standard deviation of tariff rates
   b) Regulatory trade barriers
      i) Non-tariff trade barriers
      ii) Compliance costs of importing and exporting
   c) Black-market exchange rates
   d) Controls of the movement of capital and people
      i) Foreign ownership/investment restrictions
      ii) Capital controls
5) **Regulation**
   a) Credit market regulations
      i) Ownership of banks
      ii) Private sector credit
      iii) Interest rate controls/negative real interest rates
   b) Labor market regulations
      i) Hiring regulations and minimum wage
      ii) Hiring and firing regulations
      iii) Centralized collective bargaining
      iv) Hours regulations
      v) Mandated cost of worker dismissal
      vi) Conscription
   c) Business regulations
      i) Administrative requirements
      ii) Bureaucracy costs
      iii) Starting a business
      iv) Extra payments/bribes/favoritism
      v) Licensing restrictions

2 Correlation analysis

Factor analysis is based on the covariance relationship between variables [3]. For this reason, we will now provide correlation analysis. Correlation matrices have been calculated for the original 40 (33) variables, then for 24 (18), and finally for the five aggregated variables. Some variables should be excluded from the analysis. Namely, variables 1C (Government enterprises and investment) and 3D are ordinal by their nature. Component 1D (Top marginal tax rate) is an average of two ordinal sub-components, 1Di and 1Dii. Indicator 4C (black-market exchange rates) was excluded for the same reason as for the 42-variable model (almost all countries have the same values). Because variables 5A (Credit market regulations) and 5B (Labor market regulations) are partly made from the ordinal variables (5Ai, 5Aiii and 5Biv, 5Bvi accordingly) and contain a summary error, they were also excluded from the analysis. As a result, 18 components remain for the following analysis.

From the correlation matrix for the 33 variables (part of which is presented in Figure 2 below) it is clear that variables 2A (Judicial independence), 2B (Impartial courts), 2C (Protection of property rights), and 2H (Reliability of police) form one group together with variables 5Cii (Bureaucracy costs) and 5Civ (Extra payments/bribes/favoritism). There is a strong and significant positive correlation between these variables. It is conceivable that these variables should form one factor in factor analysis. The correlation is especially high (0.983) between variables 5Cii and 5Civ. Which is, of course, natural: these variables are tied together – if bribes and favoritism exist, it means that additional bureaucracy costs presumably also exist, and vice versa. Besides, data for variables 5Cii and 5Civ have the same source based on subjective research (people estimating the existence of extra payments/bribes/favoritism in their industry and how substantial their impact is). What is more, there are seemingly significant correlations between 5Cii and 2A (0.906); between 5Cii and 2B (0.859); and between 5Cii and 2C (0.912), and 2H (0.891). In the same way, 5Civ has high correlation coefficients with 2A (0.889), 2B (0.827), 2C (0.904), and 2H (0.897).

	1A	1B	2A	2B	2C	2D	2E	2F	2G	2H	2I	3A	3B	3C	4Ai	4Aii	4Aiii	4Bi
1A	1	.616	-.605	-.466	-.573	-.613	-.710	-.408	-.172	-.602	-.478	-.163	-.082	-.428	-.178	-.294	.203	-.400
1B	.616	1	-.316	-.100	-.347	-.618	-.629	-.351	-.137	-.440	-.527	-.385	-.299	-.433	-.383	-.339	.072	-.323
2A	-.605	-.316	1	.870	.889	.529	.661	.360	.169	.830	.515	.021	.095	.403	.293	.306	-.196	.628
2B	-.466	-.100	.870	1	.882	.379	.555	.394	.122	.783	.494	-.039	.029	.335	.172	.224	-.141	.598
2C	-.573	-.347	.889	.882	1	.560	.697	.415	.151	.879	.589	.083	.154	.565	.374	.375	-.178	.700
2D	-.613	-.618	.529	.379	.560	1	.637	.409	.228	.596	.479	.254	.046	.465	.300	.559	.005	.546
2E	-.710	-.629	.661	.555	.697	.637	1	.604	.205	.775	.772	.188	.172	.487	.277	.386	-.133	.536
2F	-.408	-.351	.360	.394	.415	.409	.604	1	.403	.460	.528	.104	.074	.292	.194	.396	-.068	.322
2G	-.172	-.137	.169	.122	.151	.228	.205	.403	1	.195	.120	.060	-.120	.102	.397	.417	-.023	.113
2H	-.602	-.440	.830	.783	.879	.596	.775	.460	.195	1	.733	.130	.121	.580	.376	.397	-.083	.676
2I	-.478	-.527	.515	.494	.589	.479	.772	.528	.120	.733	1	.199	.096	.430	.252	.303	-.035	.676
3A	-.163	-.385	.021	-.039	.083	.254	.188	.104	.060	.130	.199	1	.123	.306	.298	.261	.025	.187
3B	-.082	-.299	.095	.029	.154	.046	.172	.074	-.120	.121	.096	.123	1	.176	.139	-.024	-.029	.102
3C	-.428	-.433	.403	.335	.565	.465	.487	.292	.102	.580	.430	.306	.176	1	.435	.407	-.134	.462
4Ai	-.178	-.383	.293	.172	.374	.300	.277	.194	.397	.376	.252	.298	.139	.435	1	.507	-.011	.404
4Aii	-.294	-.339	.306	.224	.375	.559	.386	.396	.417	.397	.303	.261	-.024	.407	.507	1	.390	.491

Figure 2 Part of Pearson's correlation coefficient matrix of the 33 variables

3 Factor analysis

Similar to the correlation analysis, the factor model [4] was tested on the original 40 (33) variables, then on 24 (18), and finally on the five aggregated variables. We expect classic factor model

$$X_j = \sum_{r=1}^{q} \lambda_{jr} f_r + \varepsilon_j, \; j = 1, \ldots p. \tag{1}$$

The obtained solution with four factors is shown in Figure 3. This model explains 72.594 % of the total variance. All other criteria (KMO, eigenvalues, extracted communalities and scree-plot) also confirm that this solution is appropriate. The two-factor model explains 57.165 % of the original variability (Figure 4). Almost all variables (except for 1A and 1B, which lie in the third quadrant) are situated in the first quadrant. Component 2B (impartial courts) is located very close to the first quadrant, but already lies in the fourth one; as well as 3A (money growth) which lies in the third one, but very close to the axis and to the north-east quadrant. The first factor explains

the legal system and property rights together with the trade and business barriers; while the second factor explains monetary policy in the country and military interference with the rule of law and politics.

	Component			
	1	2	3	4
2B - Impartial courts	.931	.159	-.075	.063
2C - Protection of property rights	.904	.318	.149	-.065
2A - Judicial independence	.875	.316	.006	.002
5C - Business regulations	.868	.275	.237	.158
2H - Reliability of police	.798	.469	.160	-.033
4B - Regulatory trade barriers	.714	.224	.543	-.106
1B - Transfers and subsidies	-.009	-.824	-.285	.234
2E - Integrity of the legal system	.504	.773	.059	-.013
1A - Government consumption	-.386	-.722	.018	.008
2I - Business costs of crime	.431	.675	.039	-.037
2D - Military interference in rule of law and politics	.329	.643	.319	.040
2F - Legal enforcement of contracts	.268	.622	.062	.369
4A - Tariffs	.090	.038	.820	.107
4D - Controls of the movement of capital and people	.474	.029	.554	.213
3A - Money growth	-.149	.337	.525	-.301
3C - Inflation: most recent year	.407	.372	.423	-.257
2G - Regulatory restrictions on the sale of real property	.051	.240	.365	.732
3B - Standard deviation of inflation	.045	.160	.140	-.621

Figure 3 Rotated Component Matrix for the 18-variable EFW model for the year 2014 (Varimax rotation)

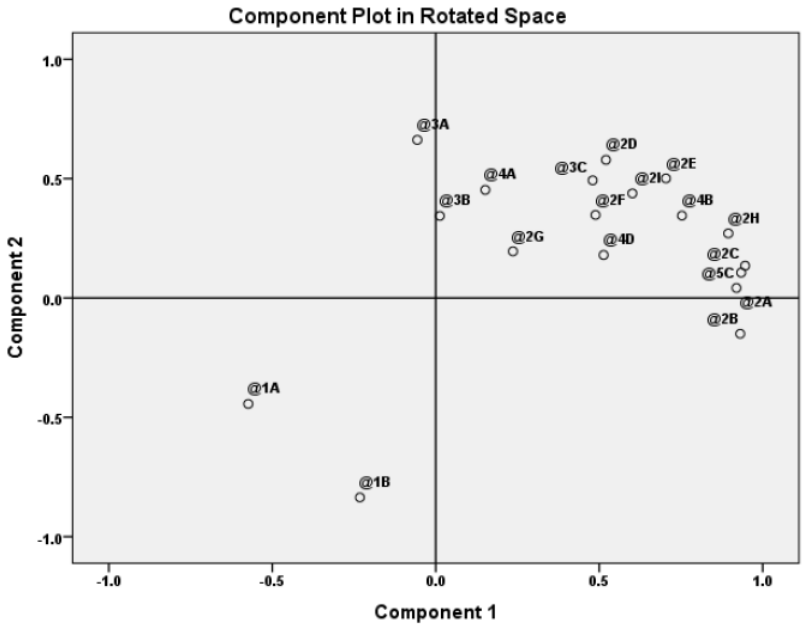


Figure 4 Rotated Component Matrix for the 18-variable EFW model for the year 2014 (Varimax rotation)

Very similar results (found by varimax rotation method) were found by using the equamax rotation: all the variables gathered in the same factors. The only difference is that indicator 26 5Cvi is more correlated with the second factor instead of the first. Otherwise, only loadings change their value. The quartimax rotation, however, gives different results, which are difficult to interpret.

4 Hierarchical cluster analysis applied to the countries

In this part of the paper, Hierarchical Cluster analysis is applied to a sample of countries to explore whether they can be organized into well-interpretable groups [5]. Firstly, the data set with 95 countries is analyzed with the original 42-index model. Secondly, the analysis is applied on the adjusted model with 88 countries and 33 indices. The same clustering algorithm is then applied to these 88 countries with the aid of the factor variables' values – once for four and once for two factor variables. Finally, all the achieved results are compared. The clustering sequence for the Ward method is visualized with the aid of a dendogram, shown in Figure 5.

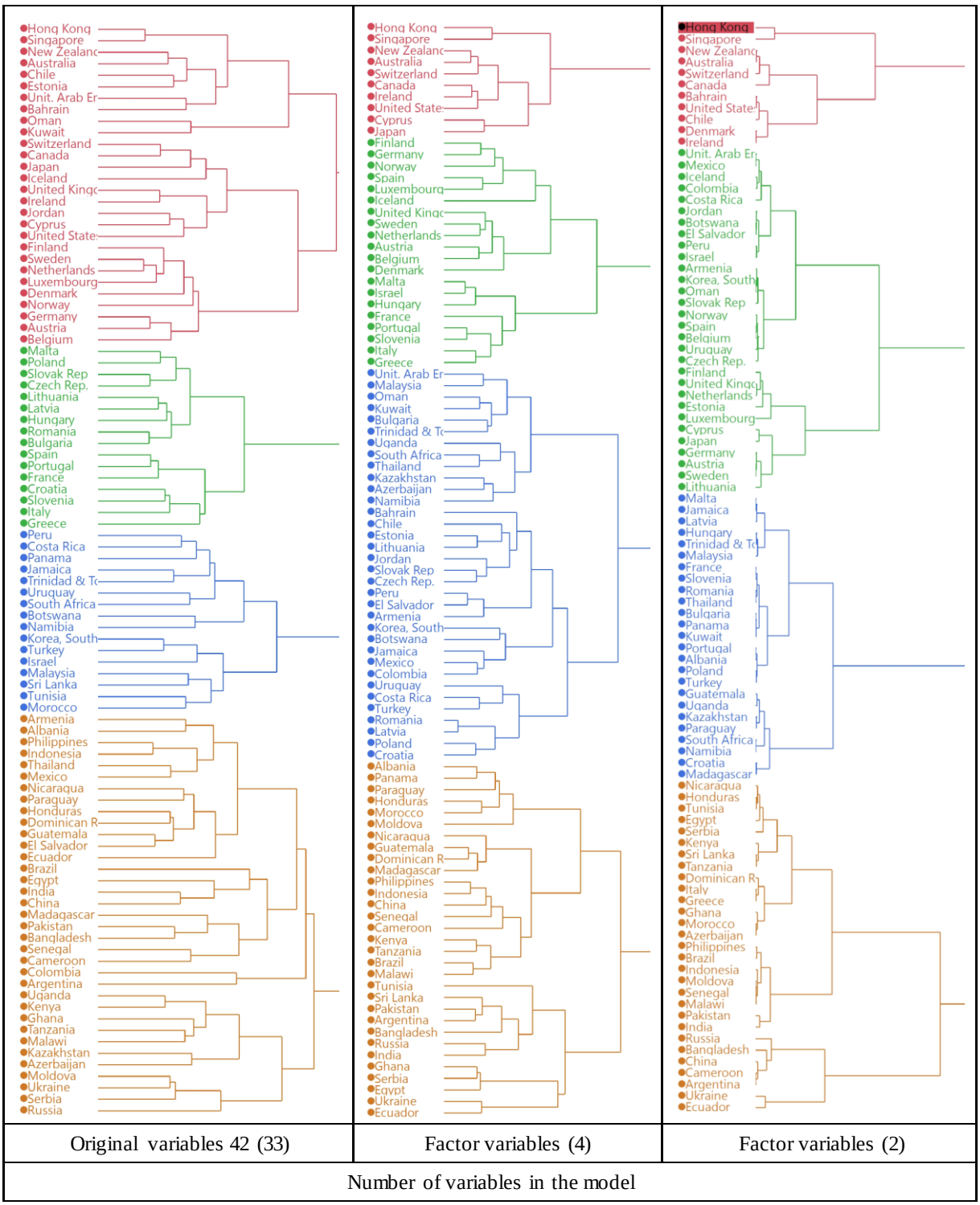


Figure 5 The clustering sequence for the Ward method

Let us point out an interesting aspect: when the number of variables is reduced, the number of countries in the "red" cluster (representing the highest value of the index of economic freedom) goes down. And complementary to that, the size of the "brown" cluster (countries with a low value of the index of economic freedom) grows in those circumstances. When the original variables are used, for example, the U.S.A. is in the "green" cluster, while it moves to the "red" cluster if the factor variables are used.

References

[1] The Wall Street Journal and The Heritage Foundation. (2014). *Index of Economic Freedom*. [Online], http://www.heritage.org/index/. 2014.
[2] Fraser Institute: *Economic Freedom of the World project*. [Online]. http://www.freetheworld.com/. 2014.
[3] Hair, J. F., Anderson, R.E., Tatham, R. L. and Black, W.C. *Multivariate Data Analysis*, Fifth Edition. New Jersey: Prentice Hall, 1998.
[4] Comrey, A. L. and Lee, H. B. *A First Course in Factor Analysis*, Second Edition. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc., 1992.
[5] Hatcher, L. A.: *Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute Inc., 1994.

Multiple-objective Optimization for Solving the Resource Allocation Problem

Nenad Vunjak[1], Jelena Birovljev[2], Otilija Sedlak [3], Zoran Ćirić[4]

Abstract. Just as in many other areas, project management makes use of mathematical modeling in cases such as setting strategic goals, formulation of strategies, selection of human resources and realization of the chosen strategy and control. Criteria and restrictions of alternatives are also encompassed in the space of uncertainty and indeterminacy.

Resource allocation problem (RAP) is the process of allocating resources among the various projects or business units for maximization of profit or minimization of cost. This paper deals with human resource allocation. We start from a set of individuals who were tested for certain qualities by the experts. Marks given by the experts need not to be exact numbers, but qualitative values, such as: absolutely true, highly true, moderately true etc. Methodology which allows us to model indeterminacy is fuzzy sets theory which is particularly well designed for dealing with non-probabilistic uncertainties. We expand the grading system and, indirectly, coefficients of model constraints which become fuzzy.

Keywords: resource allocation problem, marks given method.

JEL Classification: C6
AMS Classification: 91B32

1 Introduction

RAP - Resource allocation problem is the process of allocating resources among the various projects or business units for maximization of profit or minimization of cost (Lin and Gen, 2008). The process of the RAP is focused on finding an optimal allocation of limited resources to a certain number of tasks while controlling for the given resource constraint. Resource may be any entity used to accomplish a goal, e.g. a person, an asset, material etc. This paper deals with human resource allocation. We start from a set of individuals who were tested for certain qualities by the experts. The novelty lies in the fact that marks given by the experts need not to be exact numbers, but qualitative values, such as: absolutely true, highly true, moderately true etc. Furthermore, if experts do have certain hesitations it will be allowed to choose more than one '*mark*', thus avoiding to be artificially precise.

Methodology which allows us to model indeterminacy is fuzzy sets theory which is particularly well designed for dealing with non-probabilistic uncertainties. We expand the grading system and, indirectly, coefficients of model constraints which become fuzzy. On the other hand, we allow each candidate to be placed at the position best suited for his qualities, which is not necessarily the position he/she would apply for in the first place.

2 Theoretical Background

The concept of hesitant fuzzy sets (HFS)is a new one, dating back to Torra [19]. HFS are a generalization of intuitionistic fuzzy sets of Atanassov and Gargov [1] who allowed each fuzzy membership function to have a degree of indeterminacy. Torra defines a HFS through a function that returns a set of membership values for each element in the domain. The applications of HFS followed various directions, mostly in generalization of the existing results dealing with (ordinary) fuzzy sets. Rodríguez et al. [14] focused on group decision making (GDM) and hesitant situations that could arise from multiple opinions. They have shown how to generate comparative linguistic expressions by using a context-free grammar. Rodríguez et al. [13] extended context-free grammar to that defined in Rodríguez et al. [14], and further expanded the topics in GDM in terms of expressing and modeling doubt between different linguistic terms, requiring richer expressions to express the experts' knowledge more accurately.

On the other line of reasoning, but also tackling decision making processes, Xia and Xu [20] presented an extensive study on hesitant fuzzy information aggregation techniques and their application in decision making with

[1] Universityof Novi Sad/Faculty of Economics Subotica, Subotica, Serbia, dekanat@ef.uns.ac.rs.
[2] Universityof Novi Sad/Faculty of Economics Subotica, Subotica, Serbia, birovlevj@ef.uns.ac.rs.
[3] Universityof Novi Sad/Faculty of Economics Subotica, Subotica, Serbia, otilijas@ef.uns.ac.rs.
[4] Universityof Novi Sad/Faculty of Economics Subotica, Subotica, Serbia, zoran.ciric@ef.uns.ac.rs.

anonymity. They have developed certain hesitant fuzzy operational rules based on the interconnection between the hesitant fuzzy set and the intuitionistic fuzzy set, as well as a series of aggregation operators for dealing with various situations.

3 Methodology and the Model

We shall present in detail the proposed methodology and provide the readers with the model for human resources allocation to different positions in marketing sector. When evaluating candidates, HR management is confronted with a significant number of tradeoffs over a diverse range of criteria.

Multiple-objective optimization problems have gained momentum in researchers' community with various backgrounds since early 1960. In the realm of multiobjective optimization problem, multiobjective functions are optimized simultaneously. Therefore, there does not necessarily exist a solution that is best with respect to all objectives due to the incommensurability and confliction among objectives. A solution may be the best in one objective but underperforming in another. The so called Pareto optimal solutions are a set of solutions for the multiple-objective case which cannot simply be compared with each other, i.e. no improvement is possible in any objective function without sacrificing at least one of the other objective functions.

Fuzzy mathematical programming (FMP) was developed in order to adequately solve optimization problems which poses non stochastic indeterminacies. Inuiguchi [7] classifies FMP in three categories based on the types of indeterminacies: 1. FMP with ambiguities, also known as flexible programming, developed by Bellman and Zadeh [3], allows fuzzy preferences for the decision maker. 2. FMP with vagueness deals with fuzzy coefficients in constraints and goals. Dubois and Prade [4]were the first researchers who investigated systems of linear equations with fuzzy coefficients and suggested possible applications of fuzzy mathematical programming. Dubois [5] observed fuzzy coefficients as possibilistic distributions of coefficient values, hence the name possibilistic programming for the second type of FMP. Following the early work of Dubois, we shall apply possibility theory in our linear model. Finally, third type, robust programming combines fuzzy coefficients and ambiguous preferences of the decision makers. Negoita et al. [11] were the first to formulate such a linear programming problem.

Managing and modeling of uncertainty by different forms of information, used by experts to provide their preferences, can be done in various ways, including utility vectors, fuzzy preference relations, linguistic variables, interval values, multiplicative preference relations, hesitant fuzzy sets.

Hesitant fuzzy sets are a generalization of 'ordinary' fuzzy sets and will be used in possiblistic programming. For every element x from the domain X, membership function of a HFS A is given by $h(x) = A_x$, where A_x is any (finite) subset of [0,1]. Suppose that the experts deciding on candidate's communication skills should choose a grade from the set of linguistic descriptions, L = {none, very low, low, medium, high, very high, absolute}.Of course, if unsure of the exact grade, an expert has a possibility to opt for two or more grades[5]. Grades can be naturally ordered, and modeled by fuzzy sets, as shown in Figure 1.

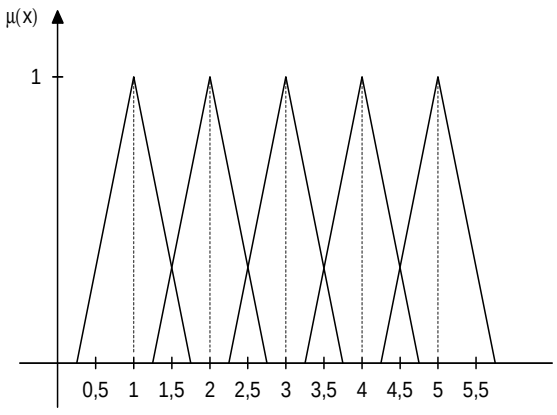


Figure 1

Choosing more than one grade would result in 'merging' the graphs of chosen grades into one convex fuzzy set. It is done by means of envelope which is a linguistic interval whose limits are minimal and maximal linguistic term, as shown in Figure 2.

[5] Choosing all options is equivalent to total absence of knowledge.

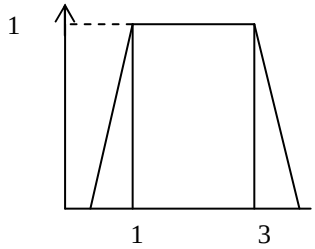


Figure 2

We observe only symmetrical marks, bounded with linear functions, though different methods of forming asymmetrical functions exist and can be found in Inuiguchi et al. [8]. We also refer an interested reader to Klir and Yuan [9] for any comprehensive and in-depth reading on fuzzy sets and fuzzy logic.

3.1 The Model

Let us observe a company 'X' consisting of S sectors (Manufacturing, Marketing, R&D, etc.) Each sector consists of up to P positions (e.g. Marketing sector embodies marketing vice president, marketing assistant, customer service representative, shipping clerk, etc.), i.e. P = maxP_i, where P_i is the number of positions in sector i, i=1,2,…S. Each position employs up to K employees, where K being maxK_{ij}, and K_{ij} is the number of employees on the position j of the sector i (e.g. there are two positions for marketing VP, 3 for marketing assistant, 6 for customer service representative, 2 for shipping clerk, etc.)

There is a total of N candidates, each being tested for C characteristics, regardless of the position he/she apply for. The program chooses based on the test results the position one is suited best for, and that doesn't necessarily mean is the position one applied for. Characteristics include experience, organizational skills, dependability, computer knowledge, contacts, formal education, etc. These characteristics are measured by both numerical values and experts' opinions in terms of hesitant fuzzy sets, and their systematization with graphical expressions is given below:

Experience is divided into three broad intensity groups: high (corresponds to 15+ years of experience), medium (6–15 years), and low (up to 5 years). For each broad group experts are asked to grade the applicant in terms how satisfactory (in terms of Figure 1 and 2) his actual achievements are, being in a certain intensity group. Similar reasoning goes for following characteristics: quality contacts and leadership, which are also divided into high, medium, and low, with further sub-divisions by the experts (Figure 3). Organizational and computer skills, as well as dependability are solely divided into broad groups, without additional experts' opinion. Finally, education is divided into PhD, MSc, BSc, and secondary, while the experts give opinion on the quality of the school attended, thus allowing for slight differences within the same educational level.

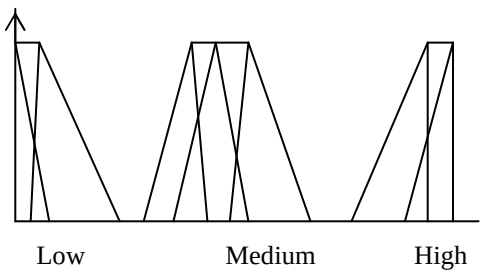


Figure 3

Let us denote byx_{ijk} a binary variable which has a value 1 if the candidate k, k=1…N is chosen for the position i, i=1…P in sector j, j=1…S, otherwise, x_{ijk} has a value 0. Natural constraints imposed on x's are following:

$$\sum_{i=1}^{P}\sum_{j=1}^{S} x_{ijk} = 1, \text{ for every k=1…N.} \quad (1)$$

One candidate can be employed at exactly one position.

$$\sum_{i=1}^{P}\sum_{j=1}^{S}\sum_{k=1}^{N} x_{ijk} \le N \tag{2}$$

Total number of candidates exceeds number of positions.
Each sector is choosing candidates in order to maximize it's manpower:

$$\sum_{i=1}^{P}\sum_{j=1}^{S}\sum_{k=1}^{N} c_{ijk}\, x_{ijk} \tag{3}$$

Coefficients c_{ijk} in the linear expression (3) are fuzzy numbers of triangular, or trapezoidal shape and are obtained through summation of grades for each position.

Analytical and graphical expression of a trapezoidal number A = (a, b, α, β) is given by following membership function and in Figure 4:

$$\mu_A(x) = \begin{cases} 1 - \dfrac{a-x}{\alpha}, & a-\alpha \le x \le \alpha \\ 1, & a \le x \le b, \\ 1 - \dfrac{x-b}{\beta}, & b \le x \le b+\beta \end{cases}$$

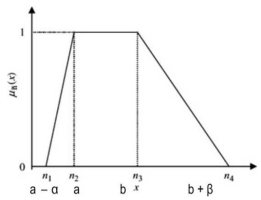


Figure 4

The linear combination of fuzzy trapezoidal numbers, which is present as a constraint expression as well as a goal function, is again a number of the same shape. If we denote by (a_{ijk}, b_{ijk}, α_{ijk}, β_{ijk}) coefficient c_{ijk}, the total manpower TM, defined at (3) is given by:

$$(\Sigma_{i,j,k}\, a_{ijk} x_{ijk}, \Sigma_{i,j,k}\, b_{ijk} x_{ijk}, \Sigma_{i,j,k}\, \alpha_{ijk} x_{ijk}, \Sigma_{i,j,k}\, \beta_{ijk} x_{ijk}) = (A(x), B(x), \alpha(x), \beta(x))$$

The possibilistic linear function is not uniquely determined, therefore its maximization, as well as constraints of the form (A(x), B(x),α(x),β(x))≥ Mmay not be completely meaningful. For overcoming this problem we introduce a so called 'fuzzy inequality', ρ(A, B), where A and B are two fuzzy sets. If the possiblistic distribution μA of a possibilistic variable α, the measurements of possibility and necessity of an event that α belongs to the set B are given by (Dubois, 1987):

$$Poss_A(B) = \sup_x \min(\mu_A(x), \mu_B(x)) \tag{4}$$

$$Nes_A(B) = \inf_x \max(1 - \mu_A(x), \mu_B(x)) \tag{5}$$

where μ_B is a membership function of a set B. $Poss_A(B)$ estimates the level up to which it is possible that possibilistic variable α belongs to the set B. on the other side, the expression $Nes_A(B)$ measures the level of necessity up to which is certain that α belongs to the set B. Let α be a possibilistic variable and B = (-∞, g], a crisp set. One could easily obtain the following: Pos(α ≤ g) = sup{ $\mu_A(x)$ | x≤ g} and Nes(α ≤ g) = 1 - sup{ $\mu_A(x)$ | x> g}, which is exhibited in Figure 5.Similarly, for B = [g, +∞) we obtain following equations presented in Figure 6:

Pos(α ≥ g) = sup{ $\mu_A(x)$ | x≥ g} and Nes(α ≥ g) = 1 - sup{ $\mu_A(x)$ | x < g}.

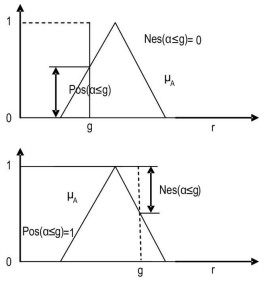


Figure 5

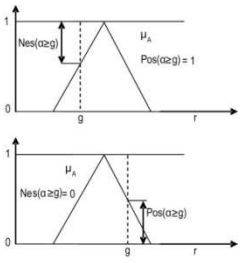


Figure 6

The constraints on the number of people in each position are given by:

$\sum_{i,j} x_{ijk} = p_{ij}$, where i is the number of sectors, i=1,…S and j is the number of positions in each sector, j=1,…P.

Maximization of the total human potential of the company is done by maximizing the right spread of the possibilistic distribution, allowing for the best possible marks to be incorporated in the optimal team:

$$Maximize \sum_{i,j,k} \left(\beta_{ijk} + b_{ijk}\right) x_{ijk} \tag{6}$$

One of the possible restrictions is a funding restriction. Each employee has suggested his expected and minimal wage, w_e and w_{min} respectively. We would like not to break funds F with high necessity, of e.g. h = 0.8 or h = 0.9. One easily concludes from Figure 5 that this constraint is equivalent to:

$$\frac{\sum x_i w_{e_i} - F}{\sum x_i \left(w_{e_i} - w_{\min i}\right)} \leq 1 - h \tag{7}$$

Other restrictions are possible, namely if we want to impose a minimal mark value in the certain job group, e.g. averagecomputer skills among marketing assistants should be at least M. Since we observe all the marks as possibilistic distribution functions, we can pose this requirement differently, namely, we could demand that the minimal average computer skill is necessarily greater than M, where we use high necessity measure of e.g. h = 0.8 or h = 0.9, as presented in Figure 6. The set of restrictions is given by:

$$\frac{M - \sum a_i x_i}{\sum a_i x_i} \leq 1 - h \tag{8}$$

4 Conclusion

We tried to present a possible application of hesitant fuzzy sets in the field of human resources allocation. Certain qualities of the applicants can not be measured precisely, and an expert's opinion is needed, which as well may not be precise, or the expert might have doubts or hesitate to give a final mark. Hesitant fuzzy sets are designed to perfectly modelthese situations. We suggested possible objective functions, as well as an array of possible constraints.

There were no restrictions on number of positions one applicant could apply. We tested the model on 30 applicants to Marketing sector of a company wanting to expand to 14 new positions. Results show that insignificant loss in total human potential, measured by sum of marks in all observed characteristics, can lead to substantial savings. Possible limitation might be the omission of interaction among cooperative individuals. When choosing a team of people it is relevant to measure a level of synergy within a team and test whether a team contributes more (or less) than the sum of independent performances of each team member. However, the accent was on presenting a new method in decision making and mark giving, but this interesting topic will be fully investigated in our future work.

References

[1] Atanassov, K.,Gargov, G.: Interval valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 1989, 31(3), 343–349.

[2] Ballesteros-Pérez, P., González-Cruz, C., Fernández-Diego, M.: Human resource allocation management in multiple projects using sociometric techniques, *International Journal of Project Management*, 2012, 30(8), 901-913.

[3] Bellman, R. E., Zadeh, L. A.: Decision-making in a fuzzy environment. *Management Sci.*, 1970(17), 141-164.

[4] Dubois, D., Prade, H.: Systems of linear fuzzy constraints, *Fuzzy Sets and Systems*, 1980(3), 37- 48.

[5] Dubois, D.: *Linear programming with fuzzy data, in: J.C. Bezdek (Ed.), Analysis of Fuzzy Information 3: Applications in Engineering and Science*, CRC Press, Boca Raton, FL, 1987, 241-263.

[6] Filho, C.F.F.C., Rocha, D.A.R., Costa, M.G.F., Pereira, W.C.A.: Using Constraint Satisfaction Problem approach to solve human resource allocation problems in cooperative health services, *Expert Systems with Applications*, 2012, 39(1), 385-394.

[7] Inuiguchi, M., Sakawa, M., Kume, Y.: The usefulness of possibilistic programming in production planning problems, *Int. J. Prod. Economics*, 1994(33), 45-52.

[8] Inuiguchi, M., Tanino, T., Sakawa, M.: Membership function elicitation in possibilistic programming problems, *Fuzzy Sets and Systems*, 2000, 111(1), 29-45.

[9] Klir, G., Yuan, B.: *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall PTR, 1995.

[10] Lin, C-M., Gen, M.: Multi-criteria human resource allocation for solving multistage combinatorial optimization problems using multiobjective hybrid genetic algorithm, *Expert Systems with Applications*, 2008(34), 2480–2490.

[11] Negoita, C.V., Minoiu, S., Stan, E.: On considering imprecision in dynamic linear programming.Economic Comput, *Economic Cybernet. Stud. Res*. 1976(3), 83-95.

[12] Osman, M.S., Abo-Sinna, M.A., Mousa, A.A.: An effectivegenetic algorithm approach to multiobjective resource allocationproblems, *Applied Mathematics and Computation*, 2005(163), 755–768.

[13] Rodríguez, R.M.,Martínez, L., Herrera, F.: A group decision making model dealing with comparative linguistic expressions based on hesitant fuzzy linguistic term sets, *Information Sciences*, 2013(241), 28–42.

[14] Rodríguez, R.M.,Martínez, L., Herrera, F.: Hesitant fuzzy linguistic term sets for decision making, *IEEE Transactions on Fuzzy Systems*, 2012, 20(1), 109–119.

[15] Saaty, T.L. (1996). Decision Making with Dependence and Feedback: the Analytic Hierarchy Process.RWS Publications, 4922 Ellsworth Avenue,Pittsburgh PA 15213.

[16] Saaty, T.L., Vargas, L.G., Dellmann, K.: The allocation of intangible resources: the analytical hierarchy process and linear programming, *Socio-Economic Planning Sciences*, 2007(37), 169-184.

[17] Saaty, T.L., Peniwatib, K., Shanga, J.S.: The analytic hierarchy process and human resource allocation: Halfthe story, *Mathematical and Computer Modelling*, 2007(46), 1041-1053.

[18] Silva, L.C., Costa, A.P.C.S.: Decision model for allocating human resources in information system projects, *International Journal of Project Management*, 2013, 31(1), 100-108.

[19] Torra, V.: Hesitant fuzzy sets, *International Journal of Intelligent Systems*, 2010(25), 529–539.

[20] Xia, M., Xu, Z.: Hesitant fuzzy information aggregation in decision making, *International Journal of Approximate Reasoning*, 2011(52), 395–407.

Analysis of the availability of Internet access for young people in schools of Silesia Province based on spatial statistics

Katarzyna Warzecha[1], Andrzej Wójcik[2]

Abstract. Computerization and in particular access to the Internet at the present time is very important for the development of young people and more interesting education methods. The aim of the article is to analyze the availability of Internet access for young people in Silesian schools, using spatial statistics. The study will be based on data from the Local Data Bank of Polish CSO (GUS). The taxonomic analysis of the development of Internet access in Silesia will allow to assess whether there is a discrepancy among districts concerning the studied phenomenon. In our research we will identify groups of districts with similar level of the availability of the Internet access for young people in schools. We are going to verify the hypothesis that the availability of the Internet access depends on the district where the school is located. For this purpose the spatial statistics will be used – Moran's measures of local and global autocorrelation. The calculations will be made in R Cran and Microsoft Excel.

Keywords: Internet, young people, spatial statistics.

JEL Classification: C49
AMS Classification: 62H30

1 Introduction

In the modern world the use of so-called modern means of communication, which include, among others, computers, tablets or mobile phones often with access to the Internet is massive and common behavior of young people. Computerization, and especially access to the Internet nowadays is very important because if offers the opportunity to send and receive an unlimited amount of information, images, sounds, as well as to communicate by means of all kinds of social networks. The Internet is a source of entertainment and knowledge for many young people and it contributes to their development and for teachers it can be a tool for interactive education of youth. However, the excessive use of the Internet (more and more often observed among teenagers who spend many hours online on social networks, in chat rooms or playing computer games) can be harmful for their health and can lead to addiction. Loss of control over the use of the Internet may contribute to the neglect of education, the lack of physical activity and recreation, the neglect of personal hygiene, the isolation and the worsening of family relationships [4]. Therefore, young people should be informed that beside many advantages the Internet can be a source of many threats (juvenile gambling, addiction to playing computer games, being online on social networks for many hours, uploading many photos on blogs which can be used improperly, cyberbullying). Parents, guardians and teachers should teach young people the proper use of the Internet as a modern mean of communication.

The main aim of this article is to analyze the availability of Internet access in schools for the Silesian young people. The district division will be taken into consideration and the spatial statistics will be used in the study. The data derived from the Local Data Bank of Polish CSO (GUS) from 2003, 2008 and 2012 will be used in the study as well. The values of Hellwig's taxonomic measure of development will enable to rank districts of Silesia Province according to the availability of Internet access in schools. Groups of districts with similar level of the availability of Internet access in schools will be identified in the research. The spatial statistics (the Moran's local and global autocorrelation measures) will be used to verify the hypothesis of the influence of geographical location of district on the availability of the Internet access. The calculations will be made in R Cran and Microsoft Excel.

[1] University of Economics in Katowice, Department of Econometrics, 1 Maja 50, 40-287 Katowice, Poland, andrzej.wojcik@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Econometrics, 1 Maja 50, 40-287 Katowice, Poland, warzecha@ue.katowice.pl.

2 Spatial statistics: local and global measures

Spatial statistics are convenient and effective way of testing the existence of spatial autocorrelation processes. The spatial autocorrelation indicates that nearby geographical observations are more similar to each other than distant observations [5]. If positive autocorrelation occurs in certain area, this means that there is a spatial cluster with high or low values of observed variables. It indicates that areas with high values of a given variable are clustered with other high value areas, and the areas with low values of a variable are clustered with other low value areas. In case of negative autocorrelation, the high value areas are neighboring to low value areas and vice versa, creating the alternating areas with dissimilar values of a variable (so called checkerboard). The lack of spatial autocorrelation indicates spatial randomness, which means that the high and low values of observed variable are distributed independently [7].

The measures of global autocorrelation (the Moran's *I* statistics) and local autocorrelation (local Moran's I_i statistics) were taken into consideration in this research. The global autocorrelation results from the existence of correlation within the whole spatial unit (global measure is the one-number indicator of autocorrelation or the general similarity of regions). In contrast, the local statistics calculated for every area answers the question whether the given area is similar/dissimilar to the neighboring areas [5].

The global Moran's spatial autocorrelation statistics

The global Moran's *I* statistics is used to test the existence of global spatial autocorrelation and it is defined as follows:

$$I = \frac{n \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \tag{1}$$

where: x_i, x_j are the values of variable in spatial unit i, j; $\bar{x}$ is the mean value of variable for all of the spatial units; n is the total number of spatial units that are included in the study; w_{ij} is an element of spatial weights matrix.

The basic element of spatial analysis is determining the structure of neighborhood with the use of spatial weights. The spatial weights matrix can be defined by two categories of neighbors: contiguity-based neighbors and distance-based neighbors. In the research, it is assumed that mutual interactions between districts occur if they have common borders. Therefore, the binary matrix is created (taking the value of 1 if the districts are adjacent or taking the value of 0 if the districts are not neighboring). Next, the matrix created in such a manner have to be row-standardized, that allows the comparison of results of various areas that have been analyzed.

The positive and significant values of Moran's *I* statistics indicate the existence of positive spatial autocorrelation, i.e. the similarity between the analyzed objects. The negative values of Moran's *I* statistics indicate the negative autocorrelation, i.e. the dissimilarity of analyzed objects. The values of statistic which equal 0 indicate the lack of spatial autocorrelation.

The graphic presentation of the Global Moran's *I* statistics is the scatter plot (Moran scatter plot), which shows local spatial associations (clusters), outliers and spatial instabilities [1]. The graph presents a standardized variable (here Hellwig's measure of development) in the x-axis versus the spatial lag of that standardized variable in the y-axis.

The local Moran's spatial autocorrelation statistics

The local Moran's I_i statistics allows the identification of spatial clusters and measures whether a given spatial unit is surrounded by units with similar or dissimilar values of a given variable in relation to random distribution of these values in space and it is determined by following formula [5]:

$$I_i = \frac{\left(x_i - \bar{x}\right)\sum_{j=1}^{n} w_{ij}\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{n}} \tag{2}$$

where: w_{ij} are the elements of first order row-standardized spatial weights matrix **W**, the other elements of the formula are defined as in the global Moran's *I* statistics.

Significantly negative value of the standardized local Moran I_i statistics indicates that the object i is surrounded by the spatial units (e.g. districts) with significantly dissimilar values of a given variable, which should be interpreted as the negative autocorrelation. Significantly positive value of I_i statistics indicates that the object i is surrounded by the similar neighboring spatial units and we deal with the positive autocorrelation and clustering of the spatial units.

3 Spatial analysis of Internet access for young people in Silesia Province

The aim of the this study is to analyze the availability of Internet access for young people in schools of Silesia Province. The Internet access in schools of Silesian districts will be characterized by the synthetic taxonomic measure of development proposed by Z. Hellwig (description of the method is e.g. in [2; 8; 9; 11]), which allows to sort out the analyzed objects (districts) in respect of a given phenomenon. Table 1 presents variables (stimulants and destimulants) which were chosen to describe various aspects of the availability of Internet access.

Variable and its description	Nature of variable	Coeficient of variation CV [in %]		
		2003	2008	2012
X_1 – number of students per 1 computer with Internet access which is assigned for use by students in upper secondary schools	Destimulant	30.6	46.4	24.2
X_2 – number of students per 1 computer with Internet access which is assigned for use by students in secondary schools	Destimulant	29.3	21.5	20.5
X_3 – number of students per 1 computer with Internet access which is assigned for use by students in primary schools	Destimulant	71.7	32.5	22.2
X_4 – share of upper secondary schools equipped with the computers with Internet access assigned for use by student in all of Silesian upper secondary schools	Stimulant	20.6	20.1	19.6
X_5 – share of secondary schools equipped with the computers with Internet access assigned for use by student in all of Silesian secondary schools	Stimulant	13.0	10.0	10.9

Table 1 Variables characterizing the level of Internet access for young people in Silesian schools

Destimulants were converted into stimulants. Due to the fact that our diagnostic variables are in different measure units they cannot be aggregated directly. Using the zero unitarisation method the variables were normalized according to the formula [3]:

$$z_{ij} = \frac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}} \tag{3}$$

where: $\min\limits_i x_{ij}$ – minimum of a variable x_j; $\max\limits_i x_{ij}$ – maximum of variable x_j.

Subsequently, the Hellwig's taxonomic measures of development (z_i) were calculated. They take the values within the range [0,1]. The higher value of this indicator the more favorable position of the object. On the basis of data presented in Table 2 and Figure 1, we can state that in Silesian districts ranking as regards the Internet access for young people in schools:

- in 2003 leading positions were held by districts: m. Bytom, m. Piekary Śląskie, częstochowski and closing positions were held by districts such as: m. Jaworzno, m. Sosnowiec and m. Zabrze.
- in 2008 leading positions were held by districts: częstochowski, będziński and myszkowski. Districts m. Jaworzno, m. Sosnowiec and rybnicki closed the ranking.
- in 2012 leading positions were held by districts: częstochowski, m. Piekary Śląskie and gliwicki. The closing positions were held by districts such as: m. Jastrzębie Zdrój, m. Ruda Śląska and m. Sosnowiec.

On the basis of taxonomic measure of development it is possible to divide analyzed districts into homogenous classes (i.e. with similar level of Internet access for young people in schools). The districts of Silesia Province were divided into four typological groups using the three-means method [6]:

- Class I – the high level of Internet access for young people in schools ($z_i > \overline{z_{1i}}$),
- Class II – the moderate level of Internet access for young people in schools ($\overline{z_i} < z_i \leq \overline{z_{1i}}$),
- Class III – the low level of Internet access for young people in schools ($\overline{z_{2i}} < z_i \leq \overline{z_i}$),
- Class IV – the very low level of Internet access for young people in schools ($z_i \leq \overline{z_{2i}}$).

where: $\overline{z_i}$ is a mean of the development measure of district i, $\overline{z_{1i}}$, $\overline{z_{2i}}$ are means of half of (respectively) higher and lower values of the development measure of district i.

2003			2008			2012		
District	z_i	**Rank**	**District**	z_i	**Rank**	**District**	z_i	**Rank**
m.Bytom	0.67	1	częstochowski	0.87	1	częstochowski	0.74	1
m.Piekary Śląskie	0.66	2	będziński	0.65	2	gliwicki	0.61	2
częstochowski	0.63	3	myszkowski	0.58	3	m.Piekary Śląskie	0.58	3
m.Dąbrowa Gór.	0.62	4	pszczyński	0.58	4	będziński	0.53	4
m.Świętochłowice	0.61	5	lubliniecki	0.56	5	pszczyński	0.52	5
gliwicki	0.60	6	m.Gliwice	0.55	6	myszkowski	0.49	6
m.Siemianowice Śl	0.60	7	m.Żory	0.53	7	m.Dąbrowa Gór.	0.49	7
raciborski	0.59	8	gliwicki	0.53	8	m.Bielsko-Biała	0.46	8
tarnogórski	0.56	9	m.Piekary Śląskie	0.49	9	m.Rybnik	0.45	9
m.Chorzów	0.50	10	m.Częstochowa	0.48	10	m.Chorzów	0.45	10
m.Mysłowice	0.49	11	m.Bielsko-Biała	0.47	11	m.Tychy	0.44	11
m.Jastrzębie-Zdrój	0.48	12	kłobucki	0.46	12	kłobucki	0.44	12
bieruńsko-lędziński	0.48	13	bieruńsko-lędziński	0.46	13	tarnogórski	0.43	13
cieszyński	0.46	14	m.Siemianowice Śl	0.45	14	m.Gliwice	0.41	14
m.Tychy	0.43	15	wodzisławski	0.45	15	lubliniecki	0.39	15
m.Żory	0.40	16	m.Dąbrowa Gór.	0.45	16	m.Żory	0.37	16
m.Bielsko-Biała	0.38	17	m.Świętochłowice	0.44	17	cieszyński	0.35	17
m.Rybnik	0.38	18	mikołowski	0.44	18	rybnicki	0.35	18
m.Katowice	0.38	19	raciborski	0.43	19	mikołowski	0.33	19
rybnicki	0.38	20	tarnogórski	0.43	20	zawierciański	0.32	20
m.Gliwice	0.37	21	m.Tychy	0.42	21	bielski	0.32	21
m.Częstochowa	0.34	22	m.Rybnik	0.41	22	bieruńsko-lędziński	0.31	22
będziński	0.34	23	cieszyński	0.41	23	m.Siemianowice Śl	0.30	23
żywiecki	0.34	24	żywiecki	0.40	24	żywiecki	0.30	24
pszczyński	0.32	25	m.Chorzów	0.38	25	raciborski	0.29	25
mikołowski	0.31	26	bielski	0.36	26	m.Katowice	0.28	26
myszkowski	0.29	27	m.Katowice	0.29	27	wodzisławski	0.27	27
zawierciański	0.25	28	m.Jastrzębie-Zdrój	0.21	28	m.Częstochowa	0.24	28
wodzisławski	0.24	29	m.Zabrze	0.17	29	m.Świętochłowice	0.23	29
lubliniecki	0.20	30	m.Mysłowice	0.13	30	m.Bytom	0.20	30
kłobucki	0.12	31	m.Ruda Śląska	0.12	31	m.Mysłowice	0.20	31
m.Ruda Śląska	0.11	32	m.Bytom	0.11	32	m.Jaworzno	0.17	32
bielski	0.11	33	zawierciański	0.07	33	m.Zabrze	0.10	33
m.Zabrze	0.11	34	rybnicki	0.04	34	m.Sosnowiec	0.00	34
m.Sosnowiec	0.06	35	m.Jaworzno	0.04	35	m.Ruda Śl.	-0.02	35
m.Jaworzno	-0.13	36	m.Sosnowiec	-0.01	36	m.Jastrzębie-Zdrój	-0.07	36

Table 2 The values of Hellwig's synthetic measure for districts in Silesia Province in 2003, 2008, 2012 with the division for typological groups (starting from Class I – the lightest grey to Class IV – the darkest grey)

Class I, namely the class of districts with the highest degree of Internet access for young people in schools included częstochowski district in every year submitted in the study, in years 2008 and 2012 Class I included also districts: będziński, pszczyński, myszkowski and in the year 2003 and 2012 there were included districts: m. Dąbrowa Górnicza and m. Piekary Śląskie. During years 2003-2012 the number of districts with the very high degree of Internet access for students decreased. In the analyzed group of districts there was the smallest number of students per one computer in a particular type of school and the percentage of schools equipped with computers was the highest. In częstochowski district (with the highest Hellwig's measure of development value) there were 6 students per 1 computer with Internet access in upper secondary schools and about 9 students per 1 computer in secondary and primary schools. 100% of upper secondary schools and 88.24% of secondary schools were equipped with the computers with Internet access for young people.

Class II, namely the class of districts with the moderate degree of Internet access for students in schools and Class III, or more specifically the class of districts with the low degree of Internet access for students in schools included the highest number of districts and what is more there were the biggest changes in the positions held by particular districts over years 2003-2012. Moreover, there was a significant raise in the number of districts in Class III comparing the year 2012 and the year 2008.

Class IV (the class with the lowest degree of Internet access for young people in schools) included districts: m. Ruda Śląska and m. Jaworzno in every analyzed year. In years 2008 and 2012 it included additionally districts: m. Mysłowice and m. Bytom and in the years 2003 and 2012 m. Zabrze. In this group of districts, there was the highest number of students per one computer in particular type of school and the percentage of schools equipped with computers was the smallest. In m. Jastrzębie Zdrój district (with the lowest Hellwig's measure of

development value) there were about 14 students per 1 computer with Internet access in upper secondary schools, about 17 students per 1 computer in secondary schools and about 15 students per 1 computer in primary schools. Only 59% of upper secondary schools and 75% of secondary schools were equipped with the computers with Internet access for the use of students.

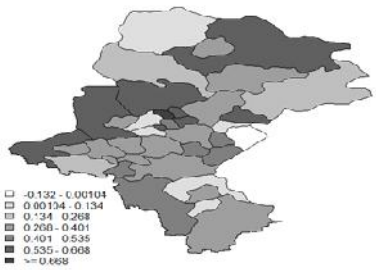
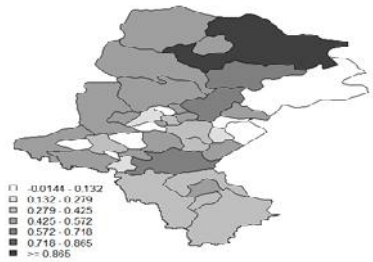
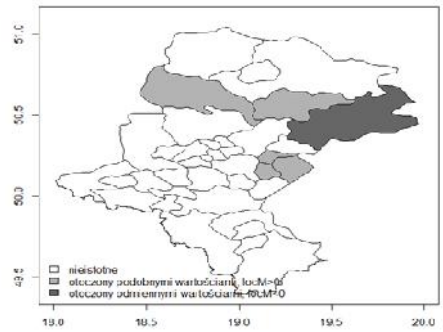
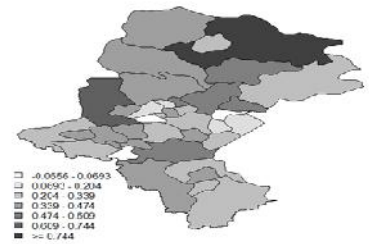
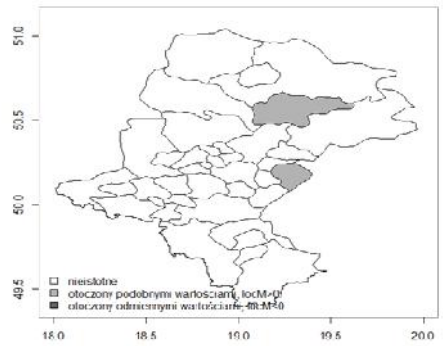
Graph of Hellwig's measure of development, 2003	Graph of significant local Moran's statisctics, 2003
	
Graph of Hellwig's measure of development, 2008	Graph of significant local Moran's statisctics, 2008
	
Graph of Hellwig's measure of development, 2012	Graph of significant local Moran's statisctics, 2012
	

Figure 1 Graphs of Hellwig's measure of development and significant local Moran's statistics in the year 2003, 2008 and 2012

In the further part of this study, it was examined, with the use of spatial statistics, whether the location of particular district in relevance to neighboring districts influences the access to Internet for young people in Silesia Province (whether the access to Internet in schools depends on the geographical position or not). The research of the spatial dependence of districts in Silesian Province was conducted with the assumption of contiguity-based weights (according to the common borders criterion). Table 3 contains calculated values of the global Moran's *I* statistics. Only in 2008 there was the similarity between districts as regards the access to Internet for young people in schools of Silesia Province. The similarity is statistically significant at the 0.1 significance level.

Year	Global Moran's *I* statistics

	I	Ec	Var c	p-value
2003	-0,024	-0,029	0,010	0,480
2008	0,134	-0,029	0,010	0,054
2012	0,073	-0,029	0,010	0,157

Table 3 Values of the global Moran's *I* statistics for the Hellwig's synthetic measure of development

District	Local Moran's I_i statistics					
	2003		2008		2012	
	I_i	p-value	I_i	p-value	I_i	p-value
bielski	0.027	0.453	-0.048	0.517	-0.045	0.514
cieszyński	-0.118	0.586	0.012	0.461	-0.008	0.480
częstochowski	**-0.950**	**0.988**	0.610	0.059	0.508	0.094
gliwicki	-0.177	0.685	-0.239	0.753	-0.443	0.911
kłobucki	-0.063	0.525	0.525	0.155	0.388	0.221
lubliniecki	-0.005	0.516	**1.012**	**0.006**	0.318	0.198
m.Bielsko-Biała	-0.009	0.485	0.012	0.470	-0.076	0.535
m.Bytom	**0.854**	**0.017**	0.011	0.462	-0.145	0.612
m.Chorzów	0.305	0.208	0.024	0.449	-0.429	0.837
m.Częstochowa	0.006	0.480	0.702	0.140	-0.834	0.884
m.Gliwice	-0.014	0.489	-0.035	0.505	0.094	0.410
m.Jaworzno	0.580	0.133	**1.807**	**0.001**	**1.014**	**0.027**
m.Katowice	0.000	0.457	0.100	0.314	0.141	0.261
m.Ruda Śląska	-0.134	0.612	0.479	0.084	0.310	0.177
m.Zabrze	**-0.801**	**0.989**	0.114	0.336	0.083	0.369
m.Żory	-0.003	0.475	-0.197	0.660	-0.025	0.497
myszkowski	0.061	0.424	**0.811**	**0.035**	**0.818**	**0.034**
raciborski	0.120	0.375	-0.033	0.504	-0.129	0.586
zawierciański	-0.324	0.737	**-2.10**	**0.999**	-0.133	0.589
żywiecki	0.068	0.430	0.013	0.469	-0.050	0.516
będziński	-0.023	0.493	-0.120	0.626	0.153	0.261
bieruńsko-lędziński	-0.223	0.682	-0.198	0.661	0.021	0.451
m.Dąbrowa Górnicza	**-1.088**	**0.974**	-0.261	0.665	-0.284	0.681
m.Jastrzębie-Zdrój	-0.059	0.530	-0.077	0.547	-0.396	0.817
m.Mysłowice	-0.539	0.864	**1.311**	**0.002**	0.728	0.051
m.Piekary Śląskie	**1.278**	**0.003**	0.069	0.417	0.205	0.307
m.Rybnik	0.004	0.468	0.010	0.462	0.135	0.344
m.Siemianowice Śląskie	**0.804**	**0.021**	-0.006	0.478	-0.90	0.560
m.Sosnowiec	0.363	0.170	**0.806**	**0.021**	0.107	0.370
m.Świętochłowice	-0.650	0.909	-0.210	0.652	0.517	0.119
m.Tychy	-0.118	0.486	0.051	0.432	0.058	0.425
mikołowski	0.012	0.452	-0.036	0.509	-0.008	0.475
pszczyński	-0.001	0.468	0.103	0.347	-0.284	0.778
rybnicki	-0.001	0.464	-0.067	0.550	-0.011	0.478
tarnogórski	0.210	0.239	0.060	0.395	0.192	0.255
wodzisławski	-0.304	0.723	-0.199	0.643	0.212	0.302

Table 4 Local Moran's I_i statistics of Hellwig's measure of development in year 2003, 2008 and 2012

The analysis of scatter plot of the global Moran's statistics (due to limited number of pages the scatter plots were not included in this article.) revealed some outliers, i.e. the districts which are markedly different from other districts (they are called hot spots). In 2003, outlying districts were as follows: m. Jastrzębie Zdrój, m. Sosnowiec, m. Ruda Śląska, m. Częstochowa, zawierciański and częstochowski. In 2008, districts: m. Sosnowiec, m. Częstochowa, zawierciański, częstochowski and rybnicki were outliers. In 2012, outlying districts were: m. Sosnowiec, m. Jastrzębie Zdrój, m. Ruda Śląska, m. Częstochowa, zawierciański and częstochowski.

The last stage of the research was to calculate the local Moran's I_i statistics and to select those which are statistically significant (at the 0.05 significance level – values written in boldface in Table 4). For the majority of districts the local Moran's I_i statistics was statistically insignificant ($0.05 < \text{p-value} < 0.95$). The p-value lower than 0.05 indicates that the significant positive spatial autocorrelation occurs, while the p-values higher than 0.95 indicate the significant negative local autocorrelation.

Analyzing Table 4, in 2003, the local Moran's I_i statistics for districts: m. Bytom, m. Piekary Śląskie, m. Siemianowice Śląskie were significant and higher than 0 which indicates that those districts were surrounded by districts with significantly similar values of the analyzed variable. Such districts are called clusters. The sig-

nificant negative spatial autocorrelation in 2003 are for districts: częstochowski, m. Zabrze and m. Dąbrowa Górnicza. Those districts are called hot spots, they are the islands surrounded by districts which are characterized by disparate values (high or low). In 2008, clusters were created by districts: lubliniecki, m. Jaworzno, m. Mysłowice, myszkowski and m. Sosnowiec. In turn, zawierciański district was a hot spot. In 2012, only myszkowski and m. Jaworzno districts were clusters.

4 Conclusions

The use of methods of spatial statistics in the districts arrangement in Silesia Province allowed to conduct the analysis of the availability of the Internet access for young people in 2003, 2008 and 2012. According to the available subject literature the spatial methods are used more and more often in the analyses of economic and demographic processes [10; 12].

On the basis of conducted research it is possible to assert that over analyzed 10 years there was a significant improvement of the availability of Internet access for young people in schools. The most unfavorable situation is invariably in districts: m. Sosnowiec, m. Ruda Śląska, m. Jaworzno and m. Zabrze. The best conditions are in districts: częstochowski, gliwicki and m. Piekary Śląskie. The local Moran's I_i statistics indicate that there are the associations between neighboring districts concerning the level of the Internet access in schools in every analyzed year.

References

[1] Anselin, L.: Local Indicators of Spatial Association-LISA. *Geographical Analysis* **27** (1995), 93–115.

[2] Berbeka, J.: *Poziom życia ludności a wzrost gospodarczy w krajach Unii Europejskiej (The standard of living of the population and economic growth in the European Union)*. Akademia Ekonomiczna w Krakowie, Kraków, 2006.

[3] Heffner, K., Gibas, P.: *Analiza ekonomiczno-przestrzenna (Economic and Spatial Analysis)*. University of Economics in Katowice, Katowice, 2007.

[4] Jarczyńska, J., Orzechowska, A.: Siecioholizm i fonoholizm zagrożeniem współczesnej młodzieży (Internet and Smartphone Addiction as a threat for today's youth). In: *Uzależnienia behawioralne i zachowania problemowe młodzieży. Teoria Diagnoza Profilaktyka Terapia* (Jarczyńska, J. ed.). Wydawnictwo Uniwersytetu Kazimierza Wielkiego, Bydgoszcz, 2014, 121–146.

[5] Kopczewska, K.: *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran (Econometrics and Spatial Statistics with R Cran)*, CeDeWu, Warszawa, 2011.

[6] Młodak, A.: *Analiza taksonomiczna w statystyce regionalnej (Taksonomic Analysis in Regional Statistics)*. Difin, Warszawa, 2006.

[7] Suchecki, B. (ed.): *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych (Spatial Econometrics. Methods and Models of Analysis of Spatial Data)*. Wydawnictwo C. H. Beck, Warszawa, 2010.

[8] Warzecha, K.: Poziom życia ludności Polski i pozostałych krajów Unii Europejskiej – analiza taksonomiczna (The Standard of Living of thePopulation in Poland and Other European Union Countries – Taxonomic Analysis). In: *Gospodarka polska po 20 latach transformacji: osiągnięcia, problemy i wyzwania* (Pongsy-Kania, S. ed.), Instytut Wiedzy i Innowacji, Warszawa, 2009, 19–32.

[9] Wierzbicka, A.: Taxonomic Analysis of the Polish Public Health in Comparison with Selected European Countries. *Statistic in Transition New Series* **13** No. 2 (2012), 343–364.

[10] Wolny-Dominiak, A., Zeug-Żebro, K.: Spatial Statistics in the Analysis of Country Budget Incomes in Poland with the R CRAN, In: *Proceedings of 30th International Conference Mathematical Methods in Economics*, Ramik, J., Stavarek, D., Silesian University. School of Business Administration, Karvina, 2012, 992–997.

[11] Zeliaś, A. (red.): *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym (Dynamic Taxonomic Analysis of Spatial Differentiation of Standard of Living in Poland)*, Akademia Ekonomiczna w Krakowie, Kraków, 2000.

[12] Zeug-Żebro, K.: *Analiza zjawiska starzenia się ludności w ujęciu przestrzennym (The Spatial Analysis of an Aging of Population)*, Studia Ekonomiczne Uniwersytetu Ekonomicznego w Katowicach **141** (2014), 100–113.

Modeling Financial News Duration Using ACD-Models

Tadeusz Winkler-Drews[1], Bogusław Bławat[2]

Abstract. Since the Autoregressive Conditional Duration (ACD) model was first published in the seminal paper of Engle and Russell in 1998, many researcher found it very useful in modeling the financial market micro-structure. Irregular spaced data constituting orders flow can be analyzed and modeled in terms of duration by applying a waste repertoire of ACD models which emerged within the last years.

The goal of our paper is not to model the orders duration but to model the flow of the underlying financial news stream which courses the informative investors to take or change their trading positions in the order book. This way we can better understand phenomena like abnormal market volatility in reaction to the financial news publishing.

Our research is based on the millisecond time stamped news related to Warsaw Stock Exchange WIG-20 index constitutes. We found that our ACD model can be very predictive in modeling the duration of news flows around certain trading time clusters and can contribute to explaining the market micro-structure processes observed at the high frequency resolution.

Keywords: financial news duration, ACD model, Weibull distribution.

JEL classification: C32, C58, D40, G14
AMS classification: 60G55, 62P20, 90B90

1 Introduction

Thanks to algorithmic news analysis it is now possible to deliver to the trading desks very fast data stream containing sentiment signals related the securities or companies of interest. These signals applied to the trading strategy can form a very useful decision variable for the informed traders. Mitra and Mitra [12] showed that the application of news analytics in finance is likely to add value for investors trading at all frequencies from volatility-based strategies to equity trading.

The absence of new information is a very important factor in predicting of market volatility, spreads, and trading volume. Diamond and Verrecchia [6] proposed a rational expectation model with short selling constraints. They assert that, in this setting, the absence of a trade is associated with the occurrence of *bad news*. In this framework, time matters only because of the imposed short selling restrictions. Thus, the absence of a trade is informative and should be correlated with price volatility.

Easley and O'Hara [7] analyze the role of time and its relationship to information in a more general model. In this approach, informed traders increase the speed of trading since they want to exploit their informational advantage. Thus, the time between consecutive trades is negatively correlated with the magnitude of information on the market. This result leads to testable hypotheses regarding the impact of trade durations on other market micro-structure variables, such as spreads or volumes. Easley and O'Hara [4] found that because the lack of trade may signal that no new information exists, the time between trades can itself affect prices and quotes will change in the absence of trades. Also the spread will decrease the longer the time between transactions.

As we can see, the flow of information appears to be very important exogenous variable in predicting the spreads and trading volume. The question is whether we can model the new information stream

[1]Kozminski University, 57/59 Jagiellonska St., 03-301 Warsaw, Poland, tadeusz@kozminski.edu.pl
[2]Kozminski University, 57/59 Jagiellonska St., 03-301 Warsaw, Poland, bblawat@kozminski.edu.pl

coming to the market? The answer to such question seems to be positive, and by applying the standard ACD framework we are going to analyze and model the duration between two consecutive news arriving to the informed trader via the sentiment data stream.

2 The ACD financial news duration model

The research in the area of conditional durations was initiated by Wold [15]), who suggested a model for correlated durations drawing on an autoregressive structure. One of the most important models was presented by Cox [5]: namely the proportional hazard model. Eangle and Russel [8] were among the first to apply the ACD model to forecast durations between consecutive trades using the exponential and Weibull distributions. Lunde [11] extended their framework by applying the generalized gamma distribution. Letting the duration process depend on the state of the price process Bauwens and Giot [2] allowed for the asymmetry in the parameter estimation making the duration be dependent on the price movement direction.

Since than, a wast amount of ACD models arise within the market micro-structure research area. Most of them were discussed by Hautsch [10], Bauwens et al. [3], Fernandes and Grammig [9], Bauwens and Hautsch [4], and recently by Saart et al. [13], pointing the importance of the right parametrization of the model.

2.1 The model foundations

Let the observations be the arrival times of news, given in the form of a simple point process, that is $\{t_0, t_1 \ldots, t_n, \ldots\}$ and $t_0 < t_1 < \cdots < t_n < \ldots$. Following the Engle and Russel [8] notation we assume that the time dependence of news arrivals can be summarized by the function ψ which is the conditional expected duration given past information set $\mathfrak{F}$.
The property of ψ, with $x_i = t_i - t_{i-1}$, is that x_i/ψ_i is independently and identically distributed.
From such assumption can be implied that the durations will satisfy equations (1) and (2)

$$\psi_i = E[x_i|x_{i-1}, \ldots, x_1; \theta] = \Psi(x_{i-1} \ldots, x_{i-p}, \psi_{i-1}, \ldots, \psi_{i-q}; \theta), \tag{1}$$

$$f(x_i/\psi_i|x_{i-1}, \ldots, x_1; \theta) = f(x_i/\psi_i; \theta) \text{ and } iid\ \forall i. \tag{2}$$

The basic idea of the autoregressive conditional duration (ACD) model is a dynamic parameterization of the conditional mean function

$$\Psi_i := \Psi_i(\theta) = E\left[x_i|\mathfrak{F}_{t_{i-1}}; \theta\right], \tag{3}$$

where θ denotes $M \times 1$ parameter vector.
The simple specification of (3) can base on a linear parameterization of the conditional mean function

$$\Psi_i = \omega + \sum_{j=1}^{P} \alpha_j x_{i-j} + \sum_{j=1}^{Q} \beta_j \Psi_{i-j}, \tag{4}$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0$.
The unconditional mean and variance have the form

$$E(x_i) = \mu = \frac{\omega}{1 - (\alpha - \beta)}, \tag{5}$$

$$Var(x_i) = \sigma^2 = \mu^2 \frac{1 - \beta^2 - 2\alpha\beta}{1 - \beta^2 - 2\alpha\beta - 2\alpha^2}. \tag{6}$$

Equations (5) and (6) give stationarity conditions for the coefficients α and β: $\alpha + \beta < 1$ and $\beta^2 + 2\alpha\beta + 2\alpha^2 < 1$.
Equation (4) can be rewritten in terms of an intensity representation

$$\lambda(t; \mathfrak{F}_t) = \tilde{\lambda}_\epsilon \left(\frac{x(t)}{\Psi_{\tilde{N}(t)+1}} \right) \frac{1}{\Psi_{\tilde{N}(t)+1}}. \tag{7}$$

It is assumed that the standardized durations

$$\epsilon_i = \frac{x_i}{\Psi_i} \tag{8}$$

follow an i.i.d. process defined on positive support with $E[\epsilon_i] = 1$.

2.2 The Weibull ACD model (WACD)

Eangle and Russel [3] raised the problem of a flat conditional intensity as not having a good fit with some semi-parametric estimate of the baseline hazard of the data, and they therefore propose to extend the exponential ACD model by generalizing the exponential density of the standardized durations to a Weibull(1,γ) density. In this case

$$f(\psi_i) = \phi_i = \psi_i \left\{ \Gamma\left(1 + \frac{1}{\gamma}\right) \right\}^{-1}, \tag{9}$$

and the density of the standardized durations is given by

$$g\left(\frac{x_i}{\phi_i}|x_{i-1};\theta\right) = \frac{\gamma\phi_i}{x_i}\left(\frac{x_i}{\phi_i}\right)^{\gamma} \exp\left(-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right). \tag{10}$$

It is straightforward to derive from (10) the conditional density of the duration

$$g\left(x_i|x_{i-1};\theta\right) = \frac{\gamma}{x_i}\left(\frac{x_i}{\phi_i}\right)^{\gamma} \exp\left(-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right). \tag{11}$$

The conditional hazard function implied by the (11) is given by

$$\lambda_i\left(x_i|x_{i-1};\theta\right) = \phi_i^{-\gamma} x_i^{\gamma-1}\gamma. \tag{12}$$

If $\gamma = 1$ in the (12) the model reduces to the exponential model. The WACD model allows therefore for more flexibility in the conditional hazard function that is either increasing, if $\gamma > 1$, or decreasing if $0 < \gamma < 1$.

3 Data and methodology

For the model estimation we used Thomson-Reuters News Analytics dataset related to the WIG-20 constitutes from the period 2003-01-01 to 2012-06-30. This data contains millisecond timestamped news scored at an entity level across of key dimensions which provides a robust set of meta-data in more than 80 fields including: author sentiment, relevance, uniqueness, item length, intensity, topic, location of the first mention, headline text, item genre, item type and index information for cross linking-stories [14]. We limited our data set to the time period from 8:00 a.m. to 6:59 p.m concentrating around the usual Warsaw Stock Exchange trading hours.

To capture the hourly patterns of the arriving news, we have divided this period into one hour time-span. We excluded records where $t_i - t_{i-1} = 0$. The model parameters were obtained by maximizing the conditional log-likelihood function on a set of observed news arrivals $\mathbf{x}_T = (x_1, \ldots, x_T)'$

$$L\left(\mathbf{x}_T|x_1;\theta\right) = \sum_{i=2}^{T}\left[\log\left(\frac{\gamma}{x_i}\right) + \gamma\log\left(\frac{x_i}{\phi_i}\right) - \left(\frac{x_i}{\phi_i}\right)^{\gamma}\right]. \tag{13}$$

We have estimated the WACD(1,1) model separately for each hourly period obtaining the ω, α, and β ACD coefficients, and the γ parameter of the Weibull distribution.

4 Results

Table 1 summarizes results obtained during the model estimation. Both stationery conditions imposed on α and β coefficients by the equations (5) and (6) are met in each time period. The mean duration between news arrivals ranks from 63.2388 ms to 100.7940 ms exhibiting strong increase around the stock exchange closing hours. This can be explained by the well known companies behavior. Many price sensitive data are published after the stock exchange closes, especially if the news to be announced could cause decrease in the securities prices. The Weibull distribution seems to be a very good choice for fitting the real news arrival duration. Only in the period from 12:00 to 12:59 estimated β coefficient breaks the imposed non negativity condition with $\beta = -0.0500$.

time	ω	α	β	Weibull γ param.	mean duration
08:00–08:59	11.7500	0.1694	0.6716	0.9508	73.8019
	(1.2480)	(0.0118)	(0.0242)	(0.0083)	
09:00–09:59	11.2633	0.1959	0.6636	0.9774	79.4976
	(2.0672)	(0.0225)	(0.0387)	(0.0169)	
10:00–10:59	8.8045	0.1115	0.7706	0.9617	74.2580
	(3.8266)	(0.0284)	(0.0720)	(0.0227)	
11:00–11:59	16.3916	0.2065	0.5505	0.9924	67.0888
	(2.5625)	(0.0230)	(0.0501)	(0.0156)	
12:00–12:59	85.6391	0.1851	-0.0500	0.9613	98.7973
	(24.7947)	(0.0641)	(0.2389)	(0.0386)	
13:00–13:59	20.0844	0.1566	0.5936	0.9855	79.6519
	(10.2213)	(0.0479)	(0.1620)	(0.0271)	
14:00–14:59	10.5312	0.2368	0.6035	0.9840	63.2388
	(3.2677)	(0.0466)	(0.0784)	(0.0280)	
15:00–15:59	16.9697	0.1653	0.5782	0.9516	65.8663
	(4.0088)	(0.0289)	(0.0757)	(0.0203)	
16:00–16:59	8.1712	0.2126	0.7047	1.0157	100.7940
	(1.1185)	(0.0175)	(0.0237)	(0.0143)	
17:00–17:59	10.2466	0.1959	0.692	0.9875	91.2104
	(1.5413)	(0.0185)	(0.0292)	(0.0141)	
18:00–18:59	17.1884	0.2128	0.5592	0.9759	75.7572
	(2.3649)	(0.0222)	(0.0431)	(0.0146)	

Table 1 Results of the $\Psi_i = \omega + \sum_{j=1}^{P} \alpha_j x_{i-j} + \sum_{j=1}^{Q} \beta_j \Psi_{i-j}$ WACD model coefficients estimation. *P=1,Q=1.* Observation taken from the period 2003-01-01 to 2012-06-30. Standard errors in parenthesis.

5 Discussion

Estimation results achieved with the Weibull distribution will lead us to the further research with the goal to find its better parametrization in terms of distribution function. A next research direction could be conducted be by testing how the change from maximum likelihood estimation (ML) to the *estimation function* (EF) framework, as proposed for the Weibull ACD model by Allen et al. [1], would lead to improve our results.

It is to be underlined, that this research is the first attempt to model the financial news arrivals duration related to the Warsaw Stock Exchange WIG-20. As depicted on the Figure 1, which presents a small cutout from the duration prediction accuracy obtained with the means of simple WACD(1,1) model, the fitted durations can be reliable source of information about expected time elapses until next news data will arrive to the market.

This way, by taking the news arrival duration as an exogenous variable in modeling the dynamics of the limit order book, it will be possible to link the trading duration and spread or volatility with the past information flow. Especially in the area of High Frequency Trading this can be of useful application.

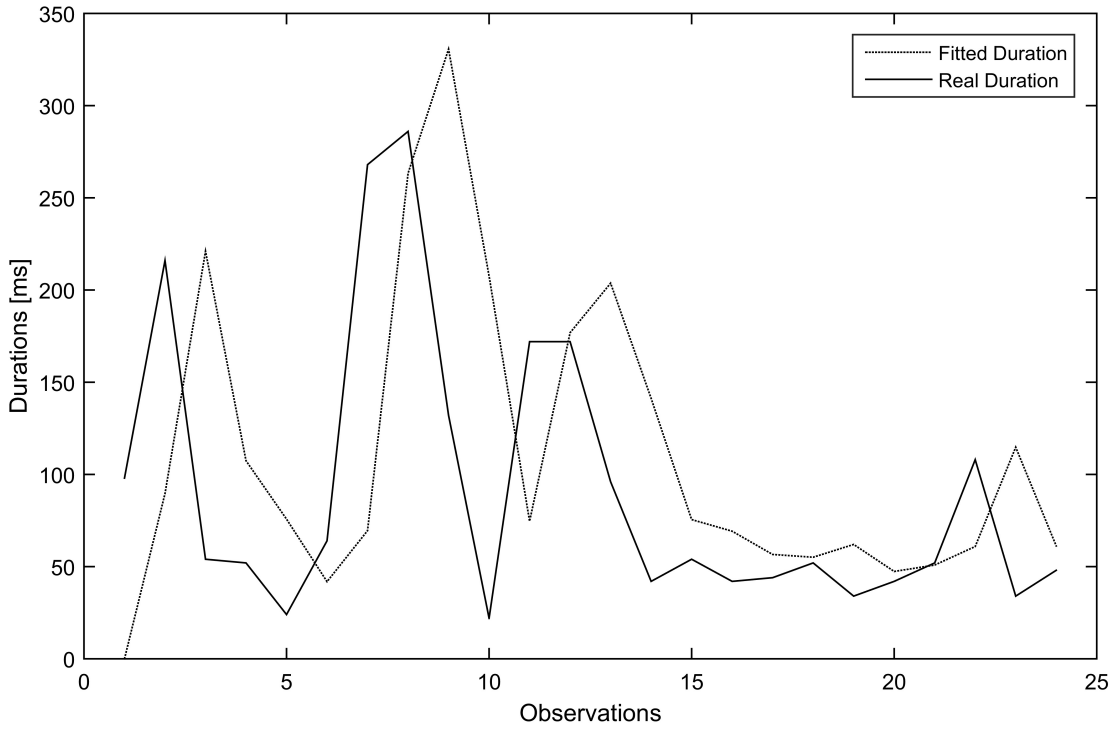


Figure 1 Real durations vs. ACD fitted durations. $N = 24$ observations taken from the period 2009-02-19 08:11:51.038 to 2009-02-19 08:11:53.148. Standard errors in parenthesis.

References

[1] Allen, D, et al.: "Estimating and simulating Weibull models of risk or price durations: An application to ACD models." *The North American Journal of Economics and Finance* **25** (2013), 214–225.

[2] Bauwens, L., and Giot P.: "Asymmetric ACD models: introducing price information in ACD models." *Empirical Economics* **28**, No. 4 (2003): 709–731.

[3] Bauwens, L., et al.: "A comparison of financial duration models via density forecasts." *International Journal of Forecasting* **20** No. 4 (2004), 589–609.

[4] Bauwens, L., and Hautsch N.: *Modelling financial high frequency data using point processes.* Springer Berlin Heidelberg, 2009.

[5] Cox, D. R., 1955, "Some statistical methods connected with series of events (with discussion)." *Journal of the Royal Statistical Society B*, **17** No. 2 (1955), 129–164.

[6] Diamond, D. W., and Verrecchia R. E.: "Constraints on Short-Selling and Asset Price Adjustment to Private Information." *Journal of Financial Economics* **18** (1987), 277–311.

[7] Easley, D., and O'Harra, M.: "Time and the Process of Security Price Adjustment." *Journal of Finance* **47**, No. 2 (1992), 577–605.

[8] Engle, R. F., and Russell, Jeffrey R.: "Autoregressive conditional duration: a new model for irregularly spaced transaction data." *Econometrica* **66**, No. 5 (1998), 1127–1162.

[9] Fernandes, M., and Grammig, J.: "A family of autoregressive conditional duration models." *Journal of Econometrics* **130** No. 1 (2006), 1–23.

[10] Hautsch, N.: *Modelling irregularly spaced financial data: theory and practice of dynamic duration models*. Springer Science & Business Media, 2004.

[11] Lunde, A. "A generalized gamma autoregressive conditional duration model." Department of Economics, Politics and Public Administration, Aalbourg University, Denmark (1999).

[12] Mitra, G, and Mitra, L., eds.: *The handbook of news analytics in finance.* Vol. 596. John Wiley & Sons, 2011.

[13] Saart, P. W., et al.: "Semiparametric autoregressive conditional duration model: Theory and practice." *Econometric Reviews* **34**, No. 6-10 (2015): 848–880.

[14] Thomson-Reuters.: *Reuters News-Scope Sentiment Engine: Guide to Sample Data and System Overview.* V3 (2009).

[15] Wold, H., "On stationary point processes and Markov chains'." *Skand. Aktuar.* **17** (1948), 229–240.

Emission management of a steel company: Fuzzy expected value approach

František Zapletal[1]

Abstract. This paper is devoted to the industrial companies profit margin optimization. Except of the common economic factors, ecological factors are taken into account here. In this paper, an emphasis is put on the emission trading and its influence on the profit margin of enterprises. Since the European companies are obliged to participate in the European emission trading system (EU ETS), this factor should not be omitted in their both planning and decision-making. Unlike other papers with this topic presented by the author (Zapletal and Němec, 2012) or (Zapletal, 2014), in which the deterministic and stochastic versions of the model were presented, in this paper, the fuzzy expected value (i.e. possibilistic mean value) approach is used. Due to the difficult predictability of some model inputs (especially the prices of allowances) caused by system dynamics, the fuzzy approach is considered to be highly suitable. The designed model is verified on the data of one steel company in the Czech Republic.

Keywords: fuzzy expected value, fuzzy optimization, emission trading, possibilistic mean value, EU ETS.

JEL classification: C61
AMS classification: 80M50

1 Introduction

Carbon emissions trading is one of the important factors which must be take into consideration by industrial companies. Therefore, the companies must face additional costs caused by legislative retstrictions. Since 2005, when the EmissionTrading Scheme of the EU (EU ETS) has been launched, prices of emission permits must be observed attentively by these companies. The impact of the emission trading on both profits and production portfolio has been already the object of research many times, see e.g. deterministic model of Zapletal and Němec [1] or stochastic programming models of Rong and Landhelma [4], Zhang and Xu [3] or Zapletal [2]. In the aforementioned stochasic models, the same variables are considered to be uncertain as in this paper (i.e. demand for production of a company and price of emission prices). But, in comparison to them, fuzzy numbers are used instead of stochastic random variables. The reason for this step is simple. Although the observations from past periods are available, the values are often influenced by external shocks (especially the legislative interventions), so the unbelief that the future development will follow the to date distriution is reasonable. An unpredictability of emission allowance prices has been confirmed e.g. by Conrad, Rittler and Rotfuß[5].

On the other hand, the use of fuzzy approach to optimize the companies' production portfolio is not really innovative. It has been already presented in [6],[7] or [8] to both single-period and multi-period production portfolio optimization. All the aforementioned authors used a possibilistic programming approach to optimize the production of companies. To the best knowledge of the author of this paper, no application of fuzzy approach to the production optimization involving the emission trading has been published so far.

The paper is organized as follows. After this short introduction, required knowledge on fuzzy numbers and their expected value are presented in section 2. In section 3, a short introduction to the emission trading in the EU and its characteristics is provided. After that, the fuzzy optimization model maximizing the profit margin of steel company, using the expected values of fuzzy numbers , is established (section

[1]VŠB- Technical university of Ostrava, Faculty of Economics, Department of Systems Engineering, frantisek.zapletal@vsb.cz

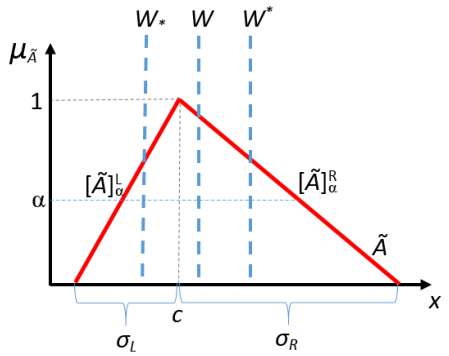


Figure 1 Triangular fuzzy number with lower, middle and upper possibilistic mean values and α-cut

4). In order to be able to verify the model, available data, their description and a way how to build the fuzzy description of chosen data are described within the fifth section. The last part (section 6) is devoted to the verification of the model established using the data of the real steel company, and the results are discussed here. In the end, the paper is concluded by overview of the results obtained and also by proposals for further research which can be done in the future.

2 Fuzzy numbers and their expected value

Despite the fact that many shapes of fuzzy numbers exist, for a sake of simplicity, only triangular fuzzy numbers (t-numbers) are considered further in the text. The generally well-known definition of a t-number is provided e.g. in [10]. In this paper, the notation shown in Figure 1 and is used. Then, a t-number $\tilde{A}(c-\sigma_L, c, c+\sigma_R)$ is given by

$$\mu_{\tilde{A}} = \begin{cases} 0, & x < c-\sigma_L \\ \frac{x-(c-\sigma_L)}{\sigma_L}, & c-\sigma_L < x \leq c \\ \frac{(c+\sigma_R)-x}{\sigma_R}, & c \leq x < c+\sigma_R \\ 0, & x > c+\sigma_R \end{cases} \tag{1}$$

with α-cuts $[[\tilde{A}]_\alpha^L, [\tilde{A}]_\alpha^R] = [c-(1-\alpha)\sigma_L; c+(1-\alpha)\sigma_R], \alpha \in [0,1]$.

2.1 Expected value of the fuzzy number

Fullér and Carlsson, see [9], presented the metric to defuzify the fuzzy numbers, i.e. transform those numbers to the real ('crisp') numbers, using the expected value (also called the possibilistic mean value) of the fuzzy number. That possibilistic mean value can be defined as a level-weighted average of the arithmetic means of all α-cuts, see (2).

$$\mathcal{W}(\tilde{A}) = \int_0^1 \alpha([\tilde{A}]_\alpha^L + [\tilde{A}]_\alpha^R)\, d\alpha = \frac{\int_0^1 \alpha \cdot \frac{[\tilde{A}]_\alpha^L + [\tilde{A}]_\alpha^R}{2}\, d\alpha}{\int_0^1 \alpha\, d\alpha}, \tag{2}$$

Except of the possibilistic mean value $\mathcal{W}$, lower and upper possibilitic mean value can be distinguished, denoted by $\mathcal{W}_*$ and $\mathcal{W}^*$, respectively (see [9]). The lower possibilistic value is nothing else but the lower possibility-weighted average of the minima of the α-cuts, see (3). Analogously, the upper possibilistic value denotes the upper possibility-weighted average of the maxima of the α-cuts, see (4). Lower and upper means can be used, for example, to express the level of optimism of a decision-maker.

$$\mathcal{W}_*(\tilde{A}) = 2\int_0^1 \alpha([\tilde{A}]_\alpha^L)\, d\alpha = \frac{\int_0^1 \alpha \cdot [\tilde{A}]_\alpha^L\, d\alpha}{\int_0^1 \alpha\, d\alpha} \tag{3}$$

$$\mathcal{W}^*(\tilde{A}) = 2\int_0^1 \alpha([\tilde{A}]_\alpha^R)\ d\alpha = \frac{\int_0^1 \alpha \cdot [\tilde{A}]_\alpha^R\ d\alpha}{\int_0^1 \alpha\ d\alpha} \tag{4}$$

It can be easily proved that the relationship between $\mathcal{W}$ and $\mathcal{W}_*$ and $\mathcal{W}^*$, respecively, is given by (5).

$$\mathcal{W}(\tilde{A}) = \frac{\mathcal{W}_*(\tilde{A}) + \mathcal{W}^*(\tilde{A})}{2} \tag{5}$$

The approach using the possibilistic means has both advantages and disadvantages. This method is as easy as possible, because no fuzzy extension of operation or relation is required. On the other hand, the concept is analogous to the mean value of the probability distribution - it is easy to solve, but unfortunatelly, the significant part of information about the fuzzy variable is lost.

Let $\tilde{A}$ be a triangular fuzzy number (see Figure 1), then, using the equations (3), (4) and (5), the values of $\mathcal{W}(\tilde{A})$, $\mathcal{W}_*(\tilde{A})$ and $\mathcal{W}^*(\tilde{A})$ can be simplified to (6).

$$\mathcal{W}(\tilde{A}) = c + \frac{\sigma_R - \sigma_L}{6},\ \mathcal{W}_*(\tilde{A}) = c - \frac{\alpha_L}{3} \text{ and } \mathcal{W}^*(\tilde{A}) = c + \frac{\alpha_R}{3}. \tag{6}$$

Due to linearity of mapping $\mathcal{W}$, equations (7) hold, which simplifies solving the linear programming problem, see [9].

$$\mathcal{W}(\tilde{A} + \tilde{B}) = \mathcal{W}(\tilde{A}) + \mathcal{W}(\tilde{B}),\ \mathcal{W}(\lambda\tilde{A}) = \lambda \cdot \mathcal{W}(\tilde{A}), \tag{7}$$

where $\lambda \in \mathbb{R}$.

3 Basic characteristics of the emission trading in the EU

The EU ETS system is the main tool of the environmental policy of the EU, established in 2005, under [11]. The main idea is that each ton of carbon dioxide released to the atmosphere by a company must be covered by one emission permit (allowance). That means that the worse technology or the higher amount of poduction, the higher impact of the system on the company, because each company with a net heat excess 20MW must participate in the system. The special type of emission permit established in the EU ETS is European Union Allowance (EUA).

The EU ETS was established as a cap-and-trade system, which means that a total amount of allowances in the system is determined by the central authority and the allowance price is generated on the market. A substantial part of allowances needed are grandfathered to companies by the EU authorities for free. The special auctions with allowances were established, where the companies can purchase emission permits in the case that they have a lack of them to cover all released CO_2 emissions released. The second possibility is a so called secondary market with emission allowances containing several stock exchanges, on which the permits can be both purchased and sold.

To get more detailed information on the EU ETS and emission trading, see e.g. [12].

4 Design of the optimization model

The model presented in this paper was established on the basis of the author's stochastic programming model, presented in [2]. The reason of establishing the model in its fuzzy version is simple. During few last years, many unpredictable shocks have occured for both model parameters, which are considered to be uncertain (fuzzy) in the model - demand for company production and emission prices. The former was distorted by the deep economic crisis, whereas the latter was influenced unexpectedly by many interventions by EU authorities in order to increase a price of allowances and thus also the system efficiency.

The model assumptions are as follows:

- total margin of the company is maximized;

- model is established as a static one - only for one period (year);
- only future demand for products and EUA price are considered to be fuzzy, all other model parameters are taken as deterministic;
- the production process is homogenous during the entire period (an amount of both emission and products is the same at any given moment);
- a company generates the profit only by selling its products or by selling unused emission allowances.

$$\begin{aligned} \max_{\boldsymbol{y}\geq 0} \quad & \boldsymbol{m}^T\boldsymbol{y} + (r - \boldsymbol{e}^T\mathbf{B}\boldsymbol{y}) \cdot \mathcal{W}(\tilde{p}) \\ \text{s. t.} \quad & \mathbf{B} = (\mathbf{E} - \mathbf{A})^{-1} \\ & \mathbf{B}\boldsymbol{y} \leq \boldsymbol{c} \\ & \boldsymbol{y} \leq \mathcal{W}(\tilde{\boldsymbol{d}}) \end{aligned} \tag{8}$$

where

- $\boldsymbol{y} \in \mathbb{R}^{n\times 1}$: the decision vector of products' sales;
- $\boldsymbol{m} \in \mathbb{R}^{n\times 1}$: the vector of profit margins on products;
- $r \in \mathbb{R}$: the scalar of EUAs allocated to the company for free;
- $\boldsymbol{e} \in \mathbb{R}^{n\times 1}$: vector of carbon coefficients indicating an amount of CO_2 released per a production unit;
- $\mathbf{E} \in \mathbb{R}^{n\times n}$: the identity matrix;
- $\mathbf{A} \in \mathbb{R}^{n\times n}$: the technological matrix of production;
- $\tilde{\boldsymbol{p}}$: the price of EUAs expressed by the fuzzy number
- $\boldsymbol{c} \in \mathbb{R}^{n\times 1}$: the vector of production capacities;
- $\tilde{\boldsymbol{d}}$: the vector of fuzzy numbers denoting the demanded amounts of products;
- $\mathcal{W}(\tilde{\cdot})$ stands for the expected value (possibilistic mean) of the fuzzy number.

The first constraint comes from the equilibrium of the Leontief's input-output model.

5 Data set

Two data types, which were required for model verification in this paper, can be distinguished - business data and data on emission prices.

All the business data (demand, prices, costs, limits and production coefficients, number of grandfathered allowances) were obtained from the researched steel company, which wants to be kept anonymous. It produces 5 iron and steel products (iron, brams, plates, profiles and cut shapes). Raw iron is used only as a raw material for manufacturing of other products and it is not sold. All the data were valid in the beginning of 2014 (more recent data are not available to the author yet). The time series of demand were taken from 2008 to 2013.

The historical prices of emissions were derived from the database of the SendeCO2 market[1]. The daily data from of the period from 1st March 2008 to 31st March 2015 were used, see Figure 2. (i.e. the period from the beginning of the second trading phase of the EU ETS).

The fuzzy numbers ($\tilde{\boldsymbol{d}}$ and $\tilde{\boldsymbol{p}}$) were established as t-numbers and in following way. The value of center is given by the current (latest) value and the spreads by min and max values, see Figure 2, in which the fuzzy number $\tilde{\boldsymbol{p}}$ is shown.

Using the equations (2), (3) and (4), lower, middle aned upper possibilistic means are calculated, see (9).

[1] Available from www.sendeco2.com

Figure 2 EUA price development with the t-number established [source: sendeco2.com]

$$\mathcal{W}(\tilde{\boldsymbol{p}}) = \mathcal{W}(2.7, 7.11, 28.3) \doteq 2.797,\ \ \mathcal{W}_*(\tilde{\boldsymbol{p}}) \doteq 5.64,\ \ \mathcal{W}^*(\tilde{\boldsymbol{p}}) \doteq 14.17 \tag{9}$$

6 Verification of the model using the data of the steel company

The results of the model (8) in three variants (using $\mathcal{W}_*, \mathcal{W}$ and $\mathcal{W}^*$) are shown in Table 1 and Table 2. It can be seen that the optimal value (profit margin) is the highest when $\mathcal{W}^*$ is used, and the lowest in the model with $\mathcal{W}_*$. This could be expected because the former counts with the most optimistic (the highest) expected demand and, despite the fact that the allowance price is the cheapest in this variant, sales have a substantially higher impact on the company profit margin in comparison to price of permits. On the other hand, costs on emission allowances are the highest for the '$\mathcal{W}^*$-model', which is not surprising again because of the fact that the production and thus also an amount of emission released is the highest. But an interesting fact is the difference between all models regarding the share of costs on allowances on the profit margin, which reach the value of 0.21% for $\mathcal{W}_*$'s, 0.69% for $\mathcal{W}$'s and 2.36% for $\mathcal{W}^*$'s, respectively. The increase in percentage change of costs on allowances is substantially higher than the increase in percentage change of profit margin, when increasing a level of decision-maker's level of optimism. Another interesting conclusion is that, regardless the model used, the company has always a lack of alowances in the end of the period due to the low amount of grandfathered allowaces.

$\mathcal{W}$ used	Iron [t]	Brams [t]	Plates [t]	Cut shapes [t]	Profiles [t]
$\mathcal{W}_*$	323,178	432,082	499840	81,667	26,667
$\mathcal{W}$	367,126	490,840	566,334	90,833	27,667
$\mathcal{W}^*$	411,075	549,597	632,828	100,000	28,667

Table 1 Results of the optimization - company production

$\mathcal{W}$ used	Total margin [1,000 EUR]	Purchased allowances [pcs]	Costs on allowances [1,000 EUR]
$\mathcal{W}_*$	38,548.15	28,372	79.35
$\mathcal{W}$	42,497.17	51,739	291.80
$\mathcal{W}^*$	05,061.02	75,106	1,064.24

Table 2 Results of the optimization - total profit margin and costs on allowances

7 Conclusions

In this paper, the first study devoted to the impact of the European emission trading on profit margin of companies using the fuzzy approach has been done and corresponding fuzzy optimization model has been established. Three variants of this model were used, reflecting different levels of decision-make's level of optimism - with the lower/"common"/upper possibilistic mean, respectively. The fuzzy numbers of allowance prices and demands were built using the spread of historical values, which filter the short-term

volatility of the values and enables the decision-maker to get know the potential effect of both uncertain factors included in the model. The case study has shown that the current values of both demand and emission prices is closer to the lower possibilistic mean than to upper one, which is partly caused by the persisting recession of the European steel sector. In case of the modeled company, a lack of allowances at the end of the period can be expected regardless the scenario considered. The interesting fact is that costs on allowances grow faster than the profit with increasing level of optimism (i.e. between lower and "common" possibilisitic mean and between "common" and upper possibilistic mean value).

The approach of possibilistic values (expected values) of the fuzzy numbers can be considered as a starting point for the further analyses, in which more advanced methods and techniques of fuzzy optimization can be used.

Acknowledgements

This paper was created with financial support of SGS project No. SP2015/117 and the Operational Programme Education for Competitiveness Project CZ.1.07/2.3.00/20.0296. The support is greatly acknowledged.

References

[1] Zapletal, F., and Němec, R.: The Usage of Linear Programming to Construction of Ecologic-economical Model for Industrial Company Profit Optimization. In: *Proceedings of 30th International Conference Mathematical Methods in Economics, 2012*, Karviná, 2012, 1010-1015.

[2] Zapletal, F.: Mean-risk Model Optimizing the Heavy Industrial Companys Profit with Respect to Environmental Aspects. In: *Proceedings of 32th International Conference Mathematical Methods in Economics 2014*, Olomouc, 2014.

[3] Zhang, B., and Xu, L.: Multi-item production planning with carbon cap and trade mechanism. *International Journal of Production Economics* **144** (2013), 118-127.

[4] Rong, A., and Landhelma, R.: CO2 emissions trading planning in combined heat and power production via multi-period stochastic optimization. *European Journal of Operational Research* **176** (2007), 1874-1895.

[5] Conrad, C., Rittler, D., and Rotfuß, W.: Modelling and Explaining the Dynamics of EUA Prices at High Frequency. *Energy Economics* **34** (2012), pp. 316-326.

[6] Wang, R. C., and Fang, H. H.: Aggregate production planning with multiple objectives in a fuzzy environment. *European Journal of Operational Research* **133** (2001), 521-536.

[7] Tang, J., Wang, D., and Fung, R.: Fuzzy formulation for multi-product aggregate production planning. *Production Planning & Control: The Management of Operations* **11** (2000), 670-676.

[8] Tang, H., and Song, G.: Optimization of Enterprise Production Based on Carbon Emission Credits Trading. In: *Proceedings of the 2nd International Conference on Green Communications and Networks*, 2012.

[9] Carlsson, C., and Fuller, R. On possibilistic mean value of fuzzy numbers. *Fuzzy Sets and Systems* **122** (2001), pp. 315-326.

[10] Ramík, J., and Vlach, M.: *Generalized Concavity in Fuzzy Optimization and Decision Analysis.* Springer, 2001.

[11] Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC.

[12] Zapletal, F., and Moravcová, E.: Analysis of current conditions within the ETS and its impact on firms in the Czech Republic. In: *Proceedings of the SMSIS conference 2013*, VŠB - Technical university of Ostrava, Valašké Meziříčí, 2013, 235-255.

The consumption-income ratio and its variability during the business cycle: a heterogeneous agent case

Jana Závacká[1]

Abstract. In this paper we use a simplified agent based model to monitor the consumption-income ratio during the business cycle. We distinguish among heterogeneous households with different incomes determined by labor structure with various tier-workers. The model is capable to generate cyclic movement of economic activity according to the level of economic sentiment in the society. We compare the consumption-income ratio and its variability during the business cycle for different tier workers. We found out that lower-income workers have in general higher consumption-income ratio. The cyclical movement of economic activity affects mainly private incomes of higher-tier workers, therefore increases the variability of their consumption-income ratio, even with the high level of savings. In addition, we found out that the consumption-income ratio for tier workers is countercyclic.

Keywords: Agent-based model, consumption-income ratio, business cycle, growth-savings paradox, economic sentiment.

JEL classification: C63, e27
AMS classification: 91b69

1 Introduction

The consumption-income ratio is usually considered in general macroeconomic theories as stable during time. The stylized fact based on macroeconomic data is that consumption spending is procyclical and less volatile than output during the business cycle (confirmed on data for example by Fiorito and Kollintzas [6]). This implies that the consumption-income ratio measured for the whole society would behave countercyclicly. On the other side, Alvarez-Cuadrado and Van Long [4] point out that there could be a difference between the cross-sectional and time-series properties of consumption data. According to Duesenberry [5] once the relative income distribution is set, there is a positive correlation between growing income (with respect to given income distribution) and saving rate. This support the idea that higher income households can have higher saving rates (e.g. lower consumption-income ratio). This growth-savings paradox is supported also by Harbaugh [9].

We used a simplified agent-based model to investigate the behavior of consumption-income ratio during the business cycle for various tier workers with heterogeneous incomes. The model simulations confirmed that there is a difference in the consumption-income ratio among different tier-workers and unemployed agents, supporting the idea that the higher income households keep the lower consumption-income ratio.

The article is organized as follows. We introduce the model structure in the section 2. The estimation results are presented in the section 3. In the section 4 we conclude.

2 The Model

The cyclical movement of economic activity in the model is based on the spread of waves of optimism and pessimism in the society introduced by Westerhoff [10]. The similar version of model without differentiated labour market was already used in Závacká [11].

[1]VŠB - Technical University Ostrava, Department of Systems Engineering, Sokolská třída 33, 701 21 Ostrava, Czech Republic, jana.zavacka1@vsb.cz

The model is based on two kinds of steps. The macroeconomic situation is evaluated during macrosteps t (we consider as one macrostep one quarter of year). Between every two macrosteps t and $t+1$ there are T microsteps τ. In this phase the spread of the optimism/stability/pessimism is driven, based on the global macroeconomic conditions from the last macrostep t. After the simulation of confidence the new state of macroeconomic situation is evaluated.

2.1 The simulation of the confidence in the society

During this simulation heterogeneous agents (consumers) form their opinions about their future incomes together with their attitude towards actual income spending. In each microstep τ there is randomly chosen one agent from the net. We are working in line with Westerhoff [10] with the net in the form of annuloid. This way we define that each consumer has exactly four neighbors (up, down, left, right). The agent i chooses to be in the macrostep t microstep τ either optimist $O^i_{t\tau}$ (with the aim of increase his consumption spending), stable $ST^i_{t\tau}$ (aimed to no change in consumption spending no matter what are expectations about his future income) or pessimist $P^i_{t\tau}$ (with the aim to decrease the consumption spending) according following probabilities

$$Prob(O^i_{t\tau}) = \begin{cases} \frac{\exp(\alpha^Y_t+\alpha^S_t+\beta^i_{t\tau})}{1+\exp(\alpha^Y_t+\alpha^S_t+\beta^i_{t\tau})} & \text{in expansion,} \\ 0 & \text{in contraction,} \end{cases} \tag{1}$$

$$Prob(P^i_{t\tau}) = \begin{cases} 0 & \text{in expansion,} \\ \frac{1}{1+\exp(\alpha^Y_t+\alpha^S_t+\beta^i_{t\tau})} & \text{in contraction,} \end{cases} \tag{2}$$

$$Prob(ST^i_{t\tau}) = 1 - Prob(O^i_{t\tau}) - Prob(P^i_{t\tau}). \tag{3}$$

The probabilities depend on two global parameters α^Y_t, α^S_t and one local parameter $\beta^i_{t\tau}$. The parameter α^Y_t express the sensitivity on the change of the economic activity Y_t and is defined as

$$\alpha^Y_t = \begin{cases} \alpha^Y & Y_t > \frac{1}{4}(Y_{t-4}+Y_{t-3}+Y_{t-2}+Y_{t-1}), \\ -\alpha^Y & Y_t \leq \frac{1}{4}(Y_{t-4}+Y_{t-3}+Y_{t-2}+Y_{t-1}). \end{cases} \tag{4}$$

α^Y in equation is income sensitivity parameter. The effect of temporal fluctuations in income is decreased by using in the comparison the average income form the last year (in line with permanent income theory of Friedman [7]).

The parameter α^S_t expresses changes in marginal propensity to consume. We distinguish among 4 cases: consumer prefers immediate consumption to accumulate savings (expansion before moment A), consumer prefers accumulate savings to immediate consumption (expansion after moment A), consumer prefers reduce immediate consumption to reduce savings (contraction before moment B), consumer prefers reduce savings to reduce immediate consumption spending (contraction after moment B.) For the simplicity we define the moment A, B as a moment when firstly all consumers became optimists, pessimists respectively. The parameter is defined as (α^S is time preference parameter)

$$\alpha^S_t = \begin{cases} \alpha^S & \text{in expansion after the moment A,} \\ -\alpha^S & \text{in contraction after the moment B,} \\ 0 & \text{else.} \end{cases} \tag{5}$$

The sentiment of the agent i is also influenced by the sentiment of its neighbors through the parameter $\beta^i_{t\tau}$, defined as

$$\beta^i_{t\tau} = \beta(\#O^i_{t\tau} + \#P^i_{t\tau} - \#ST^i_{t\tau}), \tag{6}$$

where β is a confidence sensitivity parameter, $\#O^i_{t\tau}$ the number of optimists, $\#ST^i_{t\tau}$ the number of stable agents and $\#P^i_{t\tau}$ the number of pessimists in the neighborhood of the agent i.

2.2 The model structure

We work with the simple model with M heterogeneous consumers and one firm. For the simplicity and clarity about effects of interactions among these economic subjects we do not consider any monetary or fiscal authority, nor capital. We assume that prices are constant (no inflation) and the economy is closed.

Consumers

Each consumer i in the time t offers his labor on the labor market, obtains income Y_t^i and decides about his personal demand for consumption spending ID_t^i. This demand is based on Rational Expectation Permanent Income Hypothesis (Hall, [8]), allowing for the importance of confidence confirmed by Carroll [2] and "habit" (equal to the share γ of its last consumption spending C_{t-1}^i) in line with Abel [1].

$$ID_t^i = \begin{cases} \gamma C_{t-1}^i + (1-\gamma)(1+x)\frac{1}{4}(Y_{t-3}^i + Y_{t-2}^i + Y_{t-1}^i + Y_t^i) & \text{optimist,} \\ \gamma C_{t-1}^i + (1-\gamma)\frac{1}{4}(Y_{t-3}^i + Y_{t-2}^i + Y_{t-1}^i + Y_t^i) & \text{stable,} \\ \gamma C_{t-1}^i + (1-\gamma)(1-x)\frac{1}{4}(Y_{t-3}^i + Y_{t-2}^i + Y_{t-1}^i + Y_t^i) & \text{pessimist,} \end{cases} \tag{7}$$

where x is an extrapolation constant. By summation across all agents we obtain the aggregate demand for goods AD_t

$$AD_t = \gamma C_{t-1} + (1-\gamma)(1 + \frac{x}{M}(\#O_t - \#P_t))\frac{Y_{t-3} + Y_{t-2} + Y_{t-1} + Y_t}{4}, \tag{8}$$

where $\#O_t$ is the number of optimists and $\#P_t$ the number of pessimists in the society at time t.

The structure of labor market in the model is based on the idea of Ciarli et al. [3]. We distinguish between 5 tiers of workers and the unemlployed. For each ν workers there is a need of one worker from the higher tier. Hence the total amount of workers L_t is

$$L_t = \sum_{k=1}^{5} L_t^k = L_t^1(1 + \frac{1}{\nu} + \frac{1}{\nu^2} + \frac{1}{\nu^3} + \frac{1}{\nu^4}), \tag{9}$$

where L_t^k is number of k-tier workers in time t. All agents, which are not in time t part of workers are the unemployed U_t. Each agent has an ability to work just for one type of tier. In case that firm is willing to increase the labor force, it chooses the additional workforce randomly from the workers of the proper tier. By analogy, in case the firm is willing to decrease the workforce, it is done by random choice from current workforce.

The wage structure in the model is connected with the tier structure of workers. The wage W_t^{k+1} of the tier $k+1$ is always b times higher than the wage of worker of the tier k. The highest tier workers have also a flexible part of wage, which is equal to share ω^R on the realized sale from the last period EQ_{t-1}. The unemployed agents obtain the social contribution SC_t equal to the share ω^U of the wage of worker of tier 1. This contribution is payed from the wage fund, accumulated from the taxes of the workers. The wage fund is always balanced. Thus the the personal income of each agent is defined

$$Y_t^i = \begin{cases} b^{k-1}W_t^1 & \text{for the } k\text{-tier workers, } k = 1, \dots 4, \\ bW_t^4 + \omega^R EQ_{t-1} & \text{for the } 5th\text{-tier workers,} \\ \omega^U W_t^1 & \text{for the unemployed.} \end{cases} \tag{10}$$

The wage of worker of tier-1 is then

$$W_t^1 = \frac{Y_t - \omega^R EQ_{t-1}L_t^1}{\omega^U U_t + L_t^1 + bL_t^2 + b^2L_t^3 + b^3L_t^4 + b^4L_t^5}. \tag{11}$$

In addition, we define consumption income ratio (CIR) for each worker as

$$CIR_t^i = \frac{C_t^i}{Y_t^i}. \tag{12}$$

During model simulations, we will count the average of this ratio for all tier workers together with the unemployed, in general and in four phases of business cycle.

Firm

The firm in the model forms the aggregate supply and is always trying to satisfy the aggregate demand. Hence, firm during expansion is mainly increasing the production Q_t according the level of sentiment in

society, during contraction is rather willing to offer at least the same amount as the aggregate demand from the last period, but is trying to decrease its inventory I_t by the share κ. The target production $\bar{Q}_t$ in time t is defined as

$$\bar{Q}_t = \max\{AD_{t-1} - \kappa I_t, (1 + \frac{x}{M}(\#O_t - \#P_t))Q_{t-1}\} \tag{13}$$

The production Q_t in time t depends only on the workforce of tier 1 and the marginal productivity of labor a^L. As the firm has no profit and there is no inflation, we can set the price of good equal to 1. The production costs are then in each period equal to the value of produced goods.

$$Q_t = a^L L_t^1 = Y_t, \tag{14}$$

The demand for labor force is adjusted for the level δ of rigidity and constrained from up by the total number of agents M. The workforce L_t at time t is[1]

$$L_t = \left\lfloor \min\{(\delta L_{t-1}^1 + (1-\delta)\frac{\bar{Q}_t}{a^L})(1 + \frac{1}{\nu} + \frac{1}{\nu^2} + \frac{1}{\nu^3} + \frac{1}{\nu^4}); M\} \right\rfloor. \tag{15}$$

According to the workforce L_t the workforce of tier 1 L_t^1 and the final production Q_t is determined. The aggregate supply AS_t in time t consists of the production and the inventory from the last period I_{t-1}

$$AS_t = Q_t + I_{t-1}. \tag{16}$$

The aggregate supply is compared with the aggregate demand on the market of goods. The realized sale which is also equal to the final consumption spending, is defined as

$$EQ_t = C_t = \min\{AD_t, AS_t\}. \tag{17}$$

This result from the market determines the individual consumption spending C_t^i, savings and assets of consumers in time t and inventory of firm I_t

$$C_t^i = C_t \frac{ID_t^i}{AD_t}, \tag{18}$$

$$S_t^i = Y_t^i - C_t^i, \tag{19}$$

$$A_t^i = A_{t-1}^i + S_t^i. \tag{20}$$

$$I_t = AS_t - C_t. \tag{21}$$

3 Estimation results

We ran 100 simulations of the model in program R. To separate the results from the initial setting[2] the values from first 50 macrosteps in each run were dropped. First, the average consumption-income ratio for each tier group in every macrostep in every run was counted. Then, from these results, the average consumption-income ratios for each tier group for the whole business cycle and for its initial phases for each run were counted. Next, the boxplots of these average values through all runs were investigated. The values of parameters in the model (presented in Table 1) are mainly in accordance with Westerhoff [10], Ciarli et al. [3] and other commonly used estimations.

Firstly we will concentrate on the estimation of average consumption-income ratio for initial tier workers in general. We can see the boxplots of these estimations in the Figure 1.(a). According to the results the average consumption-income ratio is mainly decreasing with the tier and smaller than 1 for all tier workers. However, the consumption-income ratio of the unemployed is higher than 1 (1.11), which means that the social contribution does not fully satisfy agent's consumption spending. We have to take into the account that the unemployed are agents from all work tier groups. Thus, in case the agent from the highest tier looses a job, the difference between his actual and past income is radical, however the difference between his past and current consumption spending is not. The group of unemployed should be than rather considered as the group of workers with temporal income dropout. The higher consumption-income ratio represents the consumption smoothing in this period.

[1] The symbol $\lfloor\rfloor$ means the closest lower integer satisfying the proportional structure given by ν for which all numbers of workers in all tiers are integers.

[2] It is at disposal upon request.

Parameter	Description	Parameter	Description
$M = 10000$	Number of consumers	$TTT = 100$	Number of simulations
$T = 10000$	Number of macrosteps	$TT = 350$	Number of microsteps
$\alpha^Y = 5$	Income sensitivity parameter	$\alpha^S = -10$	Time preference parameter
$\beta = 1$	Confidence sensitivity parameter	$x = 0.05$	Extrapolation parameter
$\gamma = 0.8$	Consumption smoothing parameter	$\delta = 0.6$	Production smoothing parameter
$\kappa = 0.7$	Melting inventory parameter	$\nu = 5$	Tier multiplier
$a_L = 1$	Marginal productivity of labor	$b = 2$	Executive wage multiplier
$\omega_R = 0.02$	High-tier profit share parameter	$\omega_U = 0.8$	Social contribution parameter

Table 1 Parameter setting.

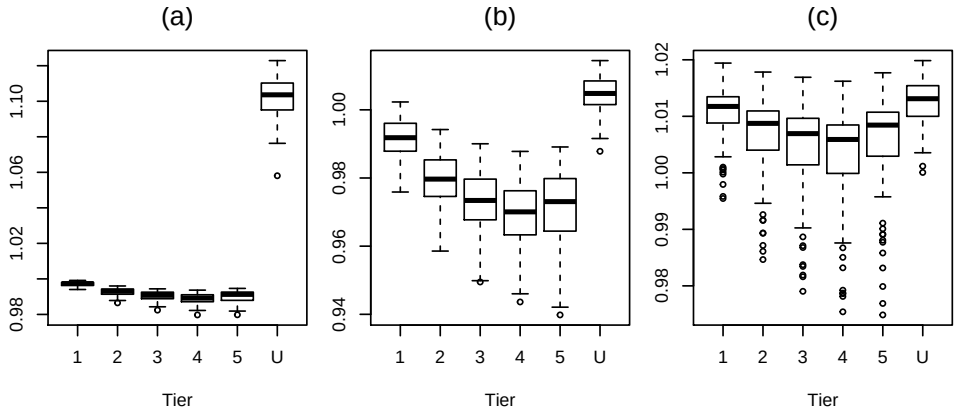


Figure 1 Average CIR during business cycle (a) and expansion before A (b), after A (c).

The consumption-income ratio in the first phase of expansion (Figure 1.(b)) is for all tier workers lower than 1. This is in accordance with the slightly increasing income of all agents. In the second phase of the cycle (Figure 1.(c)), where all the agents were optimists, the consumption-income ratio is higher than 1 nearly by all tier workers. The consumption spending is thus higher than the income by nearly all agents. This could be clearly explained by the high level of confidence in society, which together with increasing incomes create favorable conditions for consumption spending.

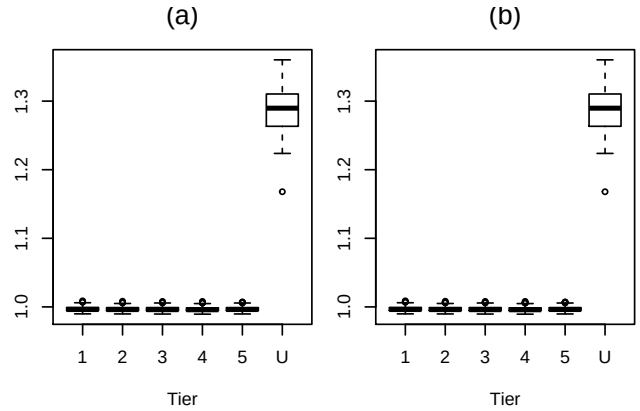


Figure 2 Average CIR during contraction before B (a), after B (b).

In case of first phase of contraction (Figure 2.(a)) the agents confidence decreases together with the willingness for consumption spending. The consumption-income ratio is close to 1 for all tier workers, representing the moderate consumption spending. In the last phase of contraction (Figure 2.(b)), when all the agents were already pessimists, the consumption-income ratios decrease by all tier groups together with the unemployed. This could be understood as the precautionary saving motive, when in the case of lower confidence agents prefer to accumulate savings for worse times.

4 Conclusion

We have simulated the behavior of heterogeneous consumers on the simplified agent-based model with differentiated labor market. According to the results we found out that the consumption-income ratios of the all work-tier groups are volatile during the business cycle and evolve procyclicly, e.g. in the opposite direction to the expected consumption-income ratio aggregated for the whole society. This could be explained by the overwhelming countercyclical effect of the consumption-income ratio of the unemployed. The unemployed agents experience an temporal income dropout and smooth their consumption because of big difference between their usual wage and obtained social contribution. Instead, the growing consumption-income ratio of all working groups during the phase of expansion is in accordance with the observable growing trend in data of spending on long-term consumption goods, as well as investment to the housing etc. It is also in line with the results of Duesenberry [5] and Harbaugh [9].

The proper understanding of the consumption spending of households during the business cycle with respect to different income groups rather than in aggregate form could be beneficial for adjustments of fiscal and monetary policies with respect to different income groups during the business cycle. This could be mainly used for evaluating the effect of proportional or progressive tax system etc. Nevertheless, we have to mention that the whole simulation was made on the simplified agent-based model without considering the capital investment and prices. The next steps in this research could be then the extension of this model and possible confirmation of these results also on more sophisticated model.

Acknowledgements

The paper was funded by Student Grant Competition of VŠB-TU Ostrava within the project SP2014/131, and within Operational Programme Education for Competitiveness - Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Abel, A.B. : Asset Prices under Habit Formation and Catching up with the Joneses, *The American Economic Review* **80** (1990), 38–42.

[2] Carroll, Ch.D., Fuhrer, J.C., and Wilcox, D.W. : Does Consumer Sentiment Forecast Household Spending? If So, Why?, *The American Economic Review* **84** (1994), 1397–1408.

[3] Ciarli, T., Lorentz, A., Savona, M., and Valente, M.: Te effect of consumption and production structure on growth and distribution. A micro to macro model. *Metroeconomica* **61** (2010), 180–218.

[4] Alvarez-Cuadrado, F., and Van Long, N. : The relative income hypothesis, *Journal of Economic Dynamics & Control* **35** (2011), 1489–1501.

[5] Duesenberry, J.S. : *Income,Saving and the Theory of Consumer Behavior*. Harvard University Press, Cambridge, MA. 1949.

[6] Fiorito, R., and Kollintzas, T. : Stylized Facts of Business Cycles in the G7 from a Real Business Cycles Perspective, *European Economic Revue* **38** (1994), 235–269.

[7] Friedman, M. : *A theory of the consumption function: a study by the National Bureau of Economic Research*. Princeton University Press, New York, 1957.

[8] Hall, R. : Stochastic Implications of the Life-Cycle-Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy* **86** (1978).

[9] Harbaugh, R. : Falling behind the Joneses: relative consumption and the growth-savings paradox, *Economics Letters* **53** (1996), 297–304.

[10] Westerhoff, F. : An agent-based macroeconomic model with interacting firms, socio-economic opinion formation and optimistic/pessimistic sales expectations, *New Journal of Physics* **12** (2010).

[11] Závacká, J. : Monetary policy during the business cycle, *Proceedings of 11th International Conference on Strategic Management and its Support by Information Systems 2015* (R. Něměc, and F. Zapletal, eds.), VŠB-Technical University Ostrava, Ostrava, 2015, 329-338.

Fiscal Policy and Economic Growth in EU countries: Evidence from Panel Data with Budget Constraints

Petr Zimčík[1]

Abstract. The aim of this paper is to find evidence whether fiscal policy can affect the economic growth based on the empirical research method. Focus is set on 25 selected member countries of the European Union. Annual government finance data from years 1995 to 2012 are used for an empirical study. Observed indicators are composition and volume of total government expenditures and through budget constrain also tax burden of individual taxes. These indicators are used within endogenous growth model together with capital stock and approximation of human capital. A panel regression with fixed effects is used as an analytic tool. Results show which composition of government expenditures is growth-friendly. Every policy maker must also take into consideration a trade-off between an increase of public expenditures and raising taxes to achieve a balanced government budget. For that reason impacts of various taxes on the economic growth were examined to find out which tax is strongly distortionary. Distrotionary taxes and unproductive govenment expenditues have a negative impact on growth, which is consistent with results in other research articles.

Keywords: fiscal policy, economic growth, EU, panel data

JEL Classification: H20, H50, O40
AMS Classification: 62M10

1 Introduction

There is still no clear consensus about an influence of fiscal policy on economic growth among the economists. Some believe that role of fiscal policy is next to zero. Others claim that a right setting of fiscal policy can positively affect a long-term growth. This general disagreement is one of the reasons why I have decided to write this research paper. The other reason is expressed by a selection of countries which are observed in the paper. All of them are members of the European Union. Although they are in a different state of supra-national integration, all member countries committed themselves to accept a common currency and therefore a common monetary policy of the EU. This fact represents an increasing role of fiscal policy in individual countries.

Fiscal policy has its own unique issues. Policy makers can choose volume of funds which will be provided to individual departments. But their decision is bound by budget constraints. To provide sufficient financial means for their designed policy they have to first collect those funds using taxes. If the policy makers cannot fulfill this equality, deficit and debt arise. This means liabilities for future decision-making. The situation which is now common in advanced economies of the European Union.

For this very reason careful settings of fiscal policy must be taken under consideration. To stimulate long-term economic growth a trade-off between expansionary government expenditure and adequate taxation must be found. Main goal of this research paper is to find impacts of different types of taxation and government expenditures on the economic growth.

2 Fiscal policy and Economic Growth

The construction of an econometric equation used for the research is based on the endogenous model of growth in Barro [1]. In this model fiscal policy can affect level of output and steady-state growth rate. To examine effects of fiscal policy we have to first divide different components of government budget same as in Kneller [2].

On the revenue side there are both distortionary and non-distortionary taxes. Difference between them is that distortionary taxes have effect on investment´s decisions of agents. In this way distortionary taxes can affect rate of the economic growth. In the contrary non-distortionary taxes shouldn´t have any effect on investment or saving rates.

[1] 1Masaryk University, Lipová 41a, Brno, Czech Republic, petr.zimcik@mail.muni.cz

Budget expenditures can be divided to productive or unproductive expenditures. This classification depends on whether they can be included in private production function or not. If they are, then we can take them as productive and they should affect the economic growth. There is of course a large debate about which particular expanses can be labeled as productive and which taxes can be marked as distortionary. Identification of individual types of taxation and spending will be made in the next chapter.

Predictions based on this model suggest that shifting from distortionary taxes towards non-distortionary taxation positively stimulate growth. Shift from productive expenditures to unproductive ones is on the other side growth-inhibiting. Furthermore increase in productive expenditure financed by rise of non-distortionary taxation should lead to enhancing the growth. The increase in unproductive government spending, which will be financed with distortionary taxes, has a significant negative effect on growth. At last there is suggested an unclear effect of financing productive expenditures with distortionary taxes and unproductive expenditure with non-distionary taxes. Some empirical review of impact of fiscal composition on economic activity can be seen in Ardagna [3].

2.1 Budget Constrains

Same as seen in both Kneller [2] and Bleaney [4] I suppose the economic growth $g_{i,t}$ at time t determined via function of conditioning non-fiscal variables $Y_{i,t}$ and a vector of examined fiscal variables $X_{j,t}$. This relation can be formally rewritten into a form of an econometric equation (1).

$$g_{i,t} = \alpha + \sum_{i=1}^{k} \beta_i Y_{i,t} + \sum_{j=1}^{m} \gamma_j X_{j,t} + \varepsilon_{i,t} \quad (1)$$

If we assume that all elements of budget as so overall budget balance are included, sum of fiscal variables will be zero:

$$\sum_{j=1}^{m} X_{j,t} = 0$$

To avoid problem with multicollinearity in estimation of equation (1), at least one fiscal element must be omitted from this sum. The omitted variable is then compensating element within the government's budget constraint. We can rewrite eq. (1) as:

$$g_{i,t} = \alpha + \sum_{i=1}^{k} \beta_i Y_{i,t} + \sum_{j=1}^{m-1} \gamma_j X_{j,t} + \gamma_m X_{m,t} + \varepsilon_{i,t} \quad (2)$$

Then if we omit $X_{m,t}$ to avoid collinearity problem, the identity

$$\sum_{j=1}^{m} X_{j,t} = 0$$

implies, that actual estimated equation is in this form:

$$g_{i,t} = \alpha + \sum_{i=1}^{k} \beta_i Y_{i,t} + \sum_{j=1}^{m-1} (\gamma_j - \gamma_m) X_{j,t} + \varepsilon_{i,t} \quad (3)$$

This changing testing null hypothesis of a zero coefficient of an element $X_{j,t}$. In this setting it is right to test hypothesis $(\gamma_j - \gamma_m) = 0$ rather than just test $\gamma_j = 0$. This is followed by change in interpretation of resultant coefficients. Coefficients of each fiscal category represent an effect of a unit change in an appropriate variable offset by unit change in omitted variable. Under the budget constrains chosen omit variable represents an implicit financing element. Of course a change of selected omitted variable is followed by a change in coefficients.

3 Methodology

For proper estimation we first need to distinguish different types of taxation and expenditures as mentioned in the previous chapter. For division of government expenditures is used Classification of the Functions of Government (COFOG). Categories of COFOG are divided into productive and unproductive based on the definition of an expenditure function. Selection was also inspired by Kneller [2] and you can see it in Table 1. Different types of taxes were also divided based on their structure and tax payers and division in Kneller [2]. You can also find it in Table 1.

Theoretical classification	Functional classification
Distortionary taxation	Personal income tax
	Corporate tax
	Social security contributions
Non-distionary taxation	Taxes on product
	Property tax
Productive expenditures	General public services
	Defence
	Public order and safety
	Economic affairs
	Environmental protection
	Housing and community amenities
	Health
	Education
Unproductive expenditures	Recreation, culture and religion
	Social protection

Table 1 Theoretical and functional classification

Regression equation follows the form of equation (3). Dependant variable is per capita GDP of specific countries. Estimation is first conducted with aggregate elements and then impacts of individual taxes are investigated.[2] In each estimation it is stated which fiscal element is omitted from regression.

4 Data

Annual data from years 1995-2012 are used for this research from 25 selected member countries of the European Union.[3] Explanatory variable is a logarithm of real GDP per capita at chained purchasing power parity in US dollars (*GDP*). Fiscal variables were used as described in Table 1. I used share of total sum of each fiscal element in respect to a nominal GDP. To complete the regression equation, approximation of human and physical capital must be done. As physical capital I used an aggregate data of a gross fixed capital formation in respect to nominal GDP (*Cap*). Share of enrollment of tertiary students on whole population was used to approximate human capital (*HCap*). Budget balance (*Budget*) and other taxes (O*ther*) were also added into every regression to satisfy budget constraints.

GDP data was taken from Penn World Table 8.1 as total real GDP of each country in mil. US $ and then with population statistic calculated to per capita GDP. All fiscal variables come from the government finance data of Eurostat.[4] Share of the fixed capital formation also comes from Eurostat data. Tertiary school enrollment data originated from Eurostat and are filled in with OECD statistics.

4.1 Stationarity testing

To achive un-biased results in regression with panel data I had to make sure that all variables have stationary time-series. Variables were tested with and ADF test by Levin [5]. Testing has showed following results. Most of variables indicated presence of unit-root in their time-series. Only variables which showed stationarity were B*udget*, *Productive* and *Corporate tax*. To resolve this problem all non-stationary variables were replaced by their first differences.[5] Additional ADF tests of these differences show that variables exhibit stationary data. Working with differences instead of original variables changes the interpretation of coefficients.

[2] Only effects of individual types of taxes are presented becuase in testing each category of government expenditures, all coeficients were statistically insignificant. So there was no point following them any further.
[3] From 28 member states of EU only Bulgaria, Croatia and Slovenia were missed out due to data unavailability.
[4] Only handicap in this data was missing values for property tax in Finland and Malta.
[5] Despite stationarity of corporate taxes, differences were used in regression to achieve consistent coefficients result with other individual taxes.

5 Empirical results

In this section of the paper are presented results of regressions based on equation (3) and data described in previous chapter. First set of regressions show aggregate fiscal variables as described as in chapter 2. In each regression one of this variable is omitted and impact of residual variables is examined. You can see results in table 2.[6] Dependant variable is d_log GDP per capita.

Omitted variable	Non-distor taxes	Unproductive	Non-distor taxes and Unproductive
Constant	−0.05745	−0.10290**	−0.08312*
	(−1.253)	(−2.160)	(−1.768)
d_Cap	0.00484***	0.00878***	0.00836***
	(4.108)	(8.667)	(8.633)
d_HCap	4.11513***	3.63176***	4.01230***
	(3.822)	(3.162)	(3.619)
Productive	0.00323*	0.00507***	0.00425**
	(1.862)	(2.802)	(2.394)
d_Unproductive	−0.01510***		
	(−4.969)		
d_Distor taxes	−0.00393	−0.00453	−0.00412
	(−1.448)	(−1.566)	(−1.470)
d_Non-distor taxes		−0.00668	
		(0.00413)	
d_Other	−0.01290	−0.01483	−0.01262
	(−1.114)	(−1.215)	(−1.058)
Budget	0.00524***	0.00680***	0.00641***
	(4.109)	(5.119)	(4.957)
R^2	0.448	0.427	0.408
F-statistic	8.78	7.91	7.73
Number of obs.	367	338	367
P-value (F)	5.40e-28	1.93e-23	5.98e-24

Table 2 Regression of aggregate fiscal variables

Method used for calculation coefficients was the panel regression with fixed-effects. This method is particularly efficient, when individual units have different features. You can then anticipate different intercept for all units. Hausmann test was executed for all three regressions and in all of them the hypothesis of common intercept was rejected. Because of that panel regression with fixed-effects is more efficient and un-biased than OLS or random effect regression.

For the interpretation of the estimates is more important the sign of a coefficient rather than the value of a coefficient itself. Negative coefficient means a negative effect on growth and vice-versa. Starting with non-fiscal explanatory variables, approximations of physical and human capital seem to be significant and positive in all three regressions. This result is consistent with theory and model in Barro [1].

The coefficients of the fiscal variables in first regression show that an increase in unproductive government expenditure financed by non-distortionary taxes is against economic growth. This finding is different compared to the theory in Barro [1], which predicted a negligible effect. In both Bleaney [4] and Kneller [2] unproductive expenditures show slightly negative but insignificant effect on the growth while funded by non-distortionary taxes.

For productive expenditures is the relation inverse. Second regression shows that an increase in productive expenditures compensated by decreasing unproductive is strongly growth-enhancing. Third regression shows an increase in productive expenditures, which is financed partly by a rise in non-distortionary taxes and partly by cuts in unproductive spending, positively stimulates growth. Effect of other taxes seems to be insignificant. Empirical results in table 3 are mostly[7] consistent with effects predicted by model in chapter 2 and with conclusions in Bleaney [4] and Kneller [2].

In next part I focus on impacts of individual taxes. Results can be seen in table 3.

[6] In brackets are listed appropriate t-statistics to each variable. Number of stars next to coefficient represents significance level. For (*) it´s significance level 10%, for (**) it is 5% and for (***) it´s 1%.
[7] Unproductive expenditures financed by non-distortionary taxation make an exception.

Omitted variable	Non-distor taxes	Unproductive	Non-distor taxes and Unproductive
Constant	−0.06381	−0.10491	−0.08462*
	(−1.413)	(−2.256)	(−1.858)
d_Cap	0.00469***	0.00750***	0.00702***
	(4.004)	(7.047)	(7.000)
d_HCap	3.64734***	3.21883***	3.42190***
	(3.419)	(2.862)	(3.167)
Productive	0.00346**	0.00513***	0.00429**
	(2.027)	(2.902)	(2.490)
d_Unproductive	−0.01171***		
	(−3.707)		
d_Income_tax	−0.01359***	−0.01280***	−0.01550***
	(−3.051)	(−2.593)	(−3.443)
d_Corporate_tax	0.00959**	0.01175**	0.01374***
	(2.088)	(2.331)	(3.032)
d_Social_contibution	−0.01148**	−0.01392***	−0.01546***
	(−2.429)	(−2.895)	(−3.307)
d_Poduction_tax		−0.00539	
		−1.292	
d_Property_tax		−0.04218*	
		(−1.815)	
d_Other	−0.01438	−0.01501	−0.01447
	(−1.259)	(−1.252)	(−1.249)
Budget	0.00530***	0.00651***	0.00617***
	(4.224)	(5.007)	(4.917)
R^2	0.505	0.479	0.448
F-statistic	8.93	8.12	8.48
Number of obs.	367	338	367
P-value (F)	1.48e-29	1.38e-25	1.56e-27

Table 3 Regression of individual tax variables

Coefficients of non-fiscal variables are very similar (sign, significance, value) to their counter-parts in table 2. Impacts of productive and unproductive expenditures are also similar comparing to the previous case. Model in chapter 2 predicted negative impacts of distortionary taxes. Results in table 3 show this significant negative impact for individual taxes and social contribution with exception of corporate taxes. This applies for all three regressions. Non-distionary taxes show a significant negative impact only in case of property tax. Taxes on production don´t show any significant impact.

In both tables relevance of budget constraints are confirmed. Variable *budget* is in all cases significant and positive. This result is similar with conclusions in Bleaney [4].

As a method for estimation was used fixed-effects panel regression. Hausmann test confirmed again that thi method produced most efficient and un-biased estimates.

6 Conclusion

Main research goal in this paper was to find impacts of different types of taxation and government expenditures on the economic growth. To analyze these impacts, fiscal variables were divided into four groups - productive and unproductive government expenditures and distortionary, non-distortionary taxes. Fixed-effects panel regression was used as a method to analyze these impacts. With budget-constraints one fiscal variable was always omitted so it won´t cause problems with multicollinearity.

Empirical results were mostly consistent with the theoretical model and relevant scientific literature. Productive expenditures were only fiscal variable to have significant positive impacts on GDP growth rate. Unproductive expenditures show exactly opposite relation. To examine effect of individual taxes change, next set of regressions were constructed. It shows that personal income taxes, social contribution and property tax are also growth-adverse. Interesting finding was surprising growth-supporting tendency of corporate taxes. Impacts of production tax and other taxes concluded to be insignificant.

Acknowledgements

This paper was created as a part of the specific research project no. MUNI/A/1223/2014 at Masaryk University.

References

[1] Barro, R.: Government spending in a simple model of endogenous growth. *Journal of Political Economy* **98** (1990), 103–117.
[2] Kneller, R., Bleaney, M., and Gemell, N.: Fiscal Policy and growth: evidence from OECD countries. *Journal of Public Economics* **74** (1999), 171-190.
[3] Ardanga, S.: Fiscal Policy Compostition, Public Debt, and Economic Activity. *Public Choice* **109** (2001), 301-325.
[4] Bleaney, M., Gemell, N., and Kneller, R.: Testing the Endogenous Growth Model: Public Expenditure, Taxation, and Growth ove the Long Run. *The Canadian Journey of Economics* **34** (2001), 36-57.
[5] Levin, A., Lin, C., and Chu, C.: Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics* **108** (2002), 1–24.

An Arrival and Departure Times Synchronization Problem

Karel Zimmermann[1]

Abstract. We consider transportation problems with transit points, in which the latest arrival times at the transit points and the earliest departure times from the transtit points are appropriately coordinated. The travelling times between departure points and transit points as well as the travellng times between the transit points and final destinations are given. The problem leads to solution of a system of equations and/or inequalities with so called $(\max,+)$ and $(\min,+)$-linear functions. Solvability as well as some other theoretical properties of the systems are studied and a solution method of the systems is proposed. Perspectives of future research in the area are outlined.

Keywords: $(\max,+)$-linear equation systems, $(\min,+)$-linear equation systems, $(\max,+)$-eigenvector space.

JEL classification: C44
AMS classification: 90B35, 90C30

1 Introduction.

Algebraic structures witha pair of operations $(\oplus,\otimes)$, which play a similar role like the addition and multiplication in the classic linear algebra appeared in the literature probably for the first time at the beginning of the sixties of the last century. Since that time, paper with various pairs of such operations have appeared, e.g. $(\max,+)$, $(\min,.)$, $(\max,\min)$. The authors consider usually Cartesian prooduct of a finite numbers of such structures and investigate properties of systems of $(\oplus,\otimes)$-linear equations and inequalities. Theoretical results were appied in various areas of operations research as e.g. machine time scheduling, discrete event systems, network reliability and others. In the present paper we investigate equation systems with $(\max,+)$-linear functions on one side and $(\min,+)$-linear function on the other side of the equations. Obtained theoretical results can be used for coordinating transportation times or machine release times in some operations research problems.

The following example shows a motivation for the research presented in this contribution.

Example 1.

We assume that passengers should be transported from places $P_j,\ j\in J\equiv\{1,\dots,n\}$ to destinations $D_k,\ k\in K\equiv\{1,\dots,n\}$ via transit points $T_i,\ i\in I\equiv\{1,\dots,m\}$. The passengers must change for another means of transport at the transit points. The time distances between T_i and P_j are equal to $a_{ij}>0$ and the time distances between T_i and D_k are equal to $c_{ik}>0$. The components of $x=(x_1,\dots,x_n)$ denote departure times at $P_j,\ j\in J$ and the components of $y=(y_1,\dots,y_n)$ denote the arrival times to destinations $D_k,\ k\in K$. We assume that the passengers can continue their journey from T_i to D_k only after all passengers who need to continue via T_i had arrived to T_i. If the passengers from P_j do not continue through T_i, we set $a_{ij}=-\infty$. The departure times z_k from T_i to D_k are under our assumption equal to $z_{ik}=y_k-c_{ik}$. The earliest feasible departure times from a fixed transit point T_i to a fixed D_k must satisfy the inequality

$$\max_{j\in J}(x_j+a_{ij})\le z_{ik}.$$

The departure times $z_{ik},\ k\in K$ coordinated with the latest arrival times to T_i must satisfy the relations

$$\max_{j\in J}(x_j+a_{ij})\le\min_{k\in K}z_{ik},$$

[1]Charles University, Faculty of Mathematics and Physics, Malostranske namesti 25, Prague 1, Czech Republic, karel.zimmermann@mff.cuni.cz

or, using the expressions of z_{ik}

$$\max_{j\in J}(x_j + a_{ij}) \le \min_{k\in K}(y_k - c_{ik})$$

We may investigate the question whether there exist **x,y** satisfying the equation system

$$\max_{j\in J}(x_j + a_{ij}) = \min_{k\in K}(y_k - c_{ik}),\ i \in I.$$

If **x,y** satisfy the equation system, no unnecessary delays between arrivals and departures in the transit points will occur. If **x,y** satisfy the equation system, all passengers arrive to their transit points sufficiently early so that they can catch their departures from transit points.

2 Notations, Preliminary Results and Problem Formulation.

Let us introduce the following notations:
R is the set of real numbers, $I = \{1,\ \dots\ m\}$, $J = \{1,\ \dots\ n\}$, **A, B** are matrices with elements $a_{ij}, b_{ij} \in R\ \forall i \in I,\ j \in J$

$$(\mathbf{A}o\mathbf{x})_i = \max_{j\in J}(a_{ij} + x_j),\ (\mathbf{B}o'\ \mathbf{y})_i = \min_{j\in J}(b_{ij} + y_j),\ i \in I,$$

$\mathbf{x}^T = (x_1,\ \dots,\ x_n)$, $\mathbf{y}^T = (y_1,\ \dots,\ y_n) \in R^n$, $\mathbf{A}o\mathbf{x} = ((Aox)_1,\ \dots,\ (Aox)_n)^T$, $\mathbf{B}o'\mathbf{y} = ((Bo'y_1,\ \dots,\ (Bo'y)_n))^T)$ (superscript T denotes transposition). In the sequel we will study equation system

$$\max_{j\in J}(a_{ij} + x_j) = \min_{j\in J}(b_{ij} + y_j),\ i \in I \tag{1}$$

or using the matrix-vector notation

$$\mathbf{A}o\mathbf{x} = \mathbf{B}o'\mathbf{y} \tag{2}$$

The set of all solutions $\mathbf{x}^T = (x_1,\ \dots,\ x_n)$, $\mathbf{y}^T = (y_1,\ \dots,\ y_n)$ of system (1) (or (2)) will be denoted $M(\mathbf{b,A,B})$. Our aim is to investigate properties of the set $M(\mathbf{b,A,B})$. For this purpose, we will instroduce some preliminary definitions and results, the proofs of which can be found in the literature (see [1], [2], [3], [4]).

Let

$$M_1(\mathbf{b,A}) = \{\mathbf{x} \in R^n \mid \mathbf{A}o\mathbf{x} = \mathbf{b}\},\ M_2(\mathbf{b,B}) = (\mathbf{y} \in R^n \mid \mathbf{B}o'\mathbf{y} = \mathbf{b}),$$
$$x(b,A)_j = \min_{k\in I}(b_k - a_{kj}),\ \hat{y}_j(b,B) = \max_{k\in I}(b_k - b_{kj}),\ j \in J, \tag{3}$$
$$\mathbf{x(b,A)} = (x_1(b,A),\ \dots, x_n(b,A))^T,\ \hat{\mathbf{y}}(\mathbf{b,B}) = (\hat{y}_1(b,B),\ \dots, \hat{y}_n(b,B))^T.$$

Definition 1.

Let $D \subseteq R^p$, $\overline{d} \in D$. Then $\overline{\mathbf{d}}$ is the maximum (minimum) element of set D if $\mathbf{d} \le \overline{\mathbf{d}}$ ($\mathbf{d} \ge \overline{\mathbf{d}}$) for any $\mathbf{d} \in D$.

Theorem 1. *([1], [2], [4])*

(a) If $M_1(\boldsymbol{b,A}) \neq \emptyset$, then $\boldsymbol{x(b,A)}$ is the maximum element of $M_1(\boldsymbol{b,A})$.

(b) If $M_2(\boldsymbol{b,B}) \neq \emptyset$, then $\hat{y}(\boldsymbol{b,B})$ is the minimum element of $M_2(\boldsymbol{b,B})$.

Definition 2.

Let $\mathbf{b} \in R^m,\ \lambda \in R,\ \mathbf{H} = ||h_{ik}||,\ i,\ k \in I$ be such that $\mathbf{H}o\mathbf{b} = \lambda o\mathbf{b}$. Then λ is called $(\max,+)$-eigenvalue and $\mathbf{b}$ $(\max,+)$- eigenvector of $\mathbf{H}$ corresponding to λ.

Since we will deal in what follows only with eigenvalues and eigenvectors in the sense of Definition 2, we will omit further the prefix $(\max,+)$.

Theorem 2. *([1], [2])*

Let $\boldsymbol{H}$ *be a real* $(m \times m)$*-matrix,* $\lambda(\boldsymbol{H})$ *be an eigenvalue of* $\boldsymbol{H}$*,* $\boldsymbol{H}^p = \boldsymbol{H} o \boldsymbol{H} \ \ldots, \ o\boldsymbol{H}$ *(*$p-$*times), let* $h_{ik}^{(p)}$ *be elements of* $\boldsymbol{H}^p$*. Then*

$$\lambda(\boldsymbol{H}) = \max_{p \in I} \max_{i \in I} (h_{ii}^{(p)}/p),$$

where we set $h_{ii}^{(1)} = h_{ii}$*.* $\lambda(\boldsymbol{H})$ *is the unique eigenvalue of* $\boldsymbol{H}$*.*

Let for any $\alpha \in R, b \in R^m$, $\alpha o \mathbf{b} = (\alpha + b_1, \ \ldots, \ \alpha + b_m)^T$ and for any pair of matrices **C, D** of equal size with elelments $c_{ik}, \ d_{ik}$, $(\mathbf{C} \oplus \mathbf{D})_{ik} = \max(c_{ik}, d_{ik})$.

Theorem 3. *([1] Theorem 4.2.4, p. 76)*

Let $\boldsymbol{H}_\lambda = (-\lambda)$ *o* $\boldsymbol{H}$*.*

$$\Gamma \boldsymbol{H}_\lambda) = \boldsymbol{H}_\lambda \oplus \boldsymbol{H}_\lambda^2 \oplus \ \ldots \ \oplus \ \boldsymbol{H}_\lambda^m.$$

Then every column of $\Gamma(\boldsymbol{H}_\lambda)$ *with zero diagonal entry is an eigenvector corresponding to* $\lambda(\boldsymbol{H})$

Remark 1.

Let us note that the problem of Example 1 may be transformed to the form (1) by setting $b_{ik} = -c_{ik}$.

Remark 2.

Note that Lemma 3 holds because all elements of **H** are under our assumptions finite. A more general version with infinite elements occuring in the matrices and and also with other algebraic structures can be found in [1], [2].

Eigenvalues and eigenvectors of a matrix are applied in the literature to obtain so called steady state solutions. If **x** denotes vector of departure or release times of a cyclicaly repeating activities, then the authors look for such release time vector **x** that the next cycle can begin at times given by $\lambda o x = \mathbf{x} \ + \lambda$ with components $x_i + \lambda$. By analogy, we will look for solutions (**x,y**) of system (1) such that $\mathbf{y} = \ \mathbf{x} \ + \lambda$, investigate properties existence conditions of such solutions.

Remark 3.

If in the motivating Example 1 the departure and arrival of the transportation is cyclically repeated in such a way that a new cycle can take place after all passengers will have reached their destinations in the preceding cycle, i.e. at times $y_1, \ \ldots, \ y_n$, then a steady state solution (**x,** $\mathbf{x} \ + \lambda$) seems to be appropriate, because it makes possible that the departures take place in a regular time intervals having a duration λ.

3 Solution Method.

Let us assume that $\mathbf{x} \in M_1(\mathbf{b},\mathbf{A}), \ \mathbf{y} \ \in M_2(\mathbf{b},\mathbf{B})$ for some $\mathbf{b} \in R^m$. Then

$$\mathbf{A} o \mathbf{x} \ = \mathbf{B} o' \mathbf{y} = \ \mathbf{b}. \tag{4}$$

and therefore according to Lemma 1

$$\mathbf{A} o \mathbf{x}(\mathbf{b},\mathbf{A}) = \mathbf{B} o' \hat{\mathbf{y}}(\mathbf{b},\mathbf{B}) = \mathbf{b}. \tag{5}$$

or componentwise

$$\max_{j \in J} [a_{ij} + \max_{k \in I} (b_k - b_{kj}) - \lambda] = b_i, \ i \in I. \tag{6}$$

so that

$$\max_{j \in J} \max_{k \in J} (a_{ij} - b_{kj} + b_k) - \lambda = b_i, \ i \in I. \tag{7}$$

and by interchaning the max-operators

$$\max_{k \in J} (\max_{j \in J} (a_{ij} - b_{kj}) + b_k) - \lambda = b_i, \ i \in I. \tag{8}$$

Let us set

$$h_{ik} = \max_{j \in J}(a_{ij} - b_{kj}),\ \forall i \in I,\ k \in I. \tag{9}$$

Finally we obtain

$$\max_{k \in J}(h_{ik} + b_k) = \lambda + b_i,\ i \in I, \tag{10}$$

or

$$Hob = \lambda ob. \tag{11}$$

Therefore b and λ satisfying relations (5), (6) must be the eigenvector and the correspondiing eigenvalue of matrix H, the elements of which are defined by (9). The following small numerical example demonstrates the theoretical results obtained above.

Example 2.

Let matrices A, B be given as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 4 & 5 \end{pmatrix}$$

Then

$$H = \begin{pmatrix} 1 & -3 \\ 4 & 0 \end{pmatrix}$$

We will compute $\lambda(\mathbf{H})$ and the corresponding eigenvector $\mathbf{b}$ using Lemma 2 , 3. We obtain:

$\lambda(\mathbf{H}) = 1$ and

$$\mathbf{H}_{\lambda(H)} = \begin{pmatrix} 0 & -4 \\ 3 & -1 \end{pmatrix}$$

We can easily verify that in this case $\Gamma(\mathbf{H}_{\lambda(H)}) = \mathbf{H}_{\lambda(H)}$ so that using Lemma 3 we obtain that the eigenvector corresponding to $\lambda(\mathbf{H})$ is $b = (0,3)^T$. Using this eigenvector, we compute $\hat{\mathbf{y}}(\mathbf{b},\mathbf{B}) = (-1,-1)^T$, $\mathbf{x} = \hat{\mathbf{y}}(\mathbf{b},\mathbf{B}) - 1 = (-2,-2)^T$. It can be easily verified by a direct computation that

$$\mathbf{A}\ o\mathbf{x}\ = \mathbf{B}o'(\hat{\mathbf{y}}(\mathbf{b},\mathbf{B})) = \mathbf{b},$$

so that $\mathbf{x},\ \hat{\mathbf{y}}(\mathbf{b},\mathbf{B})) = ((-2,-2)^T, (0,-1)^T) \in M(\mathbf{b},\mathbf{A},\mathbf{B})$.

Summarizing the computations in Example 2 we can propose the following procedure for investigating the solvability of system (1):

1 Choose $\mathbf{b}$ such that $\mathbf{B}o'\hat{\mathbf{y}}(\mathbf{b},\mathbf{B}) = \mathbf{b}$;

2 Compute matrix $\mathbf{H}$, find $\lambda(H)$, $\mathbf{H}_{\lambda(H)}$ and $\Gamma(\mathbf{H}_{\lambda(H)})$;

3 If $\mathbf{b}$ is an eigenvector of $\mathbf{H}$ (i.e. if $\mathbf{b}$ is in the eigenspace of $\mathbf{H}$), then $(\hat{\mathbf{y}}(\mathbf{b},\mathbf{B}) - \lambda(H), \hat{\mathbf{y}}(\mathbf{b},\mathbf{B}))$ solves system 1;

4 $(\hat{y}(b,B) - \lambda(H), \hat{y}(b,B))$ is a lower steady state approximation to satisfying the inequality system

$$\mathbf{A}\ o\mathbf{x} \leq \mathbf{B}o'\hat{\mathbf{y}}(b,\mathbf{B}).$$

Remark 4.

Let us note that we could proceed symmetrically and begin with the maximal solution $\mathbf{x}(\mathbf{b},\mathbf{A})$ of system $\mathbf{A}o'\mathbf{x} = \mathbf{b}$ and look for an appropriate $\tilde{\mathbf{b}}$ and λ such that $x(\tilde{\mathbf{b}}, A) = \mathbf{y} + \lambda$. We would find simillarly another matrix $\tilde{\mathbf{H}}$, and the λ would be a $(\min,+)$-eigenvalue of $\tilde{\mathbf{H}}$, i.e. $\lambda = \lambda'(\tilde{\mathbf{H}})$. The appropriate $\tilde{\mathbf{b}}$ would be any $(\min,+)$-eigenvector corresponding to $\lambda'(\tilde{\mathbf{H}})$, i.e. $\tilde{\mathbf{b}}$ would satisfy relation $\tilde{\mathbf{H}}o'\tilde{\mathbf{b}} = \lambda(\tilde{\mathbf{H}})o'\tilde{\mathbf{b}}$. After some calculations we would obtain $\tilde{\lambda}(\tilde{\mathbf{H}}) = -1$,

$$\tilde{\mathbf{H}}_{\lambda(\tilde{\mathbf{H}})} = \begin{pmatrix} 0 & -4 \\ 3 & -1 \end{pmatrix}, \quad \tilde{\Gamma}(\tilde{H}_{\lambda(\tilde{H})}) = \begin{pmatrix} 0 & -3 \\ 3 & 0 \end{pmatrix}$$

For $(\min,+)$-eigenvector $\tilde{\mathbf{b}} = (0,3)^T$ we have a solution: $(\tilde{\mathbf{x}}, \hat{\mathbf{y}}(\mathbf{b},\mathbf{B})) = ((0,0)^T, (-1,-1)^T)$.

Remark 5.

Let us note that the above results remain unchanged, if the number of components of vectors $\mathbf{x}$ and $\mathbf{y}$ in system (1) is different, we can appropriately extend the matrix with a smaller number of columns with infinite elements ∞ (if matrix $\mathbf{B}$ is extended) or $-\infty$ (in case we extend matrix $\mathbf{A}$). Note that we can replace elements ∞, $-\infty$ with sufficienty large positive or negative numbers to avoid calculations with the infinite elements.

Remark 6.

Let us note that if the eigenvalue $\lambda(\mathbf{H})$ is equal to zero and $\mathbf{b}$ is in the range of $\mathbf{B}$, then $\mathbf{x} = \hat{\mathbf{y}}(\mathbf{b},\mathbf{B})$ so that $(\hat{\mathbf{y}}(\mathbf{b},\mathbf{B}), \hat{\mathbf{y}}(\mathbf{b},\mathbf{B}))$ solves system (1) if $\mathbf{b}$ is an eigenvector of matrix $\mathbf{H}$ corresponding to $\lambda(\mathbf{H})$.

Remark 7.

Let us note that the results established above hold only under the assumption made that the system is solvable and we verify the existence of a "steady state solution" $\hat{\mathbf{y}}(b,\mathbf{B}) = \mathbf{x} + \lambda$. The obtained results say only that if such solution exists, then λ must be an eigenvalue of $\mathbf{H}$ and the right hand side b must be a corresponding eigenvector of H.

Remark 8.

Modifications and generalizations of the problems considered in this paper may be a subject of further research. If we find out that the steady state solution for a given $\mathbf{b}$ does not exist, one of the further research steps could be to find such appropriate changes of entries (i.e. of the right hand sides or matrix elements) of the problem that the steady state solution exists. We could consider problems with uncertain entries a_{ij}, b_{ij} (e.g. interval, fuzzy or probabilistic entries) or entries depending on parameters. A general solvability theory for the systems can be investigated.

Acknowledgment. Supported by the grant No. GAČR #14-02424S of the Czech Science Foundation.

References

[1] Butkovič, P.: Max-linear Systems: Theory and Algorithms, Monographs in Mathematics, Springer Verlag 2010, 271 p.

[2] Cuninghame-Green, R.A.: Minimax Algebra, Lecture Notes in Economics and Mathematical Systems, 166, Springer Verlag, Berlin, 1979.

[3] Krivulin,N.K.: Methods of Idempotent Algebra in Modelling nd Analysis of Complex Systems, S.-Peterburg 2009,255 p. (in Russian).

[4] Vorobjov, N.N.: Extremal Algebra of positive Matrices, Datenverarbeitung und Kybernetik, 3, 1967, pp. 39-71 (in Russian).

Evaluation of Economic Faculties Using DEA

Miroslav Žižka[1]

Abstract. This article aims to assess the possibility of using Data Envelopment Analysis (DEA) for the evaluation of faculties focused on teaching and research in economic fields. The analysis involved 25 faculties, both public and private. Only research oriented faculties were included into the analysis to ensure homogeneity of the investigated file. In the first part of the research, relevant indicators of inputs and outputs for evaluating teaching efficiency, research efficiency and overall efficiency were searched. The data were obtained from annual reports of faculties and other public sources. In the next part of the research, several DEA models with different inputs and outputs were created. Based on the analyses, the article suggests appropriate models for the evaluation of individual activities and overall efficiency.

Keywords: Data Envelopment Analysis, BCC model, universities, teaching efficiency, research efficiency, overall efficiency, super efficiency.

JEL Classification: C61, I23
AMS Classification: 90C90

1 Introduction

The evaluation of university's performance is one of the current issues that the Ministry of Education and Accreditation Committee deal with. Compared to businesses entities, universities show numerous specificities regarding their mission and goals. The article deals with the possibilities of using Data Envelopment Analysis (hereafter DEA) for the evaluation of the faculties of economics in the Czech Republic. This method was chosen because it is quite widely used in measuring the efficiency and performance of public institutions such as schools or hospitals. The aim of this paper is to propose a suitable model for evaluating the efficiency of the faculties of economics at research universities in the Czech Republic, and test it on real data. The article, however, does not have ambitions to objectively assess the performance of economic faculties. Such a task would require the use of more extensive data base for a longer period of time and would exceed spatial possibilities of a conference paper.

2 Literary overview

DEA model is a tool to assess the efficiency or performance of the homogeneous decision-making units (hereafter DMU). DEA typically evaluates DMUs using multiple inputs and outputs. Another significant feature of the DEA method is that the inputs and outputs are assigned different weights, for each unit separately to maximize unit efficiency [6]. DEA divides the file of examined DMUs in two parts - efficient and inefficient units. The best units are located on the so-called efficient frontier. The oldest model, called by authors' last names as CCR (Charnes, Cooper, Rhodes), was published in 1978. This model was based on the assumption of constant returns to scale. Later it was extended by the assumption of variable returns to scale. This modified model was designed by Banker, Charnes and Cooper in 1984, hence the designation of the BCC [3]. Both models can be oriented either at inputs or outputs. There are other variants of DEA models; however, their outline would go beyond the intention of this article. Those who are interested can find them, e.g. in the book [8].

Extensive research of DEA models in evaluating universities in Canada, Australia, UK, Germany and Spain was done by Rosenmayer [13]. He notes that a university evaluation as a non-profit organization is a specific matter because universities have different goals (some are more focused on preparing graduates for the labor market, while others focus primarily on basic research and others on further co-operation with industry). Kuah and Wong [10] confirm the difficulty of measuring the efficiency of universities because first, it is very difficult to express their inputs and outputs in money and second, they create different outputs using different inputs. In the educational field, the main objectives of any university are to increase the level of education of students and in the research, findings of new knowledge and its transfer into practice. As rough indicators, numbers of graduates can be used or a number of published articles and granted patents. However, as stated in [13], these indicators do not always correspond with reality because it does not affect the quality of outputs. Selected inputs and outputs used in the evaluation of university are stated in Table 1. In the literature, there are a number of other studies on the use of DEA method in evaluating the efficiency of universities, see e.g. [1], [9], [10] and [11]. Bealey created a model for simultaneous comparison of teaching and research efficiency of universities. The principle of the

[1] Technical University of Liberec, Faculty of Economics, Studentska 2, Liberec 8, miroslav.zizka@tul.cz.

model lies in the separation of indicators into two groups – on measures associated with teaching and research. For each activity, the rate of efficiency is calculated and then, the overall efficiency [1]. The specification of indicators characterizing teaching and research is easier on the output than on the input side. The educational output mainly includes graduates. The output of the research is mainly publications. On the input side, the situation is more complicated. Distribution of working capacity and related staff costs between the two activities is not easy. The Bealey's model, therefore, looks for such a split of inputs so as to achieve maximum efficiency in teaching. Subsequently, the model cross-seeks an optimal proportion of inputs in teaching to maximize efficiency in research. Kuah and Wong applied that model to the revised set of indicators. In their model, they included qualification level of academic staff on a scale of 1 (assistant with a master's degree) to 4 (professor) [10]. Nazarko and Šaparauskas first performed a correlation analysis of 5 potential inputs, 8 outputs and 2 variables characterizing the environment on a sample of 19 Polish technological universities. On the input side, they found the strongest correlation with variable of government subsidies. This value was, therefore, used as an input to a subsequent analysis of DEA. Universities were evaluated using the output-oriented CCR model because universities have only little effect on the size of subsidies [11]. Table 1 shows that the indicators used to evaluate universities are very different. Some indicators can be considered inputs and outputs. For example, the amount of grants received can be understood as an output in terms of skills and abilities of academic workers to write a project and succeed in a public competition. At the same time, grants can be considered a financial resource that should be reflected in the output of publications, patents and other research results (RIV points, RIV = Information Register of R & D Results).

Author(s), Country, Year	**Inputs**	**Outputs**
Beasley, England, 1995	General expenditures Equipment expenditure Research income	Number of undergraduates Number of taught postgraduates Number of research postgraduates Research income Assessment of research
Johnes, England, 2006	Total number of FTE undergraduate students Total number of FTE postgraduate students Total number of full-time academic staff Total depreciation and interest payable Total expenditure on libraries and IS Expenditure on central administration	Total number of first degrees awarded weighted by degree classification Total number of higher degrees awarded Value of the recurrent grant for research
Kuah and Wong, Malaysia, 2011	**Teaching efficiency:** Number of academic staff Number of taught course students Average students' qualifications University expenditures **Research efficiency:** University expenditures Number of research staffs Average research staffs' qualifications Number of research students Research grants	 Number of graduates from taught courses Average graduates' results Graduation rate (%) Graduates' employment rate (%) Number of graduates from research Number of publications Number of awards Number of intellectual properties
Nazarko and Šaparauskas, Poland, 2014	Government budget subsidy	Weighted number of full-time students Weighted number of full-time PhD students Employer hiring preferences Assessment of scholarly achievements

Table 1 Possible inputs and outputs used when evaluating universities

3 Data and methodology

The subjects of evaluation are faculties at universities that teach and research in economics, or in socio-economic fields. Since the DEA requires homogeneous units, only universities were included in the analysis. These include 22 faculties at public universities and 3 private schools that are declared as universities (research focus and doctoral studies). The research did not include professionally-oriented public and private colleges as they have different objectives. The data were collected for the year 2012 which provides maximum information available (some universities still have not published their annual reports for 2013 and also the evaluation of research results for the year 2013 is not yet known). The information on the number of students and academic staff structure were taken from the annual report of individual universities. These annual reports are handled in a unified struc-

ture according to the Ministry of Education, which ensures good comparability of information. Data on the number of graduates from the given faculties were found in the annual reports of faculties since these data are not presented in the annual reports of universities. Scoring of research results for the last five-year evaluated period (2008-12) was taken from the website database [14]. Financial volume of received research grants was elicited from the database [2]. The employment of graduates was found in the database of Education Policy Centre [12]. As an indicator, the adjunct to standardized unemployment rate of graduates at all levels of study in April 2012 was used.

The research process can be divided into the following steps:

1. **Creating a list of evaluated faculties** – the index of universities listed on the website of Ministry of Education and an overview of research organizations and their organizational units in the Research and Development and Innovation Information System of the Czech Republic were used.

2. **Definition of evaluated activities** – the efficiency of faculties was evaluated in three areas: teaching, research and overall. Assigning indicators to each model is clear from Tables 2 and 3. For each activity, technical efficiency scores were calculated (T_e – teaching efficiency, R_e – research efficiency, O_e – overall efficiency).

3. **Create a database of inputs and outputs** – data on the number of students, the number of academic staff in the structure of assistants and lecturers, assistant professors, associate professors and professors, the number of graduates, the number of RIV points, standardized unemployment rate of all the graduates and the amount of research grants were obtained from the sources mentioned above. The number of academic staff was measured using a composite indicator where the number of assistants and lecturers was assigned a weight 1, the number of assistant professors weight 2, associate professors weight 3 and professors weight 4. The source data set is too large and can be provided by the author via e-mail on request.

4. **Analysis of dependencies between inputs and outputs** – Pearson correlation coefficients among the indicators on the input side, on the output side and mutually between inputs and outputs were calculated. It is recommended [5] that input (or output) were independent.

5. **Attribution of human resources for teaching and research activities** – while most of the indicators can be assigned to teaching or research activities, the situation is more complicated concerning the number of academic staff. Academicians spend part of their capacity with both teaching and research. Beasley [1] proposes a method to determine the optimal split between the two activities. However, when testing Beasley´s model on real data from the Czech Republic, there was found a tendency to assign a large majority of human capacities to research, which does not reflect reality. In the model, there was then used a simplistic assumption that 50% of human resources (capacities) is dedicated to teaching and 50% to research.

6. **Assessment of the file size in terms of sufficient discrimination of units** – model should clearly specify a limited number of efficient units. With respect to a number of units examined (N) and the number of inputs (m) and outputs (r) the relation (1) should be fulfilled, see [4]. For most models, the number of units is sufficient, except for the models No. 3 and 12 indicated in Table 3. In these two models, the number of inputs is 4 – and outputs – 5. In this case, the max{4 . 5; 3(4 + 5)} = 27, which is more than 25 evaluated faculties. For this reason, number of inputs and outputs in the models No. 6 and 9 was reduced.

$$N \geq \max\{m.r; 3(m+r)\} \tag{1}$$

7. **Detection of returns to scale** – regression analysis to estimate the parameters α and β of logarithmized modified Cobb-Douglas production function was used, see equation (2) and (3). The dependent variable in the first model was graduates, independent variables academics and students. In the second model, the dependent variable was RIV points and independent variables academics and doctoral students. The sum of the parameters α and β in the first case is 1.19, this means that the production function presents increasing returns to scale. In the second case, the sum of α and β is 0.88, which represents decreasing returns to scale. To evaluate the efficiency of the faculties, there was primarily chosen an output oriented BCC model. The orientation of the model at outputs was chosen because the most significant inputs (number of students) may be affected in a very limited way (there are limits of the ministry). For comparison, there are also given the CCR model outputs with the assumption of constant returns to scale.

$$Y_7 = 0.065 X_3^{0.07} X_5^{1.12}, \text{ R}^2 = 0.90, \text{ F-Ratio} = 98.31, \text{ p-value} < 0.0001 \tag{2}$$

$$Y_5 = 71.91 X_3^{0.10} X_4^{0.78}, \text{ R}^2 = 0.57, \text{ F-Ratio} = 13.30, \text{ p-value} = 0.0002 \tag{3}$$

8. **Formulation of BCC model** – to minimize the objective function z (4) under constraints (5). The CCR models that differ from the BCC models only by deleting variables q from equations (4) and (5) were formulated in an analogous way. The variable q represents the difference from constant returns to scale. For each DMU,

there was created a separate model. The super-efficiency (SE) Andersen-Petersen's model was used in order to rank efficient faculties (for more details see [8]).

$$z = \sum_i^m v_j x_{jq} + q \tag{4}$$

$$\sum_i^r u_i y_{ik} - \sum_j^m v_j x_{jk} - q \le 0 \text{ , } k = 1, 2, ..., n$$

$$\sum_i^r u_i y_{iq} = 1 \tag{5}$$

$u_i \ge \varepsilon, i = 1, 2, ..., r;$ ε - *very small non- Archimedean number* (> 0)
$v_j \ge \varepsilon, j = 1, 2, ..., m,$
$q \in R$

9. **Solving models and interpretation of results –** models were solved by using the add-in Solver in MS Excel software. The super-efficiency models were solved through the DEA-Excel Solver 2014 [7]. For each field of activity and each DMU, a score of relative technical efficiency was calculated. The average score of technical efficiency and standard deviation was determined for each area.

Variable	Inputs	Variable	Outputs
X_1	no. of students in bachelor programmes	Y_1	no. of graduates of bachelor programmes
X_2	no. of students in master programmes	Y_2	no. of graduates of master programmes
X_3	weighted number of academic staff	Y_3	standardized employment of graduates
X_4	no. of students in doctoral studies	Y_4	no. of PhD graduates
X_5	no. of bachelor and master students	Y_5	no. of RIV points
		Y_6	amount of research grants
		Y_7	no. of graduates of bachelor and master studies

Table 2 Inputs and outputs used in DEA models

Model No.	Model description	Inputs	Outputs
1	BCC – model 1, teaching efficiency	X_1, X_2, $0.5*X_3$	Y_1, Y_2, Y_3
2	BCC – model 1, research efficiency	$0.5*X_3$, X_4	Y_4, Y_5
3	BCC – model 1, overall efficiency	X_1, X_2, X_3, X_4	Y_1, Y_2, Y_3, Y_4, Y_5
4 = 1	BCC – model 2, teaching efficiency	X_1, X_2, $0.5*X_3$	Y_1, Y_2, Y_3
5	BCC – model 2, research efficiency	$0.5*X_3$, X_4	Y_4, Y_5, Y_6
6	BCC – model 2, overall efficiency	X_3, X_4, X_5	Y_3, Y_4, Y_5, Y_6, Y_7
7	BCC – model 3, teaching efficiency	$0.5*X_3$, X_5	Y_3, Y_7
8	BCC – model 3, research efficiency	$0.5*X_3$, X_4	Y_5, Y_6
9	BCC – model 3, overall efficiency	X_3, X_5	Y_3, Y_5, Y_6, Y_7
10	CCR – teaching efficiency	X_1, X_2, $0.5*X_3$	Y_1, Y_2, Y_3
11	CCR – research efficiency	$0.5*X_3$, X_4	Y_4, Y_5
12	CCR – overall efficiency	X_1, X_2, X_3, X_4	Y_1, Y_2, Y_3, Y_4, Y_5

Table 3 Inputs and outputs of individual DEA models

4 Results of the analysis

Based on the correlation analysis, it was found that on the input side there is a statistically significant correlation (α = 5%) between the number of academic staff and the number of students at all levels. This dependence is stronger for students of master (r = 0.73, p-value < 0.01) and bachelor programmes (r = 0.66, p-value < 0.01) than for doctoral students (r = 0.51, p-value = 0.01). Academic staff, however, stayed in the analysis as an input since this is a crucial input that affects teaching and research activities of the faculties. On the output side, there is a strong correlation between the number of graduates of bachelor and master programs (r = 0.93, p-value < 0.01). Therefore, in the models No. 6, 7, 9, the output value of the total number of graduates with bachelor and master degree was used. Very weak and not statistically significant correlation between the number of graduates of master's degree and PhD graduates (r = 0.18, p-value = 0.38) is quite surprising. Further, there was proved relatively strong correlation between the number of doctoral graduates and the number of RIV points (r = 0.88, p-value <0.01). For this reason, only the number of RIV points on the output side was left in the model No. 8. Between inputs and outputs, there was found statistically significant, but weaker dependence between the number of master students and graduates of PhD studies (r = 0.42, p-value = 0.03). Only moderate dependence exists between the number of academic staff and RIV points (r = 0.53, p-value = 0.01). Conversely, there was no significant correlation between the amount of grants received and other input or output variable, even where it

could be expected. Factors affecting the acquisition and grant volumes are probably more complex. This was the reason why the BCC model for the research and the overall efficiency has been compiled in two versions (with variable Y_6 and without it). The employment rate of graduates is another interesting indicator that only weakly correlates with the number of students of master programmes (r = 0.45, p-value = 0.03). Dependence on other variables is statistically insignificant. The employment of graduates is probably more influenced by environmental factors (e.g. labor market situation in the region).

Technical efficiency scores for each of the activities are listed in Table 4. There is no predefined functional relationship among the score of overall efficiency and teaching scores, or scores of research efficiency. The coefficients are calculated "independently" from the overall set of inputs and outputs. Both for the BCC/CCR output oriented models it holds that the DMUs are efficient if the score of technical efficiency equals one. For inefficient units, this score is greater than 1 and it indicates the necessary amount of increase of outputs to achieve the efficient frontier. Generally, it also holds that the number of efficient units is lower in CCR models where certain increase in inputs must comply with the same multiple of outputs.

Faculty	**BCC - model 1**			**BCC - model 2**			**BCC - model 3**			**CCR**			**BCC – model 3 SE**		
University	**Te**	**Re**	**Oe**	**Te**	**Re**	**Oe**	**Te**	**Re**	**Oe**	**Te**	**Re**	**Oe**	**Te**	**Re**	**Oe**
FFU VSE	1.03	1.56	1.02	1.03	1.62	1.02	1.03	1.82	1.03	1.43	1.56	1.18			
FMV VSE	1.03	1.24	1.01	1.03	1.30	1.01	1.03	2.02	1.02	1.44	1.76	1.21			
FPH VSE	1.02	1.48	1.00	1.02	1.54	1.01	1.02	2.02	1.02	1.22	1.65	1.05			
FIS VSE	1.03	1.39	1.00	1.03	1.00	1.00	1.03	1.00	1.00	1.46	1.39	1.14		**0.65**	**0.65**
NHF VSE	1.03	1.24	1.02	1.03	1.25	1.02	1.02	1.24	1.02	1.49	1.44	1.14			
FM VSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.16	1.00	**0.00**	**0.00**	**0.00**
PEF ČZU	1.01	1.00	1.00	1.01	1.00	1.00	1.01	1.14	1.00	1.32	1.35	1.05			**0.91**
EF JČU	1.02	1.00	1.00	1.02	1.00	1.00	1.08	1.00	1.04	1.03	1.00	1.00		**0.94**	
ESF MUNI	1.02	1.02	1.00	1.02	1.06	1.00	1.02	1.32	1.01	2.16	1.03	1.03			
PEF MENDELU	1.03	1.00	1.00	1.03	1.00	1.00	1.03	1.00	1.02	1.37	1.00	1.00		**0.76**	
FRR MENDELU	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.00**	**0.00**	**0.00**
MUP	1.01	1.49	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.52	1.78	1.25		**0.86**	**1.00**
FEM UO	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	1.00	**0.00**		**0.00**
OPF SLU	1.06	1.47	1.04	1.06	1.50	1.06	1.07	1.53	1.06	1.47	1.57	1.10			
EF TUL	1.00	1.33	1.00	1.00	1.31	1.00	1.00	1.31	1.00	1.00	1.39	1.00	**0.95**		**0.84**
FIM UHK	1.00	1.06	1.00	1.00	1.06	1.03	1.08	1.06	1.08	1.00	1.10	1.00			
UJAK	1.00	2.72	1.00	1.00	2.28	1.00	1.00	6.24	1.00	1.00	2.80	1.00	**0.40**		**0.40**
FSE UJEP	1.00	1.43	1.00	1.00	1.43	1.00	1.03	1.43	1.03	1.00	1.53	1.00			
FSV UK	1.03	1.00	1.00	1.03	1.00	1.00	1.02	1.00	1.00	1.76	1.00	1.00		**0.46**	**0.45**
FES UPCE	1.08	1.08	1.00	1.08	1.07	1.04	1.11	1.07	1.09	1.20	1.09	1.00			
FAME UTB	1.12	1.00	1.00	1.12	1.00	1.00	1.12	1.00	1.00	1.26	1.00	1.00		**0.89**	**0.84**
EF VSB	1.12	1.20	1.03	1.12	1.21	1.08	1.12	1.25	1.10	1.51	1.66	1.19			
VSFS	1.02	2.90	1.01	1.02	1.00	1.00	1.02	1.00	1.01	1.32	4.18	1.17		**0.86**	
FP VUT	1.07	1.64	1.06	1.07	1.64	1.06	1.07	1.64	1.06	1.57	1.75	1.24			
FEK ZCU	1.00	2.10	1.00	1.00	2.11	1.03	1.04	2.21	1.04	1.07	2.83	1.00			
Average	**1.03**	**1.37**	**1.01**	**1.03**	**1.26**	**1.03**	**1.04**	**1.49**	**1.03**	**1.30**	**1.60**	**1.07**			
SD	**0.03**	**0.50**	**0.02**	**0.03**	**0.35**	**0.04**	**0.04**	**1.04**	**0.03**	**0.28**	**0.72**	**0.09**			

Table 4 Score of technical efficiency according to individual models

Generally speaking, the efficiency of the educational activity in the surveyed faculties was quite high. In the case of BCC models, the faculties should increase their outputs on average only by about 3% to 4%, to become efficient. Model No. 7 (5 efficient units) which uses only 2 inputs and 2 outputs discriminates the best. Results of the model No. 1 are similar, but the efficient number of units is higher (8) as well as the number of inputs (3) and outputs (3). In the research results, the differences between BCC models are more distinct. On average, the faculties should increase their research outputs by 26% to 49% in order to become efficient. Model No. 2 (8 efficient units) with 2 inputs and 2 outputs discriminates the best. Similar results were obtained using a model No. 8 (9 efficient units) with the same number of inputs and outputs. Due to the fact that there is a strong relationship between the number of PhD graduates and RIV points in the model No. 2, we tend to use the model No. 8. The most difficult situation appeared when measuring the total efficiency, where a high number of inputs and outputs led to inadequate selection of efficient units. The rate of overall efficiency was high at the monitored faculties. On average, the faculties should increase their outputs by 7% in order to become efficient.

A lot of faculties are efficient in most of the models. Therefore the super-efficiency BCC model No. 3 was used to detect the best faculties. The Faculty of Management, University of Economics (FM VSE) and Faculty of Regional Development and International Studies, Mendel University in Brno (FRR MENDELU) are the best

assessed units according to the measures of teaching, research and overall efficiency. The Faculty of Economics and Management, University of Defence (FEM UO) was evaluated as the best in two of the three efficiency measures.

5 Conclusion

Based on the analyzed models, it can be stated that the DEA method appears to be a useful tool for measuring the efficiency of universities, or faculties. The results of this article should not be used to evaluate specific faculties. The reason is that the time series used is short and some variables (e.g. the amount of grants received) could have been affected by random influences. The question for discussion is also inclusion of individual inputs and outputs. **It is important to note that the above mentioned indicators measure primarily the efficiency, not the quality of education provided.** Presumably, it would be beneficial to include additional inputs of a financial nature. Unfortunately, reliable financial data on the faculty level could not be obtained. Other factors that may affect the results, and their verification is desirable for further research, are the shares of students of combined study and work capacity spent on teaching and research. In the analysis, there was included total number of students (and graduates). At some faculties, however, combined study is not implemented. At other schools, numbers of students in the combined form are even higher than the number of students in full-time study. It is a question for further research whether and to what extent students of combined study can be incorporated into the model. The results of the models were also influenced by the assumption that one half of the working capacity is dedicated to teaching and one half to the research. It is clear from the input data that some faculties are likely to be more focused on the research and some on teaching. Therefore, in the next direction of the research, it will be appropriate to investigate sensitivity of models to changes in the proportions of the working capacity spent on teaching and research.

Acknowledgements

The article was prepared with the help of institutional support for long-term conceptual development of the Faculty of Economics TUL within an internal project called Strategic Management of Company Performance.

References

[1] Beasley, J. E.: Determining Teaching and Research Efficiencies. *Journal of the Operational Research Society* **46** (1995), 441–452.
[2] *Central Register of R&D Projects* [online]. Research Development and Innovation Council, Prague, 2014-12-03. Available at http://www.isvav.cz/prepareProjectForm.do.
[3] Charnes, A., Cooper, W., Lewin, A. Y., and Seiford, L. M.: *Data Envelopment Analysis: Theory, Methodology and Applications.* Kluwer Academic Publishers, Massachusetts, 1994.
[4] Cooper, W., Seiford, L. M., and Zhu, J., eds.: *Handbook on Data Envelopment Analysis*. Springer Science and Business Media, New York, 2011.
[5] Düzakin, E., and Düzakin, H.: Measuring the Performance of Manufacturing Firms with Super Slacks based Model of Data Envelopment Analysis: An Application of 500 Major Industrial Enterprises in Turkey. *European Journal of Operational Research* **182** (2007), 1412–1432.
[6] Hajiagha, S. H. R., Akrami, H., Zavadskas, E. K., and Hashemi, S. S.: An Intuitionistic Fuzzy Data Envelopment Analysis for Efficiency Evaluation Under Uncertainty: Case of a Finance and Credit Institution. *E&M Ekonomics and Management* **16** (2013), 128–137.
[7] Jablonský, J.: *DEA-Excel Solver 2014 (a MS Excel based system for DEA models)* [online]. University of Economics, Prague, 2014-06-11. Available at https://webhosting.vse.cz/jablon/.
[8] Jablonský, J., and Dlouhý, M.: *Modely hodnocení efektivnosti produkčních jednotek.* Professional Publishing, Praha, 2004.
[9] Johnes, J.: Data Envelopment Analysis and its Application to the Measurement of Efficiency in Higher Education. *Economics of Education Review* **25** (2006), 273–288.
[10] Kuah, Ch. T., and Wong, K. Y.: Efficiency Assessment of Universities through Data Envelopment Analysis. *Procedia Computer Science* **3** (2011), 499–506.
[11] Nazarko, J., and Šaparauskas, J.: Application of DEA Method in Efficiency Evaluation of Public Higher Education Institutions. *Technological and Economic Development of Economy* **20** (2014), 25–44.
[12] *Průvodce databází o nezaměstnaných vysokoškolácích* [online]. Charles University, Prague, April 2012. Available at http://www.strediskovzdelavacipolitiky.info/default.asp?page=db.
[13] Rosenmayer, T.: Using Data Envelopment Analysis: a Case of Universities. *Národohospodářský obzor – Review of Economic Perspectives* **14** (2014), 34–54.
[14] *Výsledky Hodnocení výsledků výzkumných organizací v roce 2013* [online]. Research Development and Innovation Council, Prague, 2014-06-18. Available at http://www.isvav.cz/h13.

What's New in Maple™ 2015

Data Sets

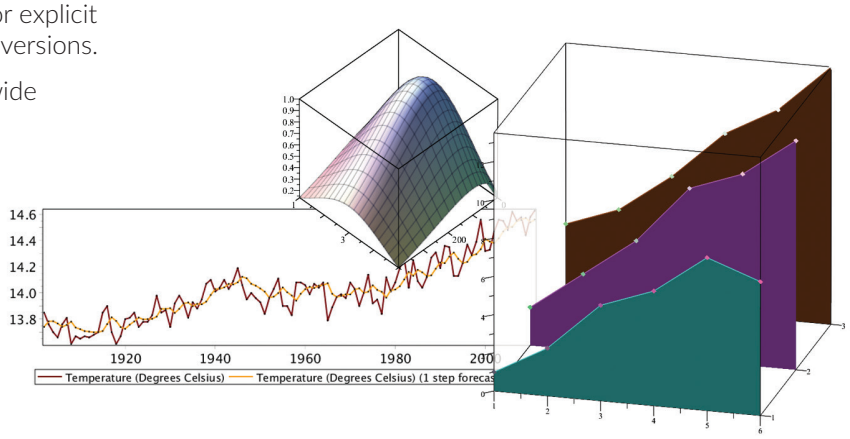
Maple™ 2015 features a powerful new infrastructure for accessing and working with millions of data sets from both built-in and online data sources.

- Easily access the entire collection of curated time series data provided by Quandl, an organization that hosts data from hundreds of sources. Through Quandl, Maple users can access a growing collection of over 12 million data sets to use in their Maple applications, free of charge.
- Select from a huge list of topics from finance, economics, and society, such as current data on stocks and commodities, foreign exchange rates, macroeconomic data on labor market indicators, population statistics, and much more.
- Use the standard Maple search box to locate available data sets, whether they are online or come from new country and city data tables built into Maple.
- Incorporate the data into your Maple computations, without the need for explicit downloading or data structure conversions.
- Easily visualize your data set in a wide variety of ways.
- Instantly create your own customized search application that makes it easy to filter your data, such as by specifying a time period and frequency, and assign the results to a variable for further manipulation.

Data Plots

Plotting data is substantially easier in Maple 2015. A comprehensive new *dataplot* command can be used to display many kinds of numerical data using a variety of 2-D and 3-D plots and animations.

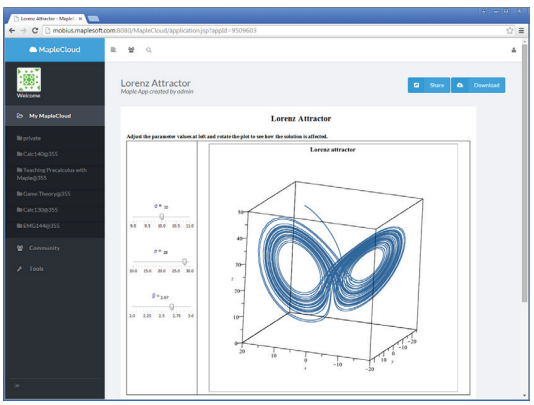
- Makes it easy to visualize data as points, surfaces, contours, density plots, bar charts, histograms, tree maps, pie charts, and more
- Supports a wide variety of plot options for customization, such as animation, symbols, color, and color palettes
- Available through both a command and from the context-sensitive menu
- Works with many different data structures, such a matrices, lists, arrays, and datasets, without the need for conversions



MapleCloud™

The MapleCloud is now accessible from a web browser, giving you a powerful new option for deploying and sharing your Maple content.

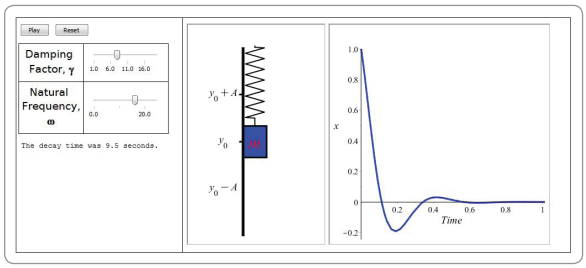
- The MapleCloud is now accessible from a web browser, from both computers and tablets, at **maplecloud.maplesoft.com**. You can browse and search collections of Maple content, read documents, and even interact with Maple applications, all without Maple.
- Your Math Apps and Maple applications can be used by anyone you choose to share them with, to explore concepts, perform computations, and visualize results. Any Maple document that makes use of interactive embedded components, such as sliders, dials, and math entry boxes, can be used and not just viewed, even by people who do not have Maple.
- You can continue to share your Maple documents in the MapleCloud at a click of a button from inside Maple. You can share your documents with members of a private group, such as the students in your class or colleagues on a research project, post them for your own personal use, or make them freely accessible by everyone.
- There are hundreds of interactive Math Apps available in the MapleCloud that explore concepts from calculus, statistics, graphing, physics, engineering, and more. You can use these Math Apps in your classes and direct your students to them for additional learning.



Math Apps

Math Apps in Maple are interactive learning tools that can be used to explore concepts and engage students. With Maple 2015, there are now more than 400 Math Apps included with Maple.

- 60 new interactive Math Apps for students and teachers to explore
- Topics taken from math, physics, statistics, chemistry, finance, and more
- Math Apps can be used from within Maple, from the free Maple Player™, and from the online MapleCloud using only a web browser

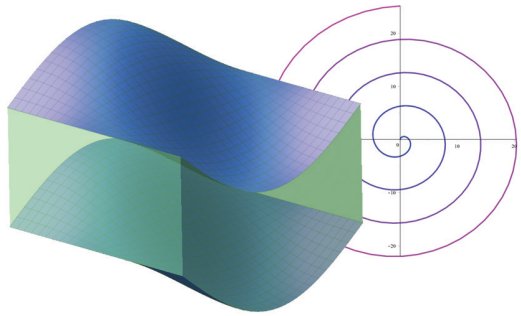


Visualization

Maple 2015 includes many new abilities and customization options for visualizations.

- New visualization tools are available for iterative maps, groups, polyhedral sets, and more.
- A new command makes it easy to shade between two curves or surfaces.
- Default ranges are now available for 3-D plots, so you no longer have to specify the range every time when using the command.
- The color scheme option, which was available for surfaces, is now available for curves, so you can specify a start and end color for your curve and Maple will modulate the color from one to the other along the length of the curve.
- Data visualization is greatly simplified with a new command that unifies point plots, surfaces, contours, density plots, bar charts, histograms, and more in a single command.

- A new line style for curves combines points with lines.
- Tree maps provide a new method of data visualization that use nested rectangles to show the magnitude of various elements in a set of data.



One-Step App Creation with *Explore*

Maple provides the *Explore* command as an easy-to-use tool for creating interactive applications, as well as interactive *Explore* functionality available from the context menu. Maple 2015 introduces significant updates and additions to the *Explore* command, including:

- Initialization code for your application can be automatically included in the start-up code of your new application when it is created in a separate document.
- Slider controls can be placed vertically or horizontally.
- Parameter names for sliders are displayed in standard math notation.
- You can customize the appearance and position of markers in a 2-D plot exploration.
- Images are now used on Play/Pause/Loop animation controls.
- Check boxes are available for parameters whose value can be true or false.
- Document properties, such as author and subject, can be set in the metadata section of your new document as part of the application.

Programmatic Content Creation

For many years, Maple has made it easy to create rich technical documents and interactive applications. Now, with Maple 2015, it is possible to generate this content programmatically as well as interactively. This ability allows you to generate documents whose content depends on the results of earlier computations, supports the creation of interactive applications without manual editing, and more.

- Programmatically generate rich technical content that includes math, text, tables, plots, sections, and more.
- Create interactive applications without ever touching the mouse.
- Use programmatic content to create entire documents or portions of documents.
- Make updates to applications quickly by modifying the defining code rather than going through the individual component properties.
- Quickly create a whole series of related content by passing in different parameters to a procedure that creates the final document or application.

Using this functionality, many Maple routines now include options to display results inside a document table for greater readability, including a new command that displays lists, matrices, and other structured data types in tables.

Interactive Components

Maple includes a wide variety of interactive components, such as buttons, sliders, and dials, that can be embedded in your worksheet to create interactive applications. Maple 2015 offers new and improved components to give you more flexibility in your application development.

- New microphone and speaker components let you capture and play back audio in your Maple application.
- New customization options include resizable sliders, angle ranges and image backgrounds on dials, hidden borders and line numbers in code edit regions, wrapping in text boxes, and multiple selection options on list boxes.

Integration

Maple 2015 can find solutions for new classes of integrals, and produces results in more compact forms.

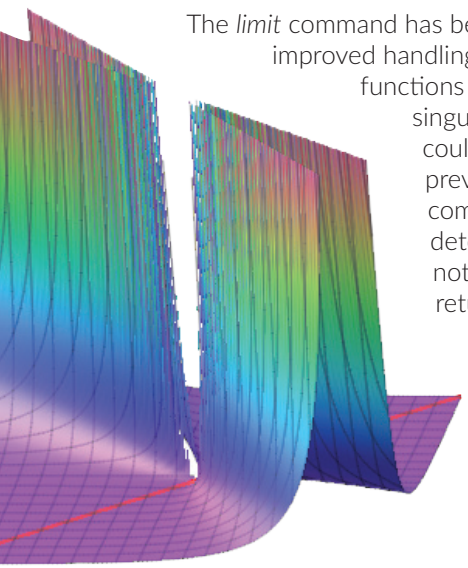
- Maple can now find analytic, closed-form solutions of new classes of indefinite integrals, including those involving inverse hyperbolic functions.
- Many integrals that used to be expressed in terms of lengthy complex sign (*csgn*) expressions are now given in more compact forms.
- Maple can compute definite integrals of non-smooth integrands, for which previous versions of Maple only could compute an indefinite integral.

Group Theory

The database of small groups included in the *GroupTheory* package has been greatly expanded to include all groups of order less than 512. In addition, you can now search both the small groups database and the transitive groups database using new search tools. Other improvements include the ability to construct and visualize Cayley graphs of small groups, construct free groups, compute exponents, element order, and complex products, and test if a group is cyclic.

Limits

The *limit* command has been enhanced for improved handling of bivariate rational functions with non-isolated singularities. Many such limits could not be determined previously, but are now computable. Maple can also determine if the limit does not exist, in which case it returns *undefined*.



Differential Equations

Maple can now find numeric solutions to delay differential initial value problems. The RKF45, CK35, and Rosenbrock methods all accommodate delay terms, and so can be used to solve ordinary differential and differential-algebraic equations with delays.

Ordinals

Maple 2015 includes a new package for computing with ordinal numbers in Cantor normal form. The Ordinals package can represent and handle all ordinal numbers that can be obtained from non-negative integers and ω, the ordinal for the standard ordering of the natural numbers $\mathbb{N}$, in a finite number of steps by ordinal addition, multiplication and exponentiation.

Physics

Maple provides a state-of-the-art environment for algebraic computations in physics, with emphasis on ensuring the computational experience is as natural as possible. The theme of the Physics project for Maple 2015 has been vector analysis, symbolic tensor manipulations, quantum mechanics, and general relativity.

- More than 400 enhancements throughout the entire package increase robustness and versatility.
- More than 100 metrics were added to the database of solutions to Einstein's equations.
- The formalism of tetrads in general relativity was implemented within Physics as a new package.
- Expanded support for commutators, anticommutators, and parametrized algebra rules aid in computations with products of non-commutative operators.
- Improved simplification algorithms and new options for automatic simplification using assumptions on variables make the system easier to work with and the results easier to interpret.

- Improved tensors are more powerful, and make handling symmetries, substitutions, and other operations more flexible and natural.
- New commands provide more tools for programming and interactive computations.
- And more!

Mathematical Functions

Maple provides a rich collection of definitions, identities, and properties for a large set of mathematical functions. Maple 2015 adds many new mathematical formulas and properties to the mathematical functions database, new conversion routines, and more tools for exploration and computation.

Polyhedral Sets

Maple 2015 has new functionality for working with polyhedral sets defined either by a set of equalities and inequalities (*H-Representation*), or by a set of vertices and rays (*V-Representation*). The added functionality lets you explore the geometric and topological properties of a set, perform standard set operations, and apply linear transformations.

Finance

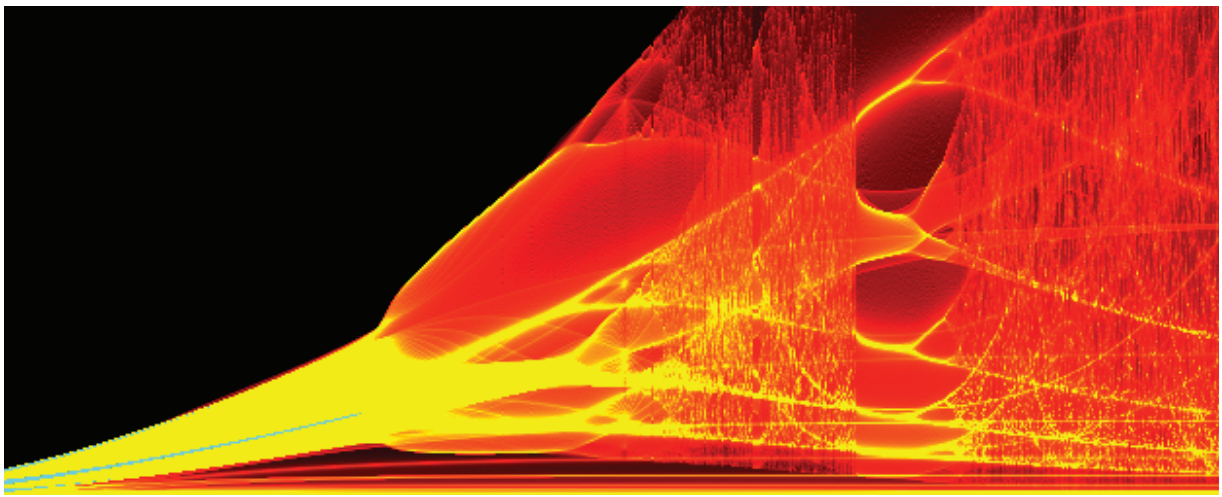
The Finance package now includes support for 10 more Greeks, quantities often used in risk management to represent the sensitivity of the price of a derivative to changes in underlying parameters on which the value of a financial instrument is dependent.

In addition, Maple can now return amortization tables in the form of an embedded table in the document, automatically generating a cleaner, easier to read result.

Iterative Maps

New iterative maps can be used to compute graphic images representing bifurcation diagrams, escape-time and related fractals, and attractors.

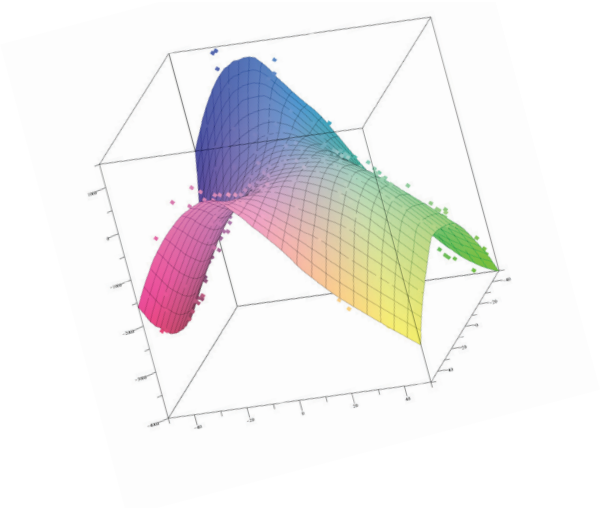
- Supports attractor, bifurcation, and escape maps
- Produces stunning visualizations
- Provides automatic parallelism that takes full advantage of the processing power of your computer for faster results
- Can be made interactive when combined with the *Explore* command, so you can see the changes in the visualization as you modify the parameters



Statistics

Maple 2015 further broadens its support for statistics with new computation and visualization abilities.

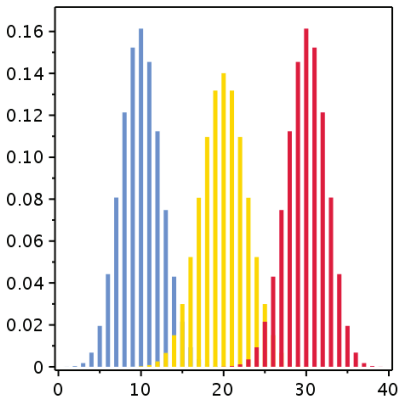
- The Lowess (locally weighted scatterplot smoothing) method for visualizing a smoothed curve or surface has been expanded to return a function whose graph is the smoothed graph, allowing programmatic manipulation and the ability to handle data points in any finite dimensions.
- The Lowess algorithm has also been improved to produce better plots and to achieve lower computation times.
- Maple can now calculate robust linear regressions using a repeated median estimator, giving less weight to outliers and better results in the presence of noise.
- A new command centers and scales a sample list or matrix, which is useful in applications such as computing standard scores.
- Visualization improvements include new tree maps, support for time series objects in bubble plots, and the new data plot command which provides a single interface to perform many statistical plots.
- Several statistics commands have been improved so the output is easier to read.
- Context-menu operations are now available for matrix data sets.
- New code generation tools for R allow you to automatically convert your Maple expressions and programs to R code.



Statistics Education

Maple 2015 further expands support for teaching and learning statistics with new Clickable Statistics tools.

- 18 new interactive Math Apps for statistics make it easy to explore confidence intervals, different distributions, chi-square tests, z-tests, stem and leaf displays, and more.
- A new palette helps students construct random variables based on different distributions.
- New commands generate probability and critical tables instantly, based on the given distribution and parameters.
- New tutors offer a command-free method for generating probability and critical tables.
- The Test Guide tutor command has been expanded to accept data samples as arguments, so the data does not need to be entered manually from inside the tutor.



Units

Computations involving units are now easier and more natural than ever.

- Keyboard shortcuts are available for inserting units.
- A new interactive tool for formatting units makes it easy to convert between both individual units and systems of measurement.

- Maple now provides the ability to work with absolute temperatures in addition to relative temperatures.

Code Generation

Code generation in Maple has been further expanded with support for R and JavaScript®. In addition, all code generation tools now include an option for the generated code to appear in a text component, making it easier to read and copy.

Import/Export

Maple 2015 makes it easier than ever to import and export data in Maple.

- New Import and Export commands provide a unified approach to all data import and export activities, replacing the need to use different commands in different situations.
- These commands can be used with any type of data, without having to specify, or even know, the format the data is stored in. They can be used with numeric and tabular data, images, audio, specialized text file formats like XML, and even special-purpose mathematical formats for linear optimization and graph theory.
- The list of supported import formats now includes JSON (JavaScript Object Notation), a human-readable, text-based standard for describing structured data.

Performance

Performance and efficiency improvements continue to be made in key areas of Maple. These improvements to fundamental operations result in faster, more efficient computations throughout the product. Improvements in Maple 2015 include:

- Maple's 'garbage collector', which is responsible for finding and reclaiming memory that is no longer needed by the evaluation engine, has been improved to perform even more operations in parallel. Greater parallelism during collection leads to better overall performance.
- Several linear algebra commands, when working with default double precision, are now faster in Maple 2015 on the 64-bit Windows platform.
- The performance when computing greatest common divisors of multivariate polynomials modulo a prime number has increased significantly, with orders of magnitude improvements in some cases.
- More special functions are now implemented in *evalhf*, for fast hardware precision evaluation.
- The evaluation of elementary and special functions in different floating point computation environments has been improved to handle branch cuts consistently.

Grid Computing

Unlimited same-machine parallel execution is built into Maple. You can spawn as many parallel processes as you want without requiring any additional toolbox or licensing. Maple 2015 makes it even easier to initiate parallel jobs with new commands that abstract away MPI-like message passing protocols. The result is a very simple and intuitive interface for running commands and dealing with data in parallel.

Language and Programming

In addition to the ability to create content programmatically, there were several other improvements to the Maple language and core functions.

- You can now easily merge two tables together.
- The keyword *double* can now be used as a synonym for *float[8]*.
- Objects can now inherit properties from other object declarations.
- You can now easily concatenate a string with a constant.
- The *rand* command now accepts floating point values for the range.

Additional Improvements

Maple 2015 includes many improvements, small and large, to enhance usability.

- New palettes are available for trigonometric and hyperbolic functions, and for constructing random variables based on statistical distributions.
- Improved 2-D Math editing now interprets the adjacent brackets in expressions such as $(x + y)(a + b)$ as implicit multiplication.
- New keyboard shortcuts support zooming, expanding and collapsing sections, and more.
- Plot thumbnails are now shown when Maple returns a plot structure, such as when you assign a plot to a variable, so you can easily verify if the result is what you intended.
- The Help Search Box can now also search for data sets and has a more compact display.
- User profiles allow you to select your system of units, set your preferred numeric formatting, and insert Maple commands into a worksheet's Startup Code region.
- Units and numeric formatting options allow you to control defaults both globally and locally.
- Documents can be saved to the MapleCloud directly from the File menu.
- Menus have been restructured to make them easier to navigate.
- Plot zooming is now easier with new zoom in/ zoom out buttons on the toolbar.
- Commands entered in Maple syntax without the terminating semi-colon no longer display a warning with the result.

CS1 | Czech Software First

Autorizovaný distributor Maplesoft Inc.

Czech First Software Company Ltd. (CS1.) was founded in 1991 in Brno and focuses its activities on consulting in the field of scientific computing, mathematical modeling and distribution of mathematical software Maple in education, research and business sectors in the Czech and Slovak Republic.

CS1. is the contractual partner of the Canadian company Maplesoft, Inc.. and an authorized distributor of its products (Maple, MapleNet, Maple TA, toolboxes, etc.) in the Czech and Slovak Republic, Slovenia, Romania and Bulgaria.

CS1. works with high and secondary schools, institutes of the Czech Academy of Sciences and design institutions on the use of computer algebra systems (symbolic computation) in college and high school mathematics teaching in mathematical modeling and scientific computing.

We participate in the preparation and organization of seminars on the use of Maple in teaching and research.

Head
Prof. RNDr. Jiří HŘEBÍČEK, CSc.
phone: (+420) 603 217 052
hrebicek@ecomanag.cz

Sales and marketing
Ing. Lenka HEFKOVÁ
phone: (+420) 733 502 577
lenka.hefkova@ecomanag.cz

Technical support
Ing. Vladimir ŽÁK
zakyn@centrum.cz

Address
Czech Software First, s.r.o.
K Západi 54, 612 00 Brno
www.maplesoft.cz

33rd International Conference
Mathematical Methods in Economics
MME 2015

Conference Proceedings

Edited by
David Martinčík
Jarmila Ircingová
Petr Janeček

Responsible editor David Martinčík

Cover layout by Hana Ovesleová

All papers passed a blind review process.
Linguistic revision of the texts was not undertaken.
Papers are sorted alphabetically by the names of their authors.

Published by University of West Bohemia, Plzeň
Univerzitní 8, 306 14 Plzeň
http://www.zcu.cz/pracoviste/vyd/
https://e-shop.zcu.cz/
vydavatel@uk.zcu.cz

First edition
Plzeň 2015

ISBN 978-80-261-0539-8

33rd international conference

Mathematical Methods in Economics

September 9-11, 2015, Cheb, Czech Republic

ISBN 978-80-261-0539-8