Conference is held under the auspices of
Department of Statistics, Cracow University of Economics, Poland
Committee of Statistics and Econometrics, Polish Academy of Sciences
Polish Statistical Association

Honorary Patronage of the Conference
Rector of the Cracow University of Economics
Prof. UEK dr hab. inż. Andrzej Chochół

THE 11TH PROFESSOR ALEKSANDER ZELIAS INTERNATIONAL CONFERENCE ON MODELLING AND FORECASTING OF SOCIO-ECONOMIC PHENOMENA

Conference Proceedings

Edited by
Monika Papież and Sławomir Śmiech

MAY 9-12, 2017
ZAKOPANE, POLAND

Editors: Monika Papież, Sławomir Śmiech

All published papers have been reviewed by two referees before publishing

The proceedings have not been amended or proofread and editors are not responsible for the language used in papers

Publishing of the proceedings is financed by the Cracow University of Economics

Suggested citation:
AUTHOR, A. (2017). Title of the paper. In M. Papież and S. Śmiech (Eds.), *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, pp. xx-xx. ISBN: 978-83-65173-85-0 (HTML).

ISBN: 978-83-65173-85-0 (HTML)

Year of publishing: 2017
Number of pages: 466

Publishing House:
Foundation of the Cracow University of Economics
Rakowicka 27, 31-510 Cracow, Poland

Programme Committee:

Chairman: **Prof. dr hab. Józef Pociecha (Chairman)**
Department of Statistics
Cracow University of Economics, Poland

Members: **Prof. dr hab. Andrzej S. Barczak**
Department of Econometrics
University of Economics in Katowice, Poland

Prof. dr hab. Czesław Domański
Department of Statistical Methods
University of Łódź, Poland

Prof. dr hab. Krzysztof Jajuga
President of the Committee of Statistics and Econometrics
The Polish Academy of Sciences
Department of Financial Investments and Risk Management
Wroclaw University of Economics, Poland

Prof. dr hab. Jacek Osiewalski
Vice President of the Committee of Statistics and Econometrics
The Polish Academy of Sciences
Department of Econometrics and Operations Research
Cracow University of Economics, Poland

Prof. RNDr. Viera Pacakova, PhD
Institute of Mathematics and Quantitative Methods
University of Pardubice, Czech Republic

Prof. Anatoliy I. Pilyavskyy
Department of Mathematics
Lviv Academy of Commerce, Ukraine

Doc. Ing. Eva Sodomova, PhD
Department of Statistics
University of Economics in Bratislava, Slovakia

Prof. dr hab. Mirosław Szreder
Department of Statistics
University of Gdansk, Poland

Organizing Committee:

Chairman:	**Prof. dr hab. Józef Pociecha (Chairman)** Department of Statistics Cracow University of Economics, Poland
Vice-chairman:	**Prof. UEK dr hab. Anna Malina** Department of Statistics Cracow University of Economics, Poland
Members:	**Roman Huptas, PhD (Scientific Secretary)** Department of Statistics Cracow University of Economics, Poland
	Prof. UEK dr hab. Barbara Pawełek Department of Statistics Cracow University of Economics, Poland
	Prof. UEK dr hab. Stanisław Wanat Department of Mathematics Cracow University of Economics, Poland
	Dr hab. Monika Papież Department of Statistics Cracow University of Economics, Poland
	Dr hab. Sławomir Śmiech Department of Statistics Cracow University of Economics, Poland
	Katarzyna Frodyma, PhD Department of Statistics Cracow University of Economics, Poland
	Jadwiga Kostrzewska, PhD Department of Statistics Cracow University of Economics, Poland
	Artur Lipieta, PhD Department of Statistics Cracow University of Economics, Poland

Conference website:
http://www.konferencjazakopianska.pl/
http://www.pliki.konferencjazakopianska.pl/proceedings_2017/

Organizer's website:
http://www.katstat.uek.krakow.pl/en/

List of Referees:

Jan Acedański, University of Economics in Katowice, Poland
Adam P. Balcerzak, Nicolaus Copernicus University in Toruń, Poland
Aleksandra Baszczyńska, University of Łódź, Poland
Jacek Białek, University of Łódź, Poland
Marcin Błażejowski, WSB University in Toruń, Poland
Justyna Brzezińska, University of Economics in Katowice, Poland
Marek A. Dąbrowski, Cracow University of Economics, Poland
Sabina Denkowska, Cracow University of Economics, Poland
Kamil Fijorek, Cracow University of Economics, Poland
Karol Flisikowski, Gdansk University of Technology, Poland
Katarzyna Frodyma, Cracow University of Economics, Poland
Krzysztof Guzik, Cracow University of Economics, Poland
Monika Hadaś-Dyduch, University of Economics in Katowice, Poland
Sergiusz Herman, Poznań University of Economics and Business, Poland
Roman Huptas, Cracow University of Economics, Poland
Alina Jędrzejczak, University of Łódź, Poland
Pavla Jindrová, University of Pardubice, Czech Republic
Lucie Kopecká, University of Pardubice, Czech Republic
Daniel Kosiorowski, Cracow University of Economics, Poland
Jadwiga Kostrzewska, Cracow University of Economics, Poland
Eva Kotlebová, University of Economics in Bratislava, Slovakia
Marta Kuc, Gdansk University of Technology, Poland
Paweł Kufel, WSB University in Toruń, Poland
Tomasz Kuszewski, Warsaw School of Economics, Poland
Łukasz Lenart, Cracow University of Economics, Poland
Paweł Lula, Cracow University of Economics, Poland
Justyna Majewska, University of Economics in Katowice, Poland
Małgorzata Markowska, Wrocław University of Economics, Poland
Jerzy Marzec, Cracow University of Economics, Poland
Błażej Mazur, Cracow University of Economics, Poland
Dominik Mielczarek, AGH University of Science and Technology, Krakow, Poland
Jacek Osiewalski, Cracow University of Economics, Poland
Monika Papież, Cracow University of Economics, Poland
Barbara Pawełek, Cracow University of Economics, Poland
Dorota Pekasiewicz, University of Łódź, Poland
Michał Bernard Pietrzak, Nicolaus Copernicus University in Toruń, Poland
Agnieszka Przybylska-Mazur, University of Economics in Katowice, Poland
Aneta Rybicka, Wroclaw University of Economics, Poland
Jerzy P. Rydlewski, AGH University of Science and Technology, Krakow, Poland
Victor Shevchuk, Cracow University of Technology, Poland
Andrzej Sokołowski, Cracow University of Economics, Poland
Erik Šoltés, University of Economics in Bratislava, Slovakia
Małgorzata Stec, University of Rzeszów, Poland
Jacek Stelmach, University of Economics in Katowice, Poland
Agata Szczukocka, University of Łódź, Poland
Sławomir Śmiech, Cracow University of Economics, Poland
Elżbieta Szulc, Nicolaus Copernicus University in Toruń, Poland
Grażyna Trzpiot, University of Economics in Katowice, Poland
Paweł Ulman, Cracow University of Economics, Poland
Agnieszka Wałęga, Cracow University of Economics, Poland
Grzegorz Wałęga, Cracow University of Economics, Poland
Stanisław Wanat, Cracow University of Economics, Poland
Justyna Wróblewska, Cracow University of Economics, Poland
Florian Zechser, FOM Hochschule, Essen, Germany

CONTENTS

PREFACE

Quantitative approach to investigation of contemporary socio-economic processes is the unique way to formulate proposals to solving global or regional economic problems. Statistical and econometric methods, models and forecasts from both, theoretical and empirical point of views, will be discuss at The 11th International Conference in honour of Professor Aleksander Zeliaś, which take place in Zakopane at 9-12 of May 2017. Participants of the conference are both well known and young scientists from Poland, Germany, Japan, Czech Republic, Slovakia and Ukraine.

The volume presented here contains selected conference proceedings, independently revised by two anonymous reviewers, among almost 60 submitted studies. Papers presented here describe a current stage of statistical theory as well as interesting applications, econometric modelling technics and data analysis applications in a variety of areas of economic processes modelling and forecasting. Conference presentations concentrate on financial problems analysis on macro- and micro-levels, threats of contemporary world, modelling and forecasting economic processes, computational tools for statistical and econometric analyses, spatial and regional modelling, risk analysis, statistical methods for business investigations.

We hope that Readers find in the collection of papers original ideas, useful methods and interesting results of empirical investigations of various socio-economical problems of Central and Central-East European countries, studied by Polish, German, Japan, Czech, Slovakian and Ukrainian researchers, which are presented at our conference.

Józef Pociecha

Overlapping generation models with housing: impact of the key parameters on the models' outcomes

Jan Acedański[1]

Abstract

Overlapping generation models (OLG) with housing are used to analyze interactions between households of different age, housing market and the main macroeconomic aggregates like production, consumption or interest rate. In the paper, we study sensitivity of these categories to changes in the main parameters of the models, especially the discount factor, the risk aversion parameter, demographic characteristics and housing market parameters. The baseline model is calibrated to match the key features of the Polish economy.

Keywords: *overlapping generations, housing, calibration, wealth, debt*

JEL classification: C63, E20

1 Introduction

Overlapping generation models developed by Diamond (1965) and Auerbach and Kotlikoff (1987) become increasingly popular recently as a tool to study interactions between macroeconomic aggregates and households in different phases of life-cycles. They are used to study the effects of fiscal and monetary policy (Auerbach and Kotlikoff, 1987; Kindermann and Krueger, 2014; Doepke et al., 2015) as well as social security on income and wealth inequality, consumption and material deprivation of households (Gertler, 1999; Hairault and Langot, 2007; Cheron et al., 2011; Acedański, 2016; Bielecki et al., 2015a, 2015b). Recently, the models are augmented with housing market (Chen, 2010; Rubaszek, 2012) because of its important role in business cycles, as shown by the last financial crisis, and the fact that majority of households' wealth is stored in real estates. Housing also plays an indispensable role in modelling households' debt as mortgage loans dominate households' liabilities in terms of their value.

Unfortunately, the popularity of the discussed models is limited to some extent by their computational complexity. Deriving household's consumption and housing decision rules under rational expectations requires solving fixed points problems with nested stochastic, dynamic programs with finite horizon. Such problems can be solved only approximately with computationally-intensive numerical procedures. The time-consuming computations result in difficulties with the correct calibration of the models' parameters as the approximations have

[1] University of Economics in Katowice, Department of Statistical and Mathematical Methods in Economics, ul. 1-go Maja 50, 40-287 Katowice, Poland, e-mail: jan.acedanski@ue.katowice.pl.

to be executed many times for different sets of parameter values to match the targeted empirical characteristics with the analyzed model.

This paper is intended to provide some form of guidance on the role of the typical OLG model’s parameters for shaping the most important characteristics like interest rate, housing characteristics and age profiles of selected variables which should facilitate the calibration process. We investigate sensitivity of the mentioned characteristics to the changes in the main parameters of the model. The baseline model’s calibration aims at matching the key characteristics of Polish economy.

The paper is organized as follows. Section 2 contains the description of the model. The baseline calibration is discussed in section 3. The results of our sensitivity checks are presented in section 4.

2 Model

We study the open economy overlapping generations model with housing closely related to Chen (2010) and Rubaszek (2012). There are four types of agents in the model: households, firms, financial intermediaries and government which are described in more details below.

We use the following notation: capital letters refer to the aggregate variables, an individual household's characteristics are described by small letters and primes denote the next-period variables.

The economy is inhabited by a continuum of households that differ in terms of age *j*, stochastic idiosyncratic labor productivity $e \in \{e_1, e_2, \ldots, e_I\}$ and asset holdings. Households store their current wealth in form of deposits a or housing h_o. They enter labor market at age 25 with no wealth, work until 64 then retire and live up to 84 at most. Because we focus on the low-frequency movements in the interest rates only twelve cohorts are considered: $j \in \{25, 30, \ldots, 80\}$. The life span is stochastic and the probability that a household at age j survives to age $j+5$ is denoted by s_j. The share of the cohorts j to $j+4$ in the total population is μ_j whereas the share of a labor productivity group $i \in \{1, 2, ..., I\}$ is denoted by μ_{ei}.

Households derive utility from consumption of nondurable goods *c* and housing services *h*. The households either own or rent houses and the flows of housing services in both cases are denoted by h_o and h_r, respectively. The momentarily utility function of a household is given by:

$$u(c,h) = \frac{(c^{\theta} h^{1-\theta})^{1-\eta} - 1}{1-\eta} \tag{1}$$

where $h = h_o$ if a household owns a house, $h = h_r$ if it rents one or $h = \underline{h}$ otherwise, where $\underline{h}$ denotes some small positive level of housing services, η is the relative risk aversion parameter and θ represents the share of nondurable goods consumption in the utility function. Households also care about next generations and derive utility from leaving bequests b in the following form:

$$u_b(b) = \kappa \frac{b^{1-\eta} - 1}{1-\eta}, \tag{2}$$

where κ governs the strength of the bequest motive.

Young households work and earn income that depends on their age and idiosyncratic labor productivity. Their net earnings are $(1-\tau)e_i\varepsilon_j W$, where τ represents the tax rate, W denotes the average wage in the economy and ε_j is deterministic, age-related productivity component. Retired households receive pensions that are constant across households and independent of their earnings' history[2]. The pension is equal to the fraction χ of the average wage in the economy. Labor-related income of a household can be written as:

$$inc = \begin{cases} (1-\tau)e_i\varepsilon_j W & \text{if} \quad j < 65 \\ \chi W & \text{if} \quad j \geq 65 \end{cases}. \tag{3}$$

A representative firm produces the final nondurable good Y using the Cobb-Douglas technology:

$$Y = K^{\alpha} L^{1-\alpha}, \tag{4}$$

where K is the aggregate physical capital and $L = \sum_{j<65} \varepsilon_j \mu_j \sum_i e_i \mu_{ei}$ is the aggregate effective labor input. The productive capital is rented from the financial intermediaries whereas labor is provided by households. Because the production sector is perfectly competitive the capital demand of the firm is set so that the marginal product of capital matches the gross interest rate $R + \delta$ and the aggregate wage is equal to the marginal product of labor:

$$W = (1-\alpha) K^{\alpha} L^{-\alpha}, \tag{5}$$

$$R + \delta = \alpha K^{\alpha-1} L^{1-\alpha} \tag{6}$$

where δ denotes the depreciation rate of physical capital. The final good can be consumed or invested in either physical capital or housing.

[2] This assumption is common in the literature and is made for computational tractability of the model.

The perfectly competitive financial sector collects deposits from households and foreign investors. The aggregate net value of foreign deposits A^* is equal to:

$$A^* = \xi(R - R^*)Y, \tag{7}$$

where R^* is the world interest rate and the parameter ξ measures the degree of the economy's openness (autarky if $\xi = 0$ and perfect international financial markets if $\xi = \infty$). This parameter determines also the sensitivity of the domestic interest rate to the world's one.

The financial sector grants mortgage loans to homeowners and buys physical capital and rental housing. Capital is then rented to the final good producers and to the households that are unable to buy their own house. It is assumed that the interest rates on deposits, mortgage loans and physical capital are equal. Solving the profit maximization problem of a representative financial intermediary and assuming that the profit in the equilibrium is zero give the rental price on housing R_r:

$$R_r = \frac{R' + \delta_r}{1 + R'} \tag{8}$$

where δ_r represents the depreciation of rental housing.

The government's role is twofold. First, it levies taxes on labor to finance pensions. The government's budget is balanced so the tax rate is equal to:

$$\tau = \frac{\chi}{L} \sum_{j \geq 65} \mu_j . \tag{9}$$

Secondly, following the literature it is assumed that the government collects bequests and distributes them equally among the living households. These lump-sum transfers are denoted by tr.

A household maximizes its expected discounted lifetime utility. It takes its current deposit stock a, housing stock h_o, idiosyncratic productivity e and age j as givens and makes a housing tenure decision first. Then, a homeowner decides on the house size, nondurable consumption level and the financial assets level (deposit or debt). A renter chooses housing services offered by the financial intermediaries, the nondurable consumption level and the deposit level. It is not allowed to borrow.

Value functions of a household that chooses to own or rent a house are denoted by V_o and V_r, respectively. The decision problem of a household that decides to own a house can be written recursively:

$$V_o(a, h_o, e, j) = \max_{c, h'_o, a'} \left\{ u(c, h_o) + \beta \left[s_j E(V(a', h'_o, e', j+1) \mid a, h_o, e, j) + (1 - s_j) u_b(b') \right] \right\} \tag{10}$$

subject to:

$$(1+R)a+(1-\delta_o)h_o+inc+tr=a'+h_o'+\phi(h_o,h_o')+c, \tag{11}$$

$$a'\geq -(1-\gamma)h_o', \tag{12}$$

$$b'=(1+R)a'+(1-\delta_o)h_o'. \tag{13}$$

where $V=\max\{V_o,V_r\}$ with V_r defined below, δ_o denotes the depreciation rate of owned housing and ϕ represents transaction costs associated with a change in the size of the owned house:

$$\phi(h,h')=\begin{cases}0 & \text{if} \quad h=h',\\ \phi(h+h') & \text{if} \quad h\neq h'.\end{cases} \tag{14}$$

Equation (11) is the budget constraint of a household. Equation (12) represents the downpayment borrowing constraint which states that a household can borrow up to $(1-\gamma)\cdot 100\%$ of the owned house value. Finally, equation (13) defines the value of a bequest.

The value function of a household which chooses to rent a house takes the following form:

$$V_r(a,h_o,e,j)=\max_{c,h_r',a'}\left\{u(c,h_r')+\beta\left[s_jE(V(a',0,e',j+1)\,|\,a,h_o,e,j)+(1-s_j)u_b(b')\right]\right\} \tag{15}$$

subject to:

$$(1+R)a+(1-\delta_o)h_o+inc+tr=a'+R_rh_r+\phi(h_o,0)+c, \tag{16}$$

$$a'\geq 0. \tag{17}$$

3 Calibration

The parameters are calibrated to match the main characteristics of Polish economy. The real risk free rate in the US is used as a proxy for the world interest rate. As already mentioned, one period in the model corresponds to five years. The values of the parameters are presented in table 1.

The steady state level of the world interest rate is set to $\overline{R}^*=3.65\%$ per annum which is equal to the mean real risk free rate in the US in period 1947–2016 according to the data provided by R. Shiller's website. The parameter ξ associated with the degree of the economy openness is set to 1.3 to match the average net foreign asset position A^*/Y of Poland which is roughly -55%.

Symbol	Description	Value
$\overline{R}^*$	Steady state level of the world interest rate (annualized)	3.65%
ξ	Openness of the economy	1.3
s_j	Survival probabilities	CSO (2016)
e, P_e	Idiosyncratic productivity shocks and transition matrix	Rubaszek (2012)
β	Discount coefficient (annualized)	0.971
η	Risk aversion	3
θ	Share of nondurable consumption in the utility function	0.75
κ	Strength of the bequest motive	15
χ	Pension replacement rate	0.6
α	Capital share in the production function	0.3
δ	Capital depreciation rate (annualized)	0.08
δ_o	Owned housing depreciation rate (annualized)	0.013
δ_r	Rented housing depreciation rate (annualized)	0.025
γ	Downpayment ratio	0.15
ϕ	Transaction costs	0.075
h_{min}	Minimum house size	0.5
h_{max}	Maximum house size	2.25
$\underline{h}$	Minimum housing consumption	$0.1 \cdot h_{min}$

Table 1. Baseline calibration of the model.

Twelve cohorts are considered. Survival probabilities are based on Polish unisex life tables published by Central Statistical Office (2016). The annual discount coefficient is equal to 0.971 to match the average interest rate spread between Poland and the US of 2.8% per annum. A fairly standard risk aversion parameter η = 3 is used. The share of nondurable consumption in the utility function is set to 0.75 following Rubaszek (2012). The parameter κ = 15 that governs strength of the leaving bequest motive is determined to match the age profiles of the fraction of households with debt and the average debt value in Poland according to data provided by the National Bank of Poland (2015), particularly in the oldest cohorts. The pension replacement rate parameter $\chi = 0.6$ matches the ratio observed in Poland in recent years. Finally, the idiosyncratic productivity levels together with the transition matrix for the shock are taken from Rubaszek (2012).

We use a standard values of the technology parameters for Poland setting the capital share $\alpha = 0.3$ and the annual capital depreciation rate $\delta = 0.08$. The depreciation rates for the housing market follow Chen (2010): $\delta_o = 0.013$ and $\delta_r = 0.025$. As a result of the difference in the depreciation rates owned housing is cheaper than rented which is another incentive for households to buy rather than rent a house.

The house sizes h that are available to buy or rent are discretized. The set of sizes takes the form $h \in \{0, h_{min}, ..., h_{max}\}$, where the sizes between h_{min} and h_{max} are equally spaced. These values together with the mortgage downpayment ratio γ, the transaction costs parameter ϕ and $\underline{h}$ are jointly determined to match the age profiles of the fraction of households with debt and the average debt value.

Table 2 contains the key model's characteristics for the baseline calibration and their empirical counterparts for Poland. The real interest rate in the model coincides with the mean real interest rate based on data provided by OECD. Average total wealth relative to mean annual income of households in the model exceeds the value observed in the data. However, the specification of the model does not allow to simultaneously match the interest rate and the wealth level as they are tightly linked. Similar result is observed for housing wealth. The model overestimates the homeownership rate but the data collected by NBP (2016) does not account for social housing as pointed out by Rubaszek (2012). The fraction of households in debt and the average value of debt, where the latter is expressed relative to households' annual income, in the model are slightly lower compared to the data. Finally, the model generates unrealistically low wealth inequality. This results from the assumptions that households enter the labor market with no wealth and receive equal pensions. Moreover, the low number of cohorts limits the possibility to generate substantial wealth inequality.

Characteristics	Source	Data	Model
Real interest rate [%]	OECD (2016)	6.4	6.4
Average total wealth	NBP (2015), CSO (2016a)	6.8	8.6
Homeownership rate [%]	NBP (2015)	77.4	90.9
Average housing wealth	NBP (2015), CSO (2016a)	5.4	6.5
Fraction of households in debt [%]	NBP (2015)	37.0	33.9
Average debt	NBP (2015), CSO (2016a)	0.84	0.67
Gini coefficient for wealth	NBP (2015)	0.579	0.368

Table 2. The model's fit for the baseline calibration.

4 Results

In table 3, we report changes in the mean values of the selected characteristics caused by changes in the parameters of the model. We separately study the effects for eight different parameters. For each parameter, we always consider two reasonable alternative values: the lower and the higher than the baseline one. The rental housing depreciation rate δ_r, transaction costs ϕ and the economy openness parameter ξ have limited impact on the considered variables. The effects generated by the other parameters are discussed in more details below.

Calibration	**Interest rate**	**Wealth**	**Homeown. rate**	**Housing wealth**	**Househol. in debt**	**Average debt**	**Gini for wealth**
Baseline	**6.4**	**8.6**	**90.9**	**6.5**	**33.9**	**0.67**	**0.368**
$\alpha = 0.25$	-0.7	-4.7	5.5	-1.5	5.1	4.5	-0.3
$\alpha = 0.4$	1.6	8.1	-19.2	1.5	-11.6	35.8	8.2
$\beta = 0.961$	0.8	-7	-6.4	-3.1	-1.1	23.9	2.4
$\beta = 0.981$	-0.7	5.8	3.1	3.1	0.8	-9	-0.5
$\delta_r = 0.02$	-0.1	0	0.3	0	0.2	-4.5	0
$\delta_r = 0.03$	-0.1	0	0.5	0	0.4	-4.5	-0.3
$h_{min} = 0.25$	-0.6	-4.7	9.1	-10.8	-4.2	-49.3	5.2
$h_{min} = 0.75$	0.1	1.2	-23.9	3.1	-3.1	11.9	8.4
$\gamma = 0.1$	0.3	-1.2	-5.1	0	-2	43.3	6
$\gamma = 0.2$	-0.1	0	0.5	-1.5	0	-7.5	-0.5
$\phi = 0.1$	-0.1	0	0.8	-1.5	0.6	-3	-0.8
$\phi = 0.05$	0	0	-0.1	1.5	0	1.5	0.8
$\theta = 0.65$	1.4	12.8	5.5	26.2	4.7	50.7	0.5
$\theta = 0.85$	-1.6	-14	-17.7	-30.8	-13.6	-62.7	7.6
$\xi = 1$	-0.2	-1.2	1.6	0	1.9	4.5	-0.3
$\xi = 2$	0.3	1.2	-0.8	0	-1.3	-6	0

For the percentage variables (interest rate, homeownership rate, households in debt) simple differences in percentage points from the baseline value ($\Delta X = X - X_b$, where X_b denotes a value for the baseline calibration) are reported; for the other characteristics relative percentage differences are calculated ($\Delta X_{rel} = 100 \cdot (X / X_b - 1)$).

Table 3. Results of the simulations under different parametrizations.

The capital share α plays the important role for almost all studied characteristics. It is positively related to the interest rate, wealth and wealth inequality. The rise in capital share significantly decreases the homeownership rate and the fraction of households in debt. Also the average value of debt is significantly nonlinearly affected by the changes in α.

The role of the discount factor β for macroeconomic characteristics is well known in the literature. It is inversely related to the interest rate and wealth inequality. Additionally, we show that the lower discount factor significantly increases average debt and reduces the homeownership rate. The latter characteristic is also sensitive to the changes in the minimum house value h_{min} which also affects housing wealth, the average debt level and, in a nonlinear fashion, the Gini coefficient for wealth. The downpayment constraint γ plays the important role for the mean value of debt. The looser constraint the higher the debt is which also leads to higher wealth inequality.

Finally, almost all characteristics are sensitive to changes in the share of nondurable consumption in the utility function θ. The rise in θ reduces wealth, homeownership rate, housing wealth, fraction of households in debt as well as the mean level of debt. It also increases wealth inequality considerably.

Conclusion

In the paper, we considered the standard OLG model with housing for Polish economy. We studied sensitivity of the selected characteristics to small changes in the key parameters to identify those who have the largest impact on the outcomes. We found that the share of nondurable consumption in the utility function plays the important role in shaping both the macro and the housing characteristics. Furthermore, the capital share and the minimum house size turned out to be important as well.

References

Acedański, J. (2016). Youth unemployment and welfare gains from eliminating business cycles – the case of Poland. *Economic Modelling*, 57, 248-262.

Auerbach, A., & Kotlikoff, L. (1987). *Dynamic fiscal policy*. Cambridge, MA: Cambridge University Press.

Bielecki, M., Goraus, K., Hagemejer, J., Makarski, K., & Tyrowicz, J. (2015a). Small assumptions (can) have a large bearing: evaluating pension system reform with OLG models. *Economic Modelling*, 48, 210-221.

Bielecki, M., Tyrowicz, J., Makarski, K., & Waniek, M. (2015b). Inequality in an OLG economy with heterogeneity within cohorts and pension systems. Working Papers 16/2015 (164), University of Warsaw, Faculty of Economic Sciences.

Chen, K. (2010). A life-cycle analysis of social security with housing. *Review of Economic Dynamics*, 13(3), 597-615.

Cheron, A., Hairault, J.O., & Langot, F. (2011). Age-dependent employment protection. *Economic Journal*, 121(557), 1477-1504.

CSO. (2016a). *Households' budgets in 2015*. Warsaw, Central Statistical Office.

CSO. (2016b). *Life expectancy in 2015*, Warsaw, Central Statistical Office.

Diamond, P. (1965). National debt in a neoclassical growth model. *American Economic Review*, 55(5), 1126-1150.

Doepke, M., Schneider, M., & Selezneva, V. (2015). Distributional effects of monetary policy. Hutchins Center Working Paper No. 14.

Gertler, M. (1999). Government debt and social security in a life-cycle economy. *Carnegie-Rochester Conference Series on Public Policy*, 50(1), 61-110.

Hairault, J.O., & Langot, F. (2007). Inequality and social security reforms. *Journal of Economic Dynamics and Control*, 32(2), 386-410.

Kindermann, F., & Krueger, D. (2014). High marginal tax rates on the top 1%? Lessons from a life cycle model with idiosyncratic income risk. NBER Working Papers 20601.

NBP. (2015). *Households' wealth in Poland*, Warsaw, National Bank of Poland (in Polish).

OECD. (2016). https://data.oecd.org/ (access: 10.12.2016)

Rubaszek, M. (2012). Mortgage down-payment and welfare in a life-cycle model. *Bank and Credit*, 43(4), 5-28.

Digital economy in Polish regions. Proposal of measurement via TOPSIS with generalized distance measure GDM

Adam P. Balcerzak[1], Michał Bernard Pietrzak[2]

Abstract

Investing in digital infrastructure and building effective digital economy is currently considered as a basic condition for keeping international competitiveness of developed economies. In the case of developing countries in some economic models it is considered as a factor that can help to avoid middle income trap. From the regional perspective developing digital economy can support the process of convergence and closing development gap between the regions. As a result, a comparative research concerning the development level of digital economy is an important scientific task. In this context, the aim of the article is to assess and compare the development level of digital economy in Polish regions (NUTS 1). The digital economy is commonly considered as a multiple-criteria phenomenon. Thus, an approach based on TOPSIS method with application of generalized distance measure GDM was used in the analysis. Six diagnostic variables concerning digital infrastructure and level of its utilization were used. The research was conducted for the years 2012-2015 with application of Eurostat Data. The conducted analysis confirmed relatively quick progress in the field of building digital economy obtained by Polish regions.

Keywords: *digital economy, multiple-criteria analysis, TOPSIS, generalized distance measure GDM*

JEL Classification: P25, C38

1 Introduction

Supporting development of a digital economy both at national and regional level is currently considered as a basic policy objective for all European governments (Balcerzak, 2016; Ciburiene, 2016; Kondratiuk-Nierodzińska, 2016; Kordalska and Olczyk, 2016; Kryk, 2016; Balcerzak and Pietrzak, 2016a, 2016b; Pietrzak and Balcerzak, 2016a, 2016b; Pohulak-Żołędowska, 2016; Shuaibu and Oladayo, 2016; Zemtsov et al., 2016; Żelazny and Pietrucha, 2017). In the case of countries that face the problem of closing technological and development gap, it can be a factor helping to avoid a middle income trap. From the regional perspective it can support the development of regions that are peripheral both form socio-economic and geographic perspective. In this context the aim of the article is to assess and compare the development level of digital economy in Polish regions (NUTS 1) in the years 2012-2015.

[1] Corresponding author: Nicolaus Copernicus University/Department of Economics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: adam.balcerzak@umk.pl.

[2] Nicolaus Copernicus University/Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: michal.pietrzak@umk.pl.

The digital economy is often treated as a multiple-criteria phenomenon. As a result, in the research TOPSIS method with application of generalized distance measure GDM is used. In the research a hypothesis stating that in Poland at the regional level an improvement of the level of development of digital economy can be seen.

2 TOSPIS method with generalized distance measure GDM

As it was stated the subject of the research is a multiple-criteria problem. Thus, in the case of regional research it can be analyzed with the application of taxonomic measure of development (*TMD*) method originally proposed by Hellwig (see Balcerzak, 2016). In order to apply this method an analyzed phenomenon is divided into a set of economic aspects, where every separate aspect describes different part of phenomenon under consideration. For every aspect a set of diagnostic variables is chosen, which should describe the aspect. Additionally, the variables should be characterized with high information quality. Next based on the diagnostic variables a synthetic taxonomic measure of development is assessed (Pietrzak et al., 2013; Balcerzak, 2016; Balcerzak and Pietrzak, 2016a; Łyszczarz, 2016; Małkowska and Głuszak, 2016; Pietrzak and Balcerzak, 2016b). The obtained taxonomic measure of development covers the impact of all determinants of a given phenomenon and enables a synthetic assessment of its level. The procedure proposed by Hellwig is based on the comparison of the object to pattern of development (also often named in the literature as a positive ideal solution), which is set based on a maximum value of diagnostic variable in the case of stimulants. This procedure was extended with the proposal of comparing the objects also to anti-pattern of development (named as a negative ideal solution). As a result of this extension TOPSIS method was popularized in the literature (see Balcerzak, 2016; Walesiak, 2016).

A key factor in the case of assessing *TMD* measure is a choice of metric which is used for estimating the distance of the objects from pattern and anti-pattern of development. Generalized distance measure GDM introduced by Walesiak is considered as a metric, which can be applied for the variables measured on the ratio scale, interval scale, the ordinal scale or the nominal scale (see Walesiak, 2016). In the economic research variables based on the ordinal scale are commonly used. In order to use properly that kind of variables in multiple-criteria analysis the application of generalized distance measure GDM is necessary. As a result the main advantage of generalized distance measure GDM is its especially high application universality for the variables measured in different scales.

The procedure of assessing of *TMD* based on TOPSIS method with generalized distance measure GDM is described by Walesiak (2016). The values of generalized distance measure

GDM from pattern of development (GDM_{it}^{P}) and anti-pattern of development (GDM_{it}^{AP}) are assessed with equations 1 and 2 (see Walesiak, 2016)

$$GDM_{it}^{P} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ijt} - P_{kj})(P_{kj} - z_{ijt}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l \neq i,k}}^{n}(z_{ijt} - z_{ljt})(P_{kj} - z_{ljt})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ijt} - z_{ljt})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(P_{kj} - z_{ljt})^2\right]^{\frac{1}{2}}}, \tag{1}$$

$$GDM_{it}^{AP} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ijt} - AP_{kj})(AP_{kj} - z_{ijt}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l \neq i,k}}^{n}(z_{ijt} - z_{ljt})(AP_{kj} - z_{ljt})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ijt} - z_{ljt})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(AP_{kj} - z_{ljt})^2\right]^{\frac{1}{2}}} \tag{2}$$

where $i, l = 1, \ldots, n$ – number of the object, k – number of pattern of development and anti-pattern of development, $j = 1, \ldots, m$ – number of variable, Z_{ijt} – normalized diagnostic variable, P_j – pattern of development, AP_j – anti-pattern of development.

The GDM measure given in equation 1 and 2 should be applied for the variables measured on the interval scale or the ratio scale. There is also a version of GDM measure for the variables measured on the ordinal scale (Walesiak, 2016).

3 Assessment of level of digital economy in Poland at regional level

The empirical aim of the article was to assess the level of development in Poland at NUTS 1 level. The research was conducted for the years 2012 and 2015. In the research a set of diagnostic variables suggested by Eurostat as potential measures of digital economy were used (see table 1). The data for the variables is available in the Eurostat service: http://ec.europa.eu/eurostat.

In order to obtain *TMD* for digital economy we applied TOPSIS method with generalized distance measure GDM and constant patterns and anti-patterns of development for both years of the analysis, which was described in the previous section. In the research standardization based on arithmetic average and standard deviation was used. The estimation was conducted in the R-Cran software – Package ‘clusterSim’, Searching for Optimal Clustering Procedure for a Data Set, R package version 0.45-1.

Digital economy	
X_1 – Individuals who ordered goods or services over the internet for private use	stimulant
X_2 – Individuals who have never used a computer	dis-stimulant
X_3 – Households with access to the internet at home	stimulant
X_4 – Individuals who accessed the internet away from home or work	stimulant
X_5 – Individuals who used the internet, frequency of use and activities	stimulant
X_6 – Households with broadband access	stimulant

Table 1. Diagnostic variables for digital economy.

Digital economy in Polish regions							
	2012				**2015**		
Region	TMD	Rank	Class	Region	TMD	Rank	Class
Central Region	0.649	1	1	Central Region	0.975	1	1
Southern Region	0.464	2	1	South-western Region	0.928	2	1
Northern Region	0.167	3	2	Southern Region	0.808	3	2
North-western Region	0.135	4	2	North-western Region	0.806	4	2
South-western Region	0.101	5	2	Southern Region	0.733	5	3
Eastern Region	0.030	6	3	Eastern Region	0.649	6	3

Table 2. Ranking and grouping of regions based on the level of digital economy.

Next based on the diagnostic variables given in table 1 the level of digital economy in Polish NUTS1 was assessed. Based on the obtained value of TMD measure a ranking of regions for both years was given. Additionally, based on the natural breaks method, which consists of minimization of variance for objects from the chosen subsets and maximization of variance between the subsets, the regions were grouped to one of three classes, where the 1 class was grouping the regions with highest level of development of digital economy and 3 class the once with its lowest level. The results are given in table 2 and figure 1.

The obtained results indicate that during the analysed period the level of digital economy in Polish regions was significantly improved. It can be seen in the increased values of TMD in the year 2015 compared to the year 2012. This indicates high dynamics of the process of digitalisation of Polish economy. In this context, it remains an open question what is a maximum saturation point for the economy in terms of digitization?

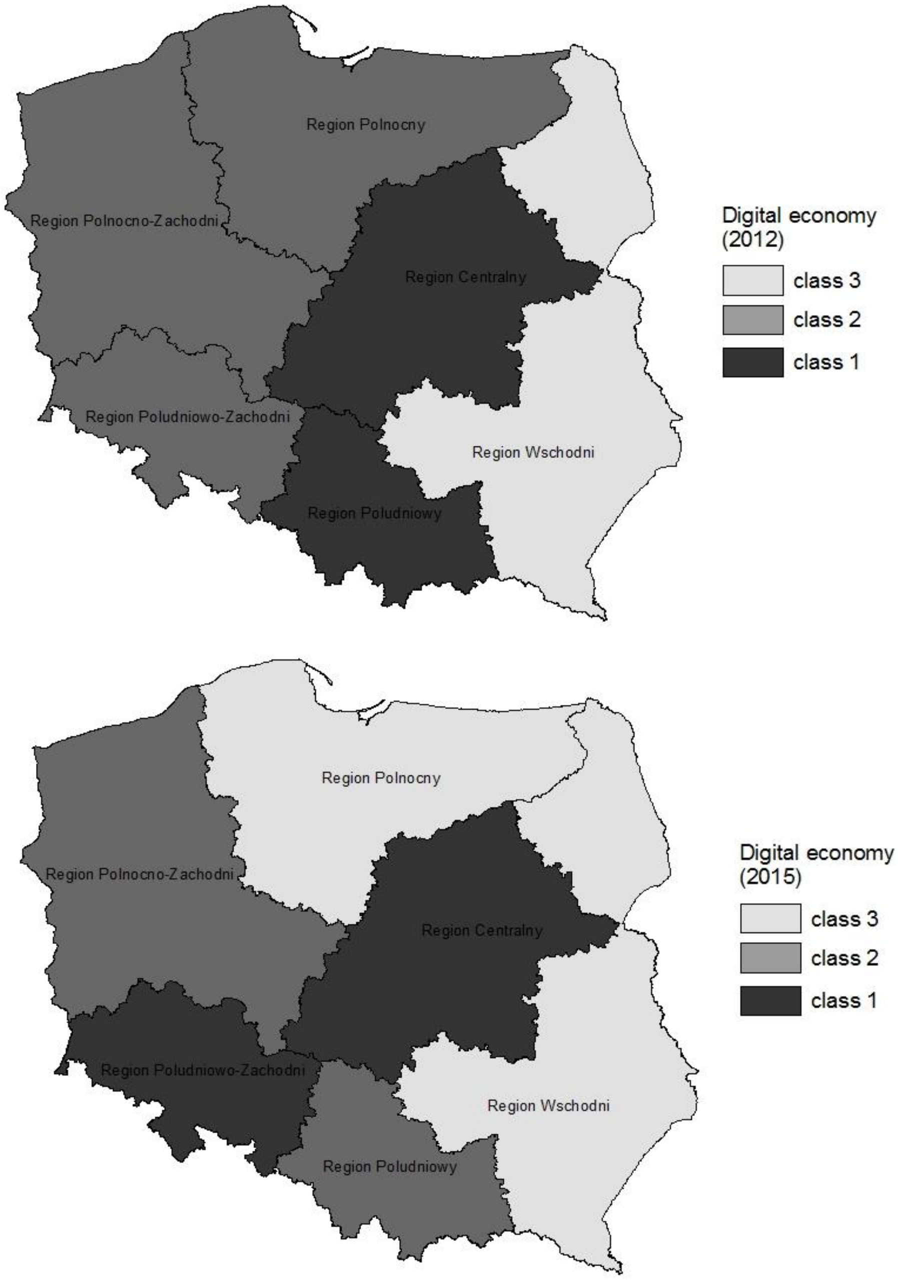


Fig. 1. The level of digital economy in Poland in the year 2012 and 2015.

In the first class grouping the regions with the highest level of digital economy development in the year 2012 one could find Central Region and Southern Region, whereas in the year 2015 these were Central Region and South-western Region. Thus, in both years

Central Region, which covers the most developed form the socio-economic perspective Mazovia Voivodeship with the capital city of the county, is the most digitalised region.

In the 3 class grouping the regions with the lowest level of digital economy development in the year 2012 one could find only Eastern Regions. In the year 2015 in the third class there could be seen two regions Eastern Region and Northern Region.

In the 2 class with average level of analyzed phenomenon in the year 2012 there were: Northern Region, North-western and South-western Regions and in the year 2015 these were two regions: Northern Region and North-western Region. A relatively small increase of the level of TMD for digital economy in the case of Northern Region should be stressed, which in the year 2015 was degraded form the 1 to the 2 class, and relatively quick speed of development of digital economy in the case of South-western Region, which in the year 2015 was promoted from class 2 to 1. These dynamics can confirm that the investment in digital economy can be used by relatively undeveloped regions in order to improve their growth potential in relatively short time.

Conclusions

The research presented in the article was devoted to the multiple-criteria analysis of development of digital economy at regional level in Poland. The research was conducted at NUTS 1 level in the years 2012 and 2015. In the analysis TOPSIS method with application of generalized distance measure GDM was used. The conducted analysis confirmed the research hypothesis, according to which a quick improvement of the level of digital economy in Poland at regional level can be recorded.

The conducted research also shows that the improvements in the case of level of development of digital economy can obtained relatively quickly by less developed regions, which can support their convergence potential with the developed regions.

References

Balcerzak, A. P. (2016). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.

Balcerzak, A. P. & Pietrzak, M. P. (2016a). Application of TOPSIS Method for Analysis of Sustainable Development in European Union Countries. In Loster, T. & Pavelka, T. (eds.). *The 10th International Days of Statistics and Economics. Conference Proceedings.* September 8-10, 2016. Prague: Libuse Macakova, Melandrium, 82-92.

Balcerzak, A.P. & Pietrzak, M. B. (2016b). Dynamic Panel Analysis of Influence of Quality of Human Capital on Total Factor Productivity in Old European Union Member States. In Kliestik, T. (eds.). *16th International Scientific Conference Globalization and Its Socio-Economic Consequences*. University of Zilina, The Faculty of Operation and Economics of Transport and Communication, Department of Economics, 5th – 6th October 2016. (Part I.). Zilina: University of Zilina, 96-103.

Ciburiene, J. (2016). The Evaluation of Economic Development Index of Poland and Lithuania in the Context of the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 437-449.

Kondratiuk-Nierodzińska, M. (2016). New Knowledge Generation Capabilities and Economic Performance of Polish Regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 451-471.

Kordalska, A., & Olczyk, M. (2016). Global Competitiveness and Economic Growth: A One-Way or Two-Way Relationship?. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(1), 121-142.

Kryk, B. (2016). Accomplishment of the European Union Lifelong Learning Objectives in Poland. *Oeconomia Copernicana*, 7(3), 389-404.

Łyszczarz, B. (2016). Public-private Mix and Performance of Health Care Systems in CEE and CIS Countries. *Oeconomia Copernicana*, 7(2), 169-185.

Małkowska, A., & Głuszak, M. (2016). Pro-investment Local Policies in the Area of Real Estate Economics - Similarities and Differences in the Strategies Used by Communes. *Oeconomia Copernicana*, 7(2), 269-283.

Pietrzak, M. B. & Balcerzak, A. P (2016a). Quality of Human Capital and Total Factor Productivity in New EU Member States. In T. Loster & T. Pavelka (eds.). *The 10th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016*. Prague: Libuse Macakova, Melandrium, 1492-1501.

Pietrzak. M. B. & Balcerzak, A. P. (2016b). Socio-economic Sustainability in Poland. SEM Analysis at Regional Level. In Kliestik, T. (eds.). *16th International Scientific Conference. Globalization and Its Socio-Economic Consequences*. University of Zilina, The Faculty of Operation and Economics of Transport and Communication, Department of Economics, 5th – 6th October 2016. (Part IV.).Zilina: University of Zilina, pp. 1704-1712.

Pietrzak M. B, Wilk J., & Siekaniec M. (2013). The impact of metropolitan areas on internal migrations in Poland. The case of southern regions. In Papież M. & Śmiech S. (eds.). *Proceedings of the 7th Professor Aleksander Zelias International Conference on*

Modelling and Forecasting of Socio-Economic Phenomena. Cracow: Foundation of the Cracow University of Economics, 124-132.

Pohulak-Żołędowska, E. (2016). Innovation in Contemporary Economies. *Oeconomia Copernicana*, 7(3), 451-466.

Shuaibu, M., & Oladayo, P. T. (2016). Determinants of Human Capital Development in Africa: A Panel Data Analysis. *Oeconomia Copernicana*, 7(4), 523-549.

Walesiak, M. (2016). *Uogólniona miara odległosci GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wrocław: Wydawnictwo UE.

Zemtsov S. P., Pavlov, P. N., & Sorokina, A. V. (2016). Specifics of Cluster Policy in Russia. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 499-536.

Żelazny R., & Pietrucha J. (2017). Measuring Innovation and Institution: the Creative Economy Index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1).

Generalization of the geo-logarithmic price index family

Jacek Białek[1]

Abstract

The paper presents a proposition of the geo-logarithmic index family generalization. It is shown that the Fisher and Walsh price indices are particular cases of this family and, under some assumptions, that the Törnqvist index can be its approximation. Due to the fact that the above-mentioned indices are superlative, they are best proxies for the Cost of Living Index (COLI). In practice, the Consumer Price Index (CPI), which is calculated by using the Laspeyres formula, is a measure of inflation but according to the economic approach in price index theory, a well-defined index should satisfy the Laspeyres-Paasche bounding test. In the simulation study we verify this test in the case of the generalized geo-logarithmic index family. The general conclusion is that values of almost all indices from the mentioned family are in the interval with bounds determined by the Laspeyres and Paasche price indices.

Keywords: *CPI, COLI, the Laspeyres index, geo-logarithmic price index family*

JEL Classification: E1, E2, E3

1 Introduction

The literature on axiomatic index theory is very wide (Krstcha, 1988; Balk, 1995). From a theoretical point of view, a well-constructed index should satisfy a group of postulates (tests) coming from the axiomatic index theory. A system of minimum requirements of an index comes from Marco Martini (1992b). According to the above-mentioned system a price index should satisfy at least three conditions: *identity*, *commensurability* and *linear homogeneity* (von der Lippe, 2007).

As it is known, all geo-logarithmic indices are proportional, commensurable and homogeneous, together with their cofactors (Martini, 1992a). Geo-logarithmic price indices satisfying the axioms of monotonicity, basis reversibility and factor reversibility have been investigated in the paper of Marco Fattore (2010). In the mention paper it is shown that the superlative Fisher price index does not belong to this family of indices. This paper presents a proposition of the geo-logarithmic index family generalization. It is shown that the Fisher and Walsh price indices are particular cases of this family and, under some assumptions, that the Törnqvist index can be its approximation. In the simulation study according to the economic approach in the price index theory we verify the Laspeyres-Paasche bounding test (von der Lippe (2007)) in the case of the generalized geo-logarithmic index family.

[1] University of Lodz, Department of Statistical Methods, 90-255 Łódź, 3/5 POW Street, e-mail: jbialek@uni.lodz.pl.

2 Geo-logarithmic price index family

Let us consider a group of N commodities observed at times s, t and let us denote[2]:

$p_s = [p_{s1}, p_{s2},..., p_{sN}]'$ - a vector of prices at time s;

$p_t = [p_{t1}, p_{t2},..., p_{tN}]'$ - a vector of prices at time t;

$q_s = [q_{s1}, q_{s2},..., q_{sN}]'$ - a vector of quantities at time s;

$q_t = [q_{t1}, q_{t2},..., q_{tN}]'$ - a vector of quantities at time t.

Let us denote by $\tau(x, y)$ the logarithmic mean of two positive real numbers x and y, *i.e.*

$$\tau(x, y) = \frac{x - y}{\ln(x) - \ln(y)}, \tag{1}$$

if $x \neq y$ and $\tau(x, y) = x$ otherwise (Carlson, 1972).

For $x, y \in [0,1]$, let q^x and q^y be two vectors, whose components are defined as follows

$$q_i^x = q_{ti}^x q_{si}^{1-x},\ q_i^y = q_{ti}^y q_{si}^{1-y},\ \text{for } i = 1,2,..., N \tag{2}$$

and let

$$w_{si}^x = \frac{p_{si} q_i^x}{\sum_{i=1}^{N} p_{si} q_i^x}, \tag{3}$$

$$w_{ti}^y = \frac{p_{ti} q_i^y}{\sum_{i=1}^{N} p_{ti} q_i^y}. \tag{4}$$

The geo-logarithmic, or the P_{xy}, family is the class of price indices defined by Fattore (2006)

$$P_{xy}(q_s, q_t, p_s, p_t) = \prod_{i=1}^{N} \left(\frac{p_{ti}}{p_{si}}\right)^{\nu_i^{xy}} \tag{5}$$

where weights ν_i^{xy} are as follows

$$\nu_i^{xy} = \frac{\tau(w_{si}^x, w_{ti}^y)}{\sum_{j=1}^{N} \tau(w_{sj}^x, w_{tj}^y)}. \tag{6}$$

It is easy to verify that (Fattore, 2010)

$$P_{La} = \frac{\sum_{i=1}^{N} p_{ti} q_{si}}{\sum_{i=1}^{N} p_{si} q_{si}} = P_{00}, \tag{7}$$

[2] The time moment s is considered as the *basis*, i.e. the reference situation, for the comparison.

$$P_{Pa} = \frac{\sum_{i=1}^{N} p_{ti} q_{ti}}{\sum_{i=1}^{N} p_{si} q_{ti}} = P_{11}, \tag{8}$$

$$P_{W} = \frac{\sum_{i=1}^{N} p_{ti} \sqrt{q_{si} q_{ti}}}{\sum_{i=1}^{N} p_{si} \sqrt{q_{si} q_{ti}}} = P_{0.5\ 0.5} \tag{9}$$

where P_{La}, P_{Pa} and P_W denote the Laspeyres, Paasche and Walsh price indexes respectively. It is a very desirable observation that not only Laspeyres and Paasche indices (used in practice) are particular cases of the geo-logarithmic price index family but also the superlative Walsh price index. Nevertheless, as it was mentioned before, the most important, superlative Fisher price index does not belong to this family.

3 Generalization of the geo-logarithmic price index family

Similarly to (2), (3), (4) and (6) let us denote by

$$q_i^{Ax} = q_{ti}^{x} q_{si}^{1-x},\ q_i^{Ay} = q_{ti}^{y} q_{si}^{1-y}, \tag{10}$$

$$q_i^{Bx} = q_{ti}^{1-x} q_{si}^{x},\ q_i^{By} = q_{ti}^{1-y} q_{si}^{y}, \tag{11}$$

$$w_{si}^{Ax} = \frac{p_{si} q_i^{Ax}}{\sum_{i=1}^{N} p_{si} q_i^{Ax}},\ w_{ti}^{Ay} = \frac{p_{ti} q_i^{Ay}}{\sum_{i=1}^{N} p_{ti} q_i^{Ay}}, \tag{12}$$

$$w_{si}^{Bx} = \frac{p_{si} q_i^{Bx}}{\sum_{i=1}^{N} p_{si} q_i^{Bx}},\ w_{ti}^{By} = \frac{p_{ti} q_i^{By}}{\sum_{i=1}^{N} p_{ti} q_i^{By}}, \tag{13}$$

$$v_{Ai}^{xy} = \frac{\tau(w_{si}^{Ax}, w_{ti}^{Ay})}{\sum_{j=1}^{N} \tau(w_{sj}^{Ax}, w_{tj}^{Ay})},\ v_{Bi}^{xy} = \frac{\tau(w_{si}^{Bx}, w_{ti}^{By})}{\sum_{j=1}^{N} \tau(w_{sj}^{Bx}, w_{tj}^{By})}, \tag{14}$$

for $i = 1,2,..., N$, $x, y \in [0,1]$.

Under significations (10)-(14) we define the new class of price indices ($\widetilde{P}_{xyz}$) as follows

$$\widetilde{P}_{xyz} = \{\prod_{i=1}^{N} (\frac{p_{ti}}{p_{si}})^{v_{Ai}^{xy}}\}^{z} \{\prod_{i=1}^{N} (\frac{p_{ti}}{p_{si}})^{v_{Bi}^{xy}}\}^{1-z},\ z \in [0,1]. \tag{15}$$

Obviously, these two price indices (defined inside curly brackets in (15)) satisfy the Martini's minimal requirements. The first one is identical with P_{xy} index (for fixed values of x and y) and his axiomatic properties were proved by Fattore (2010). The proof of the same group of axioms in the case of the second price index (being inside curly bracket on the right side of (15)) would be analogous. Thus, since a formula (15) is a geometric mean of two general indices satisfying Martini's requirements, we can conclude that each index from $\widetilde{P}_{xyz}$ also satisfies *identity, commensurability* and *linear homogeneity* (Białek, 2012). Moreover, it holds that

$$\widetilde{P}_{xy1} = \widetilde{P}_{1-x\ 1-y\ 0} = P_{xy}\,, \tag{16}$$

$$\widetilde{P}_{001} = \widetilde{P}_{110} = P_{La}\,, \tag{17}$$

$$\widetilde{P}_{111} = \widetilde{P}_{000} = P_{Pa}\,, \tag{18}$$

$$\widetilde{P}_{\frac{1}{2}\frac{1}{2}\frac{1}{2}} = P_W\,, \tag{19}$$

and, what is more interesting, the Fisher price index belongs to the considered class since

$$\widetilde{P}_{00\frac{1}{2}} = \widetilde{P}_{11\frac{1}{2}} = P_F\,. \tag{20}$$

The relation (16) means that the P_{xy} family is a special case of the $\widetilde{P}_{xyz}$ family. It can be also proved[3] that the following approximation holds

$$\forall i \in \{1,2,...,N\}\ \ q_{si} \approx q_{ti}\ \wedge\ w_{si}^{x} \approx w_{ti}^{y}\ \ \Rightarrow\ \ \widetilde{P}_{xyz} \approx P_T \tag{21}$$

where P_T denotes the Törnqvist price index defined as

$$P_T = \prod_{i=1}^{N} \left(\frac{p_{ti}}{p_{si}}\right)^{\frac{w_{si}^{0}+w_{ti}^{1}}{2}}\,. \tag{22}$$

4 Geo-logarithmic price index family and its generalization vs the Laspeyres-Paasche bounding test

The *Consumer Price Index* (CPI) is commonly used as a basic measure of inflation. The index approximates changes in the costs of household consumption assuming the constant utility (COLI, *Cost of Living Index*). In the so called economic approach, upper and lower bounds for the COLI are provided by the Laspeyres and Paasche price index formulas. These bounds are obtained under the assumption about cost minimizing behavior. If the price index value is within these bounds then we say that this price index satisfies the Laspeyres-Paasche bounding test belonging to the group of mean value tests (von der Lippe, 2007).

[3] The proof of this approximated will be presented in the extended version of the paper.

4.1 Simulation study

Let us take into consideration a group of $N=12$ components, where prices and quantities are normally distributed as follows: $p_i^\tau \sim N(p_{i0}^\tau, v_i^\tau p_{i0}^\tau)$, $q_i^\tau \sim N(q_{i0}^\tau, u_i^\tau q_{i0}^\tau)$, where $\tau = s,t$, $N(\mu,\sigma)$ - denotes the normal distribution with the mean μ and the standard deviation σ, v_i^τ - denotes the volatility coefficient of the $i-$th price at time τ, i.e. $v_i^\tau = D(p_i^\tau)/p_{i0}^\tau$, u_i^τ - denotes the volatility coefficient of the $i-$th quantity at time τ, i.e. $u_i^\tau = D(q_i^\tau)/q_{i0}^\tau$. Before generating prices and quantities we generated values of volatility coefficients using uniform distributions, i.e. $v_i^\tau \sim U(0,v^\tau)$ and $u_i^\tau \sim U(0,u^\tau)$. Expected values of prices and quantities are described by following vectors

$$P_0^t = [1000,\ 1700,\ 500,\ 3.2,\ 105,\ 1150,\ 1000,\ 1600,\ 500,\ 4.2,\ 110,\ 1100]';$$

$$P_0^s = [900,\ 1600,\ 460,\ 3,\ 100,\ 1000,\ 900,\ 1530,\ 480,\ 4,\ 100,\ 1000]';$$

$$Q_0^t = [200,\ 200,\ 3000,\ 500,\ 340,\ 700,\ 800,\ 500,\ 3000,\ 500,\ 340,\ 700]';$$

$$Q_0^s = [350,\ 550,\ 5000,\ 710,\ 350,\ 890,\ 850,\ 600,\ 5000,\ 700,\ 550,\ 800]'.$$

In our experiment we are going to control values of volatility coefficients of prices and quantities by setting values of v^s, v^t, u^s, u^t. We consider here only two cases[4]: Case 1 (volatilities of price and quantity processes are small and the quantity response on price changes is quite normal); Case 2 (volatilities of prices and quantities are large, the quantity response on price changes is strongly fluctuated). For each case we generate values of vectors of prices and quantities in $n=1000$ repetitions. Let us denote for fixed values of x and y and for each of $k-th$ repetition: $\Delta 1_k = (P_{xy} - \min(P_{La}, P_{Pa}))_k$, $\Delta 2_k = (\max(P_{La}, P_{Pa}) - P_{xy})_k$, $\Delta 3_k = (\widetilde{P}_{xy\frac{1}{2}} - \min(P_{La}, P_{Pa}))_k$, $\Delta 4_k = (\max(P_{La}, P_{Pa}) - \widetilde{P}_{xy\frac{1}{2}})_k$. The simulation results are presented in Tab. 1 and Tab. 2.

[4] Other cases will be presented in the extended version of the paper.

Statistics	$x = 0.05$ $y = 0.05$	$x = 0.25$ $y = 0.25$	$x = 0.5$ $y = 0.5$	$x = 0.75$ $y = 0.75$	$x = 0.95$ $y = 0.05$	$x = 0.95$ $y = 0.95$	$x = 0.05$ $y = 0.05$
Mean (P_{xy})	1.101	1.097	1.102	1.087	1.086	1.103	1.101
(Std. Dev. $_{P_{xy}}$)	(0.038)	(0.035)	(0.038)	(0.030)	(0.031)	(0.037)	(0.038)
Mean ($\widetilde{P}_{xy0.5}$)	1.103	1.097	1.102	1.089	1.086	1.104	1.103
(Std. Dev. $_{\widetilde{P}_{xy0.5}}$)	(0.038)	(0.034)	(0.038)	(0.031)	(0.031)	(0.038)	(0.038)
$card\{k : \Delta_{1k} < 0\}$	28	26	17	21	15	16	28
$card\{k : \Delta_{2k} < 0\}$	20	19	22	23	12	30	20
$card\{k : \Delta_{3k} < 0\}$	7	25	17	15	15	2	7
$card\{k : \Delta_{4k} < 0\}$	4	14	22	17	9	4	4

Table 1. Verifying the Laspeyres- Paasche bounding test for P_{xy} and $\widetilde{P}_{xy0.5}$ - Case 1 ($v^s = v^t = 0.05$; $u^s = u^t = 0.05$).

Statistics	$x = 0.05$ $y = 0.05$	$x = 0.25$ $y = 0.25$	$x = 0.5$ $y = 0.5$	$x = 0.75$ $y = 0.75$	$x = 0.95$ $y = 0.05$	$x = 0.95$ $y = 0.95$	$x = 0.05$ $y = 0.05$
Mean (P_{xy})	1.066	1.143	1.103	1.043	1.067	1.143	1.066
(Std. Dev. $_{P_{xy}}$)	(0.127)	(0.140)	(0.126)	(0.130)	(0.134)	(0.134)	(0.127)
Mean ($\widetilde{P}_{xy0.5}$)	1.082	1.148	1.103	1.023	1.067	1.131	1.082
(Std. Dev. $_{\widetilde{P}_{xy0.5}}$)	(0.124)	(0.141)	(0.126)	(0.137)	(0.134)	(0.132)	(0.124)
$card\{k : \Delta_{1k} < 0\}$	28	28	39	44	21	34	28
$card\{k : \Delta_{2k} < 0\}$	22	25	32	30	33	31	22
$card\{k : \Delta_{3k} < 0\}$	3	26	39	34	19	3	3
$card\{k : \Delta_{4k} < 0\}$	7	20	32	21	27	5	7

Table 2. Verifying the Laspeyres- Paasche bounding test for P_{xy} and $\widetilde{P}_{xy0.5}$ - Case 2 ($v^s = v^t = 0.2$; $u^s = u^t = 0.2$).

Conclusions

The proposed and wide class of price indices ($\widetilde{P}_{xyz}$) has similar axiomatic properties as the geo-logarithmic price index family and in particular, each index from this family satisfies the Martini's minimal requirements. Moreover, the "ideal", superlative[5] Fisher price index belongs to the proposed family. In the simulation study we observe that indices P_{xy} and $\widetilde{P}_{xy0.5}$ may differ strongly for $x \neq 0.5$ or $y = 0.5$, i.e. we observe some differences[6] between expected index values (arithmetic means) calculated for their generated values. The precision of estimation of P_{xy} and $\widetilde{P}_{xy0.5}$ indices, i.e. the standard deviations of their generated values, are comparable with respect to size and they don't seem to depend on x and y. This is a practical conclusion: even if fluctuations of prices and quantities are large we observe similar volatility among price indices[7] from the same general class of indices. Finally, the most crucial difference between compared general class of indices is that the *probability*[8] of satisfying the Laspeyres-Paasche bounding test is bigger in the case of $\widetilde{P}_{xy0.5}$ index (it is much bigger for small (near 0) and big (near 1) values of x and y. In other words, we observe relatively fewer cases when the value of $\widetilde{P}_{xy0.5}$ index is outside of the interval determined by the Laspeyres and Paasche price indexes in comparison to analogical cases for P_{xy} formula (see Tables 1-2).

References

Balk, B. M. (1995). Axiomatic Price Index Theory: A Survey. *International Statistical Review,* 63, 69-93.

Białek, J. (2012). Proposition of the general formula for price indices. *Communications in Statistics: Theory and Methods,* 41(5), 943-952.

Carlson, B. C. (1972). The logarithmic mean. *Amer. Math. Monthly*, 79, 615- 618.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4, 114-145.

[5] The term "superlative" comes from Diewert (1976). For more details see also Hill (2006).
[6] As a rule the difference between the above-mentioned indices is no more than 1 p.p.
[7] Obviously, in the simulation study we treat price indices as some random variables.
[8] The above-mention *probability* is estimated as a ratio of the number of generated cases when the considered price index fulfills the Laspeyres-Paasche bounding test and the total number of repetitions.

Fattore, M. (2006). On the monotonicity of the geo-logarithmic price indexes. *Società Italiana di Statistica*.

Fattore, M. (2010). Axiomatic properties of geo-logarithmic price indices. *Journal of Econometrics*, 156 (2), 344-353.

Hill, R. J. (2006). Superlative index numbers: not all of them are super. *Journal of Econometrics*, 130, 25-43.

Krstcha, M. (1988). Axiomatic characterization of statistical price indices. *Physica-Verlag, Heidelberg*.

Lippe, P. V. (2007). Index Theory and Price Statistics. *Peter Lang.*

Martini, M. (1992a). I numeri indice in un approccio assiomatico. *Giuffré: Milan.*

Martini, M. (1992b). General Function of Axiomatic Index Numbers. *J. Ital. Statist. Soc.*, 3, 359-376.

Multivariate statistical analysis of information society in Poland

Justyna Brzezińska[1]

Abstract

Widespread access to high-speed Internet, user-friendly public e-services, increasing digital competence of the society are main goals for the following years due to the latest reports published by the Central Statistical Office of Poland. They are included, inter alia, in the Operational Programme Digital Poland. This technological development is also connected with the development of economic areas and public services. The rapidly increase of significance of information and electronic services, and thus, of the application of information and communication technologies (ICT), in surrounding economy, public administration (central and local government) and in the everyday life of citizens has triggered a new transformation trend – a transformation towards the information society. This term describes a society for which the processing of information with the use of ICT solutions creates significant economic, social and cultural value.

In this paper we present the current state, main aspects, vision and mission of the information society in Poland, as well as statistical analysis of information society in Poland using multivariate statistical methods. All calculations based on data from Central Statistical Office are conducted in R software.

Keywords: *information society in Poland, multivariate statistical analysis, categorical data analysis, R software*
JEL Classification: O35, C30, C39
AMS Classification: 62P20, 52H25

1 Introduction

The Information Technology (IT) has a great impact on people, society, and economy. We present briefly some facts on information society in 2015 in Poland that will help to understand the speed and impact of Information Technology on economy (Kondratiuk-Nierodzińska 2016, Piech 2007). Information society and problems related to information society in modern word was also mentioned by Stephanidis (1998), Britz (2008), Sourbati (2011), Pohle (2015). From 2012 the percentage of large companies with access to the Internet was close to 100%, which indicates saturation phenomena in this group entities. In 2015 access to the Internet had 92.7% of the companies using mostly broadband (91.9%). Mobile broadband was used by over 61.5% of companies. The rate of enterprises with their own website in 2015 amounted to 65.4%. Nearly two-thirds of companies used their own website to present catalogs of products and services they serve. In 2014, every fifth company

[1] University of Economics in Katowice, Faculty of Management, Department of Economic and Financial Analysis, 1 Maja 50, 40-287 Katowice, e-mail: justyna.brzezinska@ue.katowice.pl.

consisted of orders via computer networks, and every tenth received orders via the Internet .In 2015 almost half of large enterprises used the social media . The services of cloud computing in 2015 has used one quarter of large enterprises (source: www.gus.pl).

The financial perspective for the period 2014 to 2020 has opened up new possibilities for supporting the development of the information society. Widespread and easy access to the Internet, user-friendly public e-services, increasing digital competence of the society are aims for the following years. Information and communication technologies (ICT) have bigger and bigger influence on our daily life. In particular, the Internet is creating a new layer of participation of individuals in the societal and economic life.

In this paper we present the comprehensive description of the dynamic change of the information society' in Poland, as well as multivariate statistical analysis of the strengths and weaknesses of the information society's situation in Poland using up-to-date factual data. We also present the application of advanced statistical R software with the use of multivariate statistical analysis methods for the comparative analysis of information society in Poland.

2 Facts and figures on the ICT sector and products

Based on reports published by the Central Statistical Office (www.stat.gov.pl) in the year 2014 the number of enterprises hiring 10 or more persons in the ICT sector amounted to 2146 (6.3% increase in comparison to the previous year) among which 89.1% offered ICT services. Almost three quarters of ICT service enterprises provided IT services. In comparison with 2011 the number of ICT enterprises was increasing systematically and was higher by 24.5% (of which service enterprises by 29.2%). The number of persons employed in the sector amounted to 196.4 thousand (an increase by 6.5% compared to the previous year and 10.7% compared to 2011) with persons hired in ICT services constituting over three quarters. IT services were also the field of activity in which enterprises hired the biggest number of persons of all employed in ICT services (66.1%). The value of net revenues from sales in the ICT sector increased by 8.8% in comparison with 2011 and amounted to over PLN 132 billion in 2014 (a slight decrease was only noted in 2013 compared to 2012 – 0,7%). Services, in particular telecommunications, had the biggest contribution in generating revenues of the ICT sector. In 2014 ICT manufacturing enterprises earned almost two thirds of revenue from export sales while ICT service enterprises – only 15.4%. In 2011-2013 these revenues were decreasing systematically in ICT manufacturing enterprises but in 2014 they recorded an annual growth of 8.3%. In service enterprises revenues from exports were increasing continually. The biggest share in these revenues had enterprises providing IT services (in

2014 – 63.5%). An increase of expenditures on R&D in the ICT sector was noted in the years 2011-2014 (by PLN 553 million). Enterprises offering ICT services incurred about 90% of expenditures on R&D in each surveyed year.

3 Information society

The Government of the Republic of Poland, having in mind the good of Poland and its inhabitants, is striving to ensure rapid and sustainable economic growth and social development that will improve the living conditions of our citizens. One of the key stimulants of economic growth is citizens' ability to acquire, accumulate and use information as a result of the dynamic development of Information and Communication Technologies (ICT). The importance of this factor for economic growth is confirmed by various research projects, which conclude that ICT account for approximately a quarter of the GDP growth and 40% of the productivity growth in the European Union. The rapidly increasing significance of information and electronic services, and thus, of the application of ICT, in the economy, public administration (central and local government) and in the everyday life of citizens has triggered a new transformation trend – a transformation towards the "information society". The term "information society", as adopted for the purposes of this paper, is defined as a society for which the processing of information with the use of ICT solutions creates significant economic, social and cultural value. This Strategy is sectoral and, as such, defines the vision and mission for the development of the information society in Poland until 2013. Within each of its three areas – Human, Economy and State – it maps out strategic directions and determines the objectives that should be accomplished in order to achieve the desired development status for the information society in Poland in 2013. The creation of the Strategy was preceded by a series of extensive consultations with experts who represented organizations and institutions that are most competent to express views on the issue of information society development.

It is possible to distinguish five definitions of an information society, each of which presents criteria for identifying the new. These are: technological, economic, occupational, spatial and cultural. These need not be mutually exclusive, though theorists emphasise one or other factors in presenting their particular scenarios. However, what these definitions share is the conviction that quantitative changes in information are bringing into being a qualitatively new sort of social system, the information society. In this way each definition reasons in much the same way: there is more information nowadays, therefore we have an information society. As we shall see, there are serious difficulties with this ex post facto reasoning that argues a cause from a conclusion (Webster 2006). Research on the links between the diffusion of Information and Communication Technologies (ICTs) and social and economic development has been undertaken for decades. Evidence of links between social and digital engagement, particularly with respect to the Internet, has been the focus of many studies conducted by academic as well as government institutions. These studies have shown consistently that individuals who have access to ICTs, from the telephone to the Internet, tend to have more schooling, higher incomes, and higher status occupations than do those who do not have access. To analyze this study in the next part of this paper we present statistical analysis based on data from the Central Statistical Office on information society in Poland.

4 The survey results and application in R

In this part of the paper we present multivariate statistical analysis of information society in Poland with the use of R software. We use the data from the report published by the Central Statistical Office entitled *Information Society in Poland*. We present analysis with the use of data on the number of enterprises and employees in the ICT sector in 2011-2014 (table 1). First, we present correspondence analysis for number of enterprises in the ICT sector in 2011-2014. The analysis is based on data from the Central Statistical Office on the number of enterprises in the ICT sector in 2011-2014.

Specification	2011	2012	2013	2014
ICT production	245	239	225	235
ICT wholesale	190	207	230	235
Telecommunications	219	231	258	289
IT services	1070	1181	1305	1387

Table 1. Number of enterprises in the ICT sector. Source: Central Statistical Office.

We present results of correspondence analysis for data presented in table 1.

```
Principal inertias (eigenvalues):
           1        2       3
Value      0.001629 9.3e-05 7e-06
Percentage 94.22%   5.38%   0.4%
```

The percentage being explained by first dimension is 94.22%, and for the second dimension is 5.38%. These two dimension explain 99.6% of the total inertia. Total inertia is

0.0017 showing that there is a very weal association between two variables: year and specification. Row and column masses, chi-square distance and inertia for each of the category of row and column are presented below.

Rows:

	ICT Production	ICT wholesale	Telecommunications	IT services
Mass	0.121869	0.111283	0.128712	0.638136
ChiDist	0.107785	0.015622	0.030357	0.016220
Inertia	0.001416	0.000027	0.000119	0.000168
Dim. 1	-2.669651	0.091573	0.490427	0.394954
Dim. 2	0.207310	-1.431880	2.366810	-0.267273

Columns:

	2011	2012	2013	2014
Mass	0.222566	0.239866	0.260522	0.277046
ChiDist	0.062238	0.023124	0.032682	0.040783
Inertia	0.000862	0.000128	0.000278	0.000461
Dim. 1	-1.532660	-0.524810	0.765233	0.966059
Dim. 2	0.664981	-0.877842	-1.068300	1.230398

We can display the result of a correspondence analysis in the form of perception map (fig. 1).

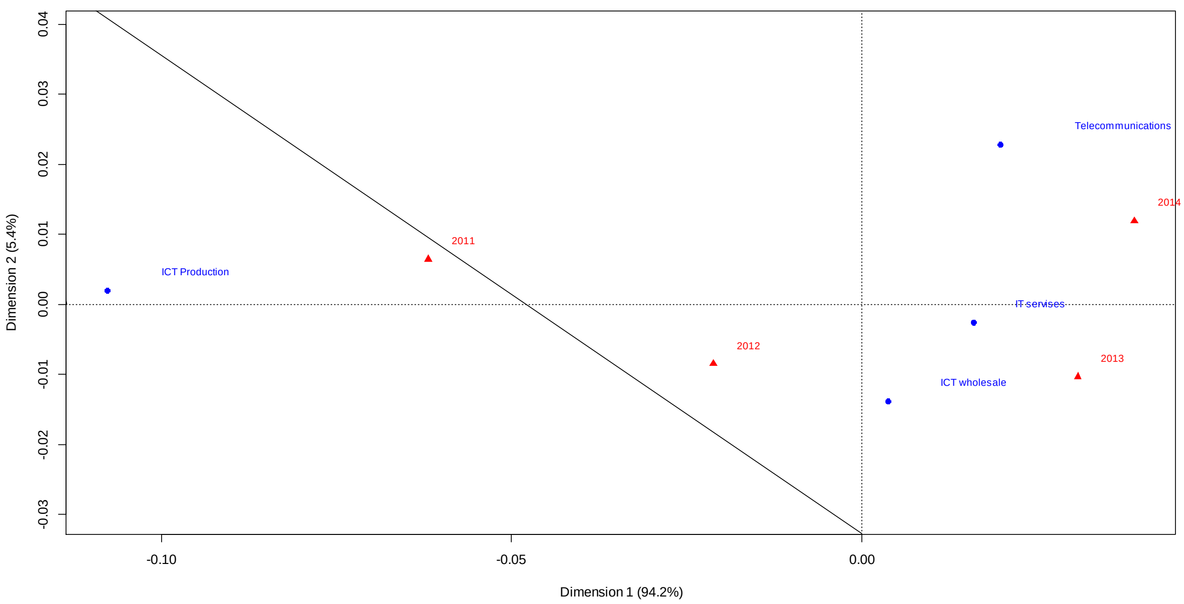


Fig. 1. Two-dimensional perception map for correspondence analysis for the number of enterprises in the ICT sector.

Looking at the graph above (figure 1) we see, the ICT production is situated very close to the year 2011. ICT wholesales and IT services are related with years 2012 and 2013. Finally, telecommunications is situated very near to the year 2014. This graphical presentation may suggest that there is a trend moving ICT production from 2011 to the telecommunications services in 2014.

To present graphically a tree diagram for categories of rows for the number of enterprises in the ICT sector (ITC Production, ICT wholesale, Telecommunications, and IT services) we will apply agglomerative hierarchical clustering. Following is a dendrogram of the results of running these data through the Ward clustering algorithm.

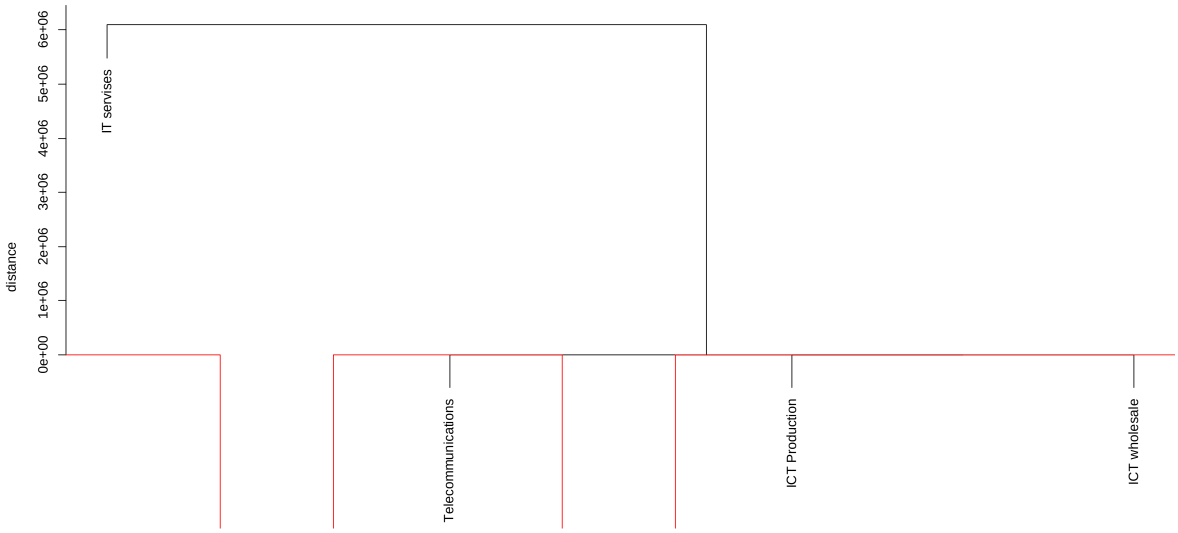


Fig. 2. Dendrogram for the number of enterprises in the ICT sector using Ward method.

We can see, that there are three clusters of ICT specifications separated. One cluster contains two categories: ICT Production and ICT wholesale, second cluster contains only one category - Telecommunications, and last cluster containing also one category is IT services. The analysis is based on data from the Central Statistical Office on the number of employees in the ICT sector in 2011-2014. The correspondence analysis was also conducted for the number of employees in the ICT sector (table 2).

Specification	2011	2012	2013	2014
ICT production	44930	41150	36892	39337
ICT wholesale	10363	10598	11372	11496
Telecommunications	46516	43890	42634	41786
IT services	75539	85178	93422	103739

Table 2. Number of employees in the ICT sector. Source: Central Statistical Office.

The percentage being explained by first dimension is 97.72%, and for the second dimension is 2.27%. These two dimension explain 99.99% of the total inertia. Total inertia is 0.0067 showing that there is a very weal association between two variables: year and specification. Row and column masses, chi-square distance and inertia for each of the category of row and column are presented below.

```
 Principal inertias (eigenvalues):
           1        2        3
Value      0.006547 0.000152 1e-06
Percentage 97.72%   2.27%    0.01%
```

We can display the result of simple correspondence analysis in the form of perception map.

```
 Rows:
        ICT Production ICT wholesale Telecommunications IT servises
Mass          0.219680      0.059321           0.236622    0.484377
ChiDist       0.099928      0.023115           0.076194    0.080015
Inertia       0.002194      0.000032           0.001374    0.003101
Dim. 1       -1.216980      0.110858          -0.921128    0.988340
Dim. 2        1.378733     -1.703701          -1.282066    0.209650

 Columns:
             2011      2012      2013     2014
Mass     0.240035  0.244729  0.249471 0.265765
ChiDist  0.122501  0.028559  0.054708 0.089983
Inertia  0.003602  0.000200  0.000747 0.002152
Dim. 1  -1.513489 -0.351812  0.631948 1.097723
Dim. 2   0.224847  0.119762 -1.578475 1.168343
```

We can display the result of a correspondence analysis in the form of perception map (fig. 3).

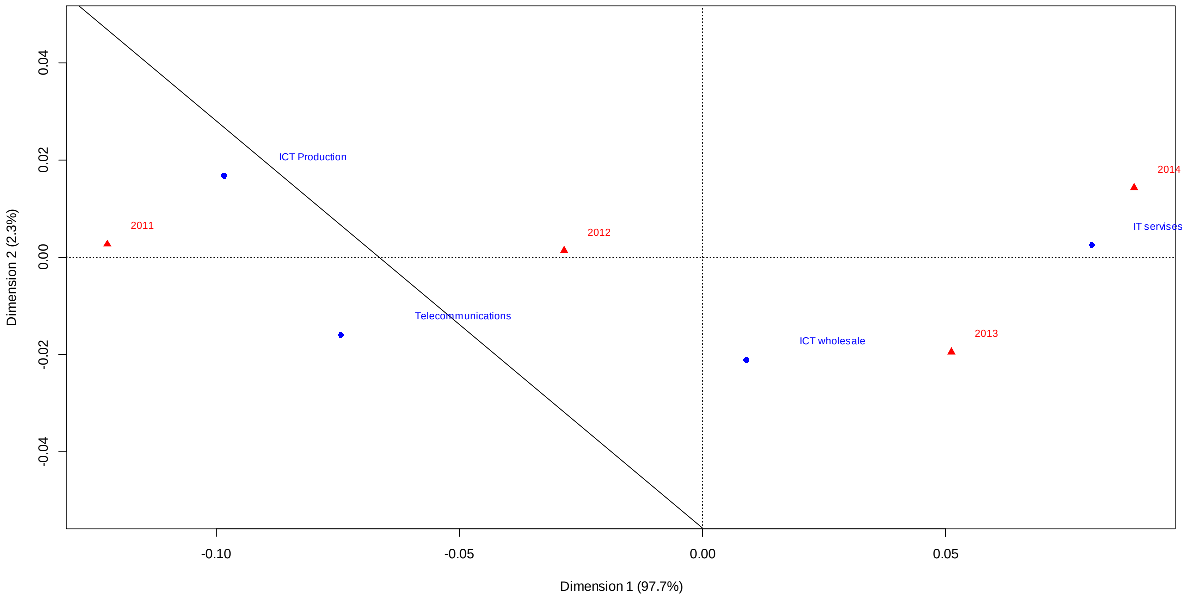


Fig. 3. Two-dimensional perception map for correspondence analysis for the number of employees in the ICT sector.

Looking at two dimensional perception map (fig. 3) we can see that the points are situated in a different way to the analysis conducted for enterprises. The ICT production is situated very close to the year 2011. Telecommunications are related with the year 2012. ICT wholesales is situated very near to the category 2013, and IT services are plotted near the year 2013.

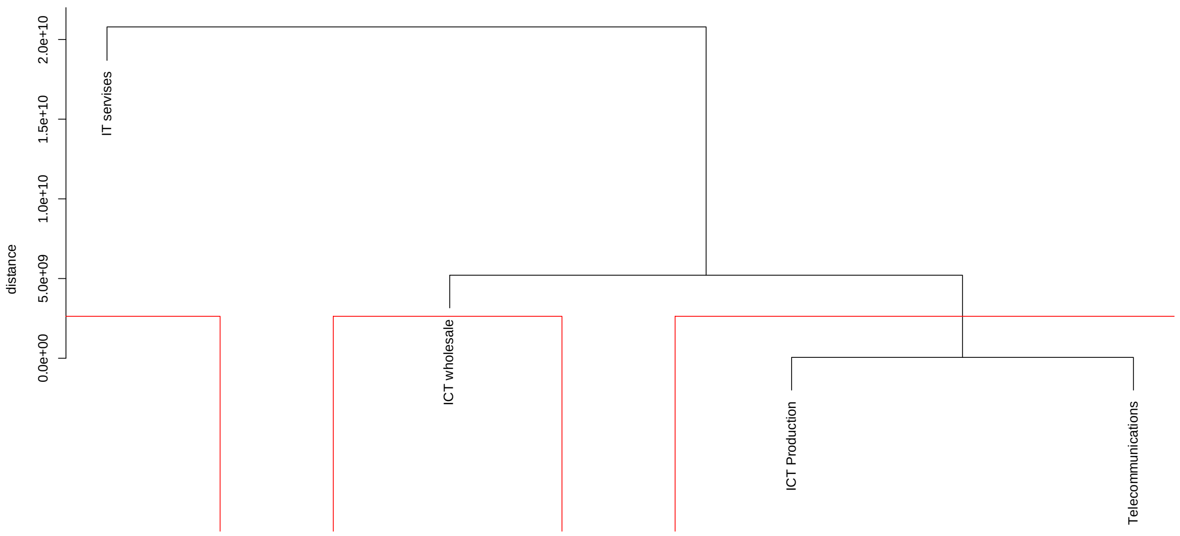


Fig. 4. Dendrogram for the number of employees in the ICT sector using Ward method.

Looking at dendrogram for the number of employees in the ICT sector (figure 4) we can see that there are three clusters separated. First cluster contains two categories: ICT Production and Telecommunications, second cluster contains one category – ICT wholesale, and last cluster contains one category – IT services. Using multivariate methods, we can see, which categories of variables analyzed belong to the cluster of objects that are similar to each other.

Conclusions

Statistical methods applied in the empirical part of this paper allow to see the deeper structure of the number of enterprises in the ICT sector, as well as the number of employees in the ICT sector in 2011-2014. The analysis is based on data from the Central Statistical Office. Statistical methods of the analysis of categorical data were applied, such as: correspondence analysis and hierarchical clustering analysis using Ward method. As a result, a perception map and dendrogram was obtained for a number of enterprises, as well as employees in the ICT sector in 2011-2014. Such methods allow to see which of the categories that were analyzed are similar to each other in the following years. The analysis allows to conclude that busing correspondence analysis for enterprises, we can distinguish three clusters: ICT production, telecommunication, and IT services and ICT wholesale in the same group. For an employee group, there are also three clusters: ICT production, telecommunication, and ICT wholesale with IT services in one cluster. Similar results were obtained in hierarchical analysis. Three clusters were separated: IT services, telecommunication, and the last cluster containing ICT production and ICT wholesale for enterprises data. For an employee’s data there are also three clusters, however with different categories: IT services, ICT production, and in the last cluster ICT production with telecommunication. The analysis conducted in the paper shows different business areas that are similar in the years 2011-2014. It also shows the trend in time perspective, which area are more developed and on top in the following years and how time changes the business moves from one area to another.

References

Adams R., & Gill, S. (2007). Augmented cognition, universal access and social intelligence in the information society. In: Schmorrow, Dylan D. & Reeves, Leah M. (eds.), *Foundations of Augmented Cognition. Lecture Notes in Computer Science 4565*, Springer, Berlin, 231-240.

Britz, J. (2008). Making the global information society good: A social justice perspective on the ethical dimensions of the global information society. *Journal of the American Society for Information Science and Technology* 59(7), 1171-1183.

Kondratiuk-Nierodzińska, M. (2016). New Knowledge Generation Capabilities and Economic Performance of Polish Regions. Equilibrium. *Quarterly Journal of Economics and Economic Policy*, 11 (3), 451-471.

Piech, K. (ed.) (2007). *Knowledge and Innovation Process in Central and East European economies.* Warsaw: The knowledge Innovation Institute.

Pohle, J. (2015). UNESCO and INFOethics: Seeking global ethical values in the Information Society. *Telematics and Informatics, 32*(2), 381-390.

Sourbati, M. (2011). The digital switchover as an information society initiative: The role of public policy in promoting access to digital ICTs. *Telematics and Informatics, 28*(4), 295-302.

Stephanidis, C. (1998). Toward an information society for all: An international research and development agenda. *International Journal of Human-Computer Interaction , 10*(2), 107-134.

Webster, F. (2006). *Theories of the information society*. Third edition. Routledge / Taylor & Francis c/o Bookpoint. Retrieved February 10, 2017, from http://stat.gov.pl/en/topics/science-and-technology/information-society/information-society-in-poland-results-of-statistical-surveys-in-the-years-20102014,1,1.html

A statistically-based classification of de facto exchange rate regimes

Marek A. Dąbrowski[1], Monika Papież[2], Sławomir Śmiech[3]

Abstract

The paper offers a new de facto classification of the exchange rate regimes. The new algorithm for classifying regimes is presented and applied to advanced, emerging and developing economies in the period from 1995 to 2014. The well-known classifications were not constructed with the use of formal statistical tools except for the classification developed by Levy-Yeyati and Sturzenegger (2005). Following their approach, this paper employs several techniques of cluster analysis but with certain important differences including the separate treatment of financially open and financially closed economies. To classify the former group, the trimmed k means method is used, whereas the latter group is classified with the k-nearest neighbour method. Moreover, foreign exchange reserves and exchange rate are treated symmetrically, standardization of main variables captures changes in the international context, and the classification is a two-way classification and is more up-to-date. The comparison of our classification with three most common classifications reveals that there are differences between them stemming from the differences in methodology and period coverage. The simple measure of overall consistency for the most similar classification is well below 80% but above 60% for the least similar classification.

Keywords: *exchange rate regimes, open economy macroeconomics, cluster analysis*

JEL Classification: F33, F31, C38, C82

1 Introduction

The choice of an exchange rate regime is one of the focal issues in international macroeconomics: suffice it to say that, according to one of the most prominent hypotheses, structural flaws of the interwar gold standard made the Great Depression so severe and prolonged (Bernanke and James, 1991). More recently, Obstfeld and Rogoff (2000) explained that the exchange rate is 'the single most important relative price, one that potentially feeds back immediately into a large range of transactions.' In his survey paper, Rose (2011, p. 671) claimed, however, that 'such choices [of the exchange rate regime] often seem to have remarkably little consequence. Exchange rate regimes are flaky: eccentric and unreliable.'

We think that confusion about ramifications of exchange rate regime choices – at least part of it – stems from the difficulties economists encounter when they attempt to classify actual exchange rate regimes. On the one hand, Obstfeld and Rogoff (1995) argued that 'the spectacular expansion of world capital markets' made the fixed exchange rate a 'mirage.' On

[1] Corresponding author: Cracow University of Economics, Department of Macroeconomics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: marek.dabrowski@uek.krakow.pl.
[2] Cracow University of Economics, Department of Statistics, e-mail: papiezm@uek.krakow.pl.
[3] Cracow University of Economics, Department of Statistics, e-mail: smiechs@uek.krakow.pl.

the other hand, Calvo and Reinhart (2002) discerned the 'fear of floating' syndrome. More generally, declared (de jure) and actual (de facto) exchange rate regimes may differ: using Levy-Yeyati and Sturzenegger's (2005) terminology 'words' do not have to match 'deeds.'

Economists have been rather economical with the employment of statistical tools to classify exchange rate regimes. Out of the well-known classifications only the one developed by Levy-Yeyati and Sturzenegger (2005) was constructed with the usage of formal statistical methods (cluster analysis). We follow their approach, although we modify it in several respects. The analysis is carried out separately for financially open and financially closed economies; information conveyed in foreign exchange reserves and in exchange rates is treated symmetrically; standardization is done for each year separately; the exchange rate regime classification is in principle a two-way classification; and it is also more up-to-date.

2 Empirical strategy and data

Empirical strategy employed consists of seven steps. First, it is important to identify the reference currency. For example, the exchange rate of Danish krone against the US dollar is highly volatile, but it is fixed against the euro. Thus, in order to establish the reference currency, we followed the approach similar to McKinnon and Schnabl's (2004) and, using regression analysis, compared variability of exchange rates of a given currency against the US dollar, euro, Japanese yen, and British pound (in some cases we included Australian dollar, South African rand, Indian rupee and SDR).

Second, the country-year observations were split into two groups with respect to the openness to capital flows. The rationale behind the split is based on the macroeconomic trilemma stating that if the capital account is closed, monetary authority can retain monetary autonomy even though it engages in stabilization of the exchange rate. Moreover, focusing on financially open countries makes the problem of differences between market and official exchange rates less important (see Reinhart and Rogoff, 2004, and Shambaugh, 2004). The country-year observations above the median of Chinn-Ito index of capital openness were defined as financially open (Chinn and Ito, 2006).

Third, all variables used to build the classification were standardized. Standardization was constrained in two ways: it was applied to the set of observations selected in the previous step, and performed for each year separately. The reason behind the latter constraint was that the behaviour of economies in the face of global shocks is different from that in normal times.

Fourth, the cluster analysis was used to detect homogeneous groups of country-year observations. As there were quite a few outlying observations (mainly isolated outliers according to García-Escudero et al. 2010), i.e. country-years with an extremely large change of exchange rates or foreign exchange reserves, we used trimmed k-means method. The trimmed k-means method allows for removing a certain fraction of the "most outlying" data, and, this way, a strong influence of outlying observations can be avoided and robustness naturally arises. The trimming approach to clustering was proposed in Cuesta-Albertos et al. (1997) and Gallegos (2002). Trimming can also be used to highlight interesting anomalous observations.

Fifth, one of the clusters obtained grouped quite a few observations with low variability of both exchange rate and foreign exchange reserves. These are characteristic for calm times. Following Levy-Yeyati and Sturzenegger (2005), we called this group 'inconclusives' and applied the k-means algorithm to such observations. The objective was to isolate peggers from floaters in this group and to extend the groups obtained in the previous step.

Sixth, the k-nearest neighbour method was used to classify countries that were relatively closed to capital flows. The method is based on finding the k nearest objects in a reference set and taking a majority vote among the classes of these k objects. Clusters obtained in the fourth step were used as the reference set, and thus a country closed to capital flows in a given year was classified to the most frequently represented category in the closest neighbourhood.

Seventh, the country-year observations that were not classified as either peggers or floaters were added to one of these groups. When the average absolute monthly change in the exchange rate was less than 0.01% (larger thresholds were also considered), a country-year was considered to peg its currency. Additionally, some countries 'under pressure' were reclassified as peggers/floaters if in the adjacent years they pegged/floated their currencies.

The sample covered 183 countries analysed in the period 1995-2015, i.e. a maximum of 3,843 country-year observations. The classification was based on five variables: capital openness index (developed by Chinn and Ito, 2006), two measures of exchange rate variability (the average absolute monthly change and standard deviation of monthly change), and two measures of foreign exchange reserves variability (the average absolute monthly change and standard deviation of monthly change). Due to limited data availability, the sample included 3,068 observations.

3 Empirical results

The examination of monthly exchange rates resulted in finding that the US dollar was by far the most prevalent reference currency – its 'share' was above 63%. The euro was found to be a base currency for slightly less than 30% of country-year observations. In eight cases a different currency was identified as a reference currency: Australian dollar for Kiribati, South African rand for Botswana, Lesotho, Namibia, Swaziland, Indian rupee for Bhutan and SDR for Libya and Myanmar. Four cases of a switch from one currency to another were observed in our sample: Algeria switched in 2003 from euro to US dollar, Lithuania and Sao Tome and Principe switched from the US dollar to euro in 2002 and 2008 respectively, and Latvia switched from the SDR to euro in 2005.

In order to obtain homogenous clusters of country-year observations with the trimmed k-means method, we made two choices: the number of clusters was set to four and the fraction of observations to be trimmed of was set to two per cent. The former choice was motivated by theoretical considerations: with basically two variables, i.e. exchange rate variability and reserves variability, out of which each can take a 'low' or 'high' value, one should expect four different clusters: 'low/low', 'low/high', 'high/low' and 'high/high.' The silhouette measure for four clusters was 0.54 and was only slightly lower than for three or two clusters (0.61 and 0.57, respectively). Fewer than four clusters, however, seemed to be rather difficult to justify from both logical (see above) and economic points of view (for instance Levy-Yeyati and Sturzenegger (2005) had five clusters, although they did not report any statistical measure for that choice). In the latter choice we followed Levy-Yeyati and Sturzenegger (2005) and trimmed two per cent of the most outlying observations.

The results of cluster analysis for financially open country-years are illustrated in Figure 1. The axes represent the first two principal components: the first one corresponds to the volatility of the exchange rate and the second one to the volatility of foreign exchange reserves. After exclusion of the outliers (34 obs.), four groups were identified. Two of them are straightforward to decipher. Peggers (green crosses) experienced low exchange rate variability and above normal variability of foreign exchange reserves (308 obs.), whereas floaters (dark blue x's) had the opposite characteristics (389 obs.). Interestingly, we isolated the group of observations with even greater exchange rate variability than that characteristic for floaters and foreign reserves variability comparable to that characteristic for peggers (blue diamonds; 81 obs.). According to Levy-Yeyati and Sturzenegger (2005) – who obtained a similar cluster – such observations constitute a group of countries under intermediate exchange rate regimes (e.g. dirty float). It seems, however, that the group includes countries that were under strong foreign exchange market pressure (if not in an overt currency crisis) rather than countries that placidly managed their exchange rates. Moreover, one would expect the managed exchange rate to display on average lower variability than the freely floating rate. This is not the case here. Thus, contrary to Levy-Yeyati and Sturzenegger (2005), we prefer to call this group 'under pressure.' The most numerous group (845 obs.), however, included country-year observations with below normal variability of both exchange rate and foreign reserves. Such characteristics are displayed by both peggers and floaters in calm times. Thus, the group consists of 'inconclusives' (red triangles), and the question about its true composition remains open. In order to narrow down the degree of inconclusiveness and in line with the methodology used by Levy-Yeyati and Sturzenegger (2005), we applied the simple k-means method (the outliers had been already excluded in the previous step) to divide this group into three categories: peggers (353 obs.), floaters (298 obs.) and deep inconclusives (194 obs.).

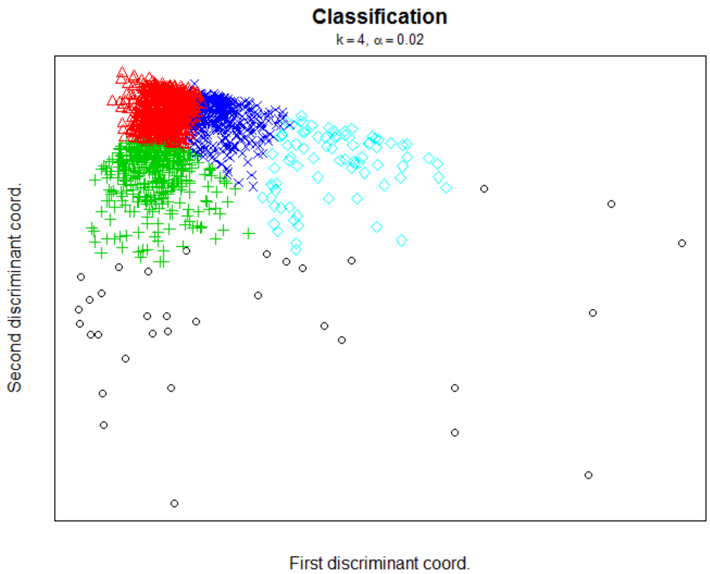


Fig. 1. Clusters of financially open country-years.

So far our procedure was directed at financially open economies which constituted 1,657 country-year observations. The remaining 1,411 observations referred to countries that were relatively closed to capital flows. As explained in the previous sections, these were classified with the k-nearest neighbour method. We tried from two to 20 neighbours and found out that the fraction of wrong classifications for a learning set (created from financially open countries) was the lowest for 13 neighbours. Thus, k parameter was set to 13 and financially closed economies were divided into peggers, floaters, inconclusives, countries 'under pressure' and outliers. The overall results of this and previous steps are reported in Table 1 in columns three and four. In the last step we moved 304 country-years from inconclusives, 'under pressure' or outliers to peggers (285 obs.) or floaters (19 obs.) if such a change was uncontroversial. Two distinct criteria were used: (1) the country-year was reclassified as a pegger if the average absolute monthly change of the exchange rate was less than 0.01% (230 obs.) or (2) the country-year 'under pressure' was reclassified as a pegger (55 obs.) or a floater (19 obs.) if in the adjacent years it belonged to such a category. The final results are tabulated in columns five and six in Table 1. Overall, we identified more peggers (54.0%) than floaters (34.4%). This result was driven by relatively low incidence of floating exchange rate regime (25.2%) and high incidence of fixed rate arrangements (61.7%) in financially closed countries. In countries with open capital account the corresponding fractions were much closer one another (42.2% and 47.4%, respectively). This finding is in line with the conjecture that can be derived from the macroeconomic trilemma: when capital flows are controlled, it is more attractive for monetary authorities to maintain *de facto* fixed exchange rate as it does not require scarifying monetary autonomy. Interestingly, the category 'under pressure' is more frequent when a country is financially closed (4.4% vs. 0.8%). This could be an indication that the effectiveness of capital controls is limited, and that such barriers do not isolate an economy from foreign exchange market pressure.

Category	**Financial openness**	**Classification after: Steps 1-6**		**Step 7**	
Peggers	open	1371	661	1656 (54.0%)	785 (47.4%)
	closed		710		871 (61.7%)
Floaters	open	1035	687	1054 (34.4%)	699 (42.2%)
	closed		348		355 (25.2%)
'Under pressure'	open	262	81	76 (2.6%)	14 (0.8%)
	closed		181		62 (4.4%)
Inconclusives	open	280	194	182 (5.9%)	125 (7.5%)
	closed		86		57 (4.0%)
Outliers	open	120	34	100 (3.3%)	34 (2.1%)
	closed		86		66 (4.7%)

Notes: 'open': Chinn-Ito index not less than 0.4, 'closed': Chinn-Ito index less than 0.4.

Table 1. Details of classification of exchange rate regimes.

Our classification as such is rather difficult to interpret. In order to shed more light on it, we compared it with other exchange rate regime classifications. Three popular classifications were taken into account. We considered the classification developed by Levy-Yeyati and

Sturzenegger (2005) since they adopted quite a similar approach to ours, i.e. they used statistical tools to distinguish alternative regimes. It was natural to take into account the classification tabulated by the IMF, because it is used in the literature as a kind of a reference point. A detailed work by Reinhart and Rogoff (2004) on exchange rate arrangements with its emphasis on market vs. official exchange rates is also quite popular in the literature on the international economics.

Examining the degree of consistency between alternative classifications, Klein and Shambaugh (2010, p.47) transformed each classification to a dichotomous division into pegs and non-pegs and then – for each pair of classifications – calculated the percentage of observations that were classified in the same way. We followed a similar, although not exactly the same, approach. First, we mapped alternative classifications into pegs and floats. In Levy-Yeyati and Sturzenegger's classification and in ours we retained the category of inconclusives and omitted the outliers. In our classification cases 'under pressure' were omitted. In two other classifications both hard and soft pegs were merged into pegs, whereas intermediate and freely floating regimes were combined into floats ('freely falling' and 'dual market in which parallel market data is missing' categories were omitted).

In Table 2 our classification is compared to the one by Levy-Yeyati and Sturzenegger (2005) for the overlapping period of both classifications, i.e. 1995-2004. The degree of consistency can be traced out on the main diagonal, whereas off-diagonal elements correspond to divergence between classifications. For example, out of 803 country-year observations recognized by our algorithm as pegs, 724 were classified in the same way by Levy-Yeyati and Sturzenegger (2005). This is more than 90%. The remaining observations were classified either as floats (51) or as inconclusives (28) – the corresponding 'shares' were 6.4% and 3.5%, respectively. There was less consistency with respect to floats: less than two-thirds of our floats were classified in the same category by Levy-Yeyati and Sturzenegger. This finding is, at least to a certain extent, the result of mapping intermediate regime into pegs. If instead they are treated as floats – i.e. in line with a dichotomous division into pegs and non-pegs used by Klein and Shambaugh (2010, p. 47) – the consistency between floats rises to 80.7% and that between pegs drops to 74.7% (not reported).

The comparison between our classification and those developed by the IMF and Reinhart and Rogoff (2004) is depicted in Table 3. The common period covered by all these classifications is from 1995 to 2010. The consistency of our classification with the IMF's one is lower for pegs (61.3%) and higher for floats (78.9%) in comparison to the consistency with Levy-Yeyati and Sturzenegger's classification. This effect tends to be even stronger if we adopt Klein and Shambaugh's (2010) mapping (54.6% and 81.0%, respectively; not reported). In turn, the consistency between our classification and the one developed by Reinhart and Rogoff (2004) was the highest for pegs (94.5%) and the lowest for floats (57.3%). Like with Levy-Yeyati and Sturzenegger's classification, however, this finding can be reversed if the alternative mapping of Klein and Shambaugh (2010) is used with coefficients 64.5% and 92.8% (not reported).

		LYS Classification			
		Inconc.	**Peg**	**Float**	∑
Our Classification	Inconc.	13 16.3%	61 76.3%	6 7.5%	80
	Peg	28 3.5%	724 90.2%	51 6.4%	803
	Float	1 0.2%	169 36.3%	296 63.5%	466

Table 2. New classification against LYS classification.

		IMF Classification			RR Classification		
		Peg	**Float**	∑	**Peg**	**Float**	∑
Our Classification	Peg	698 61.3%	441 38.7%	1139	1167 94.5%	68 5.5%	1235
	Float	164 21.1%	614 78.9%	778	325 42.7%	436 57.3%	761

Table 3. New classification against IMF and RR classifications.

The coefficient of overall consistency can be calculated as the number of observations on the main diagonal to the total number of observations. Using such a measure, we found out that our classification is the most similar to the one developed by Levy-Yeyati and Sturzenegger (76.6%), slightly less similar to the Reinhart and Rogoff's classification (75.0%), and the least similar to that of the IMF (64.1%). This result holds if we limit comparison just to pegs and floats (i.e. omit inconclusives) or/and apply Klein and Shambaugh's (2010) mapping (the relevant coefficients for this mapping were 73.3%, 70.4% and 61.1%, respectively).

Conclusion

The paper offers a new de facto classification of the exchange rate regimes adopted by both advanced economies and emerging and developing economies in the period from 1995 to 2014. We borrowed the idea of applying statistical tools, i.e. cluster analysis, to identify actual exchange rate regimes from the study by Levy-Yeyati and Sturzenegger (2005). Their study remains – to the best of our knowledge – the only one in the literature on exchange rate regimes in which cluster analysis techniques were applied. Our approach, however, differs from theirs in several respects. Its main distinctive feature is that we separated financially open countries from those that were closed to capital flows. Other differences include: 1) a symmetric treatment of foreign exchange reserves and exchange rate; 2) a standardization that provides consistency between country-years in turbulent and normal times; 3) a basically two-way classification into pegs and floats; 4) more up-to-date results.

Not surprisingly, we found that our classification is different from the one worked out by Levy-Yeyati and Sturzenegger (2005). It is also different from two other popular classifications developed by the IMF and Reinhart and Rogoff (2004). The comparison of our classification with the others is not straightforward as alternative classifications use different categories and cover different periods. A simple measure of consistency between classifications, however, revealed that our classification is the most similar to the one developed by Levy-Yeyati and Sturzenegger (2005), a bit less similar to the one worked out by Reinhart and Rogoff (2004) and the least similar to the IMF.

There are two main avenues of further research. First, our classification requires some refinements, e.g. the intermediate exchange rate regime category is missing in it. Second, the new classification can be used to establish how different (if at all) peggers are from floaters and whether Rose's (2011) scepticism about the exchange rate regime was well-founded.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Centre in Poland (grant no. DEC-2015/17/B/HS4/02681).

References

Bernanke, B., & James, H. (1991). The Gold Standard, Deflation, and Financial Crisis in the Great Depression: An International Comparison. In Hubbard, R.G. (ed.), *Financial Markets and Financial Crises*. Chicago: University of Chicago Press for NBER, 33-68.

Calvo, G. A., & Reinhart, C. M. (2002). Fear of floating. *The Quarterly Journal of Economics, 117*(2), 379-408.

Chinn, M. D., & Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics, 81*, 163-192.

Cuesta-Albertos, J. A., Gordaliza, A., & Matran, C. (1997). Trimmed k-Means: An Attempt to Robustify Quantizers. *The Annals of Statistics*, 553-576.

Gallegos, M. T. (2002). Maximum likelihood clustering with outliers. In: Jajuga K, Sokolowski A, Bock HH (eds.) *Classification, clustering and data analysis: recent advances and applications.* Springer-Verlag, Berlin, pp 247–255.

García-Escudero, L. A., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2011). Exploring the number of groups in robust model-based clustering. *Statistics and Computing, 21*(4), 585-599.

Ilzetzki, E., Reinhart, C. M., & Rogoff, K. S. (2011). The country chronologies and background material to exchange rate arrangements into the 21st century: Will the anchor currency hold. Available at personal. lse. ac. uk/ilzetzki/data/ERA-Country_ Chronologies_2011. pdf (accessed on January 31, 2017).

Klein, M. W., & Shambaugh, J. C. (2010). *Exchange rate regimes in the modern era*. MIT Press.

Levy-Yeyati, E., & Sturzenegger, F. (2005). Classifying exchange rate regimes: Deeds vs. words. *European Economic Review, 6*(49), 1603-1635.

McKinnon, R., & Schnabl, G. (2004). The East Asian dollar standard, fear of floating, and original sin. *Review of Development Economics*, 8(3), 331-360.

Obstfeld, M., & Rogoff, K. (1995). The Mirage of Fixed Exchange Rates. *Journal of Economic Perspectives*, *9*(4), 73-96.

Obstfeld, M., & Rogoff, K. (2000). The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?. *NBER Macroeconomics Annual 2000, 115(1)*, 339–90.

Reinhart, C. M., & Rogoff, K. S. (2004). The Modern History of Exchange Rate Arrangements: A Reinterpretation. *The Quarterly Journal of Economics*, *119*(1), 1-48.

Rose, A. K. (2011). Exchange rate regimes in the modern era: fixed, floating, and flaky. *Journal of Economic Literature*, *49*(3), 652-672.

Shambaugh, J. C. (2004). The effect of fixed exchange rates on monetary policy. *The Quarterly Journal of Economics*, *119*(1), 301-352.

Insulating property of the exchange rate regime in Central and Eastern European countries

Marek A. Dąbrowski[1], Justyna Wróblewska[2]

Abstract

This paper examines the insulating property of flexible exchange rate in eight middle-income Central and Eastern European economies. All of them fit to the standard of a small open economy, but have quite different exchange rate regimes. We derive long-term zero and sign restrictions on the reactions of underlying macro variables to real and nominal shocks from the stochastic macroeconomic model. For each country a Bayesian structural vector autoregression model with common serial correlations is estimated on quarterly data from the 1998-2015 period for relative GDP, interest rate differential, real exchange rate and relative price level. We use forecast error variance decompositions to identify the importance of individual shocks. Then we compare impulse response functions of relative output in order to discern the differences between floaters and peggers and in this way assess the magnitude of insulating property of flexible exchange rate. Our main finding is that though the empirical evidence is mixed, it lends non-negligible support to the hypothesis that the flexible exchange rate insulates the economy against shocks to a greater extent than the fixed exchange rate regime.

Keywords: *open economy macroeconomics, exchange rate regimes, real and nominal shocks, Bayesian structural VAR, common serial correlation*

JEL Classification: F41, C11

1 Introduction

The insulating properties of the exchange rate regime together with its impact on policy effectiveness and importance for the adjustment to trade imbalances constitute one of 'three main strands' in the literature on the choice of exchange rate regime (Ghosh et al., 2010). One of the central findings of this strand is that the floating exchange rate better insulates output against real shocks as it facilitates adjustment in the face of nominal rigidities, whereas foreign exchange reserves movements under the fixed exchange rate automatically offset nominal shocks. In a nutshell, using words of Ghosh et al. (2002), 'the relative incidence of nominal and real shocks becomes a key criterion in choosing the exchange rate regime.'

There are two main empirical approaches to study the insulating properties of the exchange rate regime. First, the relationship between volatility of output growth and the exchange rate flexibility is examined. The results are mixed: there is some evidence of greater output

[1] Corresponding author: Cracow University of Economics, Department of Macroeconomics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: marek.dabrowski@uek.krakow.pl.

[2] Cracow University of Economics, Department of Econometrics and Operations Research, e-mail: eowroble@cyf-kr.edu.pl.

volatility under pegged exchange rates than under either intermediate or floating regimes, but the result holds for advanced and developing economies and not for emerging market economies where output volatility is lower under pegged and intermediate regimes (Ghosh et al., 2010). In an earlier study Levy-Yeyati and Sturzenegger (2003) found a negative link between output volatility and exchange rate flexibility for nonindustrial countries but not for industrial ones. Similar conclusion was drawn in more recent studies by Edwards (2011) and Erdem and Özmen (2015). They found that the impact of external shocks on economic activity is less pronounced in economies under flexible exchange rate regimes.

Second, a structural vector autoregression (SVAR) is used to identify shocks hitting an economy, to assess their importance for the output variance and – in some studies – to examine output reactions to real and nominal shocks. The results are again inconclusive, e.g. the exchange rate fluctuations in Central and Eastern European countries were found to be fed by nominal shocks by Kuijs and Borghijs (2004) and Shevchuk (2014) and by real shocks by Stążka-Gawrysiak (2009) and Dąbrowski and Wróblewska (2016a, b).

The research objective of this paper is to examine whether there are discernible differences across exchange rate regimes in Central and Eastern European (CEE) economies with respect to insulating output against economic shocks. We follow the second empirical approach and construct a set of Bayesian SVAR models for each of eight CEE countries in our sample and focus on comparison of output response functions to real and nominal shocks. Our main finding is that there is weak but not negligible evidence of insulating properties of floating exchange rate regime in CEE economies. The paper is structured as follows. The next section briefly lays out theoretical issues. Empirical methodology and data are briefly presented in Section 3. Empirical results are reported in Section 4 and the last section concludes.

2 Theoretical issues

The seminal papers of Fleming (1962) and Mundell (1963), published well after the establishment of the Bretton Woods system, have reignited economists' interest in choices and consequences of exchange rate regimes. In a study summarising the state of the art on the exchange rate regimes Ghosh et al. (2002) built a simple stylized model in the spirit of the Barro-Gordon approach and demonstrated that the floating exchange rate is preferable to pegged rate if shocks are real but worse if shocks are nominal. This conclusion, however, rests on the assumption of high capital mobility (used in their formal model). If capital mobility is low, e.g. due to capital controls, aggregate demand shocks are partly offset under fixed exchange rate, whereas floating rate amplifies such shocks (Ghosh et al., 2002).

While their model is neat and elegant, it allows for two sources of shocks only: (real) productivity shock and (nominal) monetary shock. Although it is just enough to convey the main theoretical point about insulating properties of the exchange rate regime it seems too parsimonious to be used as a theoretical framework in empirical research.

The small open economy model we use in this paper is in principle taken from the study by Clarida and Galí (1994) who in turn extended the model developed by Obstfeld et al. (1985). It describes equilibria in the goods market, money market and the foreign exchange market with the conventional IS, LM and UIP relations. Due to price stickiness (PS relation) shocks hitting an economy bring about an adjustment process, so the flexible-price equilibrium is attained only in the long term. Taking into account extension put forward by Dąbrowski (2012) the model includes supply, u^s, demand, u^d, financial, u^f, and monetary, u^m, shocks. A more detailed description of the model is presented by Dąbrowski and Wróblewska (2016a).

We use the solution for the flexible-price equilibrium to derive zero and sign restrictions that will be imposed on the empirical reactions of relative output, y_t, real interest rate differential, r_t, real exchange rate, q_t, and the relative price level, p_t, to structural shocks. Vector moving representation of a SVAR model

$$\Delta z_t = C(L)u_t\,, \tag{1}$$

where $\Delta z_t = [\Delta y_t, \Delta r_t, \Delta q_t, \Delta p_t]'$ is a vector of the first differences of macroeconomic variables, $C(L)$ is a matrix of lag polynomials and $u_t = [u_t^s, u_t^d, u_t^f, u_t^m]'$ is a vector of structural shocks, can be used to succinctly write the restrictions imposed: $C_{12}(1)$, $C_{13}(1)$, $C_{14}(1)$, $C_{24}(1)$, $C_{34}(1)$ are set to zero, $C_{11}(1)$, $C_{22}(1)$, $C_{23}(1)$, $C_{32}(1)$, $C_{42}(1)$, $C_{44}(1)$ are positive, $C_{21}(1)$, $C_{31}(1)$, $C_{41}(1)$, $C_{33}(1)$ are negative and the $C_{43}(1)$ is left free, where $C_{ij}(1)$ is the ijth element of the total impact matrix $C(1)$ (see e.g. Lütkepohl, 2006).

3 Methodology and data

The basic model employed in the empirical analysis is a stable Bayesian n-dimensional VAR(k) model with a constant and non-stochastic starting points ($\Delta z_{-k+1}, \Delta z_{-k+2},\ldots,\Delta z_0$):

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + \Gamma_2 \Delta z_{t-2} + \ldots + \Gamma_k \Delta z_{t-k} + \Phi D_t + \varepsilon_t, \quad \varepsilon_t \sim iiN(0,\Sigma), \quad t = 1,2,...,T, \tag{2}$$

where Σ is a PDS covariance matrix of the Gaussian white noise process $\{\varepsilon_t\}_{t=1}^{T}$, D_t collects deterministic components and $\Gamma_1, \Gamma_2,\ldots,\Gamma_k$, Φ are the matrix/vector parameters. The Normal-Wishart prior structure is imposed on the model parameters with normal distribution for Γs centered around zero and the prior expectation of Σ is $0.01I_4$, i.e. $\Sigma \sim iW(0.01I_4, n+2)$, $(\Gamma_1, \Gamma_2,\ldots, \Gamma_k)|\Sigma, \nu_\Gamma \sim mN(0, \nu_\Gamma I_{nk}, \Sigma)$, $\Phi|\,\Sigma, \nu_\Phi \sim N(0, \nu_\Phi\Sigma)$.

Additionally, in the set of the analysed models there are also structures with the reduced rank restrictions imposed on the Γs parameters:

$$\Delta z_t = \gamma\delta_1\Delta z_{t-1} + \gamma\delta_2\Delta z_{t-2} + \ldots + \gamma\delta_k\Delta z_{t-k} + \Phi D_t + \varepsilon_t, \quad \varepsilon_t \sim iiN(0,\Sigma), \quad t = 1,2,...,T, \tag{3}$$

where $\gamma_{n\times(n-s)}$ is a matrix of full column rank. As pointed out by Engle and Kozicki (1993) in the VAR framework such restriction is equivalent to common serial correlations among the analysed processes, *s* denotes the number of these common features. Such a Bayesian VAR-CC model (Bayesian vector autoregression with common serial correlations model) has been already analysed e.g. by Dąbrowski and Wróblewska (2016a).

For each considered country we compare 20 non-nested specifications, which may differ in the number of lags ($k \in \{5,6,\ldots,9\}$) and the number of common features ($s \in \{0,1,2,3\}$). We assume equal prior probability for each of them.

The below presented results are obtained by taking the advantage of the Bayesian Model Averaging technique within the set of models with the highest posterior probability (i.e. higher than the assumed 0.05 prior probability). To impose over mentioned sign and zero restrictions, resulting from the economic model, the algorithm proposed by Arias et al. (2014) is employed.

Quarterly, seasonally adjusted data for real GDP, three-month money market interest rate, nominal exchange rate (defined as a price of domestic currency in terms of the euro), harmonised index of consumer prices for the period 1998q1-2015q4 are retrieved from the Eurostat database. They are used to construct relative real output, real interest rate differential, real exchange rate and relative price level. It was natural to choose the euro area as a reference country, so the *relative* output, for example, is the difference between the log of real GDP in a given country and the log of real GDP in the euro area.

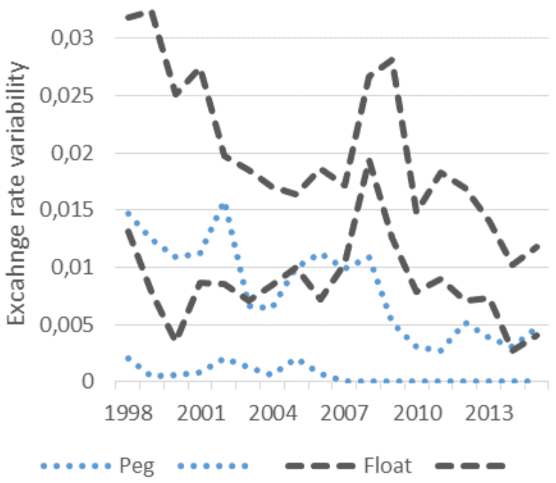
4 Empirical results

Out of eight CEE countries included in our sample only two can be classified as being at opposite poles of an exchange rate regime spectrum: Bulgaria with its currency board (adopted in 1997) and Poland with free floating (adopted in 2000). The reading of the IMF's *Annual Reports on Exchange Arrangements and Exchange Restrictions* makes it reasonable, however, to extend the group of 'pegs' to Croatia, Slovenia and Slovakia and treat the Czech Republic, Hungary and Romania as belonging to the group of 'floats.' This classification is of course imperfect since for example Hungary can be considered a soft pegger, just like Croatia, although Hungary adopted such a regime between 2002 and 2004 only. Instead of relying on one of many exchange rate regime classifications, we illustrate the degree of variability of the exchange rate in Figure 1. In the left-hand-side panel we use the average absolute monthly change of the exchange rate to compare our pegs and floats: the lines correspond to minimum and maximum averages in each group. Floats have indeed experienced a greater exchange rate variability than pegs and the explicit overlap between them can only be observed at the turn of the centuries (1998-2002) and in 2006 (this was due to a gradual appreciation of the Slovak koruna within the ERM II).

Comparison with respect to openness to capital flows, measured with the Chinn-Ito index, is depicted in the right-hand-side panel of Figure 1. All floaters recorded an above-median level of the index starting in 2002, whereas peggers were slightly lagging behind till 2005. The difference nevertheless was exclusively due to Bulgaria, so one can argue that our sample includes, by and large, countries with relatively high capital mobility. Therefore, drawing on the theory we expect the floating exchange rate to insulate output against real shocks to a greater extent than the fixed rate.

The conventional analysis of insulating properties of floating exchange rate is based on the forecast error variance decomposition as it allows to identify the proportions of variability accounted for by structural shocks. Thus, in Table 1 the sources of output and exchange rate fluctuations are depicted. It is quite clear that irrespective of the exchange rate regime real, especially supply, shocks are behind output variability. A small difference between Bulgaria and other CEE countries in this respect dissipates at longer forecast horizons and at four-year horizon the contribution of real shocks is more than 99% in all countries (results for other forecasting horizons and variables are available from the authors upon request). There is also little difference between pegs and floats with respect to the relative importance of real and nominal shocks to the real exchange rate variability. Both the floating and fixed rates are mainly driven by demand and financial shocks and the contribution of real shocks increases with the forecasting horizon, whereas that of nominal shocks goes down.

Overall, the similarity between pegs and floats can be interpreted as evidence against the view that the floating exchange rate is heavily influenced by financial shocks that are subsequently transmitted into a real economy. Instead, we observe that nominal shocks are equally important sources of real exchange rate variability in both group of countries.



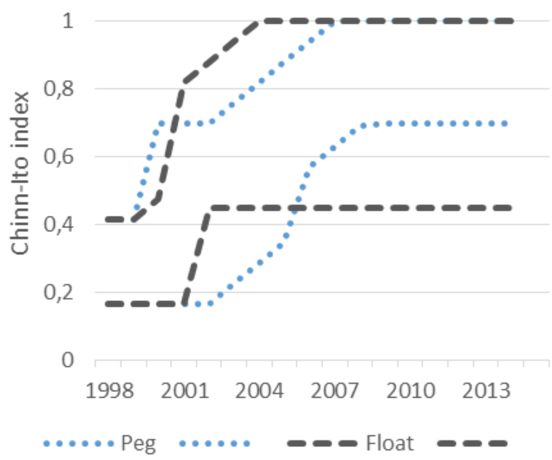


Fig. 1. Exchange rate variability and capital account openness in CEE countries.

Variable	Shock	Bulgaria	Croatia	Slovenia	Slovakia	Czech Rep.	Hungary	Poland	Romania
Relative output	supply	88.2	97.0	93.3	95.3	96.1	98.5	97.3	97.0
	demand	8.3	1.4	3.2	2.5	1.7	0.7	1.1	0.5
	financial	2.5	0.9	1.6	1.1	1.0	0.5	0.9	0.6
	monetary	1.0	0.7	1.8	1.0	1.2	0.4	0.8	1.9
Real exchange rate	supply	6.6	1.6	3.5	3.2	2.6	2.5	10.5	2.5
	demand	45.5	49.4	50.4	51.2	50.9	50.2	47.4	54.4
	financial	38.6	47.7	44.3	43.1	45.1	45.3	41.7	40.4
	monetary	9.3	1.3	1.8	2.5	1.3	1.9	0.5	2.7

Table 1. Posterior expected value of forecast error variance decomposition of the relative output and real exchange rate in CEE countries in percent (forecast horizon is one quarter).

In an attempt to examine the insulating properties of the flexible exchange rate one cannot rely on the forecast error variance decomposition only. Even though the contribution of financial shocks is similar in CEE floats and pegs, it is uncontroversial that the nominal exchange rate variability is greater under floating rate regime (see Figure 1). The important question is whether the increased exchange rate variability moderates output reactions to shocks hitting an economy. Thus, following Dąbrowski and Wrólewska (2016a) we examine the impulse response functions of the relative output to structural shocks.

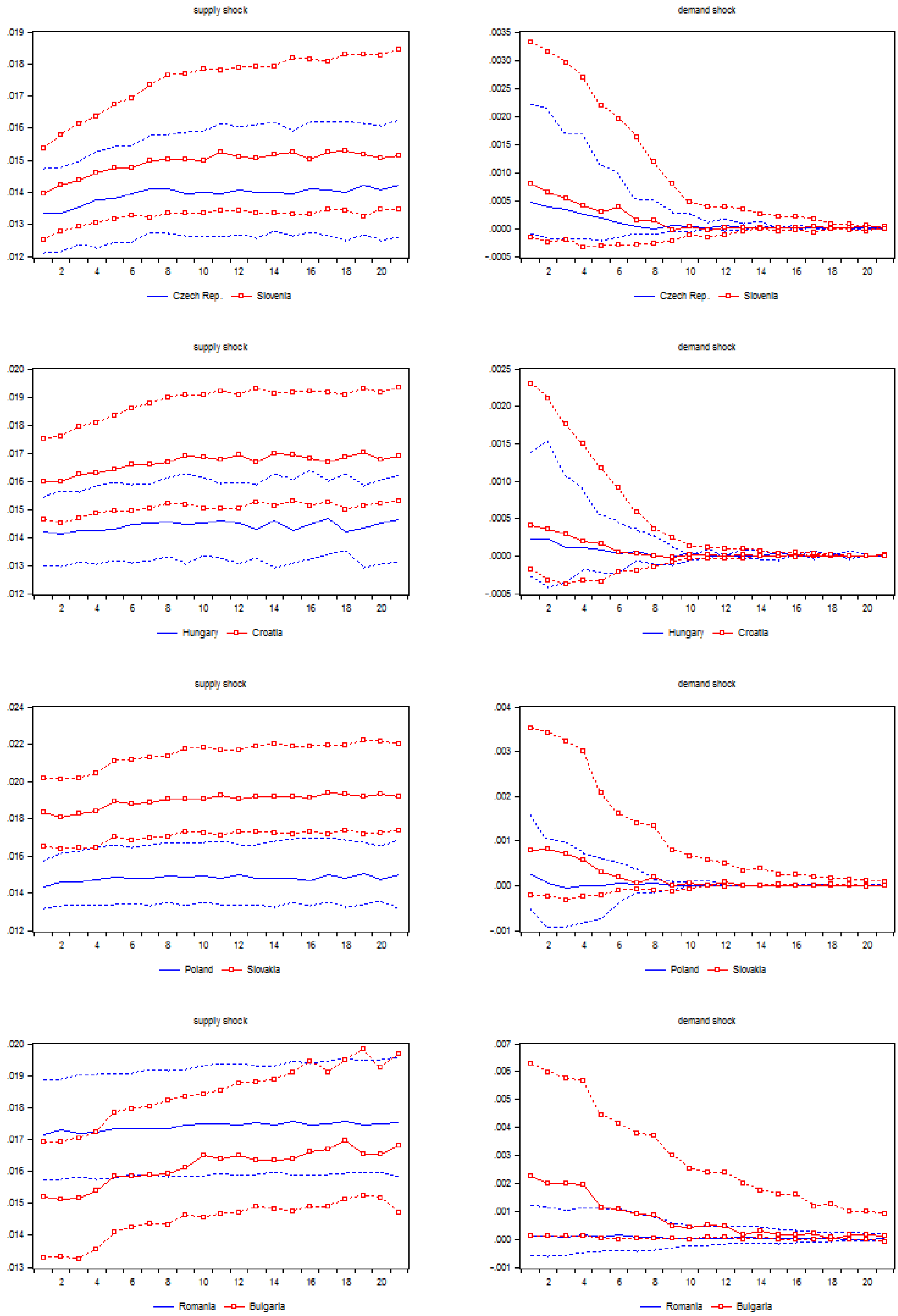


Fig. 2. Impulse response functions of the relative output in CEE countries.

In Figure 2 the median reactions of output to two real shocks are illustrated with solid lines and (the analogue of) the confidence interval, i.e. the 16th and 84th quantiles of the posterior distribution, are depicted with broken lines (results for other shocks and variables are available upon request). To keep the figure uncluttered countries are compared in pairs: one pegger (lines with squares) and one floater are presented in each row.

A closer inspection of impulse response functions presented in Figure 2 results in three observations. First, in general, there are more differences in the output reactions to supply shocks than to demand shocks. The former's contribution to output variability is much greater than that the latter's (see Table 1). This observation implies that any potential differences between pegs and floats identified below are even more important.

Second, the response of output under fixed exchange rate regime to at least one real shock is stronger than that under floating rate regime in each of our four pairs: Croatia, Slovakia and Slovenia react more intensely to a supply shock than their floating-rate counterparts, i.e. Hungary, Poland and the Czech Republic, whereas Bulgaria reacts stronger to a demand shock than Romania. This observation is in line with the view that exchange rate flexibility can be useful in insulating output against real shocks.

Third, in three pairs the median reaction of output under peg is outside the confidence interval for the corresponding reaction under floating exchange rate. This is especially the case for impulse response functions of Slovak and Polish output to supply shocks where there is no overlap between confidence intervals. The dissimilarity between reactions to a supply shock in Croatia and Hungary is smaller, although still quite pronounced. The least noticeable difference is between Bulgaria and Romania in their reactions to a demand shock: it prevails for one year only and in the long run it disappears completely (the difference in the response to a supply shock is reversed but even less distinct). In the remaining pair one can observe that the Slovenian reaction to a supply shock is a bit stronger than the Czech reaction, but both medians are within of the counterpart's confidence interval.

One can question our strategy of comparison pegs and floats in pairs arguing that the changes in pairs would result in different conclusions. To conserve space, we do not provide detailed arguments in favour of our pairs. Instead, we briefly discuss the results of comparison of output reactions to shocks in each country with analogous reactions in Poland. The latter has been chosen because its exchange rate was formally floated in April 2000 (and de facto in 1998 when the National Bank of Poland decided to refrain from foreign exchange market interventions), and out of CEE currencies it is the Polish zloty that has been floating for the longest time. Output reactions to supply shocks of all peggers turned out to be stronger except that of Slovenia which was comparable. The reactions of floaters were more diversified: stronger for Romania, comparable for Hungary and slightly weaker for the Czech Republic. The similar pattern was observed for output responses to demand shocks: stronger for peggers (except for Croatia that had a similar reaction) and comparable for floaters (except for the Czech Rep. that reacted more strongly).

Conclusion

This paper investigates the insulating properties of the floating exchange rate regime by comparing pegs and floats adopted in eight Central and Eastern European economies. We find important evidence that bolsters up the hypothesis that the flexible exchange rate regime insulates the economy against shocks to a greater extent than the fixed exchange rate regime. The caveat is that the degree of uncertainty about the results obtained is far from being small.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Centre in Poland (grant no. DEC-2012/07/B/HS4/00723).

References

Arias, J. E., Rubio-Ramirez, J. F., & Waggoner. D. F. (2014). Inference based on SVARs identified with sign and zero restrictions: theory and applications. *Board of Governors of the Federal Reserve System International Finance Discussion Papers.* pp. 1100.

Clarida, R., & Gali, J. (1994. December). Sources of real exchange-rate fluctuations: How important are nominal shocks?. In *Carnegie-Rochester conference series on public policy* (Vol. 41. pp. 1-56). North-Holland.

Dąbrowski. M. A. (2012). Exchange rate regimes and output variability in Central European countries. *Actual Problems of Economics, 2*(10), 80-91.

Dąbrowski. M. A., & Wróblewska, J. (2016a). Exchange rate as a shock absorber in Poland and Slovakia: Evidence from Bayesian SVAR models with common serial correlation. *Economic Modelling, 58*, 249-262.

Dąbrowski. M.A.. & Wróblewska. J. (2016b). Exchange rate regime and sensitivity to economic and financial shocks in Central European countries. In: *Proceedings of the 10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena (Papież. M.. and Śmiech. S.. eds.) Foundation of the Cracow University of Economics. Cracow* (pp. 26-35).

Edwards, S. (2011). Exchange-rate policies in emerging countries: Eleven empirical regularities from Latin America and East Asia. *Open Economies Review, 22*(4), 533-563.

Erdem, F. P., & Özmen, E. (2015). Exchange rate regimes and business cycles: An empirical investigation. *Open Economies Review, 26*(5), 1041-1058.

Engle, R. F., & Kozicki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics, 11*(4), 369-380.

Fleming, M. (1962). Domestic financial policies under fixed and floating exchange rates. *Staff Papers-International Monetary Fund, 9*, 369-380.

Ghosh, A. R., Gulde-Wolf, A. M., & Wolf, H. C. (2002). *Exchange rate regimes: choices and consequences*. MIT Press.

Ghosh, A. R., Ostry, J. D., & Tsangarides, C. G. (2010). *Exchange rate regimes and the stability of the international monetary system* (No. 270). International Monetary Fund.

Kuijs, L., & Borghijs, A. (2004). *Exchange Rates in Central Europe: A Blessing Or a Curse?*. International Monetary Fund.

Levy-Yeyati, E., & Sturzenegger, F. (2003). To float or to fix: Evidence on the impact of exchange rate regimes on growth. *The American Economic Review, 93*(4), 1173-1193.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Berlin and Heidelberg. Springer.

Mundell, R. (1963). Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economics and Political Science, 29*(04), 475-485.

Obstfeld, M., Cooper, R. N., & Krugman, P. R. (1985) Floating exchange rates: experience and prospects. *Brookings Papers on Economic Activity*, 1985(2), 369-464.

Shevchuk, V. (2014). Shock-absorbing properties of the exchange rates in transformation economies: SVAR estimates. In: *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena (Papież. M.. and Śmiech. S.. eds.) Foundation of the Cracow University of Economics. Cracow* (pp. 155-164).

Stążka-Gawrysiak, A. (2009). The Shock-Absorbing Capacity of the Flexible Exchange Rate in Poland. *Focus on European Economic Integration*, (4), 54-70.

GREG estimation with reciprocal transformation for a Polish business survey

Grażyna Dehnel[1]

Abstract

As a result of economic changes the scope of tasks for business statistics has expanded. There is a growing demand for short-term statistical data delivered on a monthly or quarterly basis, with improved accuracy and coherence. To meet this demand, it is necessary to develop estimation methods that can take advantage of administrative data. The purpose of these efforts is to increase the effectiveness of estimates and extend the scope of information in terms of the number of variables and cross-classifications.

The paper presents an attempt to estimate basic economic information about small enterprises by applying reciprocal transformation to one of the small area statistics methods - GREG estimation. Variables from administrative registers were used as auxiliary variables. The study was conducted for provinces cross-classified by categories of economic activity.

Keywords: *robust estimation, business statistics, outliers, GREG*

JEL Classification: C40, C51

1 Introduction

In recent years much attention has been devoted to business statistics. To meet the growing needs for information, it is necessary to conduct research aimed at expanding the scope of business statistics. The main difficulty facing official statistics in this respect is the rising nonresponse. In addition, the scope of economic information that can be collected is limited by survey costs and respondent burden resulting from statistical reporting. Under these circumstances, the growing demand for information can only be satisfied by exploiting administrative sources of data. It is expected that the adoption of new solutions will improve the efficiency of estimates and, above all, increase the number of cross-classifications available in statistical outputs (Markowicz, 2014). In the search for new approaches to estimating parameters of enterprises it is necessary to account for the specific characteristics of the target population. One characteristic feature of the population of enterprises is the presence of outliers (Schmid et al., 2016; Todorov et al., 2011). For this reason, the present article focuses on an estimation method used in small area estimation, which employs reciprocal transformation. The purpose of the study was to assess the possibility of applying

[1] Corresponding author: Poznań University of Economics and Business, Department of Statistics, Al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: grazyna.dehnel@ue.poznan.pl.

a modified generalised regression estimator (GREG) to estimate characteristics of small companies. In order to improve estimation quality, lagged auxiliary variables from administrative sources were used. Estimates were calculated using data about small companies classified into activity section: *Trade.*

2 The estimation method

One of the assumptions of a linear regression model is homoscedasticity. In the case of the population of enterprises characterised by strong asymmetry of distribution and strong variation, this assumption is often not satisfied (Zhang and Hagesaether, 2011). As a result, the variables are heteroscedastic, which results in inefficient estimates of parameters. For this reason, it is necessary to develop methods that minimize the impact of heteroscedasticity on the precision of estimates. One way of achieving this is by transforming variables. For example R. Chambers et al. (2001) propose a modification of the GREG estimator which involves an additional auxiliary variable z.

In its classic form the GREG estimator of the total of variable Y:

$$\hat{Y}_{GREG} = \sum_{i \in U} \hat{y}_i + \sum_{i \in s} w_i e_i \tag{1}$$

where $\hat{y}_i = \mathbf{x}'_i \hat{\beta}$ in a population U

s - sample size,

model parameter $\hat{\beta}$ is estimated using a modified formula which includes additional variable z (Chambers et al., 2001):

$$\hat{\beta} = \left(\sum_{i \in s} w_i \mathbf{x}_i \mathbf{x}'_i / z_i^{\gamma} \right)^{-1} \left(\sum_{i \in s} w_i \mathbf{x}_i y_i / z_i^{\gamma} \right) \tag{2}$$

where:

z, x - auxiliary variables,

γ - parameter selected depending on the degree of heteroscedasticity,

for $\gamma = 0$, estimator (1) has the classic form of the GREG estimator.

The estimator given by formula (1) can be expressed in the form which is identical to the classic formula for the GREG estimator:

$$\hat{Y}_{GREG} = \sum_{i \in s} w_i g_i y_i \ . \tag{3}$$

The only difference is the definition of weight g_i, which depends on the value of auxiliary variable x for sampled units and is defined as:

$$g_i = 1 + \left(X_d - \hat{X}_{HT,d}\right)'\left(\sum_{i \in s_d} w_i \mathbf{x}_i \mathbf{x}'_i / z_i^{\gamma}\right)^{-1} \left(\mathbf{x}_i / z_i^{\gamma}\right) \quad (4)$$

where:

d - domain,

g_i - weight of the *i*-th unit,

$\hat{Y}_{GREG,d}$ - estimate of the total in domain *d* given by the GREG estimator,

$\hat{X}_{HT}$ - direct Horvitz-Thompson (HT) estimator of the total of auxiliary variable *x*,

X - total of auxiliary variable *x*.

In the modified version of the GREG estimator proposed by R. Chambers et al. (2001), DFBETA is strongly correlated with weight g_i. This means that the value of the distance measure affects the degree to which variable *y* is modified. The DFBETA measure for *i*-th unit can be calculated using the following formula (Chambers et al., 2001):

$$DFBETA_i = \left(\sum_{i \in s} \mathbf{x}_i \mathbf{x}'_i / z_i^{\gamma}\right)^{-1} \left(\frac{\mathbf{x}_i}{z_i^{\gamma}}\right)\left(\frac{e_i}{1 - h_i}\right) \quad (5)$$

where: $$h_i = \left(\mathbf{x}_i / z_i^{\gamma}\right)'\left(\sum_{i \in n} \mathbf{x}_i \mathbf{x}'_i / z_i^{\gamma}\right)^{-1} \left(\mathbf{x}_i / z_i^{\gamma}\right) \quad (6)$$

From previous surveys it is known that the value of parameter γ is included in the interval <1,2> (Särndal, 1992), which is why the following estimators were analysed in the study:

1. $\hat{Y}_{GREG}^{0}$ - $\gamma = 0 \Rightarrow z_k^0$ an estimator based on a linear regression model under homoscedasticity (classic GREG estimator),
2. $\hat{Y}_{GREG}^{1}$ - $\gamma = 1 \Rightarrow z_k^1$,
3. $\hat{Y}_{GREG}^{1,5}$ - $\gamma = 1{,}5 \Rightarrow z_k^{1,5}$,
4. $\hat{Y}_{GREG}^{2}$ - $\gamma = 2 \Rightarrow z_k^2$.

Numbers 2-4 denote regression estimators based on a linear regression model under homoscedasticity.

Estimation quality was assessed with reference to estimates obtained using classic direct estimators: HT and the GREG estimator (Gamrot, 2014):

- Horvitz-Thompson (HT) estimator $$\hat{Y}_{HT} = \frac{N}{n}\sum_{i \in s} y_i \quad (7)$$

where : $\hat{Y}_{HT}$ - estimator of the total of variable Y,

N - general population size,

y_i - value of the variable of interest for the *i*-th unit.

- GREG estimator $\hat{Y}_{GREG} = \hat{Y}_{HT} + \left(\mathbf{X} - \hat{\mathbf{X}}_{HT}\right)'\hat{\boldsymbol{\beta}}$ (8)

where: $\hat{\mathbf{X}}_{HT} = \sum_{i\in s} w_i x_i$ - vector of direct HT estimators of auxiliary variables *x*,

$\mathbf{X}$ - vector of totals of auxiliary variables *x*,

$\hat{\boldsymbol{\beta}} = \left(\sum_{i\in s} w_i \mathbf{x}_i \mathbf{x}'_i\right)^{-1}\left(\sum_{i\in s} w_i \mathbf{x}_i y_i\right)$ is estimated using the method of weighted least squares using design weights,

w_i - design weights.

Values of the classic GREG estimator differ from estimates obtained using the estimator of interest - $\hat{Y}^0_{GREG}$. The differences are due to the fact that estimator $\hat{Y}^0_{GREG}$ does not take into account all sampled units and ignores those for which auxiliary variable 'z' is zero.

If the constant is omitted, the resulting estimator is called a ratio estimator (Hedlin, 2004).

3 Assumptions of the study

The empirical study included small companies (10-49 employees) conducting activity classified into section Trade. The response variable estimated in the model was Revenue obtained in June 2012. Information about the response variable came from the DG1 survey. The survey is conducted in the form of monthly reports submitted by all large and medium-sized enterprises and a 10% sample of small enterprises. (Dehnel, 2014). The following variables were selected as auxiliary variables 'x' and 'z': the number of employees from the social insurance register (ZUS) and revenue form the tax register for December 2011. The decision to use lagged variables was motivated by technical limitations involved in surveys conducted by the Central Statistical Office, namely the delay between the release of administrative data for purposes of official statistics. In the analysed model it is assumed that each auxiliary variable can be used both as auxiliary '*x'* and '*z'* (on condition that auxiliary 'z' cannot take zero values). Given the above, in the final approach, auxiliary variable 'z' was taken to be the number of employees.

Estimation was conducted for domains defined by cross-classifying provinces with the section of business activity according to the Polish Business Classification (PKD).

4 The method of evaluating precision

Precision and accuracy of estimates were evaluated using the bootstrap method. 1000 bootstrap samples were drawn, which were then used to estimate *Revenue* for June 2012 for domains of interests. Estimation efficiency was assessed using the coefficient of variation of the estimator (Bracha, 2004):

$$CV\left(\hat{Y}_d\right)=\frac{\sqrt{Var\left(\hat{Y}_d\right)}}{E\left(\hat{Y}_d\right)}=\frac{\sqrt{\frac{1}{999}\sum_{b=1}^{1000}\left(\hat{Y}_{b,d}-\hat{Y}_d\right)}}{E\left(\hat{Y}_d\right)}. \quad (9)$$

In order to estimate bias, it is necessary to know values of the estimated parameters for the general population. Since this information was not available in the survey it was estimated indirectly, based on data from the tax register for December 2012. It was assumed that the following relation holds true: the ratio of *Revenue* reported in tax statements submitted by companies at the domain level to the value of *Revenue* reported in the DG1 survey is constant.

$$\frac{Revenue_{XII2012}}{Revenue\_DG1_{XII2012}}=\frac{Revenue_{VI2012}}{Revenue\_DG1_{VI2012}} \quad (10)$$

$Revenue_{XII2012}$ ($Revenue_{VI2012}$) – value of revenue reported in the tax register in December 2012 (VI2012)

$Revenue\_DG1_{XII2012}$ ($Revenue\_DG1_{VI2012}$) – value of revenue reported in the DG1 survey in December 2012 (VI2012). This approach made it possible to determine approximate values of *Revenue* for June 2012.

5 Conditions of estimates and assessment of their quality

The first step of the analysis was to examine distributions of companies depending on the variables of interest. The coefficient of variation varied from 47% to 649%. The distributions were strongly asymmetric, with skewness coefficients ranging from 0.6 to 17.1.

The hypothesis of homoscedasticity was verified using the White test and the Breusch–Pagan test. For most domains of interest test results confirmed the validity of the hypothesis about the variability of the random component, see Table 1. This in turn justified the use of the above mentioned GREG estimators modified to account for variable 'z'.

The estimates were assessed both in terms of accuracy and precision. The point of reference for the assessment of precision were the estimates obtained using classic, direct estimators: the HT estimator and the GREG estimator, including its ratio estimators. Based on CV values as a measure of efficiency, it can be seen that the HT estimator exhibits the greatest degree of variation (see Table 1). Classic GREG estimators and transformed GREG estimators are characterized by less variation.

Test:	**White**		**Breusch-Pagan**	
Province	statistic	p-value	statistic	p-value
dolnośląskie	121.9	0.0000	86.7	0.0000
kujaw.-pom.	46.13	0.0000	35.05	0.0000
lubelskie	49.7	0.0000	1.95	0.3771
lubuskie	80.34	0.0000	37.33	0.0000
łódzkie	82.18	0.0000	34.9	0.0000
małopolskie	136.1	0.0000	6.36	0.0417
mazowieckie	169.7	0.0000	142.6	0.0000
opolskie	42.65	0.0000	36.43	0.0000
podkarpackie	33.91	0.0000	18.09	0.0001
podlaskie	33.33	0.0000	16.64	0.0002
pomorskie	40.15	0.0000	36.32	0.0000
śląskie	97.41	0.0000	82.98	0.0000
świętokrzyskie	23.45	0.0003	6.18	0.0456
warm.-maz.	48.97	0.0000	25.96	0.0000
wielkopolskie	121.5	0.0000	74.83	0.0000
zachodniopom.	121.1	0.0000	113.9	0.0000

Table 1. Results of the White test and the Breusch–Pagan test for heteroscedasticity for section *Trade* across provinces.

Source: based on data from the DG-1 survey.

On closer analysis, it can be seen that estimation precision of the GREG estimator depends on the sample size. It is usually bigger in domains that have more representation in the sample. In most domains of interest, the CVs of the transformed estimators, regardless of the value of the γ coefficient, are more or less similar and slightly higher than the classic GREG estimator. Moreover, it can be noticed that, as variability and asymmetry in a given domain increases, the gain in precision resulting from using each of the GREG estimators improves.

The precision of estimating *Revenue* of companies was assessed in reference to ratio estimates given by formula 10. In addition, to provide a more complete assessment, transformed estimators were compared with the HT estimators and the classic GREG estimators, see Fig. 2. The results indicate that the inclusion of variable ‘z’ in the model yields a considerable improvement in estimation precision, especially compared to the HT estimator, but also with respect to the classic GREG estimator.

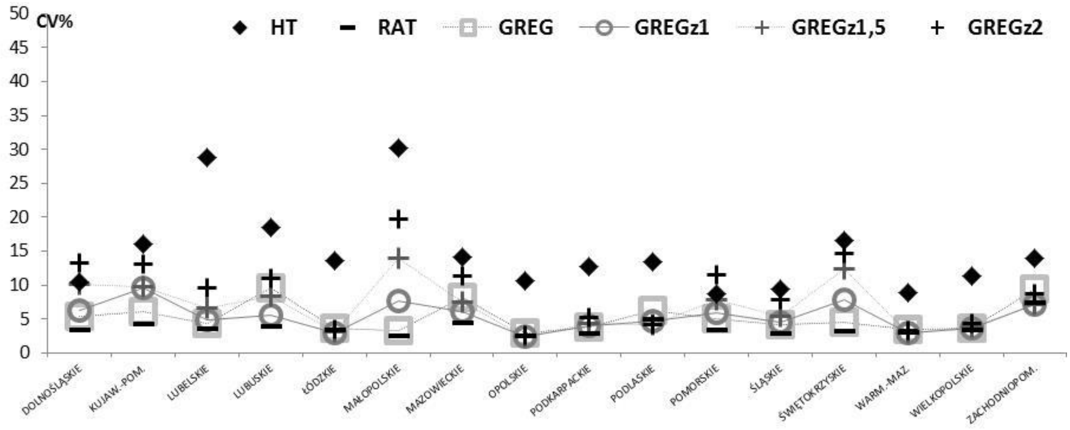


Fig. 1. Estimation precision (CV) for section *Trade* across provinces.
Source: based on data from the DG-1 survey and administrative registers.

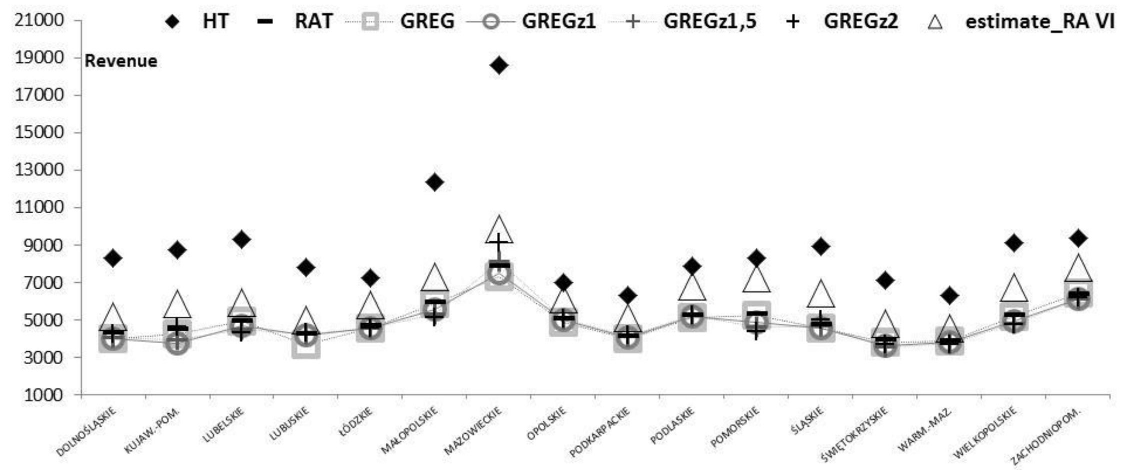


Fig. 2. Estimated revenue for June 2012 across provinces - section *Trade.*
Source: based on data from the DG-1 survey and administrative registers.

The HT estimator overestimated *Revenue* for nearly all domains of interest, while the GREG estimator tended to underestimate it. In contrast, estimates obtained by the transformed estimator are closest to the reference values. The largest discrepancies between parameter estimates produced by the different estimators can be observed for domains characterized by the largest variation and asymmetry of variables used in the model.

Figure 3 includes histograms for two selected provinces showing distributions of bootstrap estimates obtained by applying the studied estimators. The estimators using auxiliary variables from registers are less biased. As the γ parameter increases, the distribution of estimates tends to approximate the reference value marked in red.

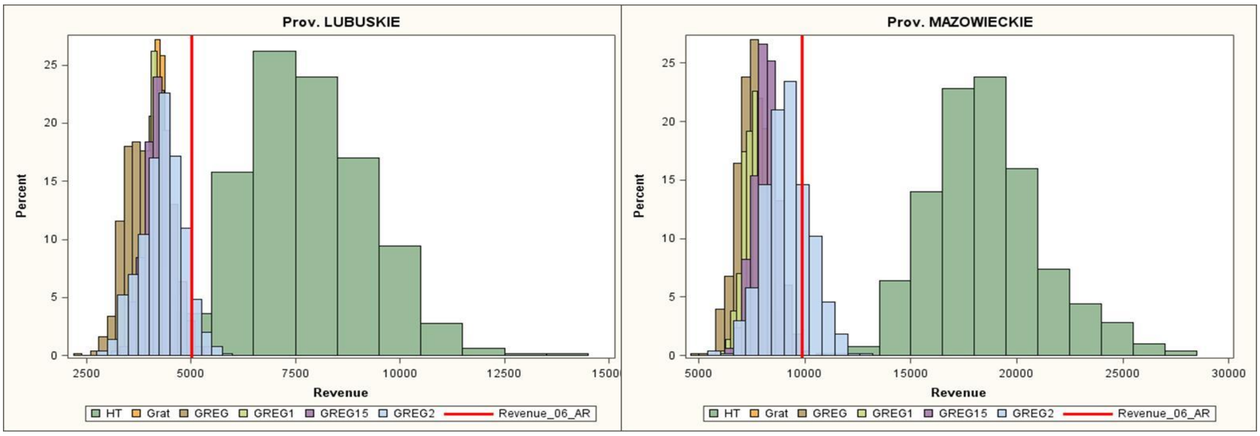


Fig. 3. Distribution of estimates for selected provinces for section *Trade.*

Source: based on data from the DG-1 survey and administrative registers.

Conclusion

The results obtained in the study lead to the following conclusions:

- the inclusion of auxiliary variables in the GREG estimator has considerably improved estimation precision compared to the HT estimator;
- the modified estimation method yields a greater gain in precision and accuracy for domains with greater variability and asymmetry;
- improvement in estimation precision for the estimators using reciprocal transformation depends on the value of the γ parameter. A considerable gain can be achieved if an appropriate model is selected for a given domain; the drawback of this approach, however, is that the application of modified GREG estimators for a very large number of domains is time-consuming and demanding.

Acknowledgements

The project is financed by the Polish National Science Centre, decision DEC-2015/17/B/HS4/00905.

References

Dehnel, G. (2014). Winsorization Methods in Polish Business Survey. *Sampling methods and estimation*, 97, 97.

Gamrot, W. (2014). Estimators for the Horvitz-Thompson statistic based on some posterior distributions. *Mathematical Population Studies*, *21*(1), 12-29.

Hedlin, D., Falvey, H., Chambers, R., & Kokic, P. (2001). Does the model matter for GREG estimation? A business survey example. *Journal of Official Statistics*, 17(4), 527-544.

Hedlin, D. (2004). Business survey estimation. *Statistiska centralbyrån*.

Lehtonen, R., Särndal, C. E., & Veijanen, A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, 7(3), 649-673.

Lehtonen, R., Särndal, C. E., & Veijanen, A. (2008, August). Generalized regression and model-calibration estimation for domains: Accuracy comparison. *In workshop on Survey Sampling Theory and Methodology* (pp. 25-29).

Markowicz, I. (2014). Business Demography-Statistical Analysis of Firm Duration. *Transformations in Business & Economics*, 13(2B), 801-817.

Rao, J. N., & Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

Schmid, T., Tzavidis, N., Münnich, R., & Chambers, R. (2016). Outlier Robust Small-Area Estimation Under Spatial Correlation. *Scandinavian Journal of Statistics*.

Todorov, V., Templ, M., & Filzmoser, P. (2011). Detection of multivariate outliers in business survey data with incomplete information. *Advances in Data Analysis and Classification*, 5(1), 37-56.

Zhang, L. C., & Hagesæther, N. (2011). A domain outlier robust design and smooth estimation approach. *Canadian Journal of Statistics*, 39(1), 147-164.

A spatio-temporal approach to intersectoral labour and wage mobility

Karol Flisikowski[1]

Abstract

The article presents the spatio-temporal approach for intersectoral labor and wage mobility. Analyses of interindustry mobility were performed with the use of general entropy mobility indices (GEMM). Spatio-temporal approach was obtained thanks to the separate measurement of spatial autocorrelation and regression for each set of sectoral wage and employment structure and was conducted in each year of the research period separately. Calculations of economic distance were based on the level of GDP, whereas in spatial regression data of previously calculated mobility indices were used. Because of the availability of homogeneous, highly aggregated sectoral data only for the period 1994-2011, the analyses were performed for 20 selected OECD countries. The use of spatial error model (SEM) and spatial lag model (SLM) improved explanatory abilities of the analysis and revealed that the higher level of interindustry wage mobility is accompanied by increased movement of labor force across sectors.

Keywords: *intersectoral mobility, spatial autocorrelation, wage mobility, labour mobility, spatial regression*

JEL Classification: J62, J21

1 Introduction

Mobility of wages and employment is an issue widely understood and analyzed. In this study, mobility is considered as a change in the structure of sectoral wages and labor force over time. This specific type of structural mobility can be characterized by a number of measures. Its choice influences their interpretation and economic sense and can be also associated with various factors of its economic environment. These include, among others: sector-specific human capital (often identified with the sectoral wages), unemployment, institutionalism, wage or income inequality. Several studies confirms the existence of clear links between employment and wage mobility at the micro-data level and form the basis for further investigations at highly aggregated sectoral level. The main objective of this paper is an attempt to aggregate and synthesizing of both mobility relationship in the form of one spatial regression model. The selection of a spatial model gives us an additional interpretability of results by implementation of weight matrix based on the economic distances. Another advantage of empirical analysis presented in this article is the form of intersectoral mobility[2].

[1] Gdansk University of Technology, Department of Statistics, Narutowicza 11/12, 80-233 Gdańsk, Poland, e-mail: karflisi@pg.gda.pl.

[2] This analysis was based on 3rd Revision of ISIC (International Standard Industrial Classification of All Economic Activities). To avoid the non-comparability of results (missing

2 Interindustry labour and wage mobility

Interindustry mobility (IM) can be understood as a cross-sectoral shift of workforce (Chiarini and Piselli, 2000) - intersectoral labor mobility (ILM). IM can also be defined as the degree of cross-sectoral shifts in wage differentiation (IWM - intersectoral mobility of wages). In majority of studies (both theoretical and empirical) researchers try to explain the determinants affecting the level of ILM and IWM. This leads to the conclusion that in studies on that relationship still many difficulties exist in explaining its cause and effect nature, so there is a presumption that a hypothetical relationship might be called as feedback.

The ratio of ILM to the level of equal pay is a very popular subject of many studies in the literature, but rarely can meet its reference to the scale of IWM. Those studies cover mostly the comparison of the ILM to the dynamics or growth of wage levels. In a study on the relationship between the mobility of employment and wage growth common conclusions can be found. It has been proved here that employment mobility leads to an increase in wages (usually from 10 to 20 percent). Slightly cautious estimates can be found in: Moore et al. (1998), McLaughlin (1990). The movers-stayers model also bind together both types of mobility (Ng and Chung, 2012) and is rooted in psychological arguments. In this model, some workers are expected to be more likely to move than others. More unstable units would therefore be less productive and would receive lower wages than others (stayers).

Other models that consider the connection between ILM and IWM are classified as static or dynamic due to the rejection of the assumptions about the dynamism of wages in the range of positions (Naticchioni and Panigo, 2004). The on-the-job search theory could therefore be classified as static, whereas current specific human capital theory as dynamic. Static models allow the inclusion of such an interdependency only in the range of the specific changes in occupation or industry, whereas dynamic models recognize changes in wages combined with shifts of resources between and in the range of the same occupation or sector. In search models it is most often indicated that shorter seniority is correlated with an increase in the level of wage mobility and that fact brings the most profitable gains in wage at the beginning of careers. The same conclusions are met by modification of that theory introduced by Burdett (1978). In human capital theory however, an inversed relationship between mobility of labor force and investments in specific job skills is indicated, but does not define clearly and precisely the relationship between ILM and IWM. It points out that the more specific the human capital transfers, the lower the expected decline in wages in relation to the expected

data, different revisions of ISIC), the empirical analyses were performed with the use of data reduced to the same time dimension (1994-2012) for 20 OECD countries.

mobility of employment. Another dynamic approach represents the theory in which the employee is looking for a job to find the best fit to his potential expectations. Many researchers believe that the worse the quality of such a match, the shorter the period of employment and the increase of wage might be related to the reward for the search for a better fit, regardless of the accumulation of specific human capital. The job-match theory does not conclude directly on the exact relationship of ILM and IWM (Naticchioni and Panigo, 2004; Göke et al., 2014). It is a theoretical model where optimum conditions for job changes determine a positive correlation between the length of employment and short-term increases in the scale of mobility. Institutional factors can affect both the shift in the structure of employment and wages growth in a number of ways. In the first case, the legal protection of employment has a significant role in the dismissal of workers and new employment for a temporary period (Baulch and Hoddinott, 2000). The more flexible the labor market, the greater the expected effects might be met (mobility can have erosive effect on wages).

3 Methodology

Interindustry mobility in majority of the empirical research is measured with the use of the individual micro-data. This entailed consequences in the application of specific statistical methods. Hence, most of the research rely on the same or very similar methodological solutions. The empirical analysis was performed in three steps for 20 selected countries (for the period: 1994-2012), which are not in every case reciprocal neighbours. Thus, it was necessary to construct a spatial weight matrix based on economic distances[3]. This kind of statistical analysis nests inside the spatial model an additional interpretation of coefficients.

First stage of analysis covers calculations of Shorrocks (1978) mobility indices (for each country, its structure of wages and labor force, keeping 2-years subperiods). The measure proposed by Shorrocks belongs to the group of generalized entropy mobility measures (GEMM) and was generalized by Maasoumi and Zandvakili (1986). They allow us to observe the degree of structural substitution between employment or wages in different periods of time. In previously conducted studies it was concluded that those indices meet most of the requirements for measurement of mobility. Let Y_{it} be the wage (or employment) for sector i in period $t=1,\dots,T$. Hence, the Shorrocks index of mobility (M) can be defined in formula (1).

$$M = 1 - \frac{I(S)}{\sum_{t=1}^{t} \alpha_t I(Y_t)} \tag{1}$$

[3] The value of real GDP was chosen for that measure.

where: $S=S_{1},\ldots,S_{n}$ is a vector of permanent or aggregated wages/employment in time T, $Y_t=Y_{1t},\ldots,Y_{nt}$ is a vector of sectoral wages/employment in time t, α_t is related weight and $I()$ stands for chosen inequality measure. This measure is a negative function of the relative stability of the distribution and shows the ratio of long-term inequality (permanent and aggregate) $I(S)$ to the short-term inequality $I(Y_t)$. The level of mobility will increase if the long-term inequalities will be reduced more than the short-term.

In the second stage of analysis, the spatial autocorrelation for previously obtained wage and employment mobility measures was checked. According to the first law of geography formulated by Tobler (1970), all objects in space (observation units) interact, and spatial interactions are greater the smaller is the distance between objects. Thus, in the analysis we must consider the spatial interactions, which may relate to the dependent variable and random component. In a situation where the value of the dependent variable in each location affect the value of this variable from other locations the so-called spatial autoregression exists. On the other hand, a case where certain spatially autocorrelated variables are omitted or cannot be considered relates to spatial autocorrelation of the random component (Rogerson, 2001). We can consider the specific relationship between the observation units (resulting from their location) thanks to the design of spatial weight matrix (Anselin, 1988; LeSage and Pace, 2008; Elhorst, 2014). It is a square matrix with $n\times n$ dimensions, which elements reflect the existing spatial structure. Specification of that matrix belongs to arbitrary decisions taken by a researcher and a choice of the alternative method of weighing is often due to the knowledge of the spatial structure of the phenomenon and links between units. It is assumed that links of spatial entities are positively affected by mutual proximity and negatively by shared distance.

Specification of spatial weight matrices is a prerequisite and the first step in the analysis of spatial autocorrelation. Among many measures used for spatial relationships testing the most commonly used is Moran's I statistic (Longley et al. 2008). This statistic is calculated based on the formula (2).

$$I=\frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}\left(z_i-\bar{z}\right)\left(z_j-\bar{z}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}\sum_{i=1}^{n}\left(z_i-\bar{z}\right)^2} \tag{2}$$

where:

n – number of observations (locations),

z_i – the observed value of the z variable for all n observations (locations),

w_{ij} – weight of spatial interactions (connections) between observations (locations) i and j.

The statistical significance of spatial autocorrelation measured by Moran's I statistic assuming null hypothesis of a random distribution of *z*-values (lack of spatial autocorrelation) is verified with the standardized normal Z_I statistic.

In the last stage of analysis, in case of spatial autocorrelation (Rogerson, 2001) two types of regression models with spatial effects were proposed[4]: SAR – spatial autoregressive model (also classified as spatial lag model – SLM) and spatial error model (SEM). The response to the negative impact of the spatial interaction to estimate the structural parameters of the OLS (ordinary least squares regression) models is an implementation to the classical form of the regression equation an additional independent variable and its parameter of ρ relating to this variable (called spatial autoregression coefficient). This variable (spatial lag) determines spatially delayed values of dependent variable, calculated as a weighted average (according to the adopted spatial weight matrix) from the value of this variable occurring in the neighborhood. We can formulate SLM in equation (3).

$$y_r = \rho\left(\sum_{s=1}^{n} w_{rs} y_s\right) + \sum_{i=1}^{k} \beta_i x_{ir} + \varepsilon_r \, . \tag{3}$$

Spatial error model (SEM) allows us to consider the spatial dependence of the sampling error (Rogerson, 2001). In this model, the overarching scheme of linear spatial autocorrelation of the random component is considered. It can be written as shown in equation (4).

$$y_r = \sum_{i=1}^{k} \beta_i x_{ir} + \varepsilon_r \, , \tag{4}$$

$$\varepsilon_r = \lambda\left(\sum_{s=1}^{n} w_{rs} \varepsilon_s\right) + u_r \, . \tag{5}$$

ε_r presented in equation (5) stands for the original random component with spatial autocorrelation (residuals from OLS regression for *r*-th location), which is a function of spatially delayed random error. The coefficient λ however, is a measure of interdependency of OLS residuals and on its basis, we can infer the existence of significant factors influencing on values of dependent variable, which were not included in the regression model.

4 Results

In the first stage of the analysis, the calculations of Shorrocks mobility indices were made (separately for labor and wage structures). In the second stage, for each subperiod and for

[4] It should be mentioned, that these are only the most popular examples of the wide range of spatial models reported in the literature multiplied with their numerous extensions and modifications.

previously calculated measures of mobility, a spatial autocorrelation Moran's measure was estimated. When spatial autocorrelation statistics are computed for variables, they assume constant variance. This is usually violated when the variables are for areas with greatly different populations. That is why the Assunção-Reis empirical Bayes standardization (Assunção and Reis, 1999) was implemented here to correct it. For each subperiod (2-years) between 1994 and 2012 negative, statistically significant ($p<0.01$) spatial autocorrelation statistics for ILM and IWM measures were obtained (from -0.2 in first subperiod to -0.27 in the last one). This was the basis for estimation of structural parameters of spatial regression models in the third stage of analysis (Rogerson, 2000).

In Table 1. the results of an estimation of linear regression models LM and regression models based on the matrix of spatial weights: SEM (spatial error model) and SLM (spatial lagged model) in two opposite subperiods are presented. The obtained results (presented in Table 1.) have correct statistical properties (LR and BP tests, significance of coefficients, Akaike criterion, R^2) and correct economic interpretation. The spatial error models (SEM) showed better performance and statistical significance of parameters than linear model (LM) and spatial lag model (SLM) in considered periods.

Interindustry labor mobility (ILM)	**LM**	**SEM**	**SLM**	**LM**	**SEM**	**SLM**
		1994-1996			**2010-2012**	
constant	0.003 (0.016)	0.003 (0.039)	-0.153 (0.017)	0.008 (0.010)	0.0087 (0.000)	0.008 (0.002)
Interindustry wage mobility (IWM)	0.319 (0.001)	0.3113 (0.000)	0.301 (0.000)	0.298 (0.001)	0.313 (0.000)	0.307 (0.000)
λ / ρ		-0.179 (0.035)	-0.153 (0.037)		-0.195 (0.013)	-0.187 (0.031)
R^2	0.536	0.538	0.553	0.626	0.664	0.632
Log-likelihood	82.680	82.710	83.032	84.838	85.672	84.979
Akaike criterion	-159.361	-159.420	-158.064	-163.678	-165.346	-157.977
LM		4.653 (0.030)	3.967 (0.045)		4.923 (0.026)	3.172 (0.074)

Table 1. Estimation of linear and spatial regression functions for intersectoral mobility (p-values in brackets).

The considered matrix of such a specific type of spatial weights led to the discovery of negative spatial autocorrelation – the intensity of interindustry labor force and wage mobility for neighboring countries (in terms of economic proximity) occurred to be completely different. What is more, statistically significant relationship between ILM and IWM was synthesized in form of one final version of regression model (SEM) and highlighted the negative value of the correlated random component. This means that specific individual effects influence the intensity of both phenomena among OECD countries. It may be a recommendation for further research in this area in order to discover the causes of such a situation.

Conclusions

In this article the problem of use of the spatial weight matrix based on the economic distance within the framework of the author's analysis of interindustry mobility phenomena was presented. The results of empirical analysis indicate that in case of the research on employment and wage mobility even studies at the most aggregate level of observation should be taken into account. Furthermore, the use of weight matrix based on the economic distance in statistical models of employment mobility greatly increases the correct interpretive impact of explanatory variable like intersectoral wage mobility, and thus significantly improves the quality of research. The higher level of interindustry wage mobility is accompanied by increased movement of labor force across sectors. Moreover, the strength of this association increased over time.

Acknowledgements

The results presented are an extension of my PhD thesis written under the aegis of Professor Stanisław Maciej Kot, to whom I would like to express sincere gratitude for scientific supervision and friendly support throughout.

References

Baulch, B., & Hoddinott, J. (2000). Economic mobility and poverty dynamics in developing countries. *The Journal of Development Studies*, 36(6), 1-24.

Burdett, K. (1978). A theory of employee job search and quit rates. *The American Economic Review*, *68*(1), 212-220.

Chiarini, B., & Piselli, P. (2000). Unemployment, wage pressure and sectoral shifts: permanent and temporary consequences of intersectoral shocks. *Journal of Policy Modeling*, 22(7), 777-799.

Elhorst, J. P. (2014). Linear Spatial Dependence Models for Cross-Section Data. In *Spatial Econometrics* (pp. 5-36). Springer Berlin Heidelberg.

Göke, S., Prinz, J., & Weimar, D. (2014). Diamonds are Forever: Job-Matching and Career Success of Young Workers. Jahrbücher für Nationalökonomie und Statistik, 234(4), 450-473.

LeSage, J. P., & Pace, R. K. (2008). Spatial Econometric Modeling of Origin- Destination Flows. *Journal of Regional Science*, 48(5), 941-967.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2006). GIS. *Teoria i praktyka. Wydawnictwo naukowe PWN, Warszawa*, 164-182.

Maasoumi, E., & Zandvakili, S. (1986). A class of generalized measures of mobility with applications. *Economics Letters*, *22*(1), 97-102.

McLaughlin, K. J. (1990). General productivity growth in a theory of quits and layoffs. *Journal of Labor Economics*, *8*(1, Part 1), 75-98.

Moore, W. J., Newman, R. J., & Turnbull, G. K. (1998). Do academic salaries decline with seniority?. *Journal of Labor Economics*, *16*(2), 352-366.

Naticchioni, P., & Panigo, D. (2004). *Employment protection, job-tenure and short term mobility wage gains-A new explanation for the Italian case* (No. 4-DEISFOL). Dipartimento di Economia, Sapienza University of Rome.

Ng, M. K., & Chung, Y. (2012). Double Mover–Stayer model on customer switching in telecommunications industry. *Naval Research Logistics (NRL)*, 59(8), 663-674.

Rogerson, P., 2001, Statistical methods for geography, SAGE Publications, London–Thousand Oaks–New Delhi.

Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 1013-1024.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*(sup1), 234-240.

The analysis of changes in the distribution of renewable energy consumption in the EU countries

Katarzyna Frodyma[1]

Abstract

The main research objective of the paper is to evaluate the changes in the distribution of renewable energy sources (RES) in the period between 1995 and 2014. The share of RES in total primary energy supply in the analysed period increase twofold (from 8% in 1995 to 16% in 2014). Thus, it might be interesting to check which types of RES develop more than others. Principal component analysis (PCA) is used to describe the share of different RES in overall renewable energy (RE) in EU member countries in the period between 1995 and 2014. The results of analysis for the year 1995 are compared with the results obtained for the year 2014, and the comparison indicates which RES gain in importance during the last 20 years. PCA also allows for describing RE development and identifying the direction of progress in this area, including progress resulting from technological advances. The main types of RES include: hydropower, wind energy, solar energy, tide, wave and ocean, biomass and renewable wastes and geothermal sources. The analysis demonstrates that three types of RES - hydropower, bioenergy and geothermal energy - predominate in analysed European countries in 1995. During the next years the share of RE in the energy mix increases, and the distribution of RES changes, with a notable increase of the share of wind and solar energy in the overall share of RES.

Keywords: *Renewable energy, European Union, PCA*

JEL Classification: C38, N740, Q2

1 Introduction

The EU dependence on import of energy sources contributes to the growing interest in renewable energy sources (hereafter RES), which is reflected in the introduction of relevant directives in the area of energy policy. The first one, issued in 2001, Directive 2001/77/EC[2], aimed at obtaining 7.5% share of energy from RES in the total gross electricity consumption in 2010 and 5.75% share of biofuels in the consumption of transport fuels. Another, Directive 2003/30/EC[3], aimed at the promotion of the use of biofuels and other renewable fuels for transport. In accordance with Directive 2009/28/EC[4], the EU member countries should

[1] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: frodymak@uek.krakow.pl.

[2] Directive 2001/77/EC of the European Parliament and of the Council of 27 September 2001 on the promotion of electricity produced from RES in the internal electricity market.

[3] Directive 2003/30/EC of the European Parliament and of the Council of 8 May 2003 on the promotion of the use of biofuels or other renewable fuels for transport.

[4] Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources.

increase their share of energy from renewable sources in the overall energy consumption. The European climate and energy package includes targets for 2020: a 20% reduction in the EU greenhouse gas emissions below the 1990 levels, raising the share of the EU energy consumption produced from RES to 20%, and a 20% improvement in the EU's energy efficiency. Additionally, at least 10% of the final consumption of energy in transport should come from RES. Although over the last 20 years the share of RES in the energy mix increases on average from 8% to almost 16%, it should be noted that the speed of development of RES is not the same in all EU countries.

Literature reports numerous studies on the development of renewable energy (for example, Sen nad Ganguly, 2017), but they often focus on individual countries (Dunjić et al., 2016, Câmpeanu and Pencea, 2014, Zakery et al., 2015) or only on electricity (Paska and Surma, 2014). Moreover, these papers do not review current changes in the structure of energy generated from renewable sources, especially with the use of methods capturing the most important factors of these changes. The study described in the paper is based on principal component analysis (PCA hereafter), which is used to describe different shares of RES in overall RE. The results of the analysis for the year 1995 are compared with the results obtained for the year 2014, which reveals which RES gain in importance during the last 20 years. PCA also allows for describing RE development and identifying the direction of progress in this area, including progress resulting from technological advances.

The aim of the paper is to evaluate the changes in the distribution of RES in the period between 1995 and 2014.

The study is conducted for 26 EU countries, excluding Malta and Cyprus, due to their almost non-existent share of RE in TPES, and the fact that they are totally dependent on import of energy sources.

Choosing the year 1995 as a reference point is dictated by the fact that it was the year in which the EU initiated legal procedures aimed at promoting RE development. That year the European Commission published *Green Paper*[5], which delineates the European Union energy policy and lists three basis targets connected with gas and electricity monopolies. It was also the year in which another official document, *White Paper. An Energy Policy for the European Union*[6], was issued by the Commission of the European communities. It contains a detailed

[5] Green Paper: For a European Union Energy Policy. European Commission, COM(94) 659, Brussels 1995.
[6] White Paper: An Energy Policy for the European Union. COM (95) 682 final, 13 December 1995.

set of regulations within the area of energy policy and states general frameworks of this policy in the EU countries (i.e. globalisation of energy markets, ecological problems, technology, institutional responsibility of the community, etc.).

The remainder of the paper is organized as follows. Section 2 briefly presents the empirical methodology used in the study, the description of the data can be found in Section 3, Section 4 reports and comments on the empirical results, and the last section contains final conclusions.

2 Methodology and data

The study compares the distribution of RES in the EU countries in the year 1995 and the year 2014. In order to capture correlations between different sources of RE, the distribution is obtained via classical principal component analysis (PCA).

PCA is normally applied as a method of variable reduction or for the detection of the structure of the relationship among variables. The information available in a group of variables is summarized by a number of mutually independent principal components. Each principal is basically the weighted average of the underlying variables. The first principal component always has the maximum variance for any of the combinations. If more than one principal component is generated, they are uncorrelated. For each principal component the eigenvalue (variance) indicates the percentage of variation in the total data explained.

The empirical analysis is conducted using the data which describe the share of RES in TPES in a sample of 26 European Union member states: Austria, Belgium, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Hungary, Greece, Germany, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom. The analysis covers the period between 1995 and 2014, and the data are obtained from the European Commission websites[7]. The study does not include Cyprus and Malta due to the fact that, in the analysed period, the share of RES in TPES in these countries is almost zero.

The dataset demonstrates the share of RES in overall RE. The main types of RES include: hydropower (HYDRO), wind energy (WIND), solar energy (SOLAR), tide, wave and ocean (TIDE), biomass and renewable wastes (BIOMASS), and geothermal sources (GEOTHERMAL).

[7] Energy datasheets: EU-28 countries (https://ec.europa.eu/energy/en/data-analysis/country, accessed on 30.10.2016 r.

3 Results

A detailed description of variables is provided in Table 1. As the table demonstrates, the share of RES in the period between 1995 and 2014 is not stable. Biomass constitutes the greatest share in overall RE: both in 1995 and in 2014 its share in the EU is on average 69%. The second most common RES is hydropower, whose share in overall RE decreases from 28% in 1995 to almost 15% in 2014. A substantial increase can be noticed in the share of wind energy, whose average share in overall RE for the EU countries increases from 0,6% in 1995 to almost 10% in 2014. The greatest, over a tenfold, growth can be noticed in the average use of solar energy: in 1995 solar energy in the EU constitutes only 0.4% of overall RE, while 20 years later - over 4%.

	Mean		**Min**		**Max**		**SD**	
Variable/year	**1995**	**2014**	**1995**	**2014**	**1995**	**2014**	**1995**	**2014**
HYDRO	0.2750	0.1463	0.0000	0.0002	0.8434	0.4278	0.2072	0.1311
WIND	0.0064	0.0993	0.0000	0.0000	0.0778	0.4599	0.0163	0.1038
SOLAR	0.0040	0.0428	0.0000	0.0000	0.0636	0.2118	0.0121	0.0520
TIDE	0.0001	0.0001	0.0000	0.0000	0.0026	0.0019	0.0005	0.0004
BIOMASS	0.6937	0.6962	0.1566	0.3828	1.0000	0.9371	0.2226	0.1623
GEOTHERMAL	0.0207	0.0154	0.0000	0.0000	0.4103	0.1975	0.0802	0.0392

Table 1. Summary statistics – the share of renewable energy sources in the renewable energy mix.

As mentioned above, the share of RES in TPES in the analysed period increases twofold (from 8% in 1995 to 16% in 2014). Thus, it might be interesting to check which types of RES develop more than others. To do so, we use PCA to describe and compare the distribution of RES. Table 4 presents the results of PCA for RES in 1995 and 2014.

In 1995 only two first principal components are sufficient to explain about 92% of variance in the original variables. The results also demonstrate that in 1995 these two principal components depend on the combinations of only three types of RES: hydropower, biomass and geothermal. In 2014, however, it is necessary to use three first principal components to explain 90% of total variance in the original variables. These principal components depend on the combinations of all types of RES, excluding tide. These results indicate that the distribution of RES in the EU countries changed dramatically over the last 20 years.

The results obtained for 1995, (Table 2) reveal that the first principal component ($PC_{95}1$) is related to two types of RES: biomass and hydropower. Hydropower is positively correlated with the first principal component ($PC_{95}1$), while biomass is negatively correlated with this component ($PC_{95}1$). It means that in 1995 the first principal component ($PC_{95}1$) divides the European countries into the ones which use hydropower but not biomass and the ones which use biomass but not hydropower as their main type of RES.

	$PC_{95}1$	**$PC_{95}2$**	**$PC_{95}3$**	**$PC_{95}4$**	**$PC_{95}5$**	**$PC_{95}6$**
HYDRO	0.675	0.471	-0.326	0.128	0.183	0.408
WIND			0.755	0.48	0.182	0.408
SOLAR			0.277	-0.851	0.182	0.408
TIDE					-0.913	0.408
BIOMASS	-0.734	0.349	-0.353	0.121	0.183	0.408
GEOTHERMAL		-0.810	-0.350	0.121	0.183	0.408
Cumulative variance	0.703	0.924	0.969	0.999	1.000	1.000

Table 2. Principal component analysis for renewable energy sources in 1995.

However, in 2014 (Table 3) the first principal component ($PC_{14}1$) differentiates the countries which either use a lot of hydropower and wind energy as their main types of RES but little biomass or the ones which use little hydropower and wind energy and a lot of bioenergy. In 1995 three types of RES - hydropower, biomass and geothermal energy - are correlated with the second principal component ($PC_{95}2$). However, hydropower and biomass are positively and geothermal energy is negatively correlated with the second principal component ($PC_{95}2$). It means that the European countries which use hydropower and bioenergy as their main type of RES and do not use geothermal energy (or vice versa) are described by the second principal component ($PC_{95}2$). However, in 2014 the highest value of the factor loadings for the second principal component ($PC_{14}2$) is obtained for hydropower and wind energy. The component also depends on the combination of solar, biomass and geothermal energy. Thus, the second principal component ($PC_{14}2$) characterizes the countries which use hydropower as their main type of RES and do not use wind energy (or vice versa). The third principal component ($PC_{14}3$) indicates these European countries which use solar as well as geothermal energy as their main types of RES and use small shares (less than average) of other RES.

	PC$_{14}$1	**PC$_{14}$2**	**PC$_{14}$3**	**PC$_{14}$4**	**PC$_{14}$5**	**PC$_{14}$6**
HYDRO	0.533	0.608	-0.355	-0.143	0.183	0.408
WIND	0.150	-0.745	-0.470		0.182	0.408
SOLAR		-0.127	0.706	-0.525	0.182	0.408
TIDE					-0.913	0.408
BIOMASS	-0.826	-0.212	-0.212	-0.120	0.183	0.408
GEOTHERMAL		0.331	0.331	0.830	0.183	0.408
Cumulative variance	0.461	0.768	0.904	0.999	1.000	1.000

Table 3. Principal component analysis for renewable energy sources in 2014.

Fig. 1 presents the results of PCA for RES in 1995 and Fig. 2 presents the results in 2014.

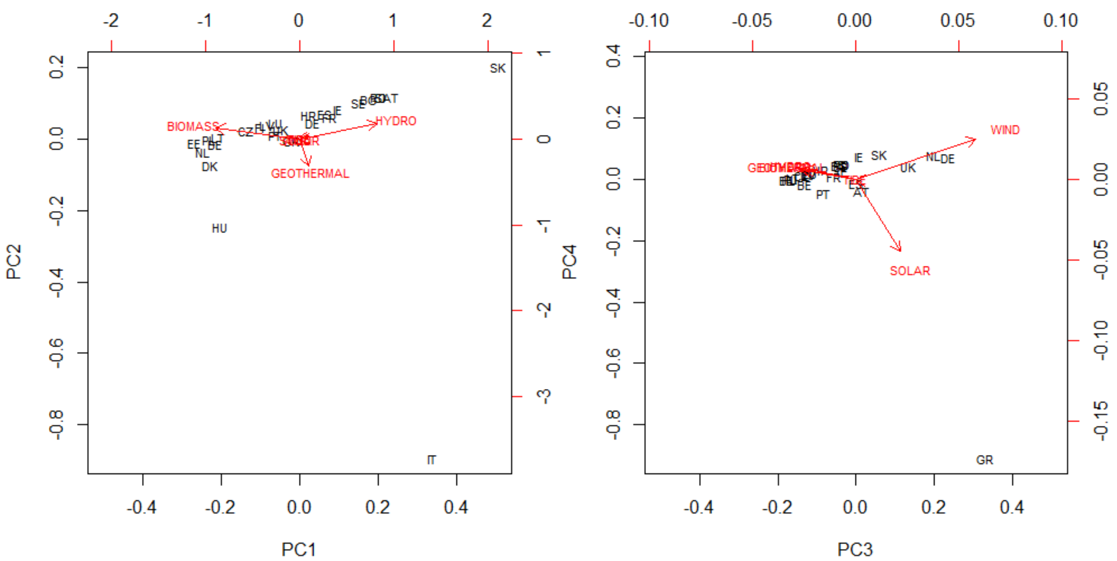


Fig. 1. The results of PCA for renewable energy sources in 1995.

The results demonstrated in Fig. 1 reveal a large concentration of most countries close to the average values of all variables. There are, however, some exceptions: Italy, Greece, and Slovakia are the countries which drive large values of the second, first and fourth principal component, respectively. In 1995 in most EU countries biomass and hydropower are the main types of RES. Slovakia and Italy has the highest share of hydropower in RES and, at the same time, a low share of biomass. Estonia, Netherlands, Poland, Denmark, Belgium, and Lithuania display a high share of biomass and a low share of hydropower. Variance of the second principal component results from a high share of geothermal energy in total RES in Italy (about 40%) and in Hungary (10%). Greece stands out from other countries due to its relatively high share of renewable energy obtained from solar energy (6% share of solar energy in RES in Greece, which is 10 times more that in Germany, the forerunner). On the other hand, Denmark is characterised by a relatively high share of wind energy (2%) in its RE mix.

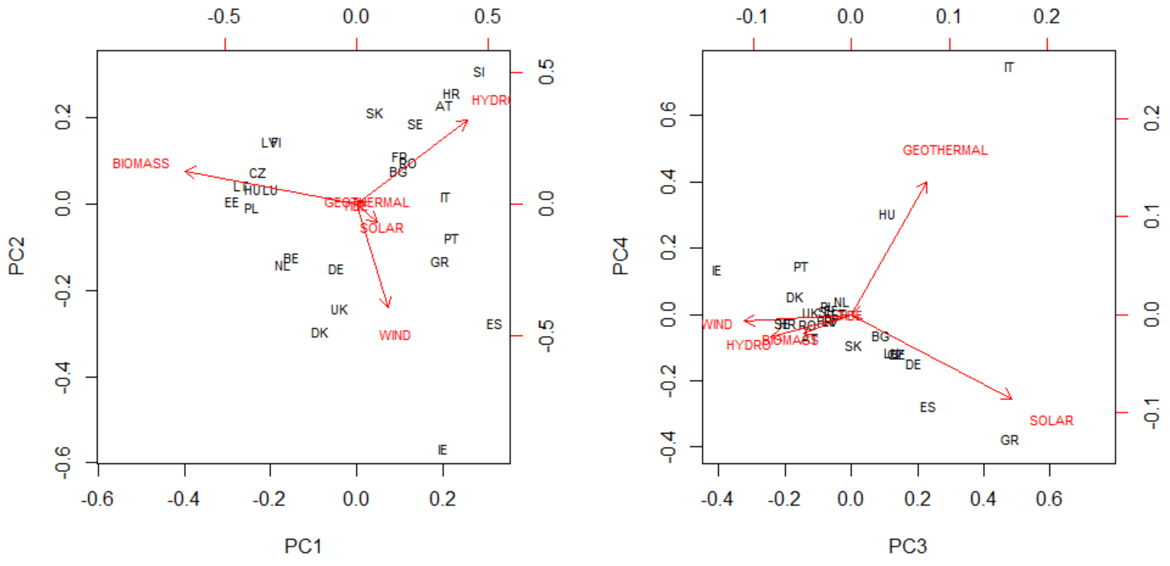


Fig. 2. The results of PCA for renewable energy sources in 2014.

Fig. 2 reveals a significantly greater diversity in the shares of different types of RES in the EU countries in 2014 than in 1995. There is a group of countries with high shares of biomass as their RES but small shares of hydropower (Estonia, Lithuania, Poland, Hungary, Czech Republic). Another group is characterised by an opposite distribution of RES, i.e. a low share of biomass and a large share of hydropower (Spain, Slovenia, Croatia, Portugal). In many European countries (Ireland, Denmark, Spain, United Kingdom) in the analysed period the share of wind energy as a renewable energy source increases. As far as solar energy is concerned, Greece maintains its leading position, while geothermal energy again is the most important energy source in Italy (in comparison to other countries).

Conclusion

The analysis reveals that RE development in the EU member countries is relatively diverse. In the analysed period all EU countries increase their shares of RES in the energy mix, however, this increase is uneven, and the shares of particular RES in particular countries are not the same.

Three types of RES - hydropower, bioenergy and geothermal energy - predominate in analysed European countries in 1995, which can be divided into the ones which use hydropower but not biomass and the ones which use biomass but not hydropower as their main type of RES. During the next years the share of RE in the energy mix increases and the distribution of RES changes, with a notable increase of the share of wind and solar energy in the overall share of RES.

The results from 1995 reveal a large concentration of most countries close to average values of all variables, while the results from 2014 reveal a significantly greater diversity in the shares of different types of RES in the EU countries than in 1995. It is possible to single out individual countries which use only one specific source of renewable energy, e.g. Italy uses geothermal energy, and Greece uses solar energy.

The changes in the distribution of renewable energy include not only the increase in the share of RE in the energy mix in the EU member states but also the kinds of RE used by them. During the last 20 years technological progress has allowed for using such renewable energy sources as wind or the sun, which, together with decreasing costs of investment in new sources is conductive to the stability of economy.

It should be mentioned that the project of the new directive regarding renewable energy sources (Renewable Energy Directive – RED II) for the period 2020 – 2030 advocates withdrawal from promoting the first generation biofuels as a renewable energy source (unless it is used in high-efficiency cogeneration), which means that biomass is treated differently than other RES.

Acknowledgements

Publication was financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Câmpeanu, V., & Pencea, S. (2014). Renewable energy sources in Romania: from a “paradise” of investors to a possible abandon or to another boom? The impact of a new paradigm in Romanian renewable sources policy. *Procedia Economics and Finance*, *8*, 129-137.

Dunjic, S., Pezzutto, S., & Zubaryeva, A. (2016). Renewable energy development trends in the Western Balkans. *Renewable and Sustainable Energy Reviews, 65*, 1026-1032.

EU Commission. (1995). Green Paper: For a European Union Energy Policy. European Commission, COM(94) 659, Brussels 1995.

EU Commission. (1995). White Paper: An Energy Policy for the European Union. COM (95) 682 final, 13 December 1995.

EU Commission. (2001). Directive 2001/77/EC of the European Parliament and of the Council of 27 September 2001 on the promotion of electricity produced from renewable energy sources in the internal electricity market. *Official Journal of the European Communities L*, *283*.

EU Commission. (2003). Directive 2003/30/EC of the European Parliament and of the Council of 8 May 2003 on the promotion of the use of biofuels or other renewable fuels for transport. *Official Journal of the European Union*, *5*.

EU Commission. (2009). Directive 2009/28/EC of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. *Official Journal of the European Union L*, *140*(5), 16-47.

Paska, J., & Surma, T. (2014). Electricity generation from renewable energy sources in Poland. *Renewable Energy*, *71*, 286-294.

Sen, S., & Ganguly, S. (2016). Opportunities, barriers and issues with renewable energy development–A discussion. *Renewable and Sustainable Energy Reviews*.

Zakeri, B., Syri, S., & Rinne, S. (2015). Higher renewable energy integration into the existing energy system of Finland–Is there any maximum limit?. *Energy*, *92*, 244-259.

Industry specifics of joint-stock companies in Poland and their bankruptcy prediction

Sergiusz Herman[1]

Abstract

The bankruptcy of companies is a characteristic of every developed market economy. The risk of bankruptcy is an object of interest for a wide group of stakeholders, including owners, employees, managers, creditors and suppliers. Negative consequences of bankruptcy led to many attempts to predict it. The one of direction of research on predicting the bankruptcy of companies is building models depending on characteristics of the industry of researched companies. Due to a difficulty of gathering a research sample that is big enough, Polish researchers rarely try to build models depending on industries. There are only two examples of authors who have compared the choice of classification for industry models and a "general" model (which does not include the characteristics of industry). There are two main aims of research. The first of them is to compare predictive ability of industry and general models error of industry and general models. The other aim was to define determinants of joint-stock company bankruptcy in particular industries. Empirical studies were conducted on 180 joint-stock companies in the Polish capital market. They represent three industries of the economy that is construction, manufacturing and trade. The calculations were performed using the bootstrapping method and the multivariate discriminant analysis.

Keywords: *companies bankruptcy, predicting, discriminant analysis, industry specifics*

JEL Classification: C530, G330

1 Introduction

Apart from, among others, using advanced statistical tools and finding new prognostic variables, the research on bankruptcy prediction also focuses on the development of models including industry characteristics of companies. E. I. Altman, a world-renowned authority on bankruptcy prediction, agrees with this approach. He believes that bankruptcy estimation models for companies should be based on financial data of companies pursuing homogeneous business activity (Altman, 1983). Due to difficulties in collecting a research sample that is large enough, Polish researchers rarely try to build industry-based models. There are only two examples of authors who compared classification adequacy of industry models and a non-industry dependent model (the "general" model) (Hołda, 2006; Juszczyk and Balina, 2014).

According to the author, due to the used split sample/holdout method, method of estimating the prediction error of the models, none of these results can unequivocally and

[1] Corresponding author: Poznań University of Economics and Business, Department of Econometrics, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: sergiusz.herman@ue.poznan.pl.

certainly answer the question whether industry models have better predictive ability compared to “general” models.

The article aims to estimate and compare predictive ability of industry and general models by using the bootstrapping method. The study enumerates financial ratios which may be deemed as determinants of bankruptcy of joint-stock companies for constructed industry models. It was verified whether determinants differ depending on the industry.

2 Research methodology

To conduct the study, the author gathered a research sample consisting of two groups: bankrupt companies in poor financial condition and healthy or non-bankrupt companies in good financial condition. Bankrupt companies were defined as the ones that were declared bankrupt by an applicable court. The sample was selected basing on information included in the Internetowy Monitor Sądowy i Gospodarczy. Financial data of the following companies was collected:

- 30 joint-stock companies from the construction industry (according to the Polish Classification of Activities – PKD 41.10-43.99z),
- 30 joint-stock companies from the manufacturing industry (PKD 10.11-33.20z),
- 30 joint-stock companies from the wholesale and retail trade (PKD 46.11-47.99z).

Each of these companies was paired with a joint-stock company in good financial condition. The companies were paired basing on: industry, the main scope of activity and asset size. Financial data of bankrupt companies was found in financial statements from a year before filing a bankruptcy petition – years 2000-2013. Financial statements of healthy companies concerned the same periods. Data was found in databases of Notoria Serwis and Bisnode Dun & Bradstreet as well as Monitor Polski B.

In empirical studies, 19 financial ratios, concerning profitability, liquidity, equity and asset structure as well as operating performance, were used (table 1). They were chosen based on literature review – these ratios are most frequently used in bankruptcy estimation models. The choice was also based on the availability of data in companies` financial statements.

The linear discriminant analysis was used to develop bankruptcy prediction models. Altman (1968) was the first person to use it for that purpose. Despite a dynamic development of statistical models, it is still vastly popular among business advisors.

Symbol	Description	Symbol	Description
ROA	Return on Assets	WPP	Cash ratio
ROE	Return on Equity	ZO	Total Debts to Assets
ZB	Gross Profitability	ZD	Long-Term Debt to Assets
ZS	Gross Profit Margin	KW	Equity to Assets Ratio
MZ	EBT to Sales Revenue	KWZ	Equity to Debt Ratio
MZ2	Net Profit Margin	RN	Accounts Receivable Turnover
MZO	Operating Margin	RZ	Inventory Turnover
KP	Working Capital to Total Assets ratio	RZob	Payables Turnover
WBP	Current liquidity ratio	Rakt	Total Assets Turnover
WSP	Quick liquidity ratio		

Table 1. Financial indicators underlying the research.

The quality of a developed classifier is defined by its ability to predict adherence of objects to defined populations. Such quality may be measured by a classifier true error rate. The world literature mentions empirical studies which aimed to compare various methods of estimating a prediction error (Wehberg and Schumacher, 2004; Braga-Neto and Dougherty, 2004; Molinaro et al., 2005; Kim, 2009). These studies concern the most popular classifiers in medicine. All of those studies bring about one common conclusion that prediction error estimators developed through the holdout method (split sample) are the most volatile of all analyzed estimators. Authors emphasize that the holdout method can be used only for large sets of data which makes it possible to define large independent training and test groups (Ripley, 1996). It is very difficult to meet this condition in the case of studies concerning the Polish capital market.

According to author`s previous studies, when predicting bankruptcy of Polish join-stock companies, prediction error estimators developed through the bootstrapping method have had the most valuable features (Herman, 2016). These methods are based on generation of B samples of $x^{*1}, x^{*2}, x^{*3},..., x^{*B}$ bootstrap type in such a way that each of them is developed by n-fold simple random sampling with replacement from the available set $\{x_1, x_2, x_3,..., x_n\}$. Next, these samples are used as training samples. Objects which were not drawn in next iterations constitute the test sample. Based on the samples, a true prediction error is estimated. It is often defined as a prediction error of a model developed basing on the entire n set and tested on a large and independent sample. In the study, the +632 estimator, proposed in 1997 by Efron and Tibshirani (1997), was used.

3 Analysis of industry differences of financial ratios

Assumptions for constructing a discriminant function concern normal distribution and equal variance in studied groups. Firstly, the hypothesis on population observation of normal distribution was verified with the use of the Kolmogorov-Smirnov test. It turned out that only three ratios (KP, ZO and KW) in the case of healthy companies and one ratio (WBP) for bankrupts have a normal distribution. Next, with the use of Levene`s test, it was assessed whether the variance of ratios was the same in both tested populations. Results of only three efficiency ratios (RN, RZ and Rakt) do not constitute the basis to reject the null hypothesis stating that populations have equal variances. According to the literature, results achieved with the use of this method are not deteriorated when assumptions concerning the linear discriminant analysis are not met (Hand, 1981; Hadasik, 1998). Therefore, despite not meeting the assumptions, the linear discriminant analysis was used in the study.

The Mann-Whitney nonparametric U test was used to verify whether values of ratios for healthy and bankrupt companies in particular industries come from the same distribution. Only in the case of four ratios in the construction industry (ZD, RN, RZ, Rakt), four ratios in the manufacturing industry (ROE, ZD, RN, RZ) and three ratios in trade (ZD, RN, RZ) with a statistical significance of 5%, the distribution of values does not differ in the studied populations. The value of statistics of manufacturing sector is much higher than other researched.

The last step of the initial analysis of financial data used later in the study was to compare distribution of values of company financial ratios between particular industries. For this purpose, the Mann-Whitney U test was used. Healthy and bankrupt companies were studied separately. In the case of healthy companies (5% of statistical significance), distribution of their values differ between industries for 9 financial ratios (WSP, WPP, ZO, KW, KWZ, RN, RZ, RZob, Rakt). When analyzing bankrupt companies, distribution of values of their financial ratios differ more frequently. Only in the case of three financial ratios (ROE, ZO and ZD, with 5% of statistical significance), distribution of values do not differ between industries. The same conclusions were reached when verifying the hypothesis on equal medians of studied financial ratios. A test based on chi-squared distribution was used for that purpose. Differences may be an indicator of a need to construct other models for predicting bankruptcy of joint-stock companies for particular industries.

4 The study of predictive ability of industry models

For constructing industry classifiers and assessing their predictability, the following assumptions were formulated:

a) models are constructed separately for 60 companies for each analyzed industry,
b) models are constructed basing on linear discriminant function and financial indicators presented in table 1,
c) before the learning process, variables strongly correlated with others (correlation coefficient higher than 0.90) are removed,
d) prediction error is assessed with the use of the bootstrapping method, the number of bootstrap samples B=50,
e) method of choosing model variables – the choice of 5 variables with the highest absolute value of the t-statistic for the test comparing an average value of financial ratios in the studied populations.

The true prediction error of industry-based classifiers was assessed with the use of the above assumptions. The results are presented in table 2.

	Construction	**Manufacturing**	**Trade**
The number of firms classified	1064	1064	1064
The number of firms correctly classified	821	950	749
non-bankrupt	395	503	394
bankrupt	426	447	355
The number of firms incorrectly classified	243	114	315
non-bankrupt	137	29	138
bankrupt	106	85	177
Prediction error	21.86%	10.45%	27.67 %

Table 2. Evaluation of classification effectiveness of industry models.

According to the analysis of data included in the table, a classifier constructed for the manufacturing industry has the lowest prediction error. Compared to the construction industry and trade, it is lower by more than 10 and 17 percentage points respectively. Prediction errors for all three industries were estimated based on the test sample which included 1064 objects. It is a result of the assumption on the bootstrap sampling and makes it possible to come up with an average prediction error for analyzed industries. The average prediction error for industry-based models is about 20.00%.

The study aims to compare the prediction ability of industry models and the general model which was constructed basing on all 180 joint-stock companies. Therefore, its prediction error needs to be estimated. In order to do that, the same b)-e) assumptions were used. The prediction error was estimated with the use of two different methods:

1) I method – all 180 companies were used at the same time,
2) II method – 100 subsamples with 60 objects were drawn from the population of 180 companies.

The results of the first method of assessing the model predictability are presented in table 3.

General model	
The number of firms classified	3278
The number of firms correctly classified	2623
non-bankrupt	1337
bankrupt	1286
The number of firms incorrectly classified	655
non-bankrupt	302
bankrupt	353
Prediction error	19.79%

Table 3. Evaluation of classification effectiveness of general model.

The estimated prediction error is lower than the average prediction error for industry models (which is 20.00%). In addition, a detailed analysis of the results of the classification of individual objects showed that the general model classifies accurately the companies from researched sectors, compared to the industry models.

The figure no. 1 shows the prediction error of general models for 100 subsequent subsamples with 60 companies (II method of assessing the prediction error) compared with the average prediction error of industry-based models.

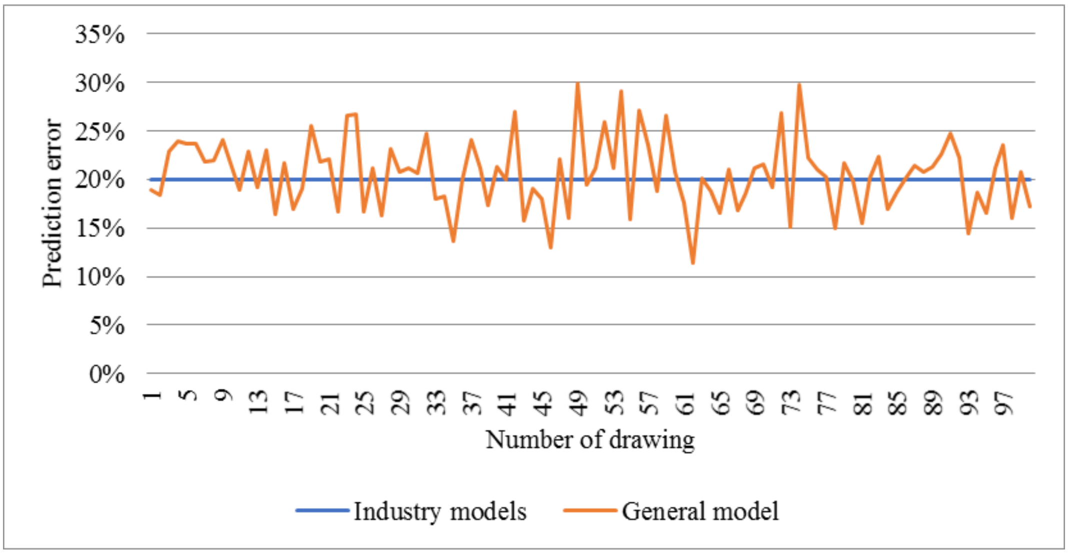


Fig. 1. Prediction error of general model for 100 another sub-samples against average prediction error of industry models.

With the use of the one sample *t*-test, the author verified the hypothesis that an average prediction error for 100 subsamples is the same as the error assessed for industry models (20.00%). The assumptions have been fulfilled in this test. The test statistic value of 1.546 indicated that, with the significance level of 5%, there is no basis to reject the null hypothesis.

The results indicate that industry-based models do not have a lower average prediction error which means that they do not have a higher predictive ability compared to the general model. If the financial ratios of companies have a desirable statistical properties - high discriminative power, as in the case for manufacturing sector, both general and industry models predicts bankruptcy correctly.

5 Determinants of joint-stock company bankruptcy in chosen industries

The previous part included a comparison of predictability of industry models and the general model which does not take into account industry characteristics. According to the approach of Efron (1983), the true prediction error of industry classifiers (models) constructed on the available sample was estimated. The form of classifiers is presented here.

The first model was constructed basing on the sample including construction companies. Two versions of the equation – non-standardized (W) and standardized (SW) – are presented below.

$$D(W) = -0{,}35 \cdot WBP + 2{,}40 \cdot MZO + 3{,}00 \cdot KP + 0{,}40 \cdot KWZ + 0{,}84 \cdot KW - 0{,}42 \qquad (1)$$

$$D(SW) = -0{,}30 \cdot WBP + 0{,}29 \cdot MZO + 0{,}66 \cdot KP + 0{,}46 \cdot KWZ + 0{,}23 \cdot KW \qquad (2)$$

The model was constructed in such a way that positive values of the discriminant function are an indicator of good condition of a company and qualify it as a healthy company. Whereas, negative values suggest that a company has a poor financial situation and is classified as a company at risk of bankruptcy. Additionally, the higher the positive values of financial ratios, the lower risk of bankruptcy. So, the increase of profitability (MZO ratio) and liquidity (KP ratio) decreases the risk of bankruptcy of analyzed companies. The situation is similar for equity and asset structure ratios. The growth of equity share – the growth of KW and KWZ ratios – leads to the lower risk of bankruptcy. In the case of WBP ratio, a negative value may raise some doubts. However, according to the literature, liquidity ratios have a certain value range deemed as correct. Too large deviations, both up and down, have a negative effect on company condition.

The second model of joint-stock company bankruptcy prediction was estimated based on financial data of entities from the manufacturing industry. The relevant equations are presented below.

$$D(W) = 1{,}29 \cdot ZB + 4{,}17 \cdot ZS - 0{,}43 \cdot ZO + 0{,}70 \cdot MZO + 0{,}32 \cdot KWZ + 0{,}34 \qquad (3)$$

$$D(SW) = 0{,}22 \cdot ZB + 0{,}49 \cdot ZS - 0{,}19 \cdot ZO + 0{,}15 \cdot MZO + 0{,}46 \cdot KWZ \qquad (4)$$

Also in this case, the profitability growth (ZB, ZS and MZO ratios) and a higher equity share in the asset and equity structure (KWZ) decreases the probability of bankruptcy. The situation is different in the case of debt – the higher the debt ratio (ZO ratio), the higher the risk of bankruptcy.

The last model was constructed basing on the sample of companies from the wholesale and retail trade. The model equation can be found below.

$$D(W) = 1{,}38 \cdot ROA + 2{,}41 \cdot MZ2 + 0{,}36 \cdot WBP - 0{,}11 \cdot WSP + 0{,}58 \cdot KWZ - 0{,}61 \qquad (5)$$

$$D(SW) = 0{,}26 \cdot ROA + 0{,}38 \cdot MZ2 + 0{,}31 \cdot WBP - 0{,}08 \cdot WSP + 0{,}51 \cdot KWZ \qquad (6)$$

Basing on the dependencies, the growth of profitability, shown by ROA and MZ2 ratios, and of equity share in the asset and equity structure decreases the risk of bankruptcy. In the case of liquidity ratios, the growth of WBP ratio is perceived as positive; whereas, it is negative for the WSP ratio.

When comparing models estimated on the basis of company data from particular industries, it can be noted that they were constructed with the use of different set of financial ratios. The KWZ (debt-to-equity) ratio is the only common feature of all three models. The adopted method of choosing variables proved that different financial ratios are useful in predicting company bankruptcy for certain industries.

The form of industry models indicates that certain industries have different determinants of joint-stock company bankruptcy in Poland. It proves that bankruptcy prediction models should be industry-based, despite the fact that it does not improve their average predictability.

Conclusion

Results presented in the article indicate that compared to general models, industry-based models do not have a lower average prediction error, which means higher predictability. Another goal of this study was to define determinants of joint-stock company bankruptcy in particular industries. The form of industry models as well as variables used for the construction of models estimating true prediction errors were taken into account. As a result, determinants of bankruptcy of joint-stock companies in Poland differ in particular industries. Therefore, it is worth to construct industry-based bankruptcy prediction models, despite not always resulting in models with higher predictability.

The results might have been affected by the fact that analyzed joint-stock companies often conducted various business activities (various sections in the classification of business activities in Poland – PKD) even though they belonged to the same industry. In further studies, it is worth to conduct analysis based on companies conducting more uniform business operations but having different legal forms.

The study may be further analyzed by the use of different methodology. In further studies, other methods for predicting company bankruptcy could be used; for instance, soft computing methods that have been popular recently.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance, 23*(4), 589-609.

Altman, E. I. (1983). *Corporate financial distress: a complete guide to predicting, avoiding, and dealing with bankruptcy*. New York: Wiley.

Braga-Neto, U. M., & Dougherty, E. R. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics, 20*(3), 374-380.

Efron, B. (1983). Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association, 78*(382), 316-331.

Efron, B., & Tibshirani, R. (1997). Improvements on Cross-Validation: The .632 Bootstrap Method. *Journal of the American Statistical Association, 92*(438), 548-560

Hadasik, D. (1998). *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*. Poznań: Akademia Ekonomiczna w Poznaniu.

Hand, D. J. (1981). *Discrimination and classification*. Chichester: Wiley.

Herman, S. (2016). Analiza porównawcza wybranych metod szacowania błędu predykcji klasyfikatora. *Przegląd Statystyczny, 62*(4)*, 449-463.*

Hołda, A. (2006). *Zasada kontynuacji działalności i prognozowanie upadłości w polskich realiach gospodarczych*. Kraków: Wydawnictwo Akademii Ekonomicznej.

Juszczyk, S., Balina, R. (2014). Prognozowanie zagrożenia bankructwem przedsiębiorstw w wybranych branżach, *Ekonomista*, *1*, 67-95.

Kim, J. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis, 53*(11), 3735-3745.

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics, 21*(15), 3301-3307.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Wehberg, S., & Schumacher, M. (2004). A Comparison of Nonparametric Error Rate Estimation Methods in Classification Problems. *Biometrical Journal, 46*(1), 35-47.

Forecasting intraday traded volume with the Weibull ACV model: an application to Polish stocks

Roman Huptas[1]

Abstract

The trading volume is one of the basic measures of intensity of trading activity and plays a crucial role in the financial market microstructure literature. The aim of the paper is to examine the out-of-sample point and density forecasting performance of the Bayesian linear Autoregressive Conditional Volume (ACV) model with Weibull distribution for the error term for intra-daily volume data. The analysis follows a rolling window scheme and is based on real-time 5-minute intra-daily traded volume data for stocks quoted on the Warsaw Stock Exchange, which is a leading stock market in Central and Eastern Europe. It is concluded that in terms of point forecasts the considered Bayesian linear ACV model significantly outperforms such benchmarks as the naïve or the Rolling Means methods. Moreover, the exponential error ACV models generate more accurate point forecasts than the structures with the Weibull distribution, but the differences between forecast errors are not in all cases statistically significant. The results obtained from analysis of density forecasts indicate that in most cases the linear ACV model with the Weibull distribution provides significantly better density forecasts as compared to the linear ACV model with exponential innovations in terms of the log-predictive score.

Keywords: *trading volume, forecasting, ACV model, market microstructure, Bayesian inference*
JEL Classification: C11, C22, C53

1 Introduction

The trading volume is one of the key characteristics of liquidity on stock markets and plays a very important role in the literature on financial market microstructure. It can be very important in order to understand stock trading and behaviour of market participants.

In existing literature on the subject a lot of attention is paid to the issue of examining the dependencies between trade size and other financial variables, such as price (Easley and O'Hara, 1987; Foster and Viswanathan, 1990 for example) or volatility (Tauchen and Pitts, 1983; Karpoff and Boyd, 1987; Andersen, 1996; Manganelli, 2005 for example). In turn, forecasting of trading volume in stock markets definitely has not been a central point of financial econometrics for years. Only few works deal exclusively with the modelling and predicting volume on stock exchanges. Kaastra and Boyd (1995) designed neural networks to forecast monthly futures trading volume. Lo and Wang (2000) applied the principal component analysis to a decomposition of the volume. Białkowski et al. (2008) presented

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: huptasr@uek.krakow.pl.

a new methodology for modelling the dynamics of intraday volume, which allows for a significant reduction of the execution risk in Volume Weighted Average Price (VWAP) orders. Brownlees et al. (2011) used a Component Multiplicative Error Model for intra-daily volumes, in which the conditional expectation of volume is the product of three multiplicative elements: a daily component, an intra-daily periodic component and an intra-daily dynamic non-periodic component. Predicting intraday trading volume is also a topic of interest of Satish et al. (2014). Finally, Ito (2016) developed and applied the Spline-DCS model to forecasting the high-frequency traded volume of selected equity and foreign currency exchange pairs.

The aim of this paper is to present results of a pilot study in which the quality of the out-of-sample point and density forecasts of intra-daily volume data generated by the linear Autoregressive Conditional Volume (ACV) model with the Weibull distribution for the error term is examined. Forecasting models are developed on the basis of the 5-minute volume data of some representative, widely traded stocks quoted on the Warsaw Stock Exchange. Considered models follow a rolling window forecast scheme. The evaluation of point forecast accuracy is performed by comparison forecast errors of the considered linear ACV models and the benchmark models (the random walk without drift and the Rolling Means technique). To compare point forecasts generated by all forecasting models analysed in the study we use standard forecast accuracy measures: the root mean square forecast error (RMSFE) and the mean absolute forecast error (MAFE), and Diebold and Mariano (1995) test. The comparison of density forecasts is obtained by using the log-predictive scores (LPSs) and the Amisano and Giacomini (2007) test.

The rest of the paper is scheduled as follows. In Section 2, we briefly present in more details the Autoregressive Conditional Volume model and the Bayesian estimation for the models under consideration, along with relevant MCMC methods is also considered. Then, Section 3 describes the forecasting performance. The results of the empirical study are discussed in Section 4. Finally, the last section presents the conclusions of the paper.

2 Research methodology

In the presented analysis the Autoregressive Conditional Volume (ACV) model suggested by Manganelli (2005) is a basic econometric tool. The ACV model is definitely a volume counterpart of the Autoregressive Conditional Duration (ACD) model, introduced in the seminal paper of Engle and Russell (1998) and applied to model the dynamics of the time intervals between successive events of the transaction process.

The general ACV model for the volume v_i , $i = 0, 1, 2,...,N$ (with N standing for the total number of observations) is defined multiplicatively as:

$$v_i = \Phi_i \cdot \varepsilon_i , \tag{1}$$

$$\Phi_i = E(v_i \mid \Im_{i-1}, \theta) = \Phi_i (v_{i-1}, \ldots, v_1; \theta) \tag{2}$$

where Φ_i represents the conditional mean value of trading volume, $\Im_{i-1}$ is the information set available at time t_{i-1}, θ is the vector of unknown parameters, and ε_i is a sequence of positive, identically and independently distributed random variables with density function $f_\varepsilon(\varepsilon_i)$ and mean value $E(\varepsilon_i) = 1$. Equations (1)-(2) formulate a very general setup that allows for a variety of specific models. In this study the following particular, linear ACV model, which remains in line with Manganelli (2005) is used:

$$v_i = \Phi_i \cdot \varepsilon_i , \tag{3}$$

$$\Phi_i = \omega + \alpha \cdot v_{i-1} + \beta \cdot \Phi_{i-1} , \tag{4}$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$. These inequality restrictions are imposed in order to ensure positive conditional volumes for all possible realizations of random variables v_i, existence of the unconditional mean of trading volume and stationarity of the model. Moreover, we assume that ε_i follows either the exponential, or the Weibull distribution, yielding ACV specifications denoted as Exp-ACV and W-ACV. The corresponding density functions (under the assumption that the error terms have unit expectation) are given by:

1. $\varepsilon_i \sim Exp(1)$:

$$f_\varepsilon(\varepsilon_i) = \exp(-\varepsilon_i),\ \varepsilon_i > 0 , \tag{5}$$

2. $\varepsilon_i \sim W(\lambda, \gamma)$ with parameter $\lambda = 1/\Gamma(1 + 1/\gamma)$:

$$f_\varepsilon(\varepsilon_i) = \frac{\gamma}{\lambda} \cdot \left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma-1} \cdot \exp\left[-\left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma}\right], \varepsilon_i > 0 , \gamma > 0 . \tag{6}$$

In order to estimate parameters of the considered ACV models, the Bayesian approach is applied. It is worth noting that the literature on the Bayesian inference in modelling and forecasting financial high-frequency data with the use of the ACD-type models is still limited (e.g. Brownlees and Vannucci, 2013; Huptas, 2014, 2016; Gerlach et al., 2016). Bayesian estimation of the ACV models outlined above requires certain prior assumptions. We assume that all parameters – whenever possible – are *a priori* independent. Moreover, in order to express the lack of prior knowledge, fairly diffuse prior distributions are assumed, so that the data dominates the inference about the parameters through the likelihood function.

Specifically, for all parameters of Equation (4) we propose the normal distributions with zero mean and standard deviation of five, adequately truncated, due to relevant restrictions imposed on the parameters in each model. For the ACV model with the Weibull innovations, the prior density for parameter γ is also specified as density of the normal distribution with zero mean and standard deviation of five adequately truncated.

The inference was conducted using MCMC techniques. The Metropolis-Hastings (MH) algorithm (Gamerman and Lopes, 2006) with a multivariate Student's *t* candidate generating distribution with three degrees of freedom and the expected value equal to the previous state of the Markov chain was used to generate a pseudo-random sample from the posterior distribution. The covariance matrix was obtained from initial cycles, which were performed to calibrate the sampling mechanism. Convergence of chain was carefully examined by starting the MCMC scheme from different initial points and checking trace plots of iterates for convergence to the same posterior. Acceptance rates were sufficiently high and always exceed 50%, indicating good mixing properties of the posterior sampler. The final results and conclusions were based on 100,000 draws, preceded by 50,000 burn-in cycles. All codes were implemented by the author and ran using the GAUSS software, version 13.0.

3 Forecasting procedure

Forecasts are determined using a rolling window prediction scheme with a fixed-length window of 4816 observations. First, we estimate ACV models and then generate one- and multi-step-ahead point and density forecasts for horizons h=2, 3, 5, 10 in each model. Then the estimation window is moved by one 5-minute bin in relation to the earlier one, the models are re-estimated and future volumes are predicted. The forecasting procedure is repeated 200 times. This means that the out-of-sample period used to evaluate forecast performance contains 200 observations.

The point forecasts are calculated as arithmetic means of draws from the predictive distributions. For assessing the quality of the point predictions, the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) are used. Further, the pair-wise Diebold and Mariano (1995) tests of equal point forecast accuracy are carried out. In this paper, to benchmark our Bayesian ACV models in terms of the one-step-ahead point forecasts we use a naïve strategy based on a random walk model and a Rolling Means (RM) method.

This research focuses also on density forecasts for traded volumes. The density forecast is a predictive density function evaluated at realization of intra-daily volume. In turn, the Bayesian density forecast is obtained as the arithmetic mean of the density forecasts corresponding to draws from the posterior distribution. Taking the logarithm of density forecast, we obtain the so-called log predictive score (LPS). To compare the performance of two different density forecasts, we calculate the difference between the corresponding LPSs and we follow the Amisano and Giacomini (2007) test.

4 Data and empirical results

The empirical analysis and forecasting study are based on 5-minute intra-daily volume data of three actively traded companies listed in the WIG20 Index of the Warsaw Stock Exchange (WSE), namely: the Polish Telecommunications (TPSA, currently Orange Polska S.A.), the PKOBP SA bank (PKOBP) and the KGHM SA company (KGHM). The data set comprises intra-daily observations spanning 60 consecutive trading days from March 23, 2009 to June 18, 2009. It must be stressed that each 5-minute bin volume was computed as the sum of all transaction volumes exchanged within the time interval covered by the bin. Additionally, the analysis covers only transactions carried out in the continuous trading phase which in the case of the WSE in 2009 fall on between 10:00 and 16:10. Taking into account that the analysis is conducted in a rolling window scheme, the first subsample starts on the first 5-minute bin of March 23, 2009 and ends on the 86th 5-minute bin of June 12, 2009. The second estimation window is moved towards the first one by one bin, so the second subsample starts on the second 5-minute bin of March 23, 2009 and ends on the first 5-minute bin of June 15, 2009. The last 200th subsample starts on the 28th 5-minute bin of March 25, 2009 and ends on the 27th 5-minute bin of June 17, 2009.

It is well known in the financial literature that trading volumes exhibit a well-pronounced intraday periodic pattern. In a way similar to Bauwens and Veredas (2004), the intraday periodic pattern for intra-daily volumes was estimated using the Nadaraya-Watson kernel estimator of regression of the trading volume on the time of the day. Following Engle and Russell (1998) the time-of-day adjusted volumes were computed by dividing plain intra-daily volumes by estimated periodic component.

Table 1 presents the values of the MAFEs for all models for TPSA, PKOBP and KGHM. It also contains the results of the one-sided Diebold-Mariano tests. Firstly, the point forecasts from the Exp-ACV models lead to the lowest MAFEs for all stocks and for all horizons. In the case of the one-step-ahead forecasts the ACV models always significantly outperform the naïve and the Rolling Means benchmarks. Moreover, the forecasts from the Exp-ACV models are significantly more accurate than the point forecasts from the W-ACV models for almost all the horizons and assets. However, it emerges that the point forecasts in both ACV specifications are not significantly different from each other in the case of KGHM for horizons h=1, 5, 10 and in the case of PKOBP for horizon h=5.

In the following part of the article, there are discussed results of the RMSFEs. These forecast accuracy measures (in levels) for all models for TPSA, KGHM and PKOBP are presented in Table 2. Furthermore, we also report the results of the Diebold and Mariano tests. An analysis of values of these measures lead to the conclusion that, similarly to the MAFEs, the smallest RMSFE errors are obtained for the Exp-ACV models for all stocks and for almost all horizons with the exception of h=5 for PKOBP. In the case of the one-step-ahead point forecasts

Model	**h=1**	**h=2**	**h=3**	**h=5**	**h=10**
TPSA					
Naïve	76009.7 *	-	-	-	-
RM	78353.5 ***	-	-	-	-
Exp-ACV	**60749.6**	**61078.8**	**58855.4**	**58738.1**	**57484.2**
W-ACV	61781.5 **	62103.1 **	59791.4 **	59878.8 ***	58521.5 ***
PKOBP					
Naïve	20259.3 **	-	-	-	-
RM	61573.3 ***	-	-	-	-
Exp-ACV	**18112.9**	**19112.9**	**19279.3**	**20082.0**	**21952.2**
W-ACV	18127.0 *	19128.3 *	19302.2 **	20088.6	21974.1 **
KGHM					
Naïve	14776.8 **	-	-	-	-
RM	19546.2 ***	-	-	-	-
Exp-ACV	**12367.1**	**12949.1**	**12668.7**	**12555.9**	**12634.4**
W-ACV	12379.6	12980.4 **	12700.8 **	12559.4	12643.6

Notes: The lowest MAFEs are shown in bold type. *, ** and *** indicate significance of the one-sided Diebold-Mariano test at the 0.1, 0.05 and 0.01 significance level, respectively. The Diebold-Mariano test of model with the lowest MAFE against a given model in a row.

Table 1. Mean absolute forecast errors for TPSA, PKOBP and KGHM.

the ACV models always significantly outperform the naïve and the Rolling Means benchmarks except for RM for TPSA. Moreover, it should be noted that, for TPSA and at all the horizons, the point forecasts from the Exp-ACV model are not significantly superior (in terms of RMSFEs) as compared with the W-ACV specification. In turn, in the case of PKOBP and KGHM forecasts generated by the Exp-ACV models are statistically more accurate than the W-ACV forecasts only at h=2, 3, 10 for PKOBP and at h=2, 3 for KGHM. For the remaining horizons the differences between root mean squared errors are statistically insignificant.

Table 3 shows the differences between the average values of the LPSs of both ACV models for all considered assets. It also presents the results of the Amisano and Giacomini tests. On the basis of the information included in Table 3, it can be noted that in the case of PKOBP and KGHM, the density forecasts from the W-ACV model are significantly superior

Model	h=1	h=2	h=3	h=5	h=10
TPSA					
Naïve	260661.6 *	-	-	-	-
RM	196736.0	-	-	-	-
Exp-ACV	**192449.6**	**190223.6**	**187175.1**	**184995.8**	**185900.5**
W-ACV	194331.8	191882.3	188354.9	185772.3	186280.8
PKOBP					
Naïve	28726.1 ***	-	-	-	-
RM	68847.2 ***	-	-	-	-
Exp-ACV	**23343.4**	**24146.4**	**24046.1**	25453.9	**26408.2**
W-ACV	23356.5	24163.6 *	24069.2 **	**25453.0**	26428.4 **
KGHM					
Naïve	24127.4 **	-	-	-	-
RM	23815.4 ***	-	-	-	-
Exp-ACV	**18517.9**	**19401.9**	**19271.1**	**19475.7**	**19890.9**
W-ACV	18539.9	19455.4 **	19322.1 *	19498.2	19898.6

Notes: The lowest RMSFEs are shown in bold type. *, ** and *** indicate significance of the one-sided Diebold-Mariano test at the 0.1, 0.05 and 0.01 significance level, respectively. The Diebold-Mariano test of model with the lowest RMSFE against a given model in a row.

Table 2. Root mean squared forecast errors for TPSA, PKOBP and KGHM.

Model	h=1	h=2	h=3	h=5	h=10
TPSA					
W-ACV vs. Exp-ACV	0.1773 *	0.0471	0.0185	0.0031	-0.0129
PKOBP					
W-ACV vs. Exp-ACV	0.0025 ***	0.0021 **	0.0023 **	-0.0049	0.0017 *
KGHM					
W-ACV vs. Exp-ACV	-0.0094 **	-0.0085 **	-0.0084 **	-0.0092 **	-0.0120 ***

Notes: Positive (negative) values indicate that the first model has a higher (lower) average LPS than the second model. *, ** and *** indicate significance of the Amisano-Giacomini test at the 0.1, 0.05 and 0.01 significance level, respectively.

Table 3. The differences between the average values of the LPSs and results of the Amisano and Giacomini test for TPSA, PKOBP and KGHM.

as compared with the Exp-ACV specification with the exception of horizon h=5 for PKOBP. Different situation can be observed in the case of TPSA. Density forecasts from the ACV model with the Weibull error distribution are qualitatively similar to density forecasts from the Exp-ACV model, that is differences between the values of the LPSs are statistically insignificant, except for h=1.

Conclusions

The objective of this study is to verify predictive performance (in terms of point and density forecasts) of the ACV model with the Weibull distribution for the error term for 5-minute intra-daily volume data. Our main findings can be summarised as follows.

It is concluded that in terms of point forecasts, MAFEs and RMSFEs the considered Bayesian linear ACV models significantly outperform such benchmarks as the naïve or the Rolling Means methods. Moreover, the exponential error ACV models generate more accurate point forecasts than the ACV specifications with the Weibull distribution, but the differences between forecast errors are not in all cases statistically significant. The comparison of the ACV models also allows us to conclude that in the case of more actively traded stocks (PKOBP and KGHM) the linear ACV model with the Weibull distribution provides significantly more accurate density forecasts than the linear ACV model with exponential innovations in terms of the log-predictive score.

The presented results are based on a pilot study taking into account only the ACV model with the Weibull distribution for innovations. In future research we will try to focus on alternative ACV model specifications with different error distributions such as the Burr or the generalized gamma distribution, longer forecast horizons and different sampling frequencies.

Acknowledgements

This work has been financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Andersen, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance, 51,* 169-204.

Amisano, G., & Giacomini, R. (2007). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics, 25(2),* 177-190.

Bauwens, L., & Veredas, D. (2004). The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics, 119,* 381-412.

Białkowski, J., Darolles, S., & Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking and Finance, 32,* 1709-1722.

Brownlees, Ch., Cipollini, F., & Gallo, G.M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics, 9,* 489-518.

Brownlees, Ch., & Vannucci, M. (2013). A Bayesian approach for capturing daily heterogeneity in intra-daily durations time series. *Studies in Nonlinear Dynamics & Econometrics, 17(1),* 21-46.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13(3)*, 253-263.

Easley, D., & O'Hara, M. (1987). Price, trade size and information in securities markets. *Journal of Financial Economics, 19,* 69-90.

Engle, R.F., & Russell, J.R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66,* 1127-1162.

Foster, D., & Viswanathan, S. (1990). A theory of intraday variations in volume, variance and trading costs in securities market. *Review of Financial Studies, 3,* 593-624.

Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press.

Gerlach, R., Peiris, S., & Lin, E.M.H. (2016). Bayesian estimation and inference for log-ACD models. *Computational Statistics, 31,* 25-48.

Huptas, R. (2014). Bayesian estimation and prediction for ACD models in the analysis of trade durations from the Polish stock market. *Central European Journal of Economic Modelling and Econometrics, 6,* 237-273.

Huptas, R. (2016). The UHF-GARCH-Type Model in the Analysis of Intraday Volatility and Price Durations – the Bayesian Approach. *Central European Journal of Economic Modelling and Econometrics, 8,* 1-20.

Ito, R. (2016). Spline-DCS for Forecasting Trade Volume in High-Frequency Financial Data. https://doi.org/10.17863/CAM.5751.

Kaastra, I., & Boyd, M. (1995). Forecasting futures trading volume using neural networks. *Journal of Futures Markets, 15,* 953-970.

Karpoff, J., & Boyd, M. (1987). The relationship between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis, 22,* 109-126.

Lo, A., & Wang, J. (2000). Trading volume: Definition, data analysis, and implication of portfolio theory. *Review of Financial Studies, 13,* 257-300.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets, 8,* 377-399.

Satish, V., Saxena, A., & Palmer, M. (2014). Predicting intraday trading volume and volume percentages. *The Journal of Trading, 9,* 15-25.

Tauchen, G., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica, 51,* 485-505.

Analysis of the properties of selected inequality measures based on quantiles with the application to the Polish income data

Alina Jędrzejczak[1], Dorota Pekasiewicz[2]

Abstract

Quantiles of income distributions are often applied to the estimation of various inequality and poverty characteristics. The most popular synthetic inequality measures, including the Gini and Pietra indices, are based on the Lorenz curve, but also simple quantile ratios or quantile dispersion ratios can be utilized to compare incomes of different population groups. Some other measures of income inequality have been constructed using differences (or ratios) between population and income quantiles. The concentration curve and corresponding synthetic concentration coefficient proposed by Zenga, are also defined in terms of quantiles. In the paper, selected inequality measures based on deciles and quintiles are considered. The main objective was to compare statistical properties of different estimation methods for quantiles, including Bernstein and Huang –Brill estimators, with the classical quantile estimator based on a relevant order statistic. Several Monte Carlo experiments have been conducted to assess biases and mean squared errors of quantile estimators for different sample sizes under the lognormal or Dagum distributions assumed as a population model. The results of the experiments have been used to the estimation of inequality measures in Poland.

Keywords: *income, inequality, estimator, bias, mean squared error*

JEL Classification: C13, C15

1 Introduction

The early development of statistical tools for income inequality measurement dates back to the end of the XIXth century, when the works of Vilfredo Pareto came out, but the literature on the evaluation and measurement of economic inequality has remarkably grown over the last few decades. For a long time income inequality measures have been used mainly for descriptive purposes. Rapid development of sample surveys after the second world war, as well as the growing demand for high-quality estimates at low levels of aggregation, made it necessary to study the sampling properties of inequality measures. Nevertheless, in many applications the estimates of inequality measures are still presented without any information about their precision, which must be the basis for further statistical inference e.g. statistical hypothesis testing and interval estimation. The problem can be neglected to some extent when

[1] Corresponding author: University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 90-214 Łódź, 41/43 Rewolucji 1905, e-mail: jedrzej@uni.lodz.pl.

[2] University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 90-214 Łódź, 41/43 Rewolucji 1905, e-mail: pekasiewicz@uni.lodz.pl.

we consider the overall population or the sample size is large enough to apply the asymptotic theory; one should be conscious however, that for heavy-tailed income distributions the sufficient sample size can be very large indeed. For some population divisions (by age, occupation, family type or geographical area) estimators of inequality measures can be seriously biased and their standard errors can be far beyond the values that can be accepted by social policy-makers for making reliable policy decisions (Jędrzejczak, 2012). The situation seems even more complicated for nonlinear sample statistics including numerous inequality indices based on quantiles.

The paper addresses the problem of statistical properties of the estimators of popular inequality measures based on quantiles. After a brief description of such measures (section 2), selected quantile estimators have been introduced (section 3). The next section comprises the results of Monte Carlo experiments which have been conducted to assess biases and mean squared errors of quantile estimators and their functions. In the last part of the paper we present the application of quantile-based inequality indices to the Polish Household Budget Survey (HBS) data.

2 Statistical inequality measures based on quantiles

Distribution quantiles of a random variable *X* which is identified as a household or personal income, or the estimators of these quantiles, have been applied to the construction of simple inequality indices as quintile dispersion ratio and decile dispersion ratio (Panek, 2011).

The quintile dispersion ratio has the following form:

$$\mathrm{W}_{20:20}^{(1)} = \frac{Q_{0.8}}{Q_{0.2}}, \tag{1}$$

where $Q_{0.8}$, $Q_{0.2}$ are quintiles, respectively, the fourth and the first.

The quintile dispersion ratio can also be defined as the ratio of the sum of incomes of the richest 20 percent of the population to the sum of incomes of the poorest 20 percent:

$$\mathrm{W}_{20:20}^{(2)} = \frac{\sum_{i \in GK_5} x_i}{\sum_{i \in GK_1} x_i}, \tag{2}$$

where GK_j is *j*-th quintile group.

The measure (2) can be interpreted as the ratio of the average income of the richest 20 percent of the population to the average income of the poorest 20 percent of the population and it is usually calculated on the basis of equivalised income.

Similar ratios can also be calculated for other quantiles, for instance deciles or percentiles (95th and 5th) of income distributions. Using the first and ninth decile we can obtain the following decile dispersion ratio:

$$\mathrm{W}_{10:10}^{(1)} = \frac{Q_{0.9}}{Q_{0.1}}, \tag{3}$$

where $Q_{0.9}$, $Q_{0.1}$ are deciles, respectively, the ninth and the first

and

$$\mathrm{W}_{10:10}^{(2)} = \frac{\sum\limits_{i \in GD_{10}} x_i}{\sum\limits_{i \in GD_1} x_i}, \tag{4}$$

where GD_j is j-th decile group.

The reciprocal of the decile dispersion ratio defined by (4) takes values from the interval [0,1] and is called the dispersion index for the end portions of the distribution:

$$\mathrm{K}_{1:10} = \frac{\sum\limits_{i \in GD_1} x_i}{\sum\limits_{i \in GD_{10}} x_i} = \frac{1}{\mathrm{W}_{10:10}^{(2)}}. \tag{5}$$

If the index $\mathrm{K}_{1:10}$ is closer to the 1, the inequality is lower (mean incomes in the extremal decile groups are the same).

The examples of more sophisticated inequality measures are Gini and Zenga indices. The popular Gini index is not considered in this paper. The synthetic Zenga index is based on the concentration curve that can be considered point concentration measure, and thus becomes sensitive to changes at every "point" of income distribution. The Zenga point measure of inequality is based on the relation between income and population quantiles (Zenga, 1990; Greselin et al. 2012):

$$\mathrm{Z}_p = \frac{x_p^* - x_p}{x_p^*} = 1 - \frac{x_p}{x_p^*}, \tag{6}$$

where $x_p = F^{-1}(p)$ denotes the population p-quantile and $x_p^* = Q^{-1}(p)$ is the corresponding income quantile. Therefore the Zenga approach consists of comparing the abscissas at which $F(x)$ and $Q(x)$ take the same value p.

Zenga synthetic inequality index is defined as simple arithmetic mean of point concentration measures $\mathrm{Z}_p, p \in \langle 0,1 \rangle$.

3 Selected quantile estimators

Let X be a continuous random variable with distribution function F and let $Q_p = F^{-1}(p)$ be the p-quantile of the random variable X, where $p \in (0,1)$. If F is continuous and strictly increasing distribution function, the p^{th} quantile always exists and is uniquely determined.

The well-known estimator of the quantile Q_p is the statistic:

$$\hat{Q}_p = F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}, \tag{7}$$

where $F_n(x)$ is empirical distribution obtaining on the basis of a n-element random sample $X_1, X_2, ..., X_n$.

The problem of quantile estimation has a very long history. In the subject literature numerous nonparametric (distribution-free) quantile estimators have been presented. Their particular expressions depend on the underlying empirical distribution function definition.

Classical quantile estimator obtained for the distribution $F_n(x) = \frac{card\{1 \leq j \leq n : x_i \leq x\}}{n}$ for $x \in R$ is defined by the following formula:

$$\hat{Q}_p = \begin{cases} X_{(np)}^{(n)}, & \text{for } np \in N, \\ X_{([np]+1)}^{(n)}, & \text{for } np \notin N, \end{cases} \tag{8}$$

where $X_{(k)}^{(n)}$ is an order statistic of rank k.

Among other estimators of quantiles Q_p, we can mention the standard estimator, Huang-Brill estimator, Harrel-Davis estimator and Bernstein estimator, to name only a few (Huang and Brill, 1999; Harrell and Davis, 1982; Zieliński, 2006).

By means of the empirical distribution *level crossing*, which has the following form:

$$F_n(x) = \sum_{i=1}^{n} w_{n,i} I_{(-\infty,x\rangle}\left(x_{(i)}^{(n)}\right), \tag{9}$$

where

$$w_{n,i} = \begin{cases} \frac{1}{2}\left[1 - \frac{n-2}{\sqrt{n(n-1)}}\right] & \text{for } i = 1, n, \\ \frac{1}{\sqrt{n(n-1)}} & \text{for } i = 2, 3, ..., n-1, \end{cases}$$

we obtain the Huang-Brill estimator of the p^{th} quantile Q_p:

$$\hat{Q}_p^{HB} = X_{(b)}^{(n)}, \tag{10}$$

where

$$b = \left[\sqrt{n(n-1)} \left(p - \frac{1}{2} \left[1 - \frac{n-2}{\sqrt{n(n-1)}} \right] \right) \right] + 2. \tag{11}$$

It can easily be noticed that for $p = 0.5$ the estimator of the quantile $Q_{0.5}$ is the order statistic $X^{(n)}_{\left(\left[\frac{n}{2}\right]+1\right)}$.

Another interesting quantile estimator is the Bernstein estimator given by:

$$\hat{Q}_p^{Brs} = \sum_{i=1}^{n} \left[\binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \right] X^{(n)}_{(i)}. \tag{12}$$

More examples of quantile estimators can be found in the papers of Pekasiewicz (2015) and Zieliński (2006).

4 Analysis of Monte Carlo experiments

The main objective of the Monto Carlo experiments conducted in the study was to assess the properties of selected estimators of quantiles. We were especially interested in their biases and sampling variances, the components of their sampling errors. The following estimators have been taken into consideration: the classical quantile estimator (8), Huang-Brill estimator (10) and Bernstein estimator (12). The estimators presenting the best performance were further applied to evaluate the quantile-based inequality measures for income distributions in Poland.

In the experiments two different probability distributions were utilized as population models: two-parameter lognormal distribution, $LG(\mu,\sigma)$, defined by the following density function:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), x > 0 \tag{13}$$

and three-parameter Dagum distribution $D(p,a,b)$, known also as the Burr type-III distribution, with the density function of the form (Kleiber and Kotz, 2003):

$$f(x) = ab^{-ap} p x^{ap-1} \left(1 + \left(\frac{x}{b} \right)^a \right)^{-p-1}, x > 0. \tag{14}$$

The sets of parameters of both theoretical distributions were established on the basis of real income data coming from Polish HBS and administrative registers, comprising large variety of subpopulations differing in the level of income inequality, which have been observed over the last two decades. The sample sizes were fixed for each variant as n=500; n=1000, n=2000. The number of repetitions of Monte Carlo experiment was N=20 000 (Białek, 2013). The simulated sample spaces were used to assess, for each estimator, its empirical bias and standard error.

Distribution	p	$\hat{Q}_p$		$\hat{Q}_p^{HB}$		$\hat{Q}_p^{Brs}$	
		BIAS	**RMSE**	**BIAS**	**RMSE**	**BIAS**	**RMSE**
$LG(8.0, 0.6)$	0.1	-0.087	3.240	0.254	3.248	0.132	3.165
	0.2	-0.079	2.718	0.139	2.726	0.108	2.669
	0.3	-0.039	2.504	0.133	2.511	0.095	2.481
	0.7	0.089	2.528	-0.082	2.521	0.042	2.469
	0.8	-0.077	2.712	-0.077	2.712	0.047	2.680
	0.9	-0.131	3.245	-0.131	3.245	0.041	3.169
$LG(8.3, 0.8)$	0.1	-0.097	4.350	0.359	4.373	0.302	4.220
	0.2	-0.088	3.581	0.195	3.592	0.177	3.571
	0.3	-0.057	3.336	0.176	3.346	0.134	3.271
	0.7	0.169	3.338	-0.061	3.324	0.108	3.280
	0.8	-0.099	3.620	-0.099	3.620	0.070	3.510
	0.9	-0.116	4.339	-0.116	4.339	0.089	4.208
$D(0.7,3.6,3800)$	0.1	-0.182	3.923	0.313	3.916	0.086	3.803
	0.2	-0.068	2.800	0.141	2.776	0.069	2.741
	0.3	-0.105	2.349	0.114	2.346	0.000	2.303
	0.7	0.010	2.054	-0.080	2.049	0.043	2.013
	0.8	-0.085	2.298	-0.078	2.287	0.032	2.256
	0.9	-0.083	2.984	0.116	2.991	0.121	2.915
$D(0.7,2.8,3800)$	0.1	-0.156	5.073	0.368	5.069	0.221	4.493
	0.2	-0.112	3.580	0.232	3.589	0.082	3.509
	0.3	-0.080	3.015	0.144	2.991	0.062	2.958
	0.7	0.137	2.652	-0.063	2.681	0.073	2.599
	0.8	-0.084	2.956	-0.077	2.935	0.069	2.900
	0.9	-0.133	3.846	-0.112	3.848	0.147	3.774

Table 1. Properties of selected quantile estimators for sample sizes n=1000.

Table 1 presents the results of the calculations for three quantile estimators: classical, Huang-Brill, and Bernstein, each of the following orders: p=0.1; 0.2; 0.3; 0.7; 0.8; 0.9. In particular, the table shows the relative biases and relative root mean squared errors (in %) of these estimators obtained for predefined population models- lognormal and Dagum - differing across the experiments in the overall inequality levels. The similar experiments for Gini and Zenga ratios were reported in Jędrzejczak (2015).

Analysing the results of the calculations it becomes obvious that the Bernstein estimator performs better than its competitors- its root mean squared errors (RMSE) are much smaller than those observed for the other quantile estimators and its relative biases (BIAS) are also smaller, especially when the quantiles of higher orders are taken into regard. The bias and RMSE of Huang-Brill estimator are similar to the respective values for the classical quantile estimator. It is worth noting that for all cases biases are rather negligible so the total errors are dominated by sampling variances. In general, the estimation errors are higher for extremal quantile orders, for the heavy-tailed Dagum model and they also tend to increase as income inequality increases. The three types of quantile estimators mentioned above were then used to the simulation study concerning income inequality measures: $W^{(1)}_{10:10}$ and $W^{(1)}_{20:20}$ given by the formulas (1) and (3). The properties of decile dispersion ratios have been demonstrated in table 2. The results obtained for quintile dispersion ratios show similar regularities.

Distribution	$\mathbf{W^{(1)}_{10:10}}$ **(stand.)**		$\mathbf{W^{(1)}_{10:10}}$ **(Huang-Brill)**		$\mathbf{W^{(1)}_{10:10}}$ **(Bernstein)**	
	BIAS	**RMSE**	**BIAS**	**RMSE**	**BIAS**	**RMSE**
$LG(8.0, 0.6)$	0.065	4.327	-0.324	4.304	0.017	4.191
$LG(8.1, 0.7)$	0.084	5.088	-0.273	5.028	0.019	4.926
$LG(8.3, 0.8)$	0.124	5.815	-0.352	5.766	0.037	5.615
$D(0.7, 3.6, 3800)$	0.162	4.702	-0.211	4.651	0.082	4.543
$D(0.8, 3.0, 3200)$	0.097	5.179	-0.266	5.193	0.039	5.009
$D(0.7, 2.8, 3800)$	0.181	6.003	-0.298	5.948	0.066	5.800

Table 2. Properties of Decile Dispersion Ratio based on quantile estimators for n=1000.

5 Application

The inequality measures based on deciles and quintiles, as well as the Zenga indices, have been applied to the inequality analysis in Poland by macroregion (NUTS1), based on HBS sample 2014. To obtain the reliable estimates of these coefficients we used the Bernstein quantile estimator which turned out to have the highest precision (tables 1 and 2). Basic characteristics of the HBS sample, divided by macroregions, are presented in table 3, while table 4 contains the results of the approximation of the empirical distributions by means of the Dagum model. We can observe very high consistency of the empirical distributions with the theoretical ones (see table 4 and figures 1 – 2).

Macroregion	Number of households	Minimum	Maximum	Average	Standard Deviation
I	8046	11.00	155017.49	4240.21	3790.53
II	7433	12.50	37152.00	3634.03	2179.59
III	6246	10.00	84032.90	3461.45	2876.23
IV	5658	3.00	43493.45	3772.15	2611.00
V	3971	1.67	37200.00	3591.07	2337.83
VI	5575	9.00	126739.54	3646.44	3225.72
Total	36929	1.67	155017.49	3755.33	2959.95

Table 3. Numerical characteristics of income in macroregions.

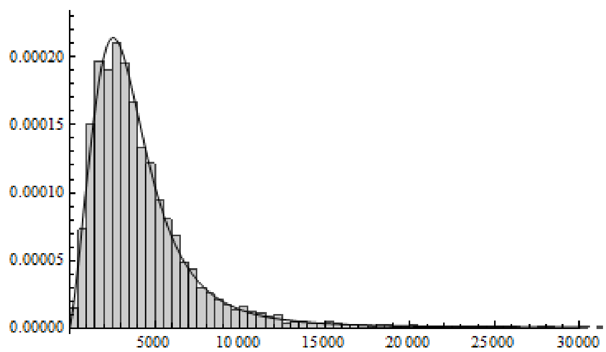


Fig. 1. Income distributions for macroregions I and fitting by means of the Dagum model.

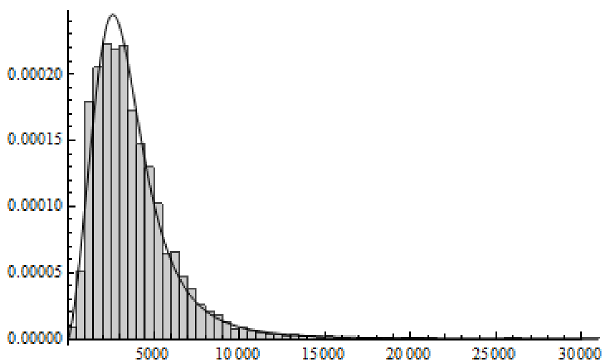


Fig. 2. Income distributions for macroregions IV and fitting by means of the Dagum model.

Macroregion	**Dagum distribution parameters**			**Overlap measure**
	p	***a***	***b***	
I	0.790	2.804	3839.630	0.982
II	0.669	3.618	3800.167	0.970
III	0.756	3.051	3286.467	0.971
IV	0.743	3.233	3687.076	0.964
V	0.722	3.301	3587.800	0.970
VI	0.718	3.158	3544.934	0.979
Total	0.747	3.125	3611.017	0.975

Table 4. Approximation of income distributions for macroregions.

The basic results of inequality analysis have been outlined in table 5. The estimated values of quintile and decile share ratios, as well as the values of synthetic Zenga inequality indices, indicate the central macroregion (I) as the one with the highest income inequality level. It is particularly evident for extremal income groups, i.e. income of the richest 10 percent of households is 12 times bigger than the income of the poorest 10 percent. The southern macroregion (II) presents the lowest values of all inequality measures except for the *K* index.

Macroregion	$\mathbf{W^{(1)}_{20:20}}$	$\mathbf{W^{(2)}_{20:20}}$	$\mathbf{W^{(1)}_{10:10}}$	$\mathbf{W^{(2)}_{10:10}}$	$\mathbf{K_{1:10}}$	**Zenga**
I	3.049	6.939	5.494	12.085	0.083	0.386
II	2.595	4.962	4.283	7.577	0.132	0.269
III	2.904	6.147	4.927	9.908	0.101	0.348
IV	2.750	5.577	4.742	8.614	0.116	0.308
V	2.789	5.375	4.536	8.172	0.122	0.295
VI	2.828	6.039	4.814	9.841	0.102	0.347
Total	2.819	5.916	4.843	9.526	0.105	0.338

Table 5. Inequality measures for macroregions.

Conclusion

Analysis of income and wage distribution is strictly connected with the estimation of inequality and poverty measures based on quantiles. Therefore, for income data coming usually from sample surveys it becomes crucial to use the quantile estimators presenting satisfying statistical properties. In the paper, the Huang-Brill and Bernstein estimators have been proposed and analysed from the point of view of their sampling errors under several income distribution models. In the simulations studies the properties of these estimators have been compared with the classical one which is most often applied in practice. The results of the calculations reveal that the Bernstein estimator performs better than its competitors- its root mean squared errors (RMSE) are much smaller than those observed for the other quantile estimators and its relative biases (BIAS) are also smaller, especially when the quantiles of higher orders are taken into regard. Consequently, the Bernstein estimator has been applied to the estimation of various inequality measures in regions NUTS 1 in Poland.

References

Białek, J. (2014). Simulation Study of an Original Price Index Formula. *Communications in Statistics, Simulation and Computation, 43(2)*, 285–297.

Greselin, F., Pasquazzi, L., & Zitikis, R. (2012). Contrasting the Gini and Zenga indices of economic inequality. *Journal of Applied Statistics, 40(2)*, 282–297.

Harrell, F. E., & Davis, C. E. (1982). A New Distribution-Free Quantile Estimator. *Biometrika*, *69*, 635–640.

Huang, M. L., & Brill, P. H. (1999). A Level Crossing Quantile Estimation Method. *Statistics & Probability Letters, 45*, 111–119.

Jędrzejczak, A. (2012). Estimation of Standard Errors of Selected Income Concentration Measures on the Basis of Polish HBS. *International Advances in Economic Research, 18(3)*, 287-297.

Jędrzejczak, A. (2015). Asymptotic Properties of Some Estimators for Gini and Zenga Inequality Measures: a Simulation Study. *Statistica & Applicazioni*, *13*, 143-162.

Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, Hoboken.

Panek, T. (2011). *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru*. Szkoła Główna Handlowa, Warszawa.

Pekasiewicz, D. (2015). *Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach społeczno-ekonomicznych*. Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Zieliński, R. (2006). Small-Sample Quantile Estimators in a Large Nonparametric Model. *Communications in Statistics Theory and Methods, 35*, 1223–1241.

Zenga, M. (1990). Concentration Curves and Concentration Indices Derived from Them. *Income and Wealth Distribution, Inequality and Poverty,* Springer -Verlag, Berlin, 94–110.

Assessment of risk factors of serious diseases in OECD countries

Pavla Jindrová[1], Lucie Kopecká[2]

Abstract

Despite a remarkable progress in health status and life expectancy in OECD countries over the past decades, there remain large inequalities across countries and also across population groups within each country. These inequalities in health status are linked to many factors, including differences in exposure to risk factors to health and in access to health care. The online OECD Health Database 2016 offers the most comprehensive source of comparable statistics on health and health systems across OECD countries. It is an essential tool to carry out comparative analyses and draw conclusion from international comparisons of health care results. The aim of this article is to use the mentioned database and apply appropriate multidimensional statistical methods to assess the risk factors in relation to mortality caused by selected serious diseases, and quantify the impact of factors such as gender, age, income inequality, costs of treatment and selected characteristics of health systems in OECD countries.

Keywords: *comparisons, factor analysis, multidimensional comparative analysis, risk factors, serious diseases*

JEL Classification: C38, I15

1 Introduction

The mission of the Organization for Economic Co-operation and Development (OECD) is to promote policies that will improve the economic and social well-being of people around the world. Today this organization focuses on helping governments around the world to re-establish healthy public finances as a basis for future sustainable economic growth.

According to *Health at a Glance 2015* (2015) people in OECD countries are living longer than ever before, with life expectancy now exceeding 80 years on average, thanks to improvements in living conditions and educational attainments, but also to progress in health care. But these improvements have come at a cost. Health spending now accounts for about 9% of GDP on average in OECD countries, and exceeds 10% in many countries. Higher health spending is not a problem if the benefits exceed the costs, but there is an evidence of inequities and inefficiencies in health systems which need to be addressed.

Nearly all OECD countries have achieved universal (or almost universal) health coverage for a core set of health services and goods. Still, inequalities in access to care exist across

[1] University of Pardubice/FEA, Institute of Mathematics and Quantitative Methods, Studentská 84, 532 10 Pardubice, Czech Republic, e-mail: Pavla.Jindrova@upce.cz.

[2] University of Pardubice/FEA, Institute of Mathematics and Quantitative Methods, Studentská 84, 532 10 Pardubice, Czech Republic, e-mail: Lucie.Kopecka1@seznam.cz.

different socio-demographic groups, including sex, age, geographic area and socio-economic status, for financial and non-financial reasons.

By using selected multidimensional statistical methods on a selected set of indicators of health status, focusing particularly on serious diseases and risk factors of health, health expenditures and health systems financing in OECD countries we attempt to identify and quantify what causes the differences in people´s health in OECD countries.

2 Data and methods

The OECD health database *OECD Health Statistics 2016* (OECD, online) offers the most comprehensive source of comparable statistics of health and health systems across OECD countries. This online database was released on June 30 and all datasets have been updated on October 12. The list of variables in OECD health statistics is very broad. The problem is missing data for some OECD countries which it is possible partly supplement from the database of World Health Organization (WHO, online).

As the basis of a multivariate statistical analysis the following indicators from the database *OECD Health Statistics 2016* have been selected:

$X1$ – Current expenditure on health, % of gross domestic product

$X2$ – Current expenditure on health, per capita, US$ purchasing power parities

$X3$ – Public expenditure on health, % of current expenditure on health

$X4$ – Public expenditure on health, per capita, US$ purchasing power parities

$X5$ – Life expectancy at birth, total population

$X6$ – Life expectancy at 65 years old, female population

$X7$ – Life expectancy at 65 years old, male population

$X8$ – Alcohol consumption, liters per population aged 15+

$X9$ – Poverty rate, 2014 or late

$X10$ – Gini coefficient

$X11$ – Neoplasms mortality

$X12$ – Malignant neoplasms of trachea, bronchus, lung mortality

$X13$ – Malignant neoplasms of colon mortality

$X14$ – Leukemia mortality

$X15$ – Malignant neoplasms of bladder mortality

$X16$ – Ischemic heart disease mortality, 2013 (or nearest year)

$X17$ – Cerebrovascular disease mortality, 2013 (or nearest year)

According to the above mentioned goals of analysis of these variables we have used the factor analysis (FA), multidimensional comparative analysis (MCA) and graphical methods.

The goal of factor analysis (Hair et al., 2007) is to characterize the *p* variables in terms of a small number of common factors. An important result of the factor analysis model is the relationship between the variances of the original variables and the variances of the derived factors. An important concept in factor analysis is the rotation of factors. In practice, the objective of all methods of rotation is to simplify the rows and columns of the factor matrix to facilitate interpretation. The Varimax criterion centres on simplifying the columns of the factor matrix.

The correlation between the original variables and the factors is shown by the factor loadings. They are the key for understanding the nature of a particular factor. Squared factor loadings indicate what percentage of the variance in an original variable is explained by a factor.

The factor scores in the results of a factor analysis procedure display the values of the rotated factor scores for each of *n* cases, in our analysis in each of 34 OECD countries. The factor score show where each country belongs to with respect to the extracted factors.

Multidimensional comparative analysis deals with the methods and techniques of comparing multi-feature objects, in our case OECD countries. The objective to establish a linear ordering among a set of objects in a multidimensional space of features, from the point of view of certain characteristics which cannot be measured in a direct way (the standard of living, public health situation ...). Application of these methods to compare health and health care in selected countries can be found for example in Pacáková et al., (2016), Pacáková et al., (2013) or Pacáková and Papoušková (2016).

At the beginning of the analysis, the type of each variable should be defined. It is necessary to identify whether the *great* values of a variable positively influence the analysed processes (such variables are called stimulants) or whether their *small* values are favourable (these are called destimulants).

$$b_{ij} = \frac{x_{ij}}{x_{\max,j}} \cdot 100 \tag{1}$$

$$b_{ij} = \frac{x_{\min,j}}{x_{ij}} \cdot 100 . \tag{2}$$

The initial variables employed in composing an aggregate measure are, usually, measured in different units. The aim to normalize them is to bring them to comparability. Normalisation is performed according to formula (1) for stimulants and to formula (2) for destimulants

(Stankovičová and Vojtková, 2007). The synthetic indicator for each country has been calculated as the average of the b_{ij} , $i = 1,\dots, n$.

3 Results and discussion

- *Results of Factor analysis*

The purpose of the analysis is to obtain a small number of factors which account for most of the variability in 17 variables. In this case, 4 factors have been extracted, since 4 factors had eigenvalues greater than 1.0. Together they account for 78.991% of the variability in the original data.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
*X*1	0.2539	0.1133	0.8779	-0.0448
*X*2	0.2726	0.0066	0.9260	0.1457
*X*3	0.2089	0.1939	-0.0288	0.7818
*X*4	0.3583	0.0032	0.7812	0.3433
*X*5	0.9097	-0.0417	0.2190	0.2316
*X*6	0.9016	-0.0708	0.1658	0.1104
*X*7	0.8508	-0.2249	0.2677	0.0598
*X*8	-0.1359	0.3003	0.2620	0.2219
*X*9	0.0802	-0.1742	-0.2042	-0.8990
*X*10	-0.1204	-0.3453	-0.1147	-0.8522
*X*11	-0.2612	0.8641	-0.0616	0.2957
*X*12	-0.1182	0.7716	0.0996	0.1547
*X*13	-0.2444	0.7495	-0.2006	0.3561
*X*14	-0.1392	0.6858	0.1673	0.0099
*X*15	0.0695	0.8633	-0.1784	0.1536
*X*16	-0.8095	0.2687	-0.1424	0.1621
*X*17	-0.5808	0.3016	-0.3667	-0.0095

Table1. Factor Loading Matrix After Varimax Rotation.

Factor loadings (Table 1) present the correlation between the original variables and the factors after Varimax rotation and they are the key for understanding the nature of a particular factor. Rotation is performed in order to simplify the explanation of the factors. Substantive interpretation of the four extracted factors is based on the significant higher loadings in Table 1. Factor 1 (*F*1), which explains 34.448% variability of the total variability in the data, has 3 significant loadings with positive signs with variables *X*5-*X*7, which are the variables of life expectancy. Factor 1 has also significant loadings with negative signs with variables *X*16 and *X*17, which are indicators of cardiovascular diseases mortality. According to *Health at a Glance 2015* all the above mentioned variables are the main indicators of health status, so *F*1 can be identified as *a Health status factor* in OECD countries. The high values of this factor mean high level of health status. Strong significant positive correlation with variables *X*11 – *X*15 is the reason that we have interpreted *Factor* 2 (*F*2) as a *Cancer mortality factor*. This factor explains 28.503% of the variability in the data. The higher the value of *F*2, the higher is the mortality from cancer. *Factor* 3 (*F*3) explains 8.695% of the variability in the data and correlates strongly with variables *X*1, *X*2 and *X*4, so we can interpret it as a *Health expenditures factor* in OECD countries. The higher are the values of *F*3 are, the higher are the health expenditures in OECD countries are, and vice versa. The Factor 4 (*F*4) explains 7.345% of the whole data variability and its significant positive correlation with variable *X*3 and significant negative correlation with variables *X*9, *X*10 is the reason that we have interpreted it as an *Economic and social situation factor*. The higher the values of *F*4 are, the better the economic and social situation of relevant OECD countries is.

Graphical display of OECD countries in a two-dimensional coordinate system with axes of the two selected factors allows us to assess quickly the level of the both factors in each OECD country, and also allows us to compare the situation in all OECD countries by these factors and to assess the causal relationship of the two selected factors.

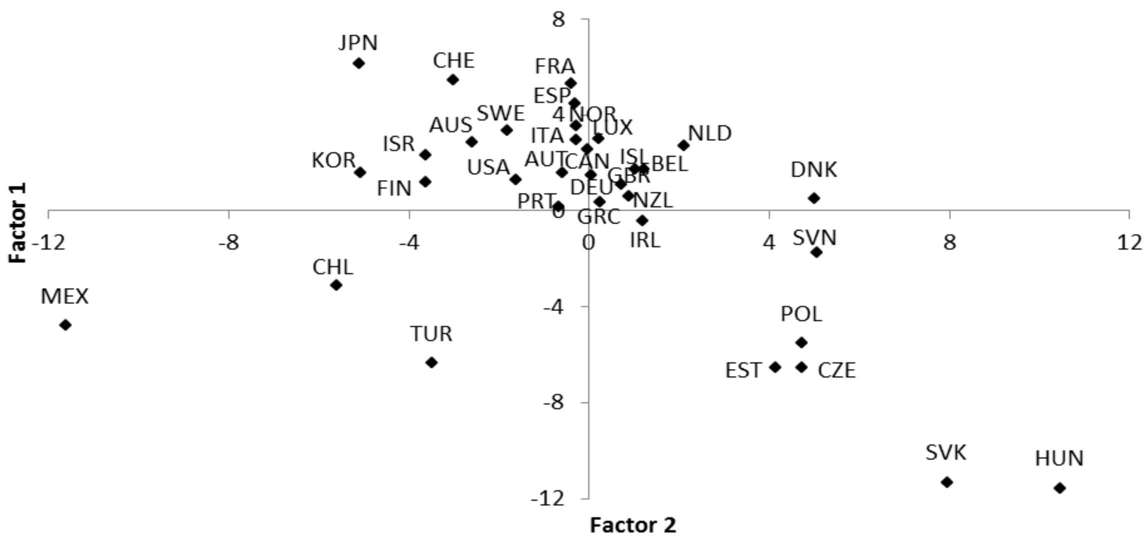


Fig. 1. Location of OECD countries in the coordinate system of factors *F*1 and *F*2.

Figure 1 presents the causal relationship of factors $F1$ and $F2$. Spearmans' rank correlation between this pair of factors is -0.4124, which means moderately strong indirect dependency. Ireland and the former socialist countries, namely Slovenia, Poland, Estonia, the Czech Republic, the Slovak Republic and Hungary, represent a group of countries with high cancer mortality and low level of health status. Besides the low mortality in Mexico, Chile and Turkey we can observe low level of health status too. Dependence of life expectancy and level of health care is confirmed also in articles by Jindrová and Slavíček (2012) and Kubanová and Linda (2014).

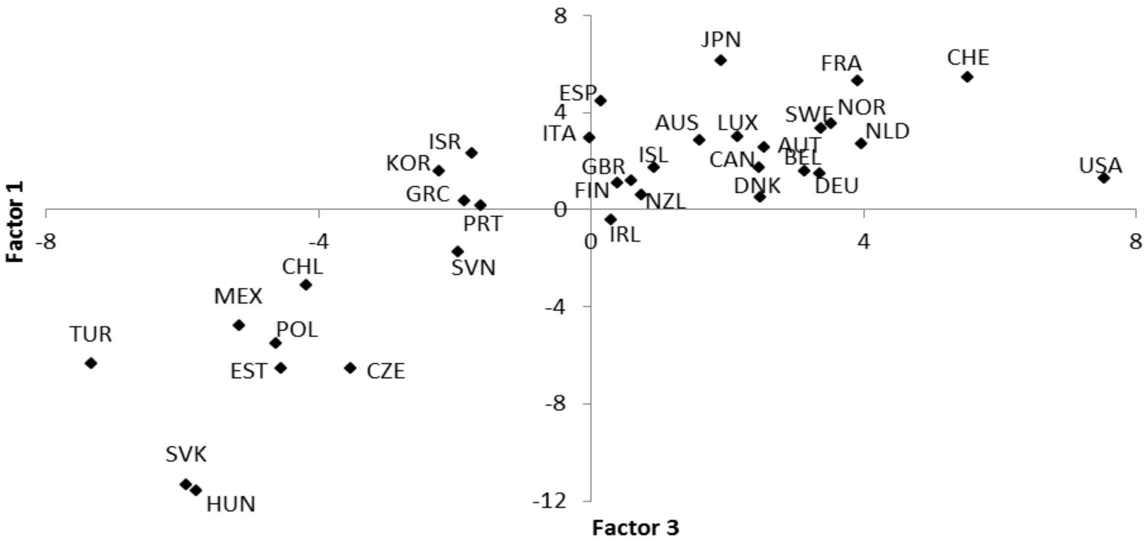


Fig. 2. Location of OECD countries in the coordinate system of factors $F1$ and $F3$

The value 0.7470 of Spearman rank correlation coefficient between factors $F1$ and $F3$ is shown in Figure 2. It is obvious that health expenditures in OECD countries considerably affect health status. In the lower left quadrant, which represents the lowest health expenditures and also the lowest health status, again the countries such as Turkey, the Slovak Republic, Hungary, Mexico, Chile, Poland, Estonia, the Czech Republic and Slovenia can be found.

It is evident that the high level of health expenditures in OECD countries does not automatically imply high level of health status (see the USA). The important thing is, of course, the efficient use of this health expenditure, but the assessment of effectiveness is not an objective of this article. According to Figure 2, we can conclude the effective use of health expenditure in Japan and Spain.

Factor 4 referring to the economic and social situation explains only 7.345% variability of the whole dataset and high values of this factor are the consequence of a high proportion of public expenditure on health from the current expenditure or the consequence of a low poverty rate. Spearman rank correlation coefficient between factors *F*1 and *F*4 is only 0.1759, but between factors *F*2 and *F*4 is higher, it equals 0.6037. The relationship of factors *F*2 and *F*4 is presented in Fig. 3.

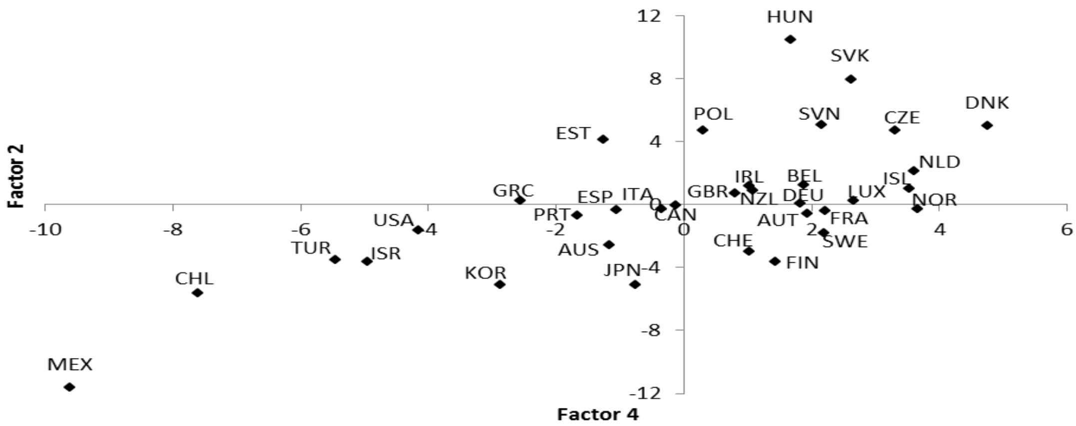


Fig. 3. Location of OECD countries in the coordinate system of the factors *F*2 and *F*4.

The fact that the economic and social situation of the inhabitants in OECD countries is an important factor of health status confirmed also the results of a self-reported health assessment in selected OECD countries (Fig. 4). A significant relationship between health and poverty and social deprivation is also presented in article by Šoltés and Šoltésová (2016).

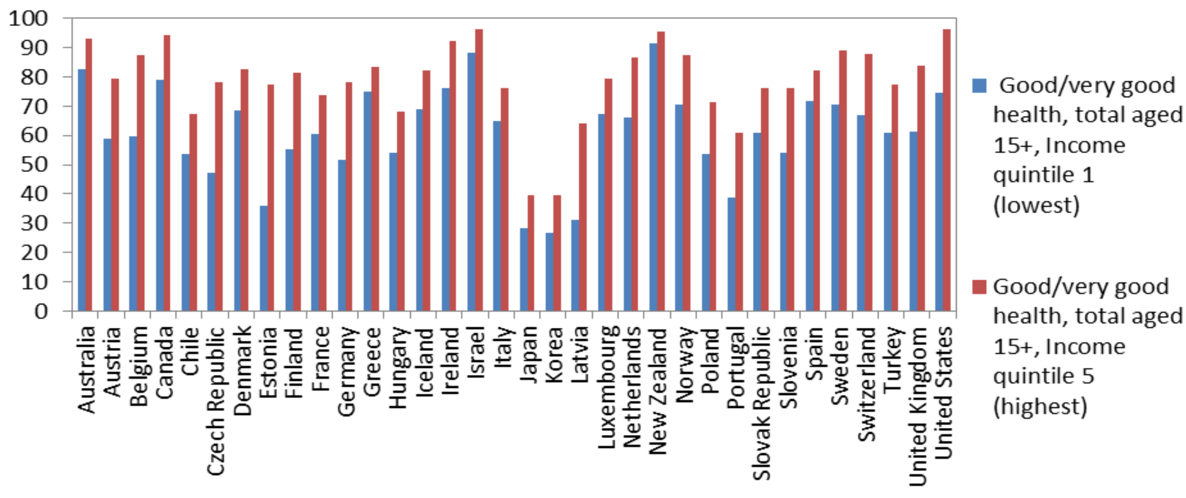


Fig. 4. Perceived health status by socio-economic situation

- *Results of multidimensional comparative analysis*

Table 2 contains the result of linear ordering of OECD countries by variables *X*1-*X*17, where variables *X*1-*X*7 are considered as stimulants and *X*8-*X*17 as destimulants with standardized values by formulas (1) and (2). Hungary, Estonia, the Slovak Republic and Poland represent the countries with the worst situation in health status and its associated indicators.

Average	Country	Rank	Average	Country	Rank	Average	Country	Rank
65.6773	CHE	1	59.6782	AUT	13	53.3999	PRT	25
65.3741	JPN	2	59.6179	BEL	14	51.7387	TUR	26
64.6189	NOR	3	59.4758	DEU	15	51.1672	SVN	27
64.6164	FRA	4	58.4798	CAN	16	50.9784	GRC	28
63.3194	NLD	5	57.6897	KOR	17	50.5093	CZE	29
62.5676	SWE	6	56.9214	AUS	18	49.9149	CHL	30
61.3194	MEX	7	56.4980	ESP	19	46.7496	POL	31
61.1696	DNK	8	55.5291	ISR	20	45.9812	SVK	32
61.0327	USA	9	55.5039	GBR	21	44.8860	EST	33
60.7910	ISL	10	54.9689	ITA	22	42.8309	HUN	34
60.4493	FIN	11	54.5284	NZL	23			
60.4416	LUX	12	53.8069	IRL	24			

Table 2. The results of multidimensional comparative analysis

- *Sex as a risk factor for serious diseases*

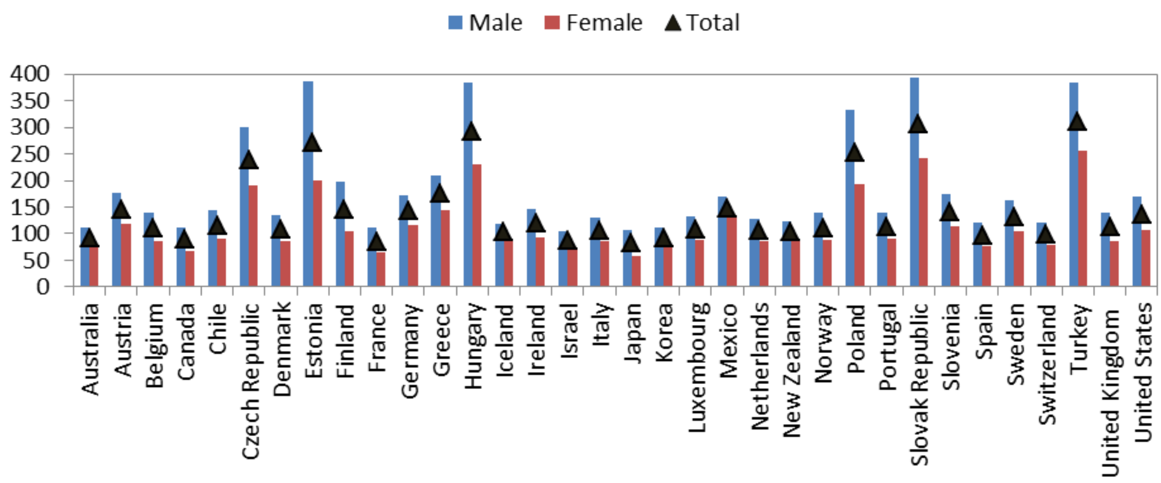


Fig. 5. Age-standardized mortality rate by cardiovascular diseases by sex (2013 or nearest year, per 100 000 population).

Cardiovascular and cancer diseases remain in the main causes of mortality in most OECD countries. Mortality from cardiovascular diseases, accounting 32.3% of all deaths in OECD countries in 2013, varies considerably across countries for both sex (Fig. 5).

Very high level of mortality from cardiovascular diseases with comparison of other countries is in the Slovak Republic, Hungary, Estonia, Poland, the Czech Republic and Turkey. The Slovak Republic and Hungary report a cerebrovascular mortality more than three times higher than that on Switzerland, Canada and France (WHO, 2017).

Cancer is the second leading cause of mortality in OECD countries, accounting for 25% of all death in 2013, up from 15% in 1960. In 2013, the average rate of mortality attributable to cancer across OECD countries was just over 200 per 100 000 population (Fig. 6).

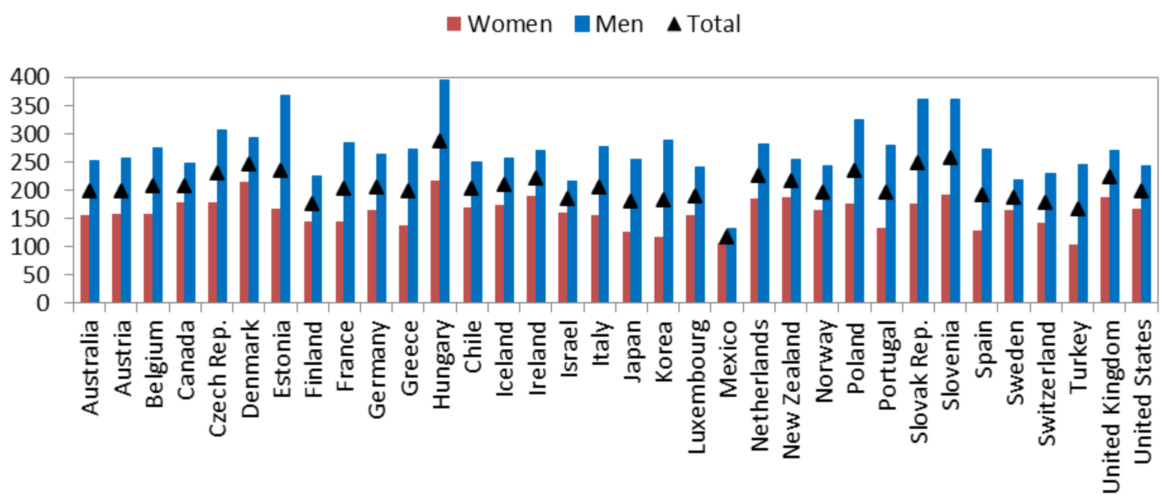


Fig. 6. Age-standardized mortality rate by cancer diseases by sex (2013 or nearest year, per 100 000 population).

Mortality due to cancer was the lowest in Mexico, Turkey, Finland, Switzerland and Japan. On the other hand, Hungary, Slovenia, the Slovak Republic and Denmark bear the highest cancer mortality burden, with the rates exceeding 240 per 100 000 population (OECD, 2015). In several countries, the death rate from cancer is as twice for men as for women (Fig. 6).

Conclusions

The results of analyses in the article confirmed the significant impact of the health expenditures on health status and cancer mortality in OECD countries. The graphic comparison of mortality from cardiovascular disease and cancer makes it evident that gender is a significant factor in mortality from these leading causes of death. The multivariate comparison of OECD countries and several other results in the article unfortunately confirm the poor state of health of the former socialist countries.

Acknowledgements

This article was supported by grant of the University of Pardubice, FEA, No. SGS_2017_022.

References

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2007). *Multivariate Data Analysis*.

Jindrová, P., & Slavíček, O. (2012). Life Expectancy Development and Prediction for Selected European Countries. In *Managing and Modelling of Financial Risks, Proceedings 6th International Scientific Conference*. Ostrava: VSB-TU Ostrava, 303-312.

Health at a glance 2015: OECD indicators. (2015). Paris: OECD Publishing.

Kubanová, J., & Linda, B. (2014). The process of population aging and its impact on country's economy. *Political sciences,Law, Finance, Economics and Tourism, 4,* 767-774.

OECD Health Statistics 2016. (n.d.). Retrieved February 15, 2017, from http://www.oecd.org/els/health-systems/health-data.htm

Pacáková, V., Jindrová, P., & Seinerová, K. (2013). Mortality Models of Insured population in the Slovak Republic. In M. Papież & S. Smiech (Eds.), *Proceedings of the 7th Professor Aleksander Zelias International conference on modelling and forecasting of socio-economic phenomena*. Cracow: Cracow University of Economics, 99-106.

Pacáková, V., Jindrová, P., & Zapletal, D. (2016). Comparison of Health Care Results in Public Health System of European Countries. In *Proceedings of the 13th International Scientific Conference European Financial Systems 2016*, Brno: MU Brno, 534-541.

Pacáková, V., & Papoušková, M. (2016). Multidimensional Comparisons of Health Systems Functioning in OECD Countries. *International Journal of Mathematical Models and Methods in Applied Sciences*, 10, 388-394.

Šoltés, E., & Šoltésová, T. (2016). *Occurrence of low work intensity in Slovakia in relation to assessment of poverty and social exclusion*. In M. Papież & S. Smiech (Eds.), *Proceedings of the 10th Professor Aleksander Zelias International conference on modelling and forecasting of socio-economic phenomena*. Cracow: Cracow University of Economics, 189-198.

Stankovičová, I., & Vojtková. M. (2007). *Viacrozmerné štatistické metódy s aplikáciami*. Bratislava: IURA Edition.

World Health Organization. (n.d.). Retrieved February 15, 2017, from http://www.who.int/en/

SVM classifiers for functional data in monitoring of the Internet users behaviour

Daniel Kosiorowski[1], Dominik Mielczarek[2], Jerzy P. Rydlewski[3]

Abstract

Novel tools offered by functional data analysis offers various methods especially adequate for monitoring phenomena appearing within the new economy, which are described by means of functions of a certain continuum. We mean here, among others, fraud detection in credit card transactions, electricity demand or the Internet traffic monitoring. This paper focuses our attention on a concept of classifier for functional data induced by the general support vector machines methodology. We, among other, study robustness, computational complexity and consistency of the classifiers using analytical as well as empirical arguments. We compare their properties with alternatives presented in the literature using real data set concerning behaviour of users of a big Internet service divided into four subservices.

Keywords: *classifier for functional objects, data depth, the Internet service, management, support vector machine classifier*

JEL Classification: C14, C22, C38

1 Introduction

Many phenomena, we dealt with in the field of economics, takes the form of a function. Note, that different economic phenomena can be formulated in terms of suitable classifier determining. Let us take, for instance, classifying a candidate for a certain position on a labour market or conducting sell/buy decisions within an algorithmic trading. Various research problems occur when functional outliers are present in the considered data set. Moreover, if we do not possess a reliable economic theory on data generating processes, which could be used for describing the economic phenomena, then functional generalizations of well-known statistical procedures are irrelevant (Ramsay & Silverman, 2005; Horvath & Kokoszka, 2012). Main aim of the paper is to propose a novel statistical methodology for classifying functional objects, that can be applied to monitor and manage the Internet users behaviour. This study shows, that a proper Support Vector Machine methods (SVM) (Schoelkopf and Smola, 2002) enable for efficient classification of functional data, appearing

[1] Corresponding author: Cracow University of Economics, Department of Statistics, ul. Rakowicka 27, 31-510 Kraków, e-mail: daniel.kosiorowski@uek.krakow.pl.

[2] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: dmielcza@wms.mat.agh.edu.pl.

[3] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland, e-mail: ry@agh.edu.pl.

in modern e-economy (Anagnostopoulos et al., 2012). It is clear, that this method can be used to classify various functional time series objects, e.g. daily concentration of dangerous particles in the atmosphere. Hence, this study proposes a novel nonparametric and robust method for phenomena classification, which may be described as functional objects. The rest of the paper is organized as follows. Section 2 sketches the basic concepts of classification in a functional setup. Section 3 introduces our procedure for classifying a functional data. Section 4 discusses the properties of the procedure through numerical simulations and tests the applicability of the proposed methodology on empirical examples (i.e., activities of Internet users). Section 5 conducts a short sensitivity analysis and comprises a brief summary.

2 Studied data

We consider data from certain Internet service. The service has been divided into four sub-services, called service 1-4, hereafter. Using classifying methods, we would like to classify a new functional object into one of the four considered services. Figure 1 presents four functional principal components for service 1(left) and service 2 (right), for functional PCA methods and its applications see Górecki and Krzyśko (2012). We have used free R packages fda and fda.usc for computations (Febrero-Bande and de la Fuente, 2012).

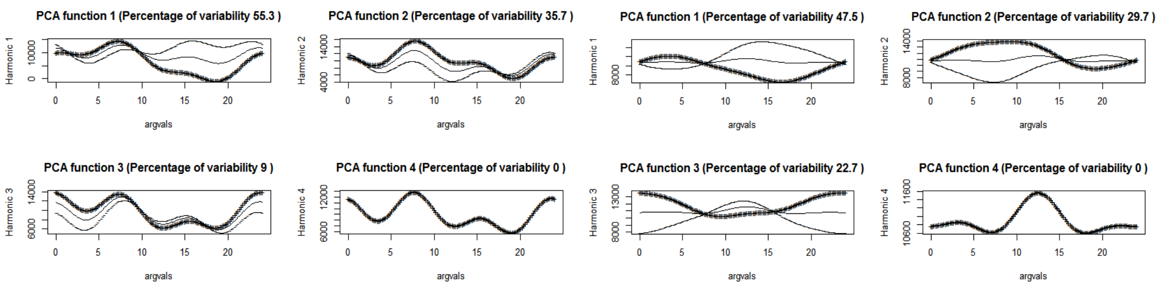


Fig. 1. Functional principal components for service 1(left) and service 2 (right).

3 Basic classification methods of functional objects

In a context of functional data classification, we are given a training sample, that is, at the beginning we are given *n* observations, namely Z_1, Z_2,…Z_n and each observation can be regarded as $Z_i=(f_i, Y_i)$ where $Y_i=0$ or $Y_i=1$, the f_i are called patterns cases, inputs or instances, while Y_i are called labels, outputs or targets. Our aim is to classify a new object *f* into one of the two labelled groups, basing on knowledge included in the training sample. In general, finding a classification rule means to obtain a partition of the feature space into nonempty disjoint subsets, which are summed up to a whole feature space. Partition subsets relate to the appropriate labels. Note, that the most natural method for classifying a functional data is the k-nearest neighbors method. The input set consists of the *k* closest to the classified object training examples in the considered feature space. Note, that many different metrics could be taken into account. The output is then the assignation to a class, which is most common among its *k* nearest neighbors. A crucial and still an opened problem is an appropriate choice of the metric defining neighborhood and the number of neighbors *k* (Ferraty and Vieu, 2006).

The second groups of methods is based on a concept of nearest centroid. The nearest centroid classifiers method assigns to observations the label of the class of training samples whose centroid is closest to the observation. As the centroid we may consider a functional mean, a functional median induced by a certain functional depth.

The third method is based on a concept of local depth for functional data (Paindaveine and Van Bever, 2013) recently used in Kosiorowski et al. (2016) and Kosiorowski et al. (2017). A depth function is a function that assigns to a point a positive number from an interval [0,1] expressing its centrality with respect to the probability measure or a sample, for details see Zuo and Serfling (2000). The local depth concept is a generalization of the concept of global depth. It takes into account the local features of the analysed data, what is especially important in a context of clustering, for example. The object, for which a depth takes a maximal value is called the median induced by this depth. Within Paindaveine and Van Bever (2013) concept, parameter of locality is denoted with β. It ranges from 0 (extreme localization) to 1 (global depth). Figure 2 shows local medians calculated for services 1-4, with locality parameter $\beta = 0,45$.

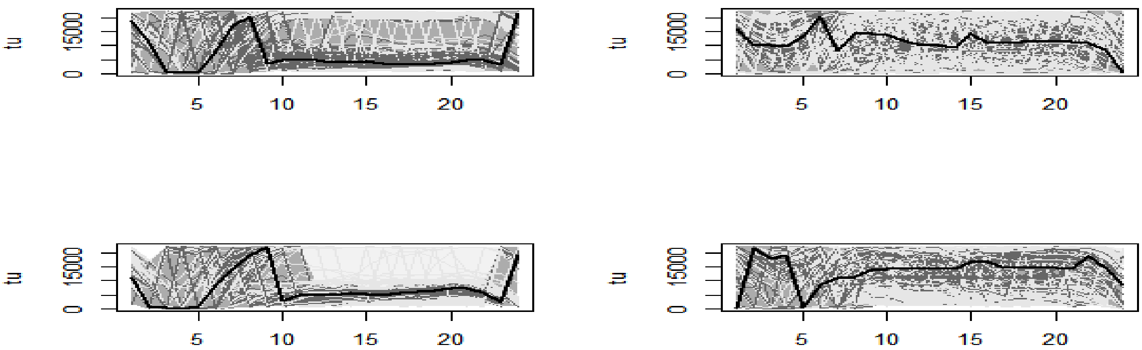


Fig. 2. Local medians calculated for services 1-4, $\beta = 0,45$.

4 Our proposals – Machine learning methods

Our first proposal requires calculation of local depth for functional data. We create then *depth versus depth* plot *(DD*-plot*)*, introduced by Liu et al. (1999). DD–plots show the values of depth function of each observation under two distributions. We use SVM methods in a two-dimensional space of DD-plots in order to divide the data into two groups. Lange et al. (2014) introduced a two-step depth transformation for functional data. Their proposition is to map the functional data into finite-dimensional location-slope space, where each functional observation is treated as a vector consisting of integrals of its levels (location) and first derivatives (slope) over and equally sized subintervals, respectively. The data are transformed then to *DD*-plots with a multivariate depth function. Finally, the data can be discriminated on the *DD*-plot with the well-known classical methods. Another possibility is to consider a depth of the functional data at the beginning of the procedure. Other methods, including Hubert et al. (2016) classifier, basing on so called bag-distance were considered in Kosiorowski and Bocian (2015). The formerly presented classifiers are characterized with a relatively large classification error. That's why we focus our attention on very promising in multivariate case SVM classifier (Schoelkopf and Smola, 2002).

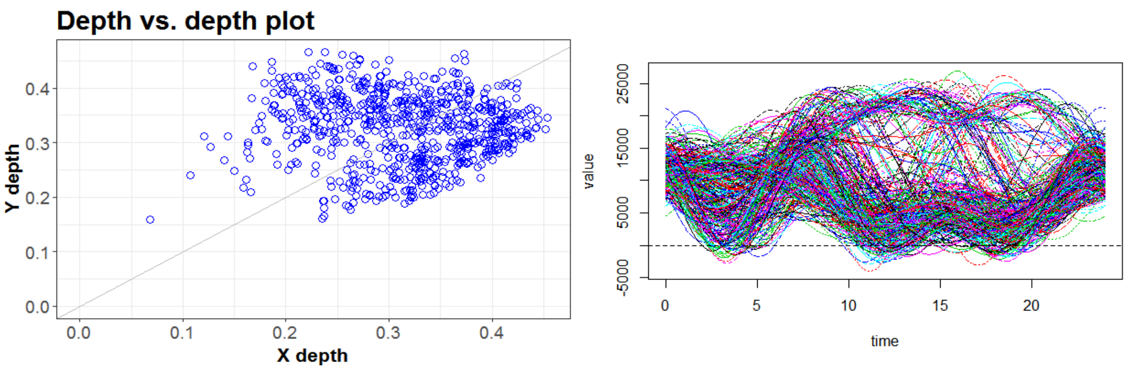
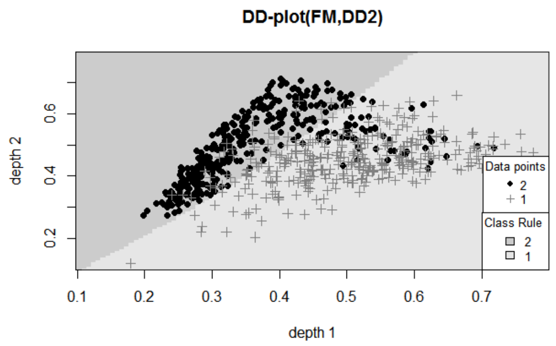
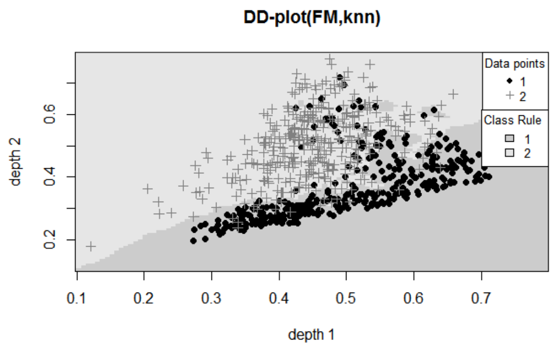
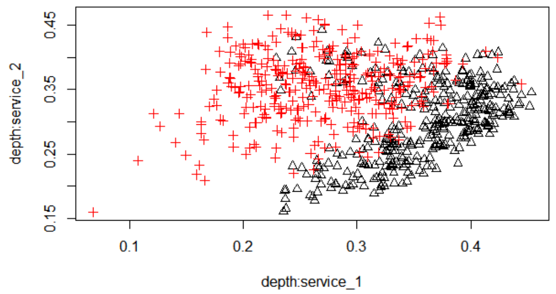


Fig. 3. DD-plot of number of users of service 1 vs. number of users of service 2, for $\beta = 0,45$ (left). Number of service 1 users (right).

Figure 4 shows the DD-plot of service 1 vs. service 2 classifiers based on Fraiman-Muniz depth. The SVM does use a second-order polynomial to make a classification. In such a case a probability of correct classification equals 0.87. Figure 5 shows the DD-plot of service 1 vs. service 2 based on Fraiman-Muniz depth. The k-nearest neighbours algorithm is used to make a classification. A correct classification probability equals 0.9148. The results show that the proposed method are a relevant methods to undertake a classification.







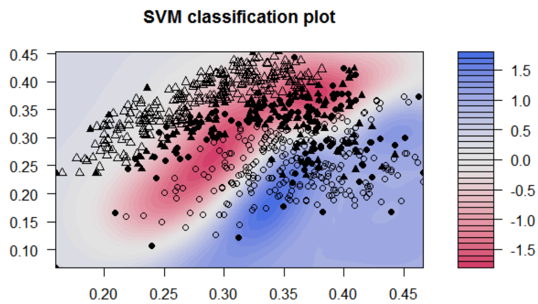


Fig. 4. (top left) DD-plot classifiers of service 1 vs. service 2 based on Fraiman-Muniz depth. The SVM does use a second-order polynomial. A correct classification probability equals 0.87.
Fig. 5. (top right) DD-plot classifier of service 1 vs. service 2 based on Fraiman-Muniz depth. The k-nearest neighbours algorithm is used. A correct classification probability equals 0.9148.
Fig. 6. (bottom left) DD-plot, calculated with modified band depth, service 1 vs. service 2.
Fig. 7. (bottom right) SVM classification used to a local DD-plot of service 1 vs. service 2.

Assume, that we have in disposal a set of points $x_{11},\dots,x_{1m},x_{21},\dots,x_{2m},x_{n1},\dots,x_{nm}$. We need to transform them into functions. In other words, in order to formulate our second proposal we set a basis in the $L^2([0,T])$ integrable space of functions, namely $\phi_1,\dots,\phi_2,\dots$ obtain a representation of functions in the assumed space. Let us consider that the sample curves (obtained from the set of points) are centered, i.e. $Ef(t)=0$, and $E\left\|f(t)\right\|^2<\infty$ and that sample curves belong to the $L^2([0,T])$ space with the usual inner product defined by $<f,g>=\int_0^T f(t)g(t)dt$. We obtain a representation:

$$f_i=\sum_{j=1}^{\infty}c_j\phi_j\ .$$

We can use spline basis, Fourier basis (which we found the most appropriate) or another properly defined basis. The Fourier basis system is the usual choice for periodic functions, while the spline basis system is a good choice for aperiodic functions. Note, that the fixed basis is usually infinite. For this reason we need to use the representation up to the K*th* number. We determine K_i for each function with AIC (Akaike information criterion) or BIC (Bayes information criterion). We choose a mode of this set, which we denote K and the number determines the degree to which the data are smoothed. Note, that finally we do not consider functions f_i but functions $\overline{f}_i$, i.e.

$$\overline{f}_i = \sum_{j=1}^{K} c_j \phi_j .$$

Coefficients of the expansion c_j are usually estimated with least squares method, i.e. we minimize the least square criterion

$$SMSSE(f \mid c) = \sum_{i=1}^{n} \left[f_i - \overline{f}_i \right]^2 .$$

Next step is to compute principal components for functions $\overline{f}_1, \overline{f}_2, ..., \overline{f}_n$. Functional principal components allow us to make a reduction of the dimension of infinitely dimensional functional data to a smaller finite dimension in an rational way. If functional observations $f_1, f_2, \ldots, f_n$ have the same distribution as a square integrable $L^2\,[0,T]$ - valued random function f, we define the functional principal components as the eigenfunctions of the covariance operator C, where $C = E[< f-\mu, \cdot > (f-\mu)]$. The eigenfunctions of the sample covariance operator $\hat{C}(x) = \frac{1}{N}\sum_{i=1}^{n} < f_i - \hat{\mu}, x > (f_i - \hat{\mu})$, where the sample mean function is $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{n} f_i(t)$, are called the empirical functional principal components of the data. Under some regularity conditions the empirical FPC estimate the FPC's up to a sign.

Empirical FPC's may be interpreted as an optimal orthonormal basis with respect to which we can expand the data. The inner product $< f_i, \hat{v}_j > = \int f_i(t)\hat{v}_j(t)dt$ is called the j*th* score of f_i. It can be interpreted as the weight of the contribution of the functional principal components $\hat{v}_j$ to the curve f_i. To sum up, we replace the actual data by the approximation

$$\overline{f}_i = \sum_{j=1}^{p} < \hat{v}_j, \overline{f}_i > \hat{v}_j .$$

We need to determine a small value of p, such that the above approximation is relevant. A pseudo–AIC method and cross–validation have been proposed for this purpose.

The last step to be done is to use SVM methods in order to decide, whether the new observation f is assigned to the group with Y_i=0 or to the second group with Y_i=1. The obtained for the new observation f (its truncated version $\bar{f}$) principal components scores, namely $<\hat{v}_j, \bar{f}>$, i=1,2,…,n, are assigned with SVM methods into one of two considered groups. This allows us to decide, to which service we can classify the new observation. Figures 8 and 9 present functional PCA for services 1 and 2.

The quality of the PC estimator may be estimated in the following way. Assume that for any $i \in \{1,\dots,n\}$ random variable f_i, which belongs to the Hilbert space, is described in a basis $\{\phi_i\}_{i=1}^{\infty}$, namely $f_i = \sum_{j=1}^{\infty} a_j^i \phi_j$. We obtain for any natural number K

$$\left\| f_i - \sum_{j=1}^{K} \bar{a}_j^i \phi_j \right\|^2 = \left\| \sum_{j=1}^{K} (a_j^i - \bar{a}_j^i)\phi_j + \sum_{j=K+1}^{\infty} a_j^i \phi_j \right\|^2 = \sum_{j=1}^{K} |a_j^i - \bar{a}_j^i|^2 + \sum_{j=K+1}^{\infty} |a_j^i|^2 \geq \sum_{j=K+1}^{\infty} |a_j^i|^2 .$$

Hence, we get $\min_{i\in\{1,\dots n\}} \left\| f_i - \sum_{j=1}^{K} \bar{a}_j^i \phi_j \right\|^2 \geq \min_{i\in\{1,\dots n\}} \sum_{j=K+1}^{\infty} |a_j^i|^2$.

Let $j_0 \in \{1,\dots n\}$ denote a natural number such, that $\sum_{j=K+1}^{\infty} |a_j^{j_0}|^2 = \min_{i\in\{1,\dots,n\}} \sum_{j=K+1}^{\infty} |a_j^i|^2$. Hence, we obtain the best estimator $\bar{f} = \sum_{j=1}^{K} a_j^{j_0} \phi_j$ and the estimation error $\left\| f_i - \sum_{j=1}^{K} a_j^{j_0} \phi_j \right\|^2 \geq \sum_{j=K+1}^{\infty} |a_j^i|^2$.

The third method we propose is based on kernel principal components (Schoelkopf and Smola, 2002). In other words, we need to study a map $\Phi : X \to H$, where H is a Hilbert space, such that $k(x,x') = <\Phi(x), \Phi(x')>$. We have to choose a kernel, where k is a real, symmetric and continuous function. Kernel function may be i.e. a Gaussian kernel or polynomial kernel. Next step is to calculate a covariance matrix $C = \frac{1}{N}\sum_{j=1}^{n} \Phi(f_j)\Phi^T(f_j)$. We create a positive definite matrix $K=(k_{ij})=(k(x_i,x_j))_{i,j}$. Consequently, we calculate kernel principal components

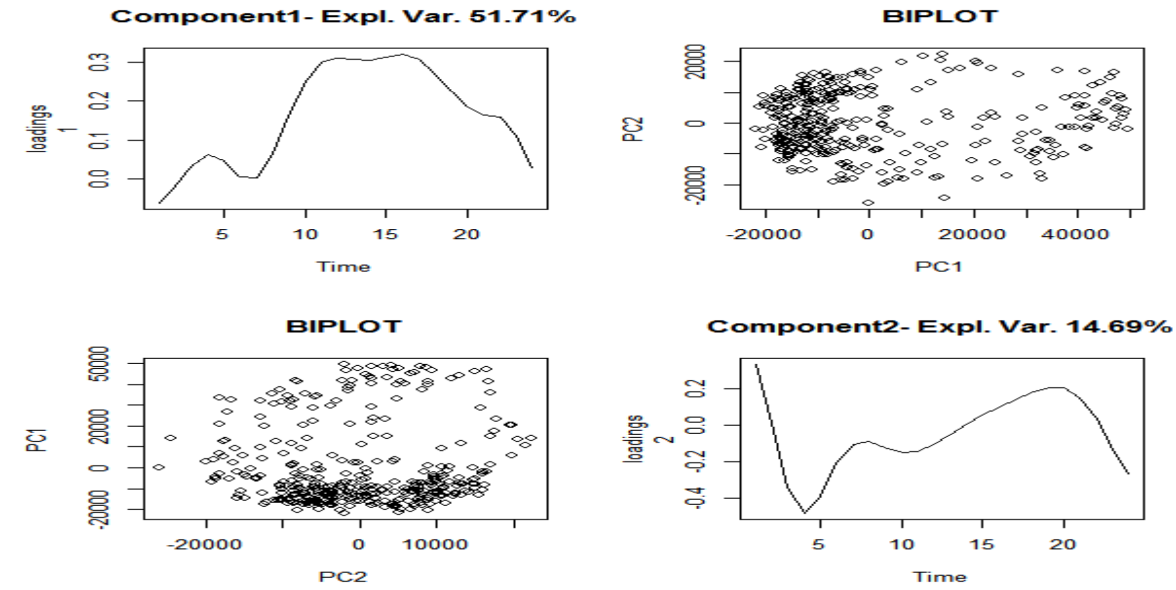


Fig. 8. Functional PCA for service 1.

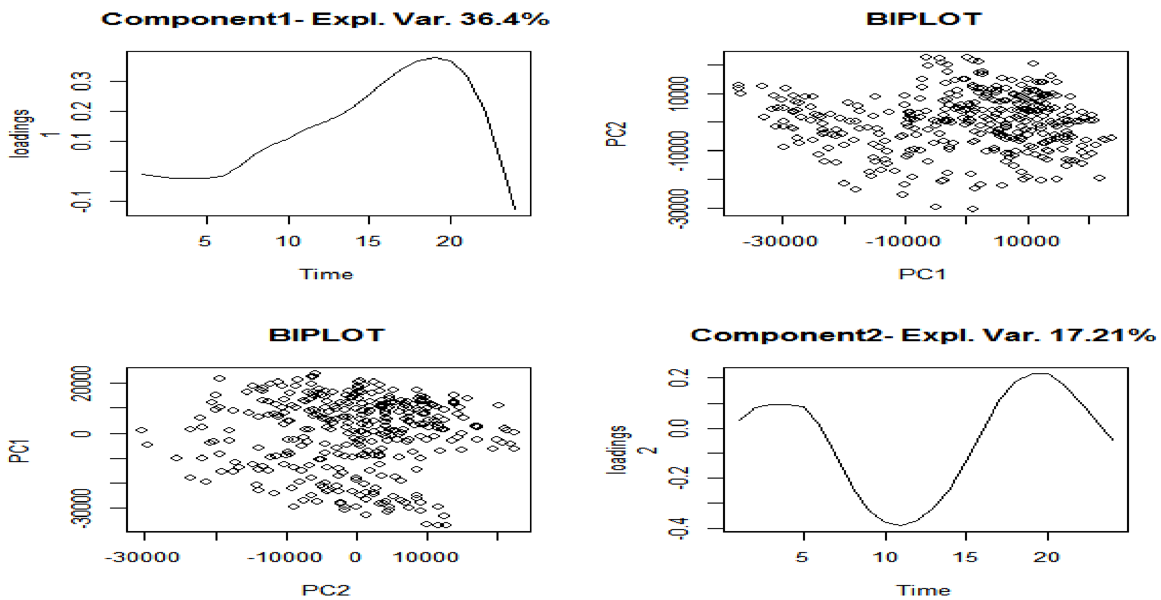


Fig. 9. Functional PCA for service 2.

from the equation λv=Cv, where λ denote eigenvalues and v denote eigenvectors. There exist α_j, j=1,2,…,n such that $v=\sum_{j=1}^{n}\alpha_j\Phi(f_j)$. In practice, we cannot calculate vectors $\{\Phi(f_i)\}$, i=1,2,....,n, because we do not know a function Φ. However, we calculate $\hat{\Phi}_i=\Phi_i-\frac{1}{n}\sum_{k=i}^{n}\Phi_k$, where $\Phi_k=\Phi(\overline{f}_k)$ and $\hat{K}=(\hat{k})_{ij}=(<\hat{\Phi}_i,\hat{\Phi}_j>)$. In this setting we obtain a formula

$\hat{K} = PKP$, where $P = (\delta_{ij} - \frac{1}{n})$ and K is a kernel matrix. Hence, we need to solve an equation $\hat{K}\alpha = \lambda\alpha$ in order to obtain principal components. Then, we do proceed as in the preceding method and we use the SVM method in order assign the new observation *f* into one of the considered two groups.

Conclusions

This paper proposes three classification methods for functional data. All of them base on SVM methodology. The first method uses DD-plots, where DD-plots have been calculated using local depth concept. The second method and the third one are using functional principal components methods, but finally SVM is used as well. The presented methodology enables monitoring phenomena appearing within the new economy, which are described by means of functions of a certain continuum. We show on a real data set, related to Internet users behaviour, promising properties of our proposal. In a future, we plan further studies of the proposals using gamma-regression (Rydlewski, 2009) and beta-regression (Rydlewski and Mielczarek, 2012) methods in SVM classifiers development.

Acknowledgements

DK thanks for the support related to CUE grant for the research resources preservation 2017.

JPR and DM thanks for the support related to AGH local grant.

References

Anagnostopoulos, C., Tasoulis, D. K., Adams, N. M., Pavlidis, N. G., & Hand, D. J. (2012). Online linear and quadratic discriminant analysis with adaptive forgetting for streaming classification. *Statistical Analysis and Data Mining,* 5, 139-166.

Febrero-Bande, M., & de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, *51*(4), 1-28.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Górecki, T., & Krzyśko, M. (2012). Functional principal component analysis. In: Pociecha, J., & Decker, R. (eds.), *Data Analysis Methods and its Applications* (pp. 71-87). Warsaw: Beck.

Horvath, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Hubert, M., Rousseeuw, P. & Segaert P. (2016). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification*. doi: 10.1007/s11634-016-0269-3

Kosiorowski, D., & Zawadzki, Z. (2014). DepthProc: An R Package for Robust Exploration of Multidimensional Economic Phenomena. Retrieved from https://arxiv.org/abs/1408.4542v7

Kosiorowski, D., & Bocian, M. (2015). Functional Classifiers in management of the Internet service. In M. Papież & S. Śmiech (eds.), *The 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow.

Kosiorowski, D., Rydlewski, J. P., & Snarska M., (2017). Detecting a structural change in functional time series using local Wilcoxon statistic. *Statistical Papers*. doi: 10.1007/s00362-017-0891-y

Kosiorowski, D., Mielczarek, D., Rydlewski, J. P. & Snarska M., (2016). Generalized Exponential smoothing in prediction of hierarchical time series. Retrieved from https://arxiv.org/abs/1612.02195v1

Lange, T., Mosler, K., & Mozharovskyi, P. (2014). Fast nonparametric classification based on data depth. *Statistical Papers*, 55(1), 49-69.

Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics*, *27*(3), 783-858.

Paindaveine, D., & Van Bever, G. (2013). "From Depth to Local Depth: a Focus on Centrality." *Journal of the American Statistical Association*, 105, 1105-1119.

Ramsay J., & Silverman, B. (2005). *Functional data analysis*. Springer.

Rydlewski, J. P. (2009). A note on the maximum likelihood estimator in the gamma regression model. *Opuscula Mathematica*, 29(3), 305-312.

Rydlewski, J. P., & Mielczarek D. (2012). On the maximum likelihood estimator in the generalized beta regression model. *Opuscula Mathematica*, 32(4), 761-774.

Schoelkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press.

Zuo, Y., & Serfling, R. (2000). General notions of statistical depth functions. *The Annals of Statistics*, 28(2), 461-482.

K-local median algorithm for functional data in empirical analysis of air pollution data

Daniel Kosiorowski[1], Ewa Szlachtowska[2]

Abstract

Novel tools offered by functional data analysis enable economists for enlarging a range of considered issues as well as for obtaining new insights into classical areas of empirical economic research. This paper presents a novel k–local functional median algorithm for functional data. Its statistical properties as well as its usefulness in analysis of real data set concerning air pollution monitoring in Malopolskie voivodeship in Poland in 2016 are shown. An implementation of the algorithm within a free R package DepthProc is indicated and its comparison with selected alternatives presented in the literature is performed.

Keywords: *air pollution monitoring, k-local functional median clustering, functional k-means clustering, functional data analysis*

JEL Classification: C14, C53, C55

1 Introduction

Pollination in Malopolska is an extremely important social problem. It influences on the social costs (such as the average length and quality of life) and economic (e.g. medical expenses, loss of tourist values). Cluster analysis for functional data representing daily trajectories of concentrations of hazardous substances (including nitrogen oxides) allows for an optimization of the environmental policy of the region.

2 K-local functional median algorithm

Main aim of the paper is to propose a novel statistical method of conducting preliminary analysis of the problem of air pollution in Cracow. For this purpose, we used two clustering algorithms for functional data: k-means algorithm and k-local functional median algorithm. The k-local functional median algorithm is our original proposal, in which we used an idea of local depth proposed in Paindaveine and Van Bever (2013) for the modified band depth proposed by Lopez-Pintado and Romo (2006) and intensively studied in Nieto-Reyes and Battey (2016) .

[1] Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, e-mail: daniel.kosiorowski@uek.krakow.pl.

[2] Corresponding author: AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, e-mail: szlachto@agh.edu.pl.

2.1 Local depth

Depths express centrality of objects with respect to samples or probability disstributions, see for example Nagy et. al. (2016). Depths describe global property of data cloud or the underlying distribution - a degree of outlyingness of a point from a center - the median. In many situations however local properties of data set are of prime importance. To these situations belong clustering issues, where multimodality of data set has to be stressed. In this context several local extensions of depths have been proposed. Our proposal base on one of them presented in Paindaveine and Van Bever (2013) for multivariate data case. For more examples on local depths we refer to Lopez-Pintado et. al. (2007).

Let $x_1(t),...,x_n(t)$ denotes a set of real functions, for simplicity let us assume that they belong to C[0,1] continuous functions defined on an interval [0,1]. A graph of a function x is a subset of $\Re^2$ defined as

$$G(x) = \{(t, t(x)) : t \in [0,1]\}. \tag{1}$$

A band in $\Re^2$ determined by k functions from a sample $x_1,...,x_n$ is defined:

$$\begin{aligned} V(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) &= \left\{(t,y) : t \in [0,1], \min_{r=1,\ldots,k} x_{i_r}(t) \le y \le \max_{r=1,\ldots,k} x_{i_r}(t)\right\} \\ &= \left\{(t,y) : t \in [0,1], y = \alpha_t \min_{r=1,\ldots,k} x_{i_r}(t) + (1-\alpha_t) \max_{r=1,\ldots,k} x_{i_r}(t), \alpha_t \in [0,1]\right\}. \end{aligned} \tag{2}$$

For any function x and set of functions $\{x_1,...,x_n\}$ an index of j functions

$$S_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_j \le n} I\left\{G(x) \subset V(x_{i_1}, x_{i_2}, ..., x_{i_j})\right\}, \tag{3}$$

$j \ge 2$, expresses a fractions of bands $V(x_{i_1}, x_{i_2}, ..., x_{i_j})$ determined by *j* different functions $x_{i_1}, x_{i_2}, ..., x_{i_j}$, covering a graph of *x*.

Definition 1: For functions $x_1,...,x_n$ the band depth of a function f equals

$$S_{n,J}(x) = \sum_{j=2}^{J} S_n^{(j)}(x), \quad J \ge 2. \tag{4}$$

In case, when $X_1,...,X_n$ are independent copies of stochastic process X, which generates $x_1,...,x_n$, population versions of depth indices are defined:

$$S^{(j)}(x) = P\left\{G(x) \subset V(X_{i_1}, X_{i_2}, \ldots, X_{i_j})\right\}, \tag{5}$$

$$S_J(x) = \sum_{j=2}^{J} S^{(j)}(x) = \sum_{j=2}^{J} P\left\{G(x) \subset V(X_{i_1}, X_{i_2}, X_{i_j})\right\}. \tag{6}$$

A function being a sample median with respect to a sample $\hat{m}_{n,J}$ is a curve which maximizes the sample depth:

$$\hat{m}_{n,J} = \arg\max_{x \in \{x_1,\ldots,x_n\}} S_{n,J}(x). \tag{7}$$

In a population case as the median we take m_J in C[0,1] which maximize $S_J()$. Unfortunatelly there are great difficulties in applications of the above concept of functional depth in case of economic time series. Trajectories of economic objects are crossing for many times what makes the band depth rather useless. Lopez-Pintado and Romo (2006) proposed much better concept of functional depth for economic applications. For any function x from a sample $\{x_1,\ldots,x_n\}$, $j \geq 2$ let

$$A_j(x) \equiv A(x; x_{i_1},\ldots,x_{i_j}) \equiv \left\{ t \in [0,1] : \min_{r=i_1,\ldots,i_j} x_r(t) \leq x(t) \leq \max_{r=i_1,\ldots,i_j} x_r(t) \right\}, \tag{8}$$

denotes a set of points in the interval [0,1], for which a function x is inside a band determined by $x_{i_1}, x_{i_2}, \ldots, x_{i_j}$.

If λ is the Lebesque's measure on the interval [0,1], $\lambda(A_j(x))$ is a fraction of time, in which the function x is inside a band.

Definition 2: A generalized band depth of a curve x is

$$GS_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_j \leq n} \lambda(A(x; x_{i_1}, x_{i_2}, \ldots, x_{i_j})), \quad j \geq 2. \tag{9}$$

A function being a sample median is defined as:

$$\hat{m}_{n,J} = \arg\max_{x \in \{x_1,\ldots,x_n\}} S_{n,J}(x). \tag{10}$$

In a population case the median m_J maximizes $S_J(\cdot)$ in C[0,1].

For further generalization of the band depth and their theoretical properties see Nieto-Reyes and Battey (2016). For any depth function D(x,P), the depth regions, $R_\alpha(P) = \{x \in L^2([0,T]) : D(x,P) \geq \alpha\}$ are of paramount importance as they reveal very various characteristics of probability distribution *P*: location, scatter, dependency structure (clearly these regions are nested and inner regions contain larger depth). When defining local depth it will be more appropriate to index the family $R_\alpha(P)$ by means of probability contents. Consequently, for any $\beta \in (0,1]$ we define the smallest depth region with P-probability equal or larger than β as

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P), \tag{11}$$

where $A(\beta)=\{\alpha\geq 0: P(R_\alpha(P))\geq\beta\}$. The depth regions $R_\alpha(P)$ or $R^\beta(P)$ provide neighborhoods of the deepest point only. However we can replace P by its symmetrized version $P_x = 1/2P^x + 1/2P^{2x-X}$. Let $D(\cdot,P)$ be a depth function. The corresponding *sample local depth function at the locality level* $\beta \in (0,1]$ is $LD^\beta(x,P^{(n)})=D(x,P_x^{\beta(n)})$, where $P_x^{\beta(n)}$ denotes the empirical measure with those data points that belong to $R_x^\beta(P^{(n)})$. $R_x^\beta(P^{(n)})$ is the smallest sample depth region that contains at least a proportion β of the 2n random functions $x_1,...,x_n,2x-x_1,...,2x-x_n$. Depth is always well defined – its an affine invariance originates from original depth. Notice for β=1 we obtain global depth, while for β≈0 we obtain extreme localization.

As in the population case, our sample local depth will require considering, for any $x \in L^2$, the symmetrized distribution P_x^n which is empirical distribution associated with $x_1,...,x_n,2x-x_1,...,2x-x_n$. Sample properties of the local versions of depths result from general findings presented in Paindaveine and Van Bever (2013). Implementations of local versions of several depths including projection depth, Student, simplicial, L_p depth, regression depth and modified band depth can be found in free R package *DepthProc* (see Kosiorowski and Zawadzki, 2014). For choosing the locality parameter β we recommend using cross validation related to an optimalization a certain merit criterion (the resolution being appropriate for comparing phenomena in terms of their aggregated local shape differences, which relies on our knowledge on the considered phenomena).

2.2 K–local functional median algorithm – our proposal

We first choose k, where k is user-specified parameter, namely, the number of cluster desired. The second parameter chosen by the researcher is the value of the parameter β (default value of β is 0.2). With the parameter β we may consider the issue at different levels, i.e. changing the value of the parameter β we can control the accuracy of the partition. Received values of local depth for all points are helpful in choosing the amount of clusters.

We consider data whose proximity measure is depth. For our objective function, which measures the quality of a clustering, we use

$$F = \sum_{i=1}^{k} \sum_{j \in C_i} LD^\beta(f_j, P_{n,i}), \tag{12}$$

where $P_{n,i}$ is empirical distribution of i-th cluster.

In first step we calculate the local depth for all points of the data set with respect to chosen values of parameters β and k. Then we are looking for k centroids $c_1,...,c_k$ satisfying

$$\sum_{i=1}^{k} LD^{\beta}(c_i, P^{(n)}) \to \max \tag{13}$$

$$\sum_{i,j=1}^{k} \| c_i - c_j \|_2 \to \max \tag{14}$$

In the second step we create new clusters in such a way that for every point we count L_2 distance from all centroids and assign the point to the closest centroid, that is,

$$Dist(f,P) = \max\{d(f,c_1), d(f,c_2), \dots, d(f,c_k)\}, \tag{15}$$

where d denotes L_2 distance.

If there are two distances with the same values, then we assign a point to clusters with a lower number. In third step for the newly formed cluster we compute the functional local median with respect to the empirical distribution of the cluster. We repeat second and third steps, until centroids do not change or until only 1% of the points change clusters. Note, robustness of the clustering procedure may be evaluated using well known measures of clustering results quality (see Walesiak and Dudek, 2015). For example small changes of input data should lead to small changes of the silhouette plot characteristics in case of robust clustering procedure. More dilemmas of robust analysis of economic data streams can be found in Kosiorowski (2016).

2.3 Trimmed k-local functional median algorithm

The user chooses the parameter γ. We calculate a measure of degree of affiliation to cluster for each observation, i.e. $d(f,c_{l(f)})=d(f,c_{(1)})$. We set obtained values descending

$$d(f_{(1)}, c_{l(f_{(1)})}) \geq d(f_{(2)}, c_{l(f_{(2)})}) \geq \dots \geq d(f_{(n)}, c_{l(f_{(n)})}). \tag{16}$$

Then we reject a proportion γ of the observation of the highest values of measure of affiliation. For this method discriminant factors can be obtained for every observation (trimmed and non trimmed) in the data set (see Fitz et. al., 2015). The quality of the assignment decision of a non trimmed observation f_i to the cluster j with $d(f_{(i)},c_{l(f(i))})$ can be evaluated by comparing its degree of affiliation with cluster j to the best second possible assignment. That is,

$$DF(f_i) = \log\left(\frac{d(f_i, c_{(2)})}{d(f_i, c_{(1)})}\right) \tag{17}$$

for f_i not trimmed. Let $f_{(1)}, \dots, f_{(n)}$ be the observations in the sample after being sorted according to their $d(f_{(i)}, c_{l(f(i))})$ values. It is not difficult to see that $f_{(1)}, \dots, f_{(\lceil \gamma n \rceil)}$ are the trimmed observations which are not assigned to any cluster. Nevertheless, it is possible to compute the degree of affiliation $d(f_{(i)}, c_{l(f(i))})$ of a trimmed observation f_i to its nearest cluster. Thus, the quality of the trimming decision on this observation can be evaluated by comparing $d(f_{(i)}, cl_{f(i)})$ to $d\left(f_{(\lceil \gamma n \rceil+1)}, c_{l\left(f_{(\lceil \gamma n \rceil+1)}\right)}\right)$ with $f_{(\lceil \gamma n \rceil+1)}$ being the non-trimmed observation with smallest value of $d(., c_{l(.)})$. That is

$$DF(f_i) = \log\left(\frac{d(f_{(i)}, c_{l\left(f_{(i)}\right)})}{d(f_{(\lceil \gamma n \rceil+1)}, c_{l\left(f_{(\lceil \gamma n \rceil+1)}\right)})}\right) \tag{18}$$

for f_i trimmed. Hence, discriminant factors $DF(f_i)$ are obtained for every observation in the data set, whether trimmed or not. Observations with small $DF(f_i)$ values (that is, values close to zero) indicate doubtful assignments or trimming decisions. Further properties and theoretical properties of the proposals may be found in Kosiorowski et. al. (2017).

3 Air pollution in Cracow

Air pollution is a very important problem in Malopolska. We use clustering algorithms for functional data to analyze air pollution in Cracow. We choose air quality data in the period from 1 to 31 December 2016 in Cracow, station Avenue Krasinski. All the considered data was taken from *Malopolskie, System monitoringu jakosci powietrza* http://monitoring.krakow.pios.gov.pl/ In the study we used the techniques used in functional data analysis. For more details we refer to Horvath and Kokoszka (2012), Ramsay et. al. (2009). We used the following packages fda.usc (see Febrero-Bande and de la Fuente, 2012) and DepthProc (see Kosiorowski and Zawadzki, 2014).

We observe that the biggest concentration of PM10 occurs in the afternoon, between the hours of 3 p.m. and 8 p.m. In contrast, the height of the concentration of nitrogen dioxide is more varied. It depends not only on time, but also depends on the day of the week.

First, the analysis was conducted for nitrogen dioxide pollution by using k-means algorithm. If we divide the observations into three clusters by k-means algorithm, then first cluster includes 1, 2, 6, 8, 9, 14, 15, 19, 22, 23, 29 December. Second cluster includes 3, 10, 11, 18, 24, 25, 26, 27, 28 December. Third cluster includes 4, 5, 7, 12, 13, 16, 17, 20, 21, 30, 31 December. In the second cluster there are weekends, holidays and the period after Christmas. Examining the first and third cluster it is difficult to find a relationship between the days of the week and the amount of pollination. It is worth noting that in the k-means algorithm, an important step is the selection of centroid. For parameter k=5 at subsequent iterations sometimes we get empty fifth cluster. The greatest concentration of suspended dust was 5, 17, 30 and 31 December, just before New Year's Eve. From the recorded data difficult to see the correctness, in which days of the week is the greatest concentration of dust. However, the lowest concentration is in the Christmas period, i.e. 23-29 of December.

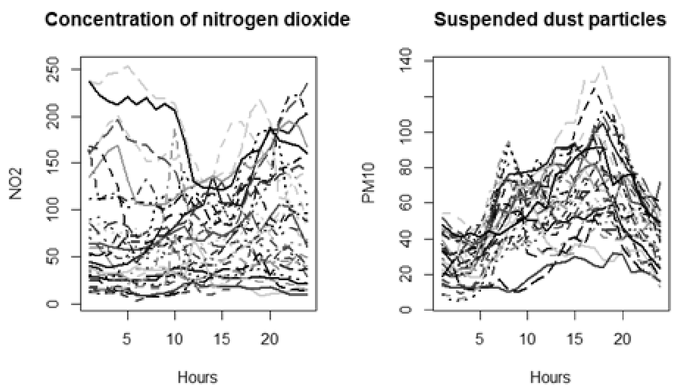


Fig. 1. The amount of nitrogen dioxide and suspended dust particles emitted into the atmosphere as air pollution in Cracow, December 2016.

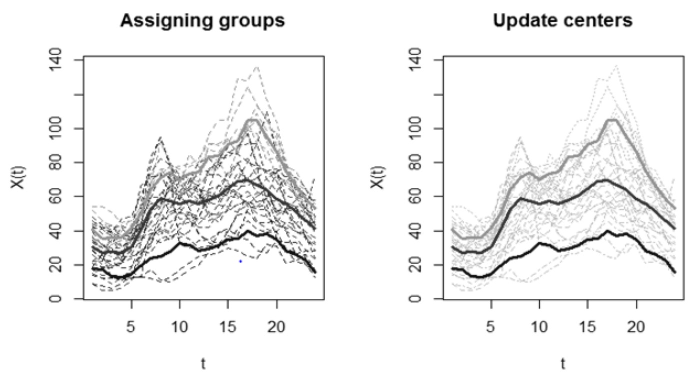


Fig. 2. Functional means for individual clusters and assigning groups for nitrogen dioxide pollution emitted into the atmosphere as air pollution in Cracow, December 2016.
Method – functional k-means algorithm, k=3.

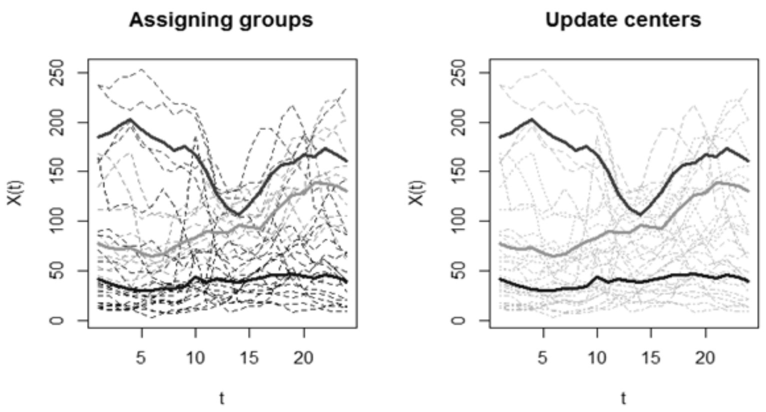


Fig. 3. Functional means for individual clusters and assigning groups for particulate matter pollution in Cracow, December 2016. Method – functional k-means algorithm, k=3.

In the next step of the analysis of concentrations of hazardous substances we used the k-local functional median. By changing input parameter k, we received the optimum division into two clusters. If we divide the observations into three clusters by k-local functional means algorithm, then first cluster includes 1, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 29, 30, 31 December. Second cluster includes 2, 10, 11, 18, 24, 25, 26, 27, 28 December. As in previous analysis in the second cluster there are weekends, holidays and the period after Christmas. While analyzing the concentration of particulate matter we obtained two clusters. First cluster includes 1, 2, 3, 9, 10, 11, 12, 15, 24, 25, 26, 27, 28, 29 December. The greatest concentration of particulate matter occurred in the period before the Christmas, as well as just before New Year's Eve.

Conclusions

In our study we faced an issue of missing data. The missing data may arise through equipment failure. They were supplemented by using the median of the observed cases on the variable in each hour. The next issue was to compare the results obtained with two clustering algorithms. They were very similar. However, because of the computational complexity of the local functional median calculation our algorithm randomly selects a sample, for which it determines centroids. Generating several times the algorithm helps us to verify the choice of number of clusters by comparing the received groups and variation in these groups.

In summary we obtain that the smallest concentration of nitrogen dioxide occurred on holidays (i.e. the first group). The largest concentration on weekdays (i.e. the second group). Just before Christmas, during the departure of the holidays nitrogen dioxide concentration remained at a medium level. Therefore it can be hypothesized that the concentration of nitrogen dioxide depends on the volume of traffic on the road. For example knowledge of the phenomenon of variability of air pollution can help in the environmental policy of the city, in the planning of free communication or traffic restrictions.

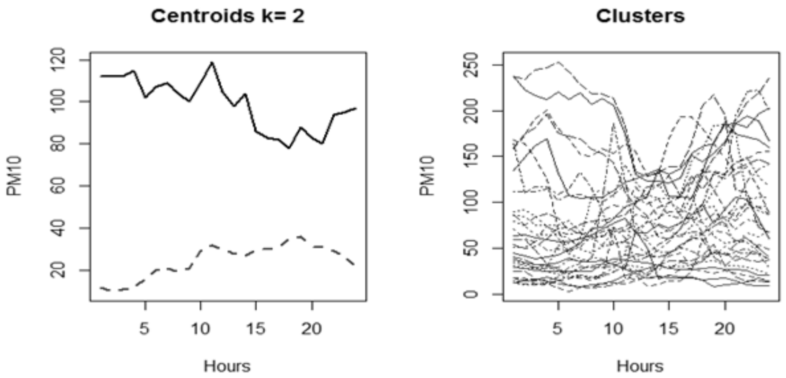


Fig. 4. Functional medians for individual clusters and assigning groups for nitrogen dioxide pollution emitted into the atmosphere as air pollution in Cracow, December 2016. Method – k-local functional median algorithm, k=2.

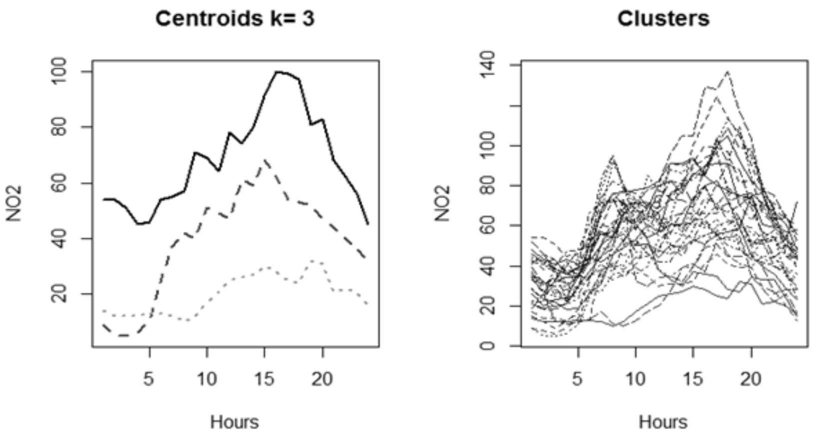


Fig. 5. Functional medians for individual clusters and assigning groups for particulate matter pollution in Cracow, December 2016. Method – k-local functional median algorithm, k=3.

Acknowledgements

DK thanks for the support related to CUE grant for the research resources preservation 2017.

References

Febrero-Bande, M., & de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software 51(4)*, 1-28.

Fritz, H., & García-Escudero, L. A., & Mayo-Iscar A. (2012). tclust: An R Package for a Trimming Approach to Cluster Analysis, *Journal of Statistical Software, 47 (12), 1–26.*

Horvath, L., Kokoszka, P. (2012). Inference for functional data with applications. *Springer.*

Kosiorowski, D., Mielczarek, D., & Szlachtowska, E. (2015). Clustering of functional objects in energy load prediction issues.*:IX International scientific conference in honour of Professor Aleksander Zelias : Zakopane, Poland, 12–15 may 2015.*

Kosiorowski, D., Szlachtowska, E., & Zawadzki Z. (2017). Selected clustering algorithms for functional data in specification of functional time series. *Submitted*

Kosiorowski, D., & Zawadzki, Z. (2014). DepthProc: An R package for robust exploration of multidimensional economic phenomena. *http://arxiv.org/pdf/1408.4542.pdf.*

Kosiorowski, D. (2016). Dilemmas of robust analysis of economic data streams. *Journal of Mathematical Sciences 1(2)*, 59-72.

Lopez-Pintado, S., & Romo, J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis 51*, 4957-4968.

Nagy, S., Hlubinka, D., & Gijbels, I. (2016). Integrated depth for functional data: Statistical properties and consistency. *ESIAM Probability and Statistics.*

Nieto-Reyes, A., & Battey, H. (2016). A topologically valid definition of depth for functional data. *Statistical Science 31(1)*, 61-79.

Paindaveine, D., & Van Bever, G. (2013). From depth to local depth: a focus on centrality. *Journal of the American Statistical Association 108(503)*, 1105-1119.

Ramsay, J., Hooker, G., & Graves, S. (2009). Functional data analysis with R and Matlab. *Springer.*

Walesiak, M., Dudek, & A. (2015). Searching for Optimal clustering procedure for a data set: the ClusterSim R package, https://CRAN.R-project.org/package=clusterSim.

Bayesian estimation of proportions in the analysis of unavailability of health care in Slovakia

Eva Kotlebová[1], Erik Šoltés[2]

Abstract

Availability of health care affects the quality of life of the population and unmet needs for health care is an important indicator of poverty or social exclusion. The paper deals with estimating the proportion of the population of the Slovak Republic who cannot afford necessary medical treatment mainly because of lack of money, but also for other reasons. In analysis, two sources of data were used: sample survey EU SILC and sample survey EHIS, both conducted by the Statistical Office of the Slovak Republic. Since these sources of information were simultaneously used, it was appropriate to apply Bayesian methods.

Keywords: *health care, EU SILC – European Union Statistics on Income and Living Conditions, EHIS – European Health Interview Survey, Bayesian estimation of the proportion*

JEL Classification: C83, I140

1 Introduction

One of the current problems the European Union is dealing with is to reduce the number of people at risk of poverty and social exclusion. To map the current state, the indicator AROPE has been developed. In addition to this indicator, an inaccessibility of health care primarily for financial reasons may be considered as a sign of poverty. It is known (see Barnay (2016)), that there exists causal relationship between health status and the probability of being employed, which strongly influences the social situation. There is no doubt that health depends (among other things) on health care (see Nováková and Chinoracká (2015)).

The level of health care in the EU is regularly monitored through a database of the World Health Organization (WHO), fixed annually by the European Health Consumer Index (EHCI). According to this criterion, the standard of health care in Slovakia has been on a declining trend since 2013. For 2016 the value of EHCI was 678, which ranks Slovakia at number 23 among the 35 European countries (according to Björnberg (2016)).

[1] Corresponding author: University of Economics in Bratislava, Faculty of Economic Informatics, Department of Statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: eva.kotlebova@gmail.com.

[2] Corresponding author: University of Economics in Bratislava, Faculty of Economic Informatics, Department of Statistics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, e-mail: erik.soltes@euba.sk.

We have dealt with those indicators of access to the health care, which result directly from the social situation of the recipients of healthcare: the inability to meet the need for health care for financial reasons, for a big distance of provider or too long waiting period. In Chaupain-Guillot and Guillot (2015), some other determinants of unmet needs for medical and dental care are analysed.) We used the database of sample surveys EU SILC and EHIS, in which indicators of unavailability of health care for the said reasons are mapped. To estimate the proportion of the population who suffer from a deficit in some of mentioned areas, we used the data from EHIS in Section 2, in which we also identified some of the factors affecting the degree of inaccessibility of healthcare. In Section 3, we specified the analysis by application of Bayesian methods. For modelling the prior distribution, data from EU SILC were used.

2 Analysis of the unavailability of health care based on a database EHIS

The legal framework for developing the European Health Interview Survey (EHIS) is the Regulation (EC) No 1338/2008 of the European Parliament and of the Council of 16 December 2008 on Community statistics on public health and health and safety at work. Detailed specification of the data and metadata to be provided is pursuant to the Commission Regulation (EU) No 141/2013 of 19 February 2013 (according to Eurostat, 2013). We used the part of the questionnaire UN (**U**nmet **n**eeds for health care), which included questions:

UN1a: Have you experienced delay in getting health care in the past 12 months because the time needed to obtain an appointment was too long?

UN1b: Have you experienced delay in getting health care in the past 12 months due to distance or transport problems?

UN2: Was there any time in the past 12 months when you needed the following kinds of health care, but could not afford it?

A. medical care

B. dental care

C. prescribed medicines

D. mental health care (by a psychologist or a psychiatrist for example).

(In following, we use the abbreviations UN1a, UN1b, UN2a, UN2b, UN2c, UN2d for the listed variables). For all questions, respondents could choose one of the three options: “yes”, “no” or “no need for care” (in UN2c the last option was “no need for prescribed medicines”).

Besides the above mentioned variables, we dealt with some other variables that might come into consideration as factors affecting the access to health care: sex, age, region, degree of urbanisation, educational attainment, self-declared labour status, status in employment, net monthly equivalised income of the household and self-perceived general health. For each of them, we conducted test for independence between them and the above listed variables from the group UN. Each test was performed twice: in the first case all three possible answers to the questions of the group UN were taken into account; in the second case we analysed only two different answers – "yes" or "not yes", so that we merged second and the third answers into one group. This was done with respect to create the indicator comparable with some of the indicators of EU SILC. No doubt, that the more valuable information is provided by the first case, as we can detect the proportion of those who could not meet needs for healthcare out of those who really needed it. Obviously, in the second case, the sample proportions are smaller: the difference depends on the number of respondents, who said they had not need the health care in particular questions (it does not hold for the question UN2d, which requires special analysis). To compare the different kinds of proportions for each question from the group UN, they are shown in Table 1.

Question	**UN1a**	**UN1b**	**UN2a**	**UN2b**	**UN2c**	**UN2d**
Proportion (1st case)	6.26 %	1.42 %	2.21 %	6.95 %	4.89 %	1.80 %
Proportion (2nd case)	4.83 %	1.06 %	1.68 %	4.95 %	3.39 %	0.29 %

Table 1. Proportions of respondents, who did not receive particular (according to the question) care out of those who needed the care (1st case) and out all respondents (2nd case). Source: EHIS 2014, authors´ own elaboration.

As we can see, the highest proportions relate to unavailability of dental care (for more details see Tchicaya and Lorentz (2014)) and too long waiting time (detailed analysis of this indicator is in Muzik and Szalayova (2013)).

It came out, that the results in detecting factors influencing proportions in both cases were not significantly different except that the last question UN2d, where in the first case the strong dependence appeared, while in the second case the p-values of the independence tests were mostly higher than 0.5. This may be caused by the content of the question – many people do not want reveal the need for mental care.

Considering the limited extent of the text, we only present the most interesting results of the first case (three different answers to questions from the group UN). Among the above listed factors, the strongest influence on the UN variables have self-perceived general health (HS1), self-perceived labour status (MAINSTAT), region (REGION), net monthly equivalised income of the household (HH INCOME) and educational attainment

(HATLEVEL). For all of them, the p-values of independence tests were less than 0.01 (for both cases). Table 2 shows the proportion of those respondents who said their health care needs have not been satisfied out of those who needed it, for each level of mentioned factors. According to http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-13-018, the meanings of factors´ variations[3] are as follows:

HS1 (self-perceived general health):

1 very good

2 good

3 fair

4 bad

5 very bad

MAINSTAT (self-perceived labour status):

10 carries out a job or profession, including unpaid work for a family business or holding, an apprenticeship or paid traineeship, etc.

20 unemployed

31 pupil, student, further training, unpaid work experience

32 in retirement or early retirement or has given up business

33 permanently disabled

35 fulfilling domestic tasks

36 other inactive person

REGION (according to NUTS II):

1 Bratislava

2 West Slovakia

3 Middle Slovakia

4 East Slovakia

HHINCOME (net monthly equivalised income of the household):

1 below 1st quintile

2 between 1st quintile and 2nd quintile

3 between 2nd quintile and 3rd quintile

4 between 3rd quintile and 4th quintile

5 between 4th quintile and 5th quintile

[3] Variations that did not occurred are not listed.

HATLEVEL (educational attainment based on ISCED-2011 classification):

0 early childhood development, pre-primary education

1 primary education

2 lower secondary education

3 upper secondary education

4 post-secondary but non-tertiary education

5 tertiary education; short-cycle

6 tertiary education; bachelor level or equivalent

In Table 2, the largest number of each row is highlighted. As we can see, the largest proportions are for each factor (except for REGION) in the most vulnerable categories: people with bad or very bad health, permanently disabled people or people fulfilling domestic tasks, people with the lowest income and people with the lowest educational attainment. All these attributes meets especially the Roma segment of the population, which accounts for approximately 7.5 percent of the population (more detailed analysis is in Bartošovič (2016) or Bobakova et. al. (2015)). As for the factor REGION, the results were surprising for us. Despite predictions that in Bratislava region live people with the highest income, it turned out that many of them cannot afford health care. As we can see from Šoltés and Gajdošík (2016), there exist some other variables that may cause such relation among regions.

Factor	**Variations of factor**						
HS1	**1**	**2**	**3**	**4**	**5**		
UN1a	3.41	4.08	8.38	*10.95*	6.29		
UN1b	0.00	0.78	1.71	2.89	*5.88*		
UN2a	1.23	0.91	2.47	5.22	*5.63*		
UN2b	3.84	4.83	8.90	*14.29*	9.09		
UN2c	1.88	2.29	4.71	*12.02*	10.69		
MAINSTAT	**10**	**20**	**31**	**32**	**33**	**35**	**36**
UN1a	5.76	7.33	2.60	6.92	9.74	*11.11*	4.22
UN1b	0.83	1.20	0.00	2.10	*3.66*	0.00	1.82
UN2a	1.06	6.31	1.11	2.23	*7.29*	2.94	1.85
UN2b	4.95	17.89	3.23	6.82	15.38	*20.00*	5.29
UN2c	2.27	13.99	1.38	4.97	*13.92*	9.38	4.17

REGION	**1**	**2**	**3**	**4**		
UN1a	7.00	5.75	4.21	*8.85*		
UN2a	*5.49*	1.48	2.06	1.83		
UN2b	*19.03*	4.11	5.05	6.75		
UN2c	*12.57*	1.98	5.36	4.56		
HHINCOME	**1**	**2**	**3**	**4**	**5**	
UN2a	*4.65*	2.65	1.80	1.32	0.74	
UN2b	*16.93*	6.81	5.32	3.65	3.48	
UN2c	*12.17*	4.10	4.09	2.54	1.81	
HATLEVEL	**1**	**2**	**3**	**4**	**5**	**6**
UN2b	*16.13*	12.01	6.62	9.30	4.31	3.50
UN2c	*14.71*	10.10	4.07	6.74	1.09	1.69

Table 2. Sample proportions (%) of respondents who´s health care needs have not been satisfied out of those respondents who needed health care.

Source: EHIS 2014, authors´ own elaboration.

3 Application of Bayesian methods for estimating the proportion of people, who are not receiving medical care for some selected reasons

Bayesian statistical inference is the alternative to the sampling-theory statistics. If there exists, besides the sample data, some additional information on the analysed variable, Bayesian approach can be used to make estimation more precise.

The main difference between the classical and the Bayesian approach is that the estimated parameter in Bayesian statistics is not an unknown constant, but it is considered as a random variable. Along with the sample data, there exists another piece of information, on the base of which the prior distribution of variable is created. Including sample data into analysis, the distribution transforms into posterior distribution, serving as a background for inference conclusions. If the prior and the posterior are of the same type, together with the sample distribution, they create *conjugate family*. For estimating the proportion π, the conjugate family binomial/beta is often used. In the case, the hyper-parameters of posterior distribution $Be(\alpha';\beta')$ satisfy (see Bolstad (2004))

$$\alpha' = \alpha + x \tag{1}$$

$$\beta' = \beta + n - x \tag{2}$$

(α, β are hyper-parameters of prior distribution and x denotes number of occurrence of event out of n trials). The Bayesian point estimation of parameter π is the posterior mean:

$$E(\pi / \mathbf{x}) = \frac{\alpha + x}{\alpha + \beta + n}. \tag{3}$$

Our goal was to estimate the proportions of people, who cannot receive the medical care, more precisely than it was in Section 2. In order to do so, we used as a prior information data from the P-file (Personal Data) of sample survey EU SILC, in which some of the questionnaire´s questions are similar to the questions in the survey EHIS. This refers to following questions:

PH040: Unmet need for medical examination or treatment,

PH050: Main reason for unmet need for medical examination or treatment,

PH060: Unmet need for dental examination or treatment,

PH070: Main reason for unmet need for dental examination or treatment.

The respondents had to choose answers “yes” or “no” to the questions PH040 and PH060 and one following options to questions PH050 and PH070:

1 Could not afford to (too expensive)

2 Waiting list

3 Could not take time because of work, care for children or for others

4 Too far to travel/no means of transportation

5 Fear of doctor/hospitals/examination/ treatment

6 Wanted to wait and see if problem got better on its own

7 Didn’t know any good doctor or specialist

8 Other reasons

Although the questions look similar to those in the EHIS questionnaire, their exact formulation and cross connection divided respondents into different groups than it is in EHIS. After logical analysis, we managed to develop an indicator that has the same content in both surveys: We monitored the proportion of respondents (out of all respondents) who did not meet the need for medical care for one of the three reasons: “Could not afford to” (1), “Waiting list” (2) and “Too far to travel/no means of transportation” (4). Each of them has strong relation to social situation of the respondent.

To create the prior distribution, we used data from EU SILC 2005 – EU SILC 2015 for evaluating corresponding proportions; they are shown in Table 3.

On the base of evaluated proportions, we created the prior beta distribution. By maximum likelihood method we estimated parameters $(\alpha = 15.466, \beta = 746.788)$ and performed Kolmogorov-Smirnov test to verify the appropriateness of the distribution. Since the p-value was more than 0.99, the distribution $Be(15.466; 746.788)$ fitted well and might be used as a prior. In the data from EHIS, there were 297 respondents ($x = 297$) out of all 5490 $(n = 5490)$, who had not met need for health care for one of the mentioned reasons. Thus, according to (1) and (2) we evaluated parameters of posterior distribution $Be(\alpha'; \beta')$: $\alpha' = 312.466$, $\beta' = 5939.788$, so according to (3) the Bayesian point estimation of the proportion is $\hat{\pi}_B = 0.0500$. This value is smaller than the point estimation based only on sample data (EHIS) 0.0541 and bigger than prior mean 0.0203. It means, that according to the data from survey EHIS, 5.41 % out the whole population is not provided medical care for one of mentioned reason. The estimation of proportion based only on the prior information (EU SILC) is smaller (2.03 %) and when both surveys are used, we get the value 5.00 %.

YEAR	(1)	(2)	(4)	(1)+(2)+(4)	SAMPLE SIZE	PROPORTION
2005	324	44	24	392	12868	0.0305
2006	278	42	33	353	12630	0.0279
2007	109	46	18	173	12570	0.0138
2008	67	71	41	179	13645	0.0131
2009	84	81	60	225	13580	0.0166
2010	80	101	47	228	13907	0.0164
2011	89	149	43	281	13261	0.0212
2012	112	135	35	282	13502	0.0209
2013	92	135	23	250	13044	0.0192
2014	114	130	32	276	13187	0.0209
2015	102	154	52	308	13535	0.0228

Table 3. Proportions of respondents with unmet need for (at least) one of the reasons: “Could not afford to”, “Waiting list” and “Too far to travel/no means of transportation”.
Source: EHIS 2014, authors´ own elaboration.

Let´s compare interval estimations (on confidence level 95 %): The classical confidence interval based on normal approximation, is (4.81 %; 6.01 %). Prior beta distribution gives interval (1.15 %; 3.14 %) and Bayesian credible interval is (4.47 %; 5.55 %). We may conclude that on confidence level 95 %, for more than 4.47 % up to 5.55 % of population the medical care is not available for lack of money, for too big distance from provider of medical care or for too long time they have to wait for the care.

Conclusion

One of the signs of poverty is unavailability of the health care for (mainly) financial reasons, but also for other reasons, which are related with social situation of people. By analysis of data from survey EHIS 2014 we estimated, that 6.26 % of those, who needed health care, don´t receive it because of too long distance from provider, 1.42 % because of long waiting period, 2.21% because they cannot afford it. Lack of money is reason, why 6.95 % of population do not receive dental care, 4.89 % cannot receive prescribed medicine and 1.80 % do not receive mental care, despite they need it. Analysis of factors influencing unavailability of health care shows, that the groups of people who are the most threatened by unmet need for health care (as well as dental care or mental care) or they cannot afford the prescribed medicines, are the most vulnerable groups: people with bad health, permanently disabled people, people with the lowest income and people with the lowest educational attainment. In many cases (see Table 2) the proportion of such people is bigger than 10 %. This is really a serious situation and the government should design such arrangements that would help to these most vulnerable groups.

In order to make estimations more precise, we analysed another source of similar data – P-files from EU SILC which was used as a prior information for Bayesian estimation. Unfortunately, the formulations of the questions are slightly different from those of survey EHIS, so we could only analyse the proportion of people, who cannot receive medical care for one of three reasons: lack of money, to long distance or too long waiting time. As the proportions in EU SILC were smaller than the proportions evaluated from EHIS, the Bayesian estimation was smaller (5.00 %), than the classical point estimation on the base of EHIS (5.41 %). If we realize that Slovakia has about 5 400 000 people, of which 84.7 percent has 15 or more years (the age of EHIS´s respondents), estimated share indicates that more than 220 000 inhabitants have no access to health care for mentioned reasons. This is a serious fact deserving attention of the competent authorities.

Acknowledgements

Realization of the paper was supported by a grant agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic VEGA. Project: VEGA no. 1/0548/16 *The progress of the SR in achieving of the strategy Europe 2020 in the area of poverty and social exclusion*.

References

Barnay, T. (2016). Health, work and working conditions: a review of the European economic literature. *The European Journal of Health Economics, 17*(6), 693-709.

Bartošovič, I. (2016). Some aspects of health status of the Gypsy population in Slovakia. *Bratislava Medical Journal - Bratislavské lekárske listy, 117*(1), 26-30.

Björnberg, A. (2016, January 26). Euro Health Consumer Index 2015. Retrieved February 8, 2017, from http://www.healthpowerhouse.com/media/prerelease/EHCI_2015_media.pdf.

Bobakova, D., Veselska, Z. D., Babinska, I., Klein, D., Geckova, A. M., & Cislakova, L. (2015). Differences between Roma and non-Roma in how social support from family and friends helps to overcome health care accessibility problems. *International Journal for Equity in Health, 14*.

Bolstad, W. M. (2004). *Introduction to Bayesian statistics*. Hoboken (New Jersey), USA: John Wiley & Sons.

Chaupain-Guillot, S., & Guillot, O. (2015). Health system characteristics and unmet care needs in Europe: an analysis based on EU-SILC data . *European Journal of Health Economics, 16*(7), 781-796.

Eurostat. (2013). *European Health Interview Survey (EHIS wave 2) - Methodological manual* (Working paper No. KS-RA-13-018-EN-N).

Muzik, R., & Szalayova, A. (2013). Measuring Waiting Times in Slovakia . In *Proceedings of the 17th International Conference on Current Trends in Public Sector Research* (pp. 66-74). Slapanice, Czech Republic: Masarykova Univ.

Nováková, M., & Chinoracká, A. (2015). Regional disparities in working life quality as a factor of the health of the Slovak population. In *Proceedings of 12th International Scientific Conference: Economic Policy in the European Union Member Countries, PTS I ans II* (pp. 597-605). Karvina, Czech Republic: Silesian Univ Opava, School Business Administration Karvina.

Šoltés, E., & Gajdošík, M. (2016). Dopad revízie deprivačných položiek na hodnotenie materiálnej deprivácie slovenských domácností na základe databázy EU SILC 2014. *Ekonomika a informatika, 14*(2), 178-196. Retrieved February 8, 2017, from http://ei.fhi.sk/index.php/EAI/article/view/75/106.

Tchicaya, A., & Lorentz, N. (2014). Socioeconomic inequalities in the non-use of dental care in Europe. *International Journal for Equity in Health, 13*.

Statistical analysis of trends and factors in cancer incidence of Ukrainian women in reproductive age

Natalya V. Kovtun[1], Zinaida O. Palian[2], Igor M. Motyzuk[3]

Abstract

The dynamics of reproductive organ cancer of women in Ukraine and their survival in the context of radiation effect of Chronobyl nuclear power station was analyzed in this paper. The attempt was made to evaluate the volume of indirect demographic losses due to cancer incidence of women of reproductive age.

Statistical data of All-Ukrainian cancer institute, National cancer-registry of Ukraine for the period of 1981-2015 and State statistics service of Ukraine was used in the research.

Keywords: *survival analysis, t-test, Levene's criterion*

JEL Classification: C120, C 430, I190

1 Introduction

Formulation of the problem. Each historical epoch is characterized by a set of factors which regulates the number and structure of the population. Life expectancy of current mankind depends to a great extent on endogenous factors, diseases of non-infectious nature, namely cancer ones are among them. In Ukraine, as well as in most industrialized countries, the main cause of women's mortality from cancer diseases is cancer of reproductive system organs. Besides direct losses of women, including fertility cohort, indirect losses occur – non-born children because of permanent or temporary loss of reproductive function as a result of the disease of this location. In view of this, it is important to study the trend of cancer incidence of reproductive organs of women in Ukraine and their survival, in particular of the consequences of the disaster of Chornobyl nuclear power station (ChNPS), and to determine the volume of indirect demographic losses due to cancer incidence of women of reproductive age.

Analysis of recent research and publications. As world scientific researches prove, in recent years there has been an increase in the frequency of multiple primary malignant

[1] Corresponding author: Taras Shevchenko National University of Kyiv, Statistics and Demography Department, Vasylkivska Street, #90. 03022 Kyiv, Ukraine, e-mail: kovtun_natali@ukr.net

[2] Taras Shevchenko National University of Kyiv, Statistics and Demography Department, Vasylkivska Street, #90. 03022 Kyiv, Ukraine, e-mail: zipalyan@ukr.net.

[3] Bogomolets National Medical University, Oncology Department, Lomonosova Street, #33/43. 03022 Kyiv, Ukraine, e-mail: 180978igor@gmail.com

growths, when two or more separate primary malignant growths, which originate from different organs, appear in the organism at the same time or one after another (Merimsky et al., 2001). Multiple primary malignant growths belong to the development of malignant affect de novo, their characteristics totally differ from primary tumor (Xu & Gu, 2014). Breast cancer (BC) constitutes 30-50% of all other primary malignant growths of women who had primary breast cancer (PBC). The risk is 2-6 times more that they will have new primary cancer in contra-lateral breast during their life time (Levi et al., 2003; Jobsen, et al., 2003; Heron, et al., 2000), which constitutes 0.3-1% annually (Jobsen, et al., 2003; Lee, et al., 1999). In the literature, there is uncertainty as to the prediction and survival of the sick with metachronous cancer processes compared with primary one-side breast cancer. In view of this, regularities of the occurrence of metachronous processes of the patients with breast tumors are studied in this paper, their consequences are estimated and predictions are suggested.

The development of multiple processes can be connected with radiation contamination as a result of the disaster at ChNPS. Extra radiation-induced cases of PBC were recorded among irradiated woman after atomic bombing of Hiroshima and Nagasaki (Xu & Gu, 2014).

The development of the cancer location depended on the age of this population group at the moment of irradiation. Higher tumor incidence was observed among the individuals irradiated at the age of 10-19. PBC incidence was lower among the irradiated people at the age of 20. Average latent period, followed by BC, did not depend on the dose and was about 18 years. World literature confirms that BS among women irradiated at the age 19-20 was higher than among those older than 30 years old (Levi et al., 2003).

The purpose of the article is to study the dynamics of cancer incidence of reproductive organs of Ukrainian women and their survival in the context of the disaster at Chronobyl nuclear power station, and to estimate the volume of indirect demographic losses due to cancer diseases of women of reproductive age.

Methodology. Methodological basis of the research includes methods of descriptive statistics and specific methods of statistical analysis which helped make statistic conclusion (Bongers et al., 2016; Palian, 2014). To get general characteristic of the patients’ survival, that constituted the basis of the sample, 2032 female-patients were included; they received specific treatment (the second surgery) in the department of breast tumor and its reconstructive surgery of the National cancer institute from the year of 2008 till December, 2015 within the randomized controlled open research aimed at studying the criteria of objective choice in surgery scope of the patients with BC. A separate group consisted of 195

(9.6%) female-patients with multiple malignant growths. The patients were divided into 4 groups: 1) with synchronous breast cancer (SBC); 2) with synchronous breast cancer and other location (SBCO); 3) with metachronous breast cancer (MBC); 4) with metachronous breast cancer and other location (MBCO).

Out of sample (195 women) by 107 patients (54.9%) experienced synchronous processes. The distribution of synchronous processes as to localization was as follows: 56% – breast cancer and 44% – combination of breast cancer with other localizations. 45% of the patients had metachronous processes, 46.6% had localization in breast, and the rest – was the combination of nosologies.

The analytical foundation of statistic conclusion contained special method of in-depth statistical analysis. Thus, a survival function, which characterizes the probability that an individual will live through the time longer than “t”, has been built to analyze survival using incomplete (censored) data. In addition, the time tables of life expectancy were developed, the line-up adjustment of the distribution of survival with simultaneous estimation of the parameters of a survival function (Kaplan-Meyer procedure) was made to compare survival in two and more groups.

To statistical test the hypothesis about the variation among average formation periods of metachronous cancer processes before and after the year of 2008, Student’s t-criterion was used; to check the hypothesis about the variation of their dispersions, Leven’s criterion and Brown-Forsyth criterion were used, they were more stable in case of the availability of possible deviations from standard distribution.

A standard data model was made in EXEL, and an analytical data model – in STATISTICA.

2 Statistical analysis of the dynamics of women’s cancer incidence in Ukraine

In current structure of cancer diseases of women in Ukraine, breast cancer takes the first place (19.5%) which causes 20% mortality of all malignant growths. Reproductive organ cancer (corpus uteri and cervix) is among six mostly widely-spread nosological forms of women’s cancer. However, the intensity of these diseases among Ukrainian women is almost the same as that in the post-soviet European countries, and the level of breast cancer (41.7 per 100.00 women, 2014 in Ukraine) is even lower than in some industrialized European countries, such us United Kingdom (95 per 100.00 women, 2012), Germany (91.6 per 100.00 women, 2012), France (89.7 per 100.00 women, 2012) (*Cancer Today*...). The latter can be connected with the fact that more women are covered with preventive examination and better quality medical diagnostics in these countries.

For a short period of time (2003-2015) the level of cancer incidence of women of reproductive age was constantly growing in Ukraine. The frequency of preliminary breast cancer increased more than by 9% (0.8% annually), reproductive organs depending on localization by 7.5% (ovary cancer) and by 21.3% (corpus uteri), respectively. However, the largest number of malignant growths identified for the first time included breast cancer. The coefficient of disease incidence in this localization exceeded three times similar indicators of the three main localizations of a reproductive system (cervix cancer, corpus uteri, ovary). As one can see from Fig. 1, the coefficients of malignant growths of reproductive organs, standardized by world age structure, show enhanced increase, beginning from the year of 2009.

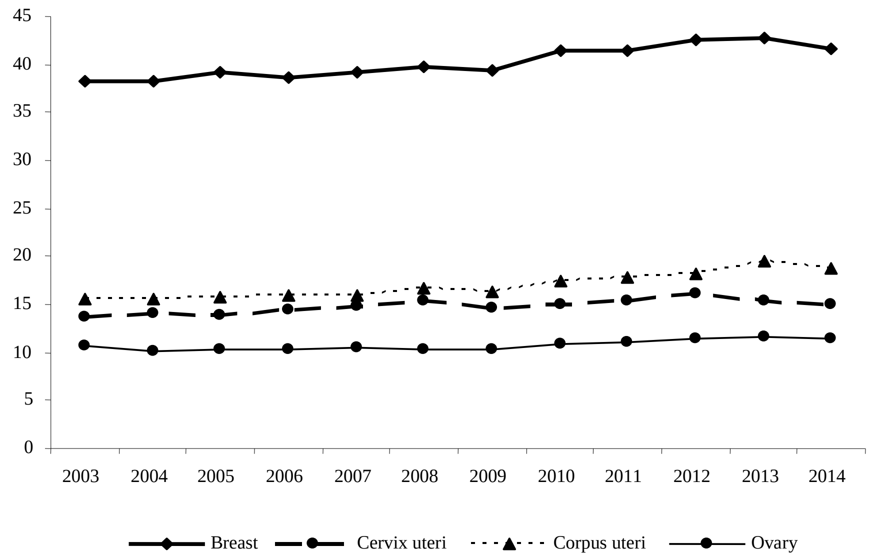


Fig. 1. Age-standardized (world standard) cancer incidence rates of female's reproductive system by sites in Ukraine (per 100,00), 2003-2014 (source: data From the Bulletin of National Cancer Registry of Ukraine).

Mostly this concerns breast and corpus uteri cancer. In 2014 the number of new incidents (per 100.00 women) decreased slightly in every localization due to the fact that the registration of women in annexed and temporarily occupied areas, traditionally with high incidence indicators of the mentioned nosological cancer forms, stopped (the Republic of Crimea, Sevastopol, Donetsk and Luhansk regions).

The indicators of disease incidence signal not only about potential mortal losses of women as a result of malignant growths. As the main nosological cancer forms of women are those of reproductive system and breast, the issue of indirect demographic losses arises – the number of unborn children because of permanent and continuous disorder of reproductive functions. According to the data of the National cancer-registry of Ukraine, during the last 10 years the highest intensity of breast cancer and that of corpus uteri and cervix was observed among women of older than reproductive age (60-69 years old). At the same time in 2004-2014 age structure of malignant growths among reproductive female-population got worse slightly. For instance, in the structure of breast cancer incidence, the proportion of young and average-age groups (25-39 years old) increased by 4 p.p.; they have shown high birth activity in recent years. Women of an early age (15-19) and those of an early reproductive age (25-29) contributed to ovary cancer incidence, its indicator increasing by 3.7 p.p. However, malignant growths of the mentioned organs do not result in permanent loss of reproductive function, which can be restored provided proper treatment is given. On the contrary, other two localizations (corpus uteri and cervix cancer) lead to permanent loss of the childbearing ability. In Ukraine, 52.8% of the sick with cervix cancer are women of average and older reproductive age (35-39 and 40-44 years old) and fertility of this age-groups increased by 2.3 and 2.5 times in 2004-2014, respectively. Totally, in 2004-2014 the annual number of new incidents of reproductive organ cancer increased by 3.5%, and within the compared territory of Ukraine (including annexed and temporarily occupied regions) this indicator increased by 4% in 2003-2013, i.e., 0.39% annually. The question arises: how big indirect demographic losses can be – the number of unborn children because of new cases of malignant growths among women of a reproductive age. To find the answer, a multiple index model of hypothetic number of unborn children was built, taking into consideration cancer incidence of women at a reproductive age, their age-specific fertility rates and age structure for the years of 2004-2014.

Table 1 shows a relative change of hypothetic number of unborn children taking into account new cases of malignant growths of all nosological forms, concerning a reproductive system, including breast cancer, and also excluding it, i.e. only reproductive organs.

In 2004-2014 total hypothetic losses of child birth due to unrealized reproductive functions by cancer-sick women increased by 89.2. If to exclude women diagnosed with breast cancer whose reproductive ability can be restored after long-term sever therapy, the scope of hypothetic losses would be increased by 80.6%. However, such high growth rates are mostly associated with general tendency of age fertility increase, typical for all Ukrainian women (+62.9%); it is during the last 11 years that reproductive activity of women at the age of 30-40 has increased by two times. The change in the age structure of a women's reproductive cohort caused the increase of potential child-birth and also hypothetic losses (+2%). Primary cancer incidence of women at a reproductive age led to the increase of hypothetic losses by 11.8% in general, and with permanent loss of reproductive functions because of cancer of a reproductive system the number of unborn children increased by 8.7%. However, a mentioned index model enables to make rough estimation of hypothetic losses of child birth based on the statistics of the annual number of identified new cases of disease. Unfortunately, the lack of the information about age structure of all registered cancer-sick women of a reproductive age does not allow estimating a true number of unborn children. While estimating the losses of child birth, it is advisable to consider age survival probability of the women with malignant growths of reproductive organs. Hence, the next step is to identify the survival parameters of cancer-sick women of a reproductive age.

Changes of hypothetic number of unborn children	**With all nosological forms of reproductive system cancer**	**Without breast cancer**
Total	189.2	180.6
by factors:		
age cancer incidence rates	111.8	108.7
age –specific fertility rates	165.8	162.9
female's age structure	102.1	102.0

Table 1. Dynamics of hypothetical number of unborn children resulted from cancer incidence of women of a reproductive age in Ukraine in 2004-2014, % (source: author's own calculations based on data From the National Cancer Registry of Ukraine).

3 The estimation of women's survival in Ukraine among the sick with malignant growths of reproductive organs

Modeling the survival character of cancer-sick women requires preliminary estimation of their survival, taking into account the fact that patients have synchronous (SMPMNs) and metachronous (MMPMNs) processes of cancer development.

According to the data of sample studies, the probability to live another 125 months for patients on SMPMNs was 0.73, whereas for patients on MMPMNs it was 0.92 ($p = 0.00162$). Thus, disease development of the patients with synchronous cancer is more aggressive and diagnostically unfavorable.

When analyzing time interval between the occurrence of the first and the second disease, it was found out that median of the patients with the first diagnosis of BC before 2009 was 605 weeks, and after 2009 it was decreased to 104 weeks (p = 0.000001). The interval decrease between the occurrences of the second disease after 2009 is explained by the fact that the average age of BC patients was 30.5 before 2009 at the time of the disaster at Chornobyl nuclear power station, after the year of 2009 – 25.5 (p=0.000798). The presented data can confirm more aggressive disease development of BC patients who received radiation at the age of 30 when the disaster happened at ChNPS.

It has been statistically proved that recent years show the decrease of the period between the first and the second disease occurrence (Table 2). Thus, the period between diagnoses when identifying the first incidence before 2008 was 605 weeks (more than 11 years), and beginning from 2008 it decreased by six times and was 94 weeks (1 year and 9 months). Such drastic changes are the grounds for unfavorable prognoses for potential enhancement of the mentioned processes in Ukraine in the near 10-15 years.

Characteristics	**Mean, weeks**		**A Number of patients**		**Standard deviation, weeks**	
	group 1*	**group 2***	**group 1***	**group 2***	**group 1***	**group 2***
Period between diagnoses	605.0**	93.9**	74	16	349.6	105.7

*) group 1 – till 2008, group 2 – 2008 and later;

**) p=0.000000 for Student's criterion, p=0.000008 for Fisher's criterion, p=0.000616 for Leven's criterion, p=0.000221 for Brown-Forsyth's criterion.

Table 2. Duration between diagnoses of patients depending on the moment of the first diagnosis (before 2008 not including it – group 1 and after 2008 including – group 2; source: own calculations based on statistics Ukrainian Institute of Cancer).

The described situation definitely influenced the patients' survival (Fig. 2). It was statistically proved (Table 3), that the probability to live through another 39 months after 101 months for the group of patients whose first diagnosis was made before 2008 was 0.959, and for those whose first diagnosis was made beginning from 2008 – 0.552.

Period when the first diagnosis was made	Metachronous processes	
	Number	Observation time 101 month
Before 2008 (group 1)	74	0.959*
2008 and later (group 2)	16	0.552*

*p=0.03232.

Table 3. Survival of patients depending on the time when the first diagnosis was made (before 2008 not including it – group 1 and after 2008 including – group 2; source: own calculations based on statistics Ukrainian Institute of Cancer).

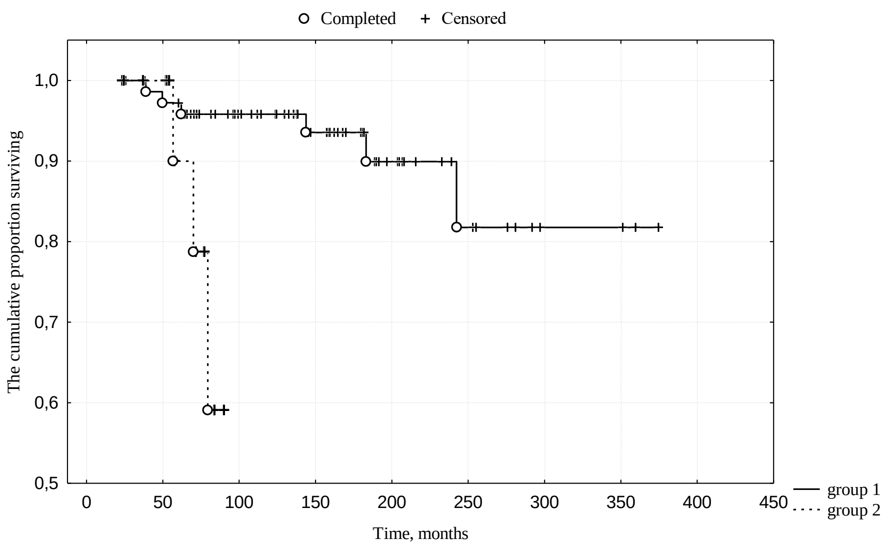


Fig. 2. Cumulative share of survived patients depending on the moment when the first diagnosis was made according to Kaplan-Meyer (before 2008 not including it – group 1 and after 2008 including – group2).

Apparently, the prediction that the situation with cancer incidence in the near 20-25 years will worsen appears to be true. However, the cause of the aggravated situation is to be searched in a complex effect of the factors: not only the consequences of the disaster at Chornobyl nuclear power station, but also worsening of the characteristics of life quality which can be seen in decreasing living standards, environmental pollution and poor health care system. All this taken together leads to the situation when a person is in permanent continuous stress which results in cancer incidence, in particular if someone is inclined to it (the availability of the first cancer diagnosis).

Besides, a more detailed disperse analysis between the periods of cancer incidence in two groups of the sick confirms that these two groups of the sick is a sample of two different general complexes which are characterized by variations. The exceeding of the variation coefficient by 100% in the group of the sick whose diagnosis was made in 2008 and later proves the availability of latent factors of subjective order which will further classify disease occurrence of Ukraine's population into primary and secondary cancer incidence.

Conclusions

Malignant neoplasms of the female reproductive system not only leading to death of women in fertility age, but also deprive them the childbearing opportunity. To assess the extent of the hypothetical loss of births, we should have proper information on the number of patients who are registered, their age distribution and the probability of surviving to next age-group in the context of each malignancies nosology.

For the last 25 years the frequency of recurrences increased and the interval between the first and the second diagnoses of reproductive organ cancer decreased by six times. For instance, in 1981-2008 the period between the first and the second disease lasted longer than 11 years, then after 2008 and till now it reduced to 2 years on the average. With the help of a survival function, it was established that the probability to live through another 3 years after 8.5 years from the moment the diagnosis of the first cancer disease was made before 2008 was 0.959, and after 2008 – only 0.562. More aggressive development of breast cancer was recorded among women who were 30 years old and were irradiated at the times of the disaster at ChNPS. With probability $p=0.000798$ it was proved that the formation factors of patients' survival before and after 30 differed greatly, which requires separate studying. We can not state for sure that the patients'age at the time of the disaster at ChNPS impacts their survival, as it is described in scientific literature mentioned before. The age is a formation criterion of two clusters of women as to survival, and statistic significance confirms the hypothesis that cause-effect factors of survival differentiation of these two groups of women have absolutely different nature.

References

Bongers, ML., de Ruysscher, D., Oberije, Cary, C., & Lambin, P. (2016). Multistate statistical modeling: a tool to build a lung cancer microsimulation model that includes parameter uncertainty and patient heterogeneity. *Medical decision making*, *36*(1), 86-100.

CANCER IN UKRAINE, 2014-2015. (n.d.). Retrieved March 23, 2017, from http://www.ncru.inf.ua/publications/BULL_17/index_e.htm

Cancer today. (n.d.). Retrieved March 23, 2017, from http://gco.iarc.fr/today/fact-sheets-cancers

Heron, D. E., Komarnicky, L. T., Hyslop, T., Schwartz, G. F., & Mansfield, C. M. (2000). Bilateral breast carcinoma. *Cancer, 88*(12), 2739-2750.

Jobsen, J., Palen, J. D., Ong, F., & Meerwaldt, J. (2003). Synchronous, bilateral breast cancer: prognostic value and incidence. *The Breast, 12*(2), 83-88.

Lee, M. M., Heimann, R., Powers, C., Weichselbaum, R. R., & Chen, L. M. (1999). Efficacy of Breast Conservation Therapy in Early Stage Bilateral Breast Cancer. *The Breast Journal, 5*(1), 36-41. doi:10.1046/j.1524-4741.1999.005001036.x

Levi, F., Randimbison, L., Te, V., & Vecchia, C. L. (2003). Prognosis of bilateral synchronous breast cancer in Vaud, Switzerland. *The Breast, 12*(2), 89-91.

Merimsky, O., Kollender, Y., Issakov, J., Bickels, J., Flusser, G., Gutman, M., Meller, I. (2001). Multiple primary malignancies in association with soft tissue sarcomas. *Cancer, 91*(7), 1363-1371.

Palian, Z. (2014). Statistical analysis and simulation of the parameters of survival. *Proceedings of the 8th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena*, 125-134.

Xu, L., & Gu, K. (2014). Clinical retrospective analysis of cases with multiple primary malignant neoplasms. *Genetics and Molecular Research, 13*(4), 9271-9284.

Spatial differences in sustainable development components in Nordic regions

Marta Kuc[1]

Abstract

Geographical proximity, common historical roots and collaboration within the Nordic Council make the Nordic countries, often wrongly treated as monoliths. However, in reality, Nordic regions differ in terms of broadly defined socio-economic development. The aim of this study is to analyze the spatial differences in sustainable development components in Nordic NUTS-3 regions in the period 2006-2014. Each sustainability dimension is measured using Pietrzak's spatial taxonomy measure of development. Analyzed problem seems to be important since Nordic countries are currently implementing the Fourth Nordic Strategies for Sustainable Development. The analysis showed that the achievement of sustainable development goals manage much better in the Swedish and Norwegian regions. The analysis showed as well that the Nordic countries are not as homogeneous as it might seem at first glance

Keywords: *sustainable development, spatial taxonomy measure of development, Nordic counties*

JEL Classification: Q01, C10, C43

1 Introduction

The concept of sustainable development links socio-economic issues with environmental problems. A widely-used definition of sustainable development is that one presented in World Conservation Strategy Living Resource Conservation for Sustainable Development (1980): 'development that meets the needs of the present without compromising the ability of future generations to meet their own needs'. According to this definition sustainable development contains two main concepts: the concept of needs and the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs. Starting from 1992 when the first United Nation Conference on Environment and Development (UNCED) was held in 1992 in Rio de Janeiro, the sustainable development concept was widely discussed in world conferences not only in Johannesburg in 2002 and in Rio de Janeiro in 2012, but furthermore in many other smaller conferences and meetings (Sustainable Development Timeline, 2012). Currently one can still see growing international interest in green sustainable economy, even thought the concept has evolved over time (Du Pisani, 2006).

[1] Gdansk University of Technology, Faculty of Management and Economics, ul. Traugutta 79, 80-233 Gdańsk, e-mail: marta.kuc@zie.pg.gda.pl.

The main goal of this study is to analyze the spatial differences in sustainable development components in Nordic NUTS-3 regions in the period 2006-2014. It is worth mentioning here what Nordic countries are: Nordic countries consist of not only Scandinavian countries (Denmark, Norway and Sweden) but as well Finland, and autonomous territories Greenland, Åland and the Faroe Islands. Nordic regions have been selected as object of interest for several reasons. The first reason is the fact, that since many years Nordic countries are leaders in the worldwide sustainable development rankings (Country Sustainability Report, 2013; SDG Index & Dashbords. A global report, 2017). The second reason is the fact, that Nordic countries cooperate in the framework of Nordic Council and Nordic Council of Ministers which are responsible for the agreements within the Nordic countries and the pursuit of the sustainable development of associated regions. The third reason is the fact that, due to their geographical proximity and common historical roots, the Nordic countries are often wrongly treated as unity. However, in reality, different regions of the Nordic countries are diverse in terms of widely understood socio-economic development. Author of this article want to show that even thought Nordic countries are leaders in sustainable development and they have common sustainability policy, there are still substantial differences among Nordic NUTS-3 regions. Therefore the simplification to perceive that all regions have equally strong sustainability level is misleading.

To analyze the spatial differences in sustainable development dimensions in Nordic NUTS-3 countries the spatial taxonomy measure of development (Pietrzak, 2014) was used. It seems that the inclusion of spatial relationships is justified because nowadays no region develops in isolation. Therefore, the situation in each region is influenced by neighbourhood. The impact of geographic surroundings is especially noticeable in the ecological dimension of sustainable development; it can be noted here quite obvious example of air pollutions that do not recognize administrative boundaries and move to the neighbouring regions along with the air movements.

The analysis was conducted for 67 NUTS-3 regions of Nordic countries (excluding: Höfuðborgarsvæði, Landsbyggð, Grønland, Føroyar, Åland, Gotland and Bornholm) in 2006-2014 period. Empirical material was taken from the national statistical offices of analysed countries.

2 Sustainable development discourses in Nordic countries

Governments of Nordic countries give high priorities to reduce pressure on the natural environment while pulling out millions of people out of poverty. It seems that sustainable development and development of the Nordic welfare model go hand in hand. Denmark, Finland, Norway and Sweden not only have their national sustainability programmes but additionally collaborate with each other on that field. In 1952, Denmark, Iceland, Norway and Sweden formed the Nordic Council, which was later joined by Finland and also by the autonomous territories Greenland, Åland and the Faroe Islands. In 1962, the Nordic countries signed the so-called 'Helsinki Treaty' (The Helsinki Treaty, 1962), which regulates cooperation between them.

The importance of sustainability issues expresses the fact that the Nordic countries are currently implementing a fourth strategy for the sustainable development of the Nordic region (A Good Life in a Sustainable Nordic Region. Nordic Strategy for Sustainable Development, 2013) which complements national sustainable development strategies. The time frame of this strategy covers the period up to 2025. In this strategy, the emphasis is on: stable, healthy and sustainable economic growth; cooperation leading to higher employment and reduction of structural unemployment; healthy and decent life on inhabitants; elimination of poverty and trafficking in human rights; strengthening the role of culture in sustainable development. Financing the implementation of fourth sustainable development strategy falls under the budget of the Nordic Council of Ministers. (Sustainable Development. New Bearings for the Nordic Countries, 2001).

The first Nordic strategy for sustainable development was adopted 16 years ago. However it was initiated by the prime ministers declaration in 1998. Thus it can be seen that the issues associated with the combination of sustainable development and welfare model have quite long traditions in the Nordic countries.

3 Empirical analysis using spatial approach

To analyze spatial differences in sustainable development among Nordic NUTS-3 regions taxonomy spatial measure of development was used (Pietrzak, 2014). There are several reasons to include spatial factors into this kind analysis. Firstly, according to Waldo Tobler: 'Everything is related to everything else, but near things are more related than distant things' (Tobler, 1970). Secondly the use of a regional dataset implies consideration of the possibility that observations may not be independent, as a result of the inter-connections between neighbouring regions (Buccellato, 2007). Thirdly, it is better to use the simplest weight matrix than assume the independence in advance (Griffith, 1996). Fourthly, the diversification of economic phenomena in an established group of regions is highly affected by the spatial conditions (Pietrzak et al., 2014a; Pietrzak et al., 2014b). In polish literature one can find different attempts to include spatial factor into taxonomy measure of development constriction (Antczak, 2013; Pietrzak, 2014, Sobolewski et al., 2014; Pietrzak, 2016). Pietrzak's (2014) proposition has the advantage over alternative methods that in that approach it is necessary to test spatial autocorrelation for each variable and it is possible to diversify the potential power of spatial interaction for each variable. There are different approaches to analyse sustainable development but as a research shows a synthetic measure is still a good solution (Vu et al., 2014; Balcerzak and Pietrzak, 2016; Pietrzak and Balcerzak, 2016).

To analyze the sustainable development in Nordic NUTS-3 regions a following set of variables was used (S - stimulant, D - destimulant):

- Social dimension: life expectancy at birth (S), upper-secondary attainment in total population (S), research and development expenditures as percentage of GDP (S), urbanisation (S).
- Environment dimension: air pollution emission (D), forestry level (S), share of renewable energy in gross energy supply (S), greenhouse gas emission (D), gross energy consumption (D)
- Economic dimension: unemployment rate (D), gender pay gap (D), Gini-coefficient (D), risk of poverty among families with children (D), GDP per capita (S).

Based on variables presented above the spatial taxonomy measure of development was calculated for each dimension of sustainable development. The taxonomy spatial measure of development according to Pietrzak (2014) was calculated as follows:

1. Testing the presence of spatial autocorrelation using Moran's I statistics:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (i = 1,.., n;\ j = 1,..., n) \qquad (1)$$

where: I - the value of Moran's I statistics; n - number of observations; w_{ij} - spatial weight matrix; x_i, x_j - the value of analysed variable in i and j objects; $\bar{x}$ - the mean average of analysed variable.

The variables for which the value of Moran's I statistic are statistically significant are included in the group of 'spatial' variables and otherwise - in the group of variables having no spatial character ('non-spatial' variables). In this research, spatial contiguity weight matrix was used, since it is a matrix that appears most frequently in the studies, taking into account the spatial relationship. These weights basically indicate whether regions share a common boundary or not.

$$w_{ij} = \begin{cases} 1, & bnd(i) \cap bnd(j) \neq \varnothing \\ 0, & bnd(i) \cap bnd(j) = \varnothing \\ 0, & i = j \end{cases}. \tag{2}$$

1 refers to the situation in which region i and j have a common boundary; 0, if not. Diagonal elements in matrix W have value equal to 0 as the object cannot be its own neighbour. Spatial weight matrix was row standardised.

2. Estimating the SAR model for each variable from ‘spatial’ group of variables (LeSage, 1999):

$$\mathrm{X_j} = \rho \mathrm{W} \mathrm{X_j} + \varepsilon \tag{3}$$

where: $\mathrm{X_j}$ - the vector of analysed j variable; ρ - the spatial autoregression parameter; W - the spatial weight matrix; ε - the spatially correlated residuals.

3. Preparing the set of diagnostic variables:

3.1. Adjusting the values of variables from ‘spatial’ group according to formula:

$$\mathrm{S}_j = (I - \rho \mathrm{W})^{-1} \mathrm{X}_j \tag{4}$$

where: S_j - the vector of spatially adjusted j variable; I - identity matrix; ρ - the spatial autoregression parameter, W - the spatial weight matrix.

3.2. Remaining unchained the values of variables from ‘non-spatial’ group.

4. Changing destimulants for stimulants and standardise variables according to Hellwig’s formula:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1,...,n;\ j = 1,...,m) \tag{5}$$

where: z_{ij} - standardised value of j variable in i object; x_{ij} - the value of j variable in i object; $\bar{x}_j$ - the mean average of j variable; s_j - the standard deviation of j variable.

5. Calculating the distance between the i object and ‘ideal’ object:

$$d_i = \sqrt{\sum_{j=1}^{m} \left(z_{ij} - \varphi_j\right)^2} \quad (i = 1,...,n;\ j = 1,...,m) \tag{6}$$

where: z_{ij} - standardised value of j variable in i object; φ_j - value of j variable in the ‘ideal’ object.

6. Calculating the spatial taxonomy measure of development (sTMD) according to formula (Pietrzak, 2014):

$$sTMD_i = 1 - \frac{d_i}{d_{i-}} \quad (i = 1,..., n) \tag{7}$$

where:

$$d_{i-} = \bar{d} + 2s_d \quad (i = 1,..., n). \tag{8}$$

$sTMD_i$ - the taxonomy spatial measure of development for the county i; d_i - the distance between object i and 'ideal' object; $\bar{d}$ - the average value of d vector ($d = d_1,..., d_n$); s_d - the standard deviation of d vector.

The higher the value of $sTMD_i$ the better from the point of view of analysed phenomena. The analysed regions were also grouped on the basis of similar values of synthetic measure. Those groups were constructed as follows:

- the highest sustainability level: $sTMD_i \geq \bar{s}TMD + sd_{sTMD}$,
- medium sustainability level: $\bar{s}TMD + sd_{sTMD} > sTMD_i \geq \bar{s}TMD$,
- low sustainability level: $\bar{s}TMD > sTMD_i \geq \bar{s}TMD - sd_{sTMD}$,
- the lowest sustainability level: $sTMD_i < \bar{s}TMD - sd_{sTMD}$.

The result of analysis are presented in Figure 1-3.

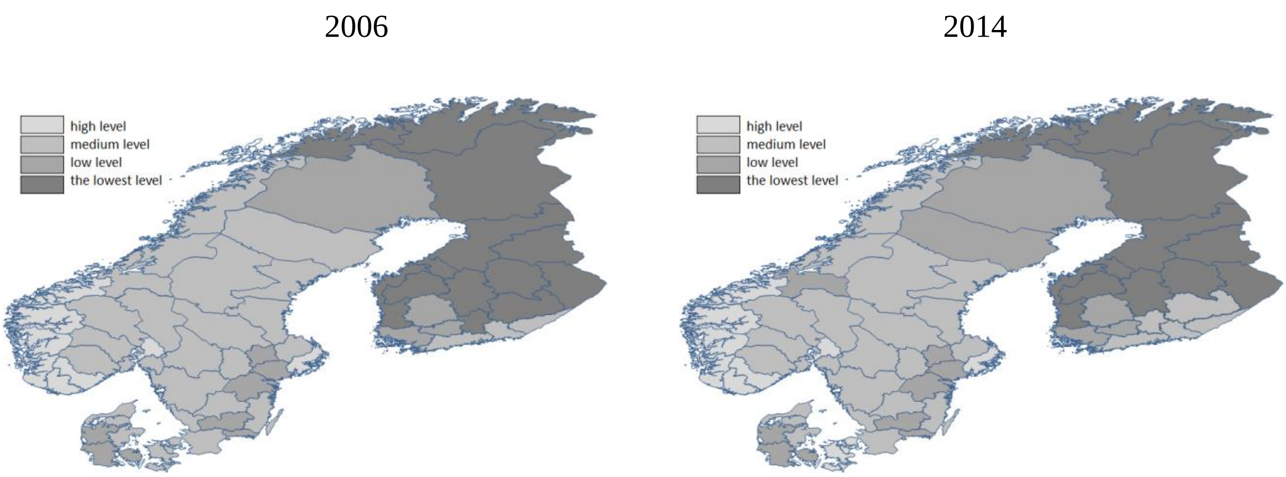


Fig. 1. Similar group of Nordic NUTS-3 regions in terms of social sustainability level in 2006 and 2014.

2006 2014

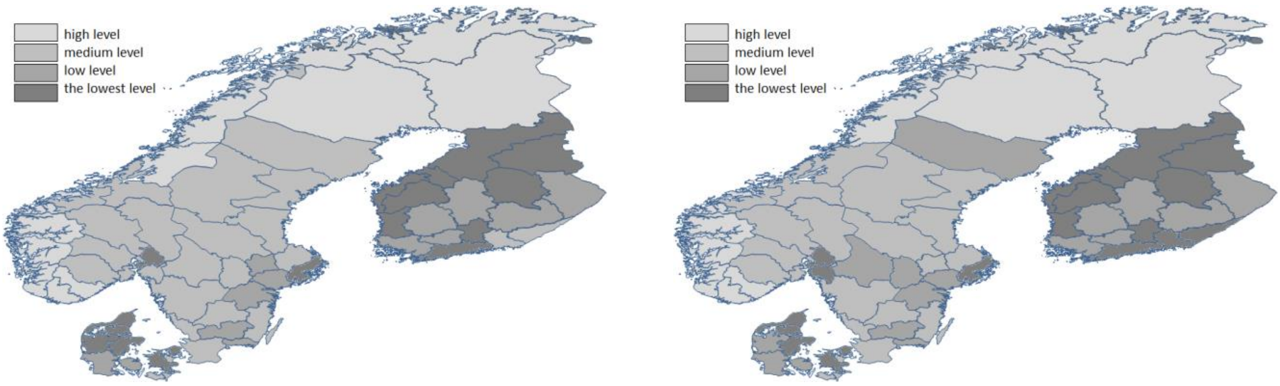


Fig. 2. Similar group of Nordic NUTS-3 regions in terms of ecologic sustainability level in 2006 and 2014.

Analysing figures 1-3 it can be seen that not only there are differences in the level of sustainability between regions but also particular regions differ in degree of balance in a particular sustainable development dimension. Northern regions show a high level of environmental sustainability, which is not surprising, since the harsh climate and low population density translates into low degree of urbanization and industrialization, which in turn is connected to the relative low air pollution emissions and energy consumption, and often with high forestry level and a small land use. On the other hand, the harsh living conditions and the lack of a developed industry results in low or very low level of sustainability due to the economic and social conditions.

2006 2014

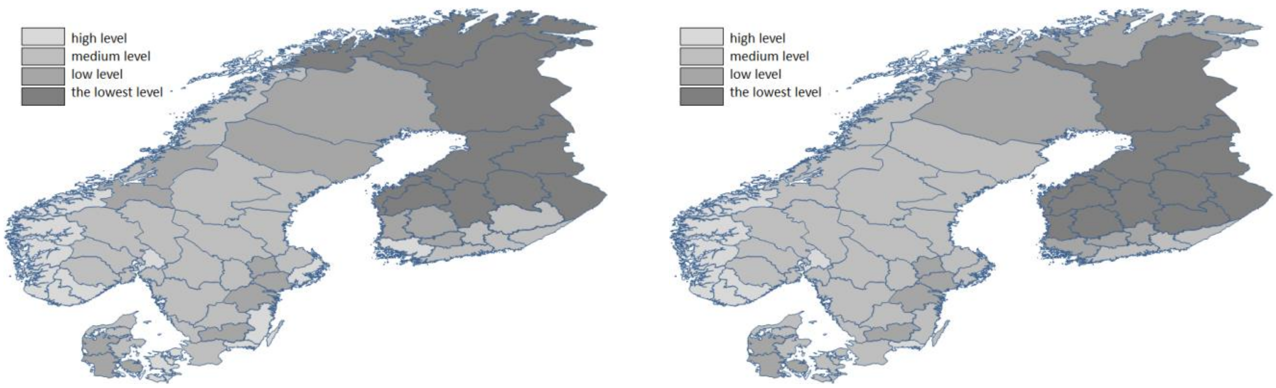


Fig. 3. Similar group of Nordic NUTS-3 regions in terms of economic sustainability level in 2006 and 2014.

The opposite of northern regions are the regions including capitals of analyzed countries, as well as northern Denmark and southern Finland. The high population density and strong industrialization place those regions as a low-balanced due to ecological conditions; however, those regions have much better sustainability level in economic and social areas. It should be also noted that, in general, Norway and Sweden realise sustainable development aim better than other countries. Swedish and Norwegian regions in the majority of cases were classified into high or medium level groups. While in most cases Finnish regions were usually classified into low or the lowest level groups.

Conclusions

In this study the spatial differences in components of sustainable development in Nordic NUTS-3 regions were analyzed. For this purpose a spatial taxonomy measure of development proposed by Pietrzak was used. The usage of this methodology seems to be justifies, as 9 out of 14 diagnostic variables showed the presence of spatial autocorrelation. The analysis showed that the achievement of sustainable development goals manage much better in the Swedish and Norwegian regions. One should also pay attention to the differences in balance in different areas of sustainable development. Central and highly industrialized regions still have problems with elements related to the protection of the environment, which is over-used to achieve social and economic objectives. While regions located in the area of the Arctic Circle realize ecologic objectives, while still exhibit difficulties in socio-economic areas.

The analysis showed that the Nordic countries are not as homogeneous as it might seem at first glance. It also indicates that despite the occupied high places in the rankings of sustainable development there is still much to do. Therefore, it should not be surprising, that in addition to the implementation of national sustainability development programs, Nordic countries have also decided to establish international cooperation in this area.

References

A Good Life in a Sustainable Nordic Region. Nordic Strategy for Sustainable Development. (2013). Retrieved June 16, 2016, from Norden website http://norden.diva-portal.org/smash/get/diva2:701472/FULLTEXT01.pdf.

Antczak E. (2013), Przestrzenny taksonomiczny miernik rozwoju. *Wiadomości Statystyczne 7/2013*, 37-53.

Balcerzak A.P. & Pietrzak M.B. (2016). Application of TOPSIS method for analysis of sustainable development in European Union countries. In: Loster, T., Pavelka, T. (eds.), *The 10th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016*, 82-92.

Buccellato T. (2007), Convergence across Russian Regions: A spatial econometrics approach. *Economics Working Paper No. 72*, 1-24.

Country Sustainability Ranking. (Rep.). (2013). Retrieved 16 February, 2015 from RobecoSAM website: http://www.robecosam.com/images/CS_Ranking_E _Rel.FIN AL.pdf

Du Pisani, J. A. (2006). Sustainable development – historical roots of the concept. *Environmental Sciences 3(2)*, 83-96.

Griffith D.A. (1996). Some guidelines for specifying the geographic weights matrix contained. In Arlinghaus, S. L. (ed.), *Practical Handbook of Spatial Statistics*. CRC Press.

Hellwig Z. (1968), Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasady i strukturę wykwalifikowanych kadr. *Przegląd Statystycznyn 5(4)*, 307-327.

LeSage J.P. (1999). The theory and practice of spatial econometrics. Retrieved from http://www.spatial-econometrics.com/html/sbook.pdf.

Pietrzak, M. B. (2014). Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych, *Przegląd Statystyczny RLX(2)*, 181-201.

Pietrzak, M.B. (2016). The Problem of Inclusion of Spatial Dependence Within the TOPSIS Method. *Mountenegrin Journal of Economics*, vol. 12, no. 3. 69-86.

Pietrzak M.B. & Balcerzak A.P. (2016). Assesment of socio-economic sustainability in the new European Union member states in the years 2004-2012, In Papież M., & Śmiech S. (eds.), *The 10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, 120-129.

Pietrzak M.B., Wilk J., Bivand R., & Kossowski T. (2014a). The Application of Local Indicators for Categorical Data (LICD) in Identifying of Spatial Dependences in the Analysis of Socio-Economic Development. *Comparative Economic Research. Central and Easter Europe 17(4)*, 203-220.

Pietrzak, M.B., Wilk J., Kossowki T., & Bivand R., (2014b). The Identification of Spatial Dependence in the Analysis of Regional Economic Development - Join-count Test Application. In Papież M., & Śmiech S., (eds.), *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, 135-144.

SDG Index & Dashbords. A global report (Rep.). (2017). Retrieved February 09, 2017, from Sustainable Development Solutions Network (SDSN) website: https://www.dropbox.com/s/gy2zmh9065v0mr5/SDG_Index_Dashboard_full.pdf?dl=0.

Sobolewski M., Migała-Warcohł A., & Mentel G. (2014), Ranking poziomu życia w powiatach w latach 2003-2012 z uwzględnieniem korelacji przestrzennych, *Acta Universitatis Lodziensis Folia Oeconomica* 6 (308), 159-172.

Sustainable Development. New Bearings for the Nordic Countries. 2001. Nordic Council of Ministers, Copenhagen.

Sustainable Development Timeline (Rep.). (2012). Retrieved February 01, 2017, from International Institute for Sustainable Development website: http://www.iisd.org/pdf/2012/sd_timeline_2012.pdf.

The Helsinki Treaty. (1962). Retrieved June 16, 2016, from Norden website: http://www.norden.org/en/om-samarbejdet-1/nordic-agreements/treaties-and-agreements/basic-agreement/the-helsinki-treaty.

Tobler W.R. (1970), A computer movie simulating urban growth in the Detroit region, *Economic Geography 46*, 234-240.

Vu J., Wu G., Zhou Q., & Li M., (2014), Spatial Variation of Regional Sustainable Development and its Relationship to the Allocation of Science and Technology Resources, *Sustainability 6(9)*, 6400-6417.

World Conservation Strategy Living Resource Conservation for Sustainable Development (Rep.). (1980). Retrieved January 04, 2017, from International Union for Conservation of Nature and Natural Resources website: https://portals.iucn.org/library/efiles/documents/wcs-004.pdf.

Source of data as a determinant of the assessment of agricultural producers' financial situation

Tomasz Kuszewski[1], Agata Sielska[2]

Abstract

The aim of the paper is to examine whether the analysis of agricultural producer's income in 2004-2013 based on the different sources of data leads to noncontradictory conclusions concerning the changes of economic situation in agriculture. Data from the Central Statistical Office of Poland, Farm Accountancy Data Network (FADN), Research Institute for Economic Development of Warsaw School of Economics (IRG SGH) and results reported in the Social Diagnosis have been used. The changes in variables reflecting the agricultural producers' financial situation have been compared on the basis of the interpretation of graphs, correlation coefficients and distance measured in the Clark's metrics. The study shows that the inference about farmers' financial situation based on data from different sources can lead to different conclusions. Results from FADN may be considered the most consistent with other databases whereas results obtained by IRG SGH may be considered the most disparate.

Key words: *value added, farms' income, agricultural indicators*
JEL Classification: Q19

1 Introduction

Until recently, lack or scarcity of the quantitative data was one of analysts' main problems. The research presented in the papers (Bailey et al., 2004; Alvarez and Arias, 2004; Charnes et al., 1973) is a good illustration of the compromises related to the scarcity of relevant data. This situation has changed and the period of rapidly developing both the data collection techniques and the methods of forecasting and classification lead to problems related to the excess of information. Nowadays, the analyst has to choose from many existing data sources, evaluate datasets' adequacy for a given problem and to ensure substantive and statistical comparability.

The aim of this paper is to illustrate and solve that dilemma. We use an example of changes in economic situation of Polish agriculture in period 2004-2013 and the influence of this changes on economic producers' opinion on their own financial situation. Polish agriculture has

[1] Corresponding author: Warsaw School of Economics, Institute of Econometrics, Al. Niepodległości 162, 02-554 Warsaw, e-mail: tomasz.kuszewski@sgh.waw.pl.
[2] Warsaw School of Economics, Department of Applied Economics, Al. Niepodległości 162, 02-554 Warsaw, e-mail: asiels@sgh.waw.pl.

undergone a substantial change since country accessed the European Union and is still an important sector of Polish economics (Piecuch, 2013; Wicka, 2012).

The following sources of data are used: official national and EU statistics i.e. Central Statistical Office of Poland (GUS) and Farm Accountancy Data Network (FADN) database respectively, data gathered by the Research Institute for Economic Development of Warsaw School of Economics (IRG SGH) and survey results reported periodically in the Social Diagnosis (DS).

We propose a set of data illustrating the changes in financial situation of Polish agricultural producers. Due to the different data sources, variables cannot be compared in a straightforward way. It is difficult to compare time series data on agricultural production based on value added created in agriculture published by GUS in current or fixed prices to the qualitative results from IRG SGH or DS surveys or to variables published by FADN. Another aspect of comparability concerns the scope of the survey. Data published by GUS concerns the whole population and result from the full statistical survey, whereas FADN data concerns whole population but results from the research conducted among the representative sample of Polish agricultural households. This representativeness concerns chosen criteria such as economic size of the farm, type of farming and location (Floriańczyk et al., 2014). Both GUS and FADN data have well defined economic categories and will be referred to as objective. FADN data and methodology is frequently used to evaluate the situation in agricultural sector in the EU countries (Galluzzo, 2015; Serences et al., 2016). IRG SGH and DS data come from survey and – in contrast to those previously mentioned – will be referred to as subjective.

We propose a method of data transformation, which ensure its comparability. In the next step of the analysis we examine the dynamics of changes in agricultural producers' financial situation on the basis of selected data sources and discuss results. We calculate distances between time series in order to evaluate the similarities in the pictures of the financial situation of Polish agricultural producers in 2004-2013.

2 Data sources and methodology

We use Central Statistical Office of Poland (GUS)[3] data for 2004-2013 on:

[3] http://stat.gov.pl; page Roczne wskaźniki makroekonomiczne, accesed in April 2016.

- gross value added in „Agriculture, forestry, hunting and fishing” section of Polish economy (in m zł, current prices; in the next sections of the article we refer to this variable as the value of gross production in agriculture);
- gross value added in „Agriculture, forestry, hunting and fishing” section of Polish economy (in fixed prices, expressed as “previous year =100”; in the next sections of the article we refer to this variable as the value of gross production in agriculture);
- real gross disposable income in households that simultaneously run business as individual farms in agriculture (expressed as “previous year =100”);
- the number of working in „Agriculture, forestry, hunting and fishing” (thousands); in the next part of the paper this value is treated as the number of working in agriculture.

We also collected data on net disposable income of individual households in agriculture in 2004 (in current prices) (GUS, 2005).

The following time series were prepared for the analysis after proper conversions:

- gross value added, in current prices per working in agriculture - GUS_WD;
- real net disposable income of households per working in individual agricultural household - GUS_DD.

The following data for 2004-2013 was collected from Farm Accountancy Data Network[4]:

- farm net value added expressed per agricultural work unit, in current prices (variable coded as SE425 in the FADN database) - FADN_WD;
- farm net income per family labour unit i.e. unpaid full time family labour, expressed in current prices (variable coded as SE430 in FADN database) - FADN_D.

Data on the economic situation in agriculture in 2004-2013 was collected from IRG SGH[5]. In this research respondents are asked about their cash income. They can choose one of the following answers: income is higher than previously, as high as previously, lower than previously, no income at wall was obtained, with weights 1; 0; -1; 0 respectively. When preparing the data for further analysis a standard procedure for calculating the balance of responses to the question in the form of a weighted average of the percentage of positive answers is applied. Therefore, quarterly data refer to the structure of responses to the questions

[4] European Commission - EU FADN, http://ec.europa.eu/agriculture/rica/database/database.cfm, accessed in April 2016.
[5] Data compiled by K. Walczyk from IRG SGH.

characterizing the financial situation. In the second step the arithmetic average of the quarterly balances is calculated and the IRG_P time series is obtained.

A survey of conditions and quality of life called Social Diagnosis is carried out since 2000. Since 2003 it is carried out in two-year intervals. Since that year the survey includes a question on the individual assessment of the quality of life. Different aspects and spheres of life are specified and the respondents are asked to evaluate their satisfaction on a scale of 1 (very satisfied) to 6 (very dissatisfied). Financial situation of the family is one of the aspects mentioned. In the research reports (Czapiński and Panek, 2009; 2011; 2013) the average of the responses to this question of all respondents can be found. On the basis of survey data[6] the average level of satisfaction from the family financial situation can be assessed for each year for respondents classified as "farmers"[7].

Average degree of satisfaction with the financial situation of the family for the families of farmers surveyed was indicated as DS_R. Because the DS research is conducted in two-year intervals, the time series of DS_R has a lower number of observations than the others. We assumed the values from 2003 survey as the value for 2004, which is the first year for other time series. In order to improve the comparability and readability of results missing values of DS_R were filled with averages of neighbouring values in figures and tables.

As previously mentioned, the nature of the collected time-series value of these indicators is not uniform. To ensure comparability and enable the combined analysis the time series of all indicators were transformed in such a way that the values of each indicator for 2004 was arbitrary set on 100 and values for 2005-2013 were calculated in proportion. Therefore we interpret only the relative changes of each indicator.

We calculated the distance, i.e. similarity, between the time series using the Clark's metrics[8]. We also study the correlation between indicators.

---

[6] Social Diagnosis. Integrated database. www.diagnoza.com, accessed 16 April 2016.
[7] Respondents indicating to usage of farm as the main source of income.
[8] Distance measured in J.P. Clark's metrics, also known as the divergence coefficient, between points x_i, x_k from R^n space is defined as: $d(x_i, x_k) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_{ij}-x_{kj}}{x_{ij}+x_{kj}}\right)^2}$. It is normalized on the [0, 1] interval and easy to interpret (Młodak, 2006). In our interpretation, the point from R^n space is the time series of a given indicator for the period 2004-2013. An alternative approach to the comparison of time series which develop according to the clear trend is to use the function

3 Results

Values of the objective indicators from GUS and FADN lead to the most optimistic assessment of the agricultural producers' situation. DS results indicate smaller improvement, whereas IRG SGH results suggest relative deterioration of the financial situation in the analysed period as compared to 2004. It can be said, that this data source seems to be burdened with the highest risk of committing an error in interpretation.

Discrepancies in the assessment of the farms situation on the basis of all indicators are presented in figure 1. If we ignore the IRG_P indicator, it can be seen that GUS_WD, FADN_D and FADN_WD achieved their final high values in 2013 by different paths. On the basis of the GUS_DD and DS_R indicators the growth can be assessed as slow.

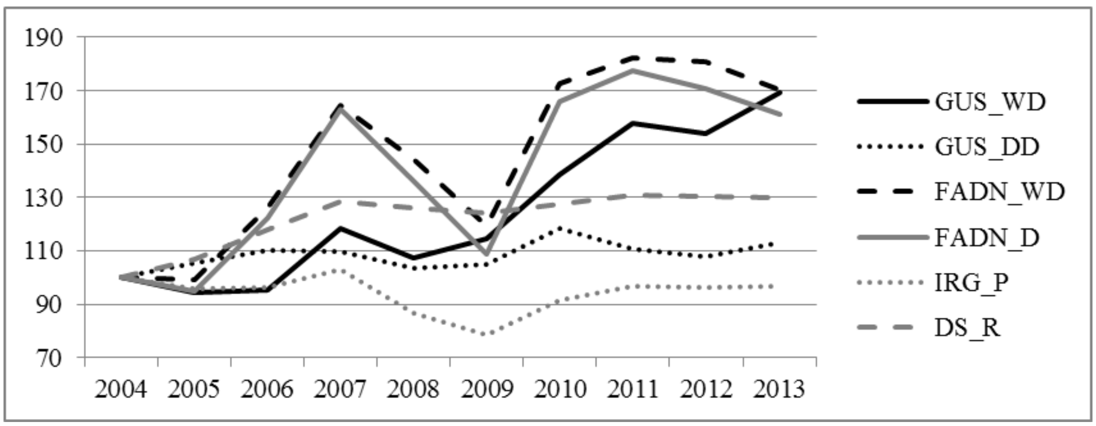


Fig. 1. Changes in values of indexes describing economic situation in 2004-2013 (2004=100).

Source: own calculations.

The indicators obtained from the FADN data show that both the value added and income developed according to the same regularities, which to some extent may result from the algorithm used to calculate the values of these variables. When FADN data are considered, it can be seen that the difference between the value added and farm income is smaller than in the case when GUS data are used (figure 2). FADN_(WD-D) is the difference between FADN_WD and FADN_D; GUS_(WD-DD) is the difference between GUS_WD and GUS_DD.

similarity measure (Dorosiewicz, Michalski, 1998). We didn't use this approach due to the small number of observations.

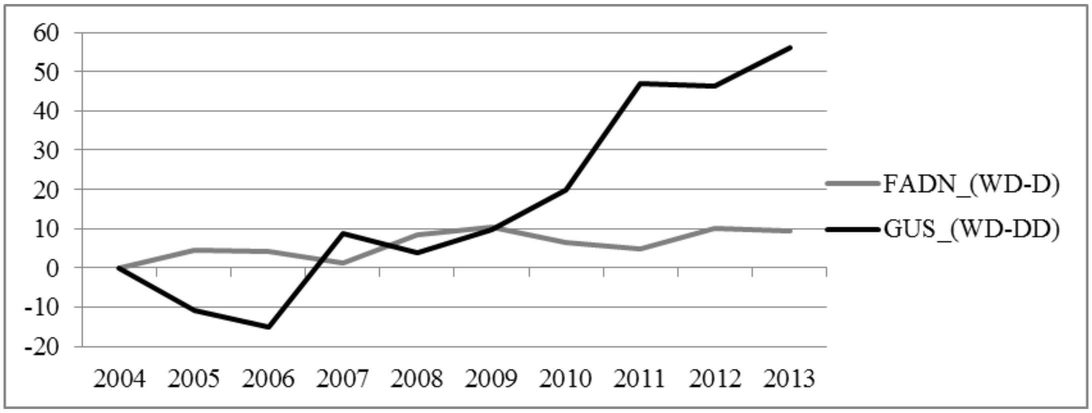


Fig. 2. Difference between the indexes of value added and income on the basis of FADN and GUS data.

Source: own calculations.

Indicator	GUS_WD	GUS_DD	FADN_D	IRG_P	DS_R	FADN_WD
GUS_WD	1.0000					
	[0.00]					
GUS_DD	0.5828	1.0000				
	[0.35]	[0.00]				
FADN_D	0.8291	0.6828	1.0000			
	[0.26]	[0.47]	[0.00]			
IRG_P	0.1228	0.1266	0.2682	1.0000		
	[0.49]	[0.19]	[0.67]	[0.00]		
DS_R	0.7285	0.6083	0.8526	-0.1803	1.0000	
	[0.24]	[0.22]	[0.31]	[0.45]	[0.00]	
FADN_WD	0.8543	0.6827	0.9942	0.1837	0.8905	1.0000
	[0.30]	[0.52]	[0.08]	[0.73]	[0.34]	[0.00]

Table 1. Correlation coefficients and Clark's distances for pairs of indexes.

Source: own calculations.

Pearson correlation coefficient was used as the first method of analysis of the time series' similarity[9] (table 1). Results of IRG research are still not similar to any of the objective variables. Results of the DS research are similar to both FADN and GUS. In brackets in table 1 we present the values of Clark's distance. Both measures are symmetrical, so part of the table is unfilled.

The evaluation of interdependence between the pairs of indicators would be fully consistent if high values of Clark's measure were accompanied by small absolute values of the correlation coefficient. As we can see, this is not the case, for example for (IRG_P, GUS_DD) pair. On the other hand, the compatibility of measures is confirmed by both interdependence measures for pair (FADN_WD, FADN_D).

Conclusions

Using only one source of quantitative data or one indicator in the economic analysis may lead to the risk of error in interpretation. Especially when analyzing changes over time we should refer to the values of several indicators, preferably those whose values derive from surveys conducted by various institutions. The fact that the original data, which is the basis for determining the values of the indicators are measured by different methods and in different units does not constitute an obstacle to a proper transformation of the indicators in order to ensure their comparability.

Comparing the indicators can help identify the study, the results of which are consistent with other sources. In our case data on the situation in agriculture collected by FADN seem to be most trustworthy, whereas the IRG SGH data can be seen as most controversial and difficult to compare with other time series.

Another issue unresolved in this study, is the choice of indicators for comparisons, in order to illustrate the state or dynamics of change. We use indicators on financial situation of the farm same as (Augustyńska-Grzymek et al., 2013), but we could as well choose the amount of free time which a family have as a prosperity measure (Bazyl, 2010).

[9] The changes are nonlinear but the Pearson correlation coefficient was used in order to show the interdependence of directions of changes.

References

Augustyńska-Grzymek, I., Skarżyńska, A., & Abramczuk, Ł. (2013). The pluriactive development of agricultural holdings in Poland with regard to the living standards of their users. *Studies in Agricultural Economics*, *115*(1), 16-24.

Alvarez, A., & Arias, C. (2004). Technical efficiency and farm size: a conditional analysis. *Agricultural Economics*, *30*(3), 241-250.

Bailey, A., Irz, X., & Balcombe, K. (2004). Measuring productivity growth when technological change is biased – a new index and an application to UK agriculture. *Agricultural Economics*, *31*(2-3), 285-295.

Bazyl, M. (2010). Czas wolny miarą dobrobytu wsi. In: *Świat widziany oczami ekonomisty*, (eds.) D. Stala, S. Sztaba, 10-11, Warszawa: Oficyna Wydawnicza Szkoły Głównej Handlowej.

Charnes, A., Cooper, C.C., & Kozmetsky, G. (1973). Measuring, monitoring and modeling quality of life. *Management Science*, *19*(10), 1172-1188.

Czapiński, J., & Panek, T., (eds.) (2009). Diagnoza społeczna 2009. www.diagnoza.com, accessed April 2016.

Czapiński, J., & Panek, T., (eds.) (2011). Diagnoza społeczna 2011. www.diagnoza.com, accessed April 2016.

Czapiński, J., & Panek, T., (eds.) (2013). Diagnoza społeczna 2013. www.diagnoza.com, accessed April 2016.

Diagnoza społeczna: zintegrowana baza danych. www.diagnoza.com, accessed April 2016.

Dorosiewicz, S., & Michalski, T. (1998). Podobieństwo funkcji w badaniach ekonomicznych (metody i przykłady)**.** *Przegląd Statystyczny*, *2*, 187-196.

European Commission - EU FADN, http://ec.europa.eu/agriculture/rica/database/database.cfm, accessed April 2016.

GUS. 2005. Rocznik Statystyczny. Warszawa. http://stat.gov.pl, accessed April 2016.

Floriańczyk, Z., Mańko, S., Osuch, D., & Płonka, R. (2014). Wyniki Standardowe 2013 uzyskane przez gospodarstwa rolne uczestniczące w Polskim FADN. Część I. Wyniki Standardowe. Warszawa: Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej - Państwowy Instytut Badawczy.

Galluzzo, N. (2015). Technical and economic efficiency analysis on Italian smallholder family farms using Farm Accountancy Data Network dataset. *Studies in Agricultural Economics*, *117*(1), 35-42.

Młodak, A. (2006). Analiza taksonomiczna w statystyce regionalnej. Warszawa: Difin.

Piecuch, J. (2013). Evolution Of Social And Economic Situation In Polish Agriculture Since The Accession To The European Union [in:] A. Eglite (ed.) *Economic Science For Rural Development: Rural Development And Entrepreneurship Marketing And Sustainable Consumption*, *32*, 54-59.

Serences, P., Cierna, Z., & Piterkova, A. (2016). The development of value added and net income of farms in Slovakia. International Scientific Days 2016. *The Agri-Food Value Chain: Challenges For Natural Resources Management And Society*, 378-388.

Wicka, A. (2012). Polish Agriculture Against Selected States of the European Union [in:] A. Zvirbule Berzina (ed.) *Economic science for rural development: resources and cooperation: 1. Resources and sustainable consumption, 2. Production and cooperation in primary and secondary agriculture*, *29*, 140-147.

Exponential smoothing models with time-varying periodic parameters

Łukasz Lenart[1]

Abstract

In the literature there is a growing number of models with time-varying in time parameters. The purpose of this article is to show the reduced form for the linear innovations model with time-varying periodic in time parameters. It will be show that when the state variables are eliminated from a linear innovations state space model with time-varying periodic in time parameters, an Periodic Autoregressive Integrated Moving Average model (PARIMA in short) with equality restrictions on parameters is obtained. This is the generalization for reduced form of the state space model with constant in time parameters. In particular, known models called are generalized to the periodic case. Finally, the real data example with macroeconomic data will be presented where the performance of competing models (based on Logarithmic Score) in pseudo-real forecasting exercise is used to assess the adequacy of a specific model.

Keywords: *state space model, exponential smoothing model, periodic ARIMA model*

JEL Classification: C22, E32

1 Introduction

Time-varying with the seasons sample autocovariance function are refereed to periodic time series. Such class of models called periodic models were introduced firstly by Hannan (1955), while in Gladyshev (1961) the Periodically Correlated time series with period T (PC(T) in short) ware defined and examined. For theory and applications for PC(T) time series see to Hurd and Miamee (2007). There are a few alternative ways to consider periodic (or seasonal) in time dynamic of parameters in time series models. The most popular is the usual ARMA model with periodic coefficients (PARMA in short). This generalization assumes periodic in time coefficients in AR and MA part, with the same period. Under appropriate regularity conditions the PARMA model is PC. The PARMA models are well examined in time and frequency domain (see for example Wyłomańska (2008)). In the same way the PARIMA and Seasonal PARIMA models can be defined. The alternative and more sophisticated models with time-varying periodic parameters are known. The applications and theoretical background concerning models with time-varying parameters can be found in Pagano (1978), Osborn (1991), Franses and Paap (2004), Burridge and Taylor (2001) and many others. The seasonal volatility (or seasonal heteroscedasticity) ware considered by Trimbur and Bell (2012), Berument and Sahin (2010), Doshi et al. (2011), Lenart (2017), and many other studies. In

[1] Cracow University of Economics, Department of Mathematics, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: lukasz.lenart@uek.krakow.pl.

Novales and Fruto (1997) the authors consider performance of periodic models over non-periodic models in forecasting. Extension of Unobserved Component Models to periodic case ware considered in Koopman and Oomos (2008), Proietti (2004) and many others.

In this paper we consider the case of periodic coefficients in linear innovations state space model. The constant parameters were examined in details in Hyndman et al. (2008). Note that modeling with periodic coefficients is effective if appropriate models and estimations procedures are considered. Additionally, good in-sample fit is not equivalent with good forecasting power. We decide to extend this class of time series to periodic case since the components of exponential smoothing models (e.g. level component, trend component, seasonal component) has an natural interpretation. By assuming periodicity for parameters the extended natural interpretation is possible (for example seasonal conditional volatility). In the first part we examine the exponential smoothing models with periodic coefficients from theoretical point of view. We show the exact reduced PARIMA forms for chosen models. In empirical part, we would like to shed light on the forecasting problem with such class of time series. We use data concerning monthly price in education COICOP in Poland (Jan. 1997 to Dec. 2015).

2 Reduced Form for the General Linear Innovations Model with periodic coefficient

We consider the following general linear innovation model of the form:

$$\begin{cases} y_t = \boldsymbol{\omega}'\mathbf{x}_{t-1} + \varepsilon_t & \text{observation equation} \\ \mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{g}_t\varepsilon_t & \text{state equation,} \end{cases} \tag{1}$$

where y_t is real-valued observed time series and $\mathbf{x_t}$ is the state vector, $\varepsilon_t \sim IID(0,\sigma_t^2)$. For the matrixes $\boldsymbol{\omega}$ and $\mathbf{F}$ we assume that are constant. For matrix $\mathbf{g}_t$ we assume that contains element being an periodic functions with the same period $T > 1$. The periodicity of variance σ_t^2 is also assumed. Hyndman et al. (2008) consider case with constant $\mathbf{g}_t$, σ_t^2 and proved that such model has reduced ARIMA form. In the next part we show that generalization to periodic case produce reduced periodic ARIMA model with period T, i.e. ARIMA model with time-varying periodic coefficients with period T.

Based on the same steps as in Hyndman et al. (2008) (page 170, we drop the details) the general linear innovation model can be reduced to Periodic ARIMA model of the general form:

$$\eta(L)y_t = \theta_t(L)[\varepsilon_t], \tag{2}$$

where the autoregressive polynomial equals

$$\eta(L) = \det(I - \mathbf{F}L) \tag{3}$$

and time-varying moving average polynomial equals

$$\theta_t(L) = \boldsymbol{\omega}' \operatorname{adj}(I - \mathbf{F}L)L[\mathbf{g}_t] + \det(I - \mathbf{F}L), \tag{4}$$

and $L^k[X_t] = X_{t-k}$ for any $k,t \in \mathrm{Z}$. Note that the polynomial θ depends on t. Hence, after this polynomial we put the argument in brackets $[\cdot]$. To clarify, we consider the following local level model of the form

$$\begin{cases} y_t = l_{t-1} + \varepsilon_t \\ l_t = l_{t-1} + \alpha_t \varepsilon_t, \end{cases} \tag{5}$$

where $\varepsilon_t \sim \mathrm{NID}(0,\sigma_t^2)$ and α_t, σ_t are periodic with period T (see Hyndman et al. 2008, page 40 in case of constant parameters). In such case, the elementary calculations give that $(1-L)y_t = (1+(\alpha_{t-1}-1)L)[\varepsilon_t]$. It means that this model has reduced PARIMA(0,1,1) model form with periodic variance of the white noise and equality restrictions on parameters.

In the next section we consider more advanced models with periodic parameters. We consider damped level model with seasonal pattern; local additive seasonal model with damped level and trend pattern and finally double damped local trend model. The extension of known seasonal $ARIMA(p,d,q)(P,D,Q)_m$ model with season m to periodic case (with period T) we denote by seasonal $PARIMA(p,d,q)(P,D,Q)_m$ with season m and period T. Note that it is natural to consider $T = m$.

2.1 Damped level model with seasonal pattern

In this section we consider model with damped level pattern and seasonal pattern with periodic coefficients of the form

$$\begin{cases} y_t = \phi l_{t-1} + s_{t-m} + \varepsilon_t \\ l_t = \phi l_{t-1} + \alpha_t \varepsilon_t \\ s_t = s_{t-m} + \gamma_t \varepsilon_t, \end{cases} \tag{6}$$

where m is the length of the season for seasonal pattern and α_t, γ_t, σ_t^2 are periodic functions at t with the same period $T > 1$. Equivalently, above model can be written in form

$$\begin{cases} y_t = \phi l_{t-1} + s_{t-m} + \varepsilon_t \\ l_t = \alpha_t (y_t - s_{t-m}) + (1-\alpha_t)\phi l_{t-1} \\ s_t = \gamma_t (y_t - \phi l_{t-1}) + (1-\gamma_t) s_{t-m}. \end{cases} \tag{7}$$

To simplify this section we assume that $T = m = 4$. But the general case is passible. In such specific case we have $\mathbf{x}_{t'} = (l_t \quad s_t \quad s_{t-1} \quad s_{t-2} \quad s_{t-3})$, $\boldsymbol{\omega}' = (\phi \quad 0 \quad 0 \quad 0 \quad 1)$, $\mathbf{g}_{t'} = (\alpha_t \quad \gamma_t \quad 0 \quad 0 \quad 0)$,

$$\mathbf{F} = \begin{pmatrix} \phi & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad I - \mathbf{F}L = \begin{pmatrix} 1-L\phi & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -L \\ 0 & -L & 1 & 0 & 0 \\ 0 & 0 & -L & 1 & 0 \\ 0 & 0 & 0 & -L & 1 \end{pmatrix}. \tag{8}$$

Elementary calculations give that $\eta(L) = \det(I - \mathbf{F}L) = (1 - L^4)(1 - L\phi)$ and

$$\mathrm{adj}(I - \mathbf{F}L) = \begin{pmatrix} 1-L^4 & 0 & 0 & 0 & 0 \\ 0 & 1-L\phi & L^3 - L^4\phi & L^2 - L^3\phi & L - L^2\phi \\ 0 & L - L^2\phi & 1 - L\phi & L^3 - L^4\phi & L^2 - L^3\phi \\ 0 & L^2 - L^3\phi & L - L^2\phi & 1 - L\phi & L^3 - L^4\phi \\ 0 & L^3 - L^4\phi & L^2 - L^3\phi & L - L^2\phi & 1 - L\phi \end{pmatrix}. \tag{9}$$

Hence (after calculations) $\boldsymbol{\omega}'\mathrm{adj}(I - \mathbf{F}L)L[\mathbf{g}_t] = (-\phi\alpha_{t-5} - \phi\gamma_{t-5})L^5 + \gamma_{t-4}L^4 + \phi\alpha_{t-1}L.$ Finally

$$\theta_t(L) = \phi(1 - \alpha_{t-5} - \gamma_{t-5})L^5 + (\gamma_{t-4} - 1)L^4 + \phi(\alpha_{t-1} - 1)L + 1. \tag{10}$$

Hence, the considered model is an seasonal PARIMA $(1{,}0{,}5)(0{,}1{,}0)_4$ model with period $T = 4$, periodic variance of the white noise and equality restrictions on parameters.

2.2 Local additive seasonal model with damped level

In this section we consider additive model with damped level, seasonal pattern, and trend pattern of the form

$$\begin{cases} y_t = \phi_t l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \\ l_t = \phi l_{t-1} + b_{t-1} + \alpha_t \varepsilon_t \\ b_t = b_{t-1} + \beta_t \varepsilon_t \\ s_t = s_{t-m} + \gamma_t \varepsilon_t, \end{cases} \tag{11}$$

where we assume that α_t, β_t, γ_t and σ_t^2 are periodic functions at t with period $T = m$. For $T = m = 4$ we have $\mathbf{x}_t' = (l_t \quad b_t \quad s_t \quad s_{t-1} \quad s_{t-2} \quad s_{t-3})$, $\boldsymbol{\omega}' = (\phi \quad 1 \quad 0 \quad 0 \quad 0 \quad 1)$, $\mathbf{g}_t' = (\alpha_t \quad \beta_t \quad \gamma_t \quad 0 \quad 0 \quad 0)$,

$$\mathbf{F} = \begin{pmatrix} \phi & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad I - \mathbf{F}L = \begin{pmatrix} 1 - L\phi & -L & 0 & 0 & 0 & 0 \\ 0 & 1-L & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -L \\ 0 & 0 & -L & 1 & 0 & 0 \\ 0 & 0 & 0 & -L & 1 & 0 \\ 0 & 0 & 0 & 0 & -L & 1 \end{pmatrix}. \quad (12)$$

Note that $\eta(L) = \det(I - \mathbf{F}L) = (1-L)\left(L^5\phi - L^4 - L\phi + 1\right) = (1-L)(1-L^4)(1-\phi L)$. In addition,

$$\begin{aligned} \boldsymbol{\omega}' \mathrm{adj}(I - \mathbf{F}L)L[\mathbf{g}_t] \; &= \phi(\alpha_{t-6} + \gamma_{t-6})L^6 + (-\phi\alpha_{t-5} - \beta_{t-5} - \phi\gamma_{t-5} - \gamma_{t-5})L^5 \\ &+ \gamma_{t-4}L^4 - \phi\alpha_{t-2}L^2 + (\phi\alpha_{t-1} + \beta_{t-1})L, \end{aligned} \quad (13)$$

which means that

$$\begin{aligned} \theta_t(L) \; &= \phi(\alpha_{t-6} + \gamma_{t-6} - 1)L^6 + (\phi - \phi\alpha_{t-5} - \beta_{t-5} - \phi\gamma_{t-5} - \gamma_{t-5} + 1)L^5 \\ &+ (\gamma_{t-4} - 1)L^4 + \phi(1 - \alpha_{t-2})L^2 + (\phi\alpha_{t-1} + \beta_{t-1} - \phi - 1)L + 1, \end{aligned} \quad (14)$$

or equivalently

$$\begin{aligned} \theta_t(L) \; &= (1-L)\left[\phi(1 - \alpha_{t-5} - \gamma_{t-5})L^5 + (\beta_{t-4} + \gamma_{t-4} - 1)L^4 + \beta_{t-3}L^3 \right. \\ &\left. + \beta_{t-2}L^2 + (\phi\alpha_{t-1} + \beta_{t-1} - \phi)L + 1\right], \end{aligned} \quad (15)$$

which means that the factor $1 - L$ is common in polynomials $\eta(L)$ and $\theta_t(L)$. Hence,

$$\eta(L) = L^5\phi_t - L^4 - L\phi_t + 1 = (1 - L^4)(1 - \phi L), \quad (16)$$

$$\begin{aligned} \theta_t(L) \; &= \phi(1 - \alpha_{t-5} - \gamma_{t-5})L^5 + (\beta_{t-4} + \gamma_{t-4} - 1)L^4 + \beta_{t-3}L^3 \\ &+ \beta_{t-2}L^2 + (\phi\alpha_{t-1} + \beta_{t-1} - \phi)L + 1. \end{aligned} \quad (17)$$

The considered model is an seasonal $\mathrm{PARIMA}(1,0,5)(0,1,0)_4$ model with period $T = 4$, periodic variance of the white noise and equality restrictions on parameters.

2.3 Double damped local trend model

Following by Hyndman et al. (2008), page 181 we consider

$$\begin{cases} y_t = \phi_1 l_{t-1} + \phi_2 b_{t-1} + \varepsilon_t \\ l_t = \phi_1 l_{t-1} + \phi_2 b_{t-1} + \alpha_t \varepsilon_t \\ b_t = \phi_2 b_{t-1} + \beta_t \varepsilon_t, \end{cases} \quad (18)$$

where α_t, β_t and σ_t^2 are periodic functions at t with period $T > 0$. We have $\mathbf{x}_t' = \begin{pmatrix} l_t & b_t \end{pmatrix}$, $\boldsymbol{\omega}' = \begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix}$, $\mathbf{g}_t' = \begin{pmatrix} \alpha_t & \beta_t \end{pmatrix}$,

$$\mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 \\ 0 & \phi_2 \end{pmatrix}, \qquad I - \mathbf{F}L = \begin{pmatrix} 1 - L\phi_1 & -L\phi_2 \\ 0 & 1 - L\phi_2 \end{pmatrix}. \quad (19)$$

Elementary calculations give that $\eta(L) = \det(I - \mathbf{F}L) = (1-\phi_1 L)(1-\phi_2 L)$. In addition,

$$\theta_t(L) \quad = \phi_1\phi_2(1-\alpha_{t-2})L^2 + (\phi_1\alpha_{t-1} + \phi_2\beta_{t-1} - \phi_1 - \phi_2)L + 1. \tag{20}$$

It means that the considered model is PARMA(2,2) model with period T, periodic variance of the white noise and equality restrictions on parameters.

3 Forecasting experiment

We consider monthly price in education in Poland (monthly rate of change, m-o-m, HICP (2015 = 100), source: Eurostat) form Jan. 1997 to Dec. 2015 (see Fig. 1). This price process is an important driver of inflation at September and October, where the peaks are observed (due to some administrative regulations). The seasonal pattern in obvious is such data (with period 12), while the trend is clearly not observed. Therefore, we propose to apply damped level model with seasonal pattern (see Section 2.1). We consider 9 different specifications for this model. For time-varying parameters: σ_t, γ_t, α_t we consider sequence of labels during one year (see details in Table 1). The same number at different months (for example 1 and 1 or 2 and 2) means the same value of parameter. Different numbers (for example 1 and 2 or 1 and 2 and 3) at months means that the values are different. For example, model M1 assumes constant parameters, while M3 assumes that σ_t and α_t are constant and γ_t has three values during year (labeled by: 1,2,3).

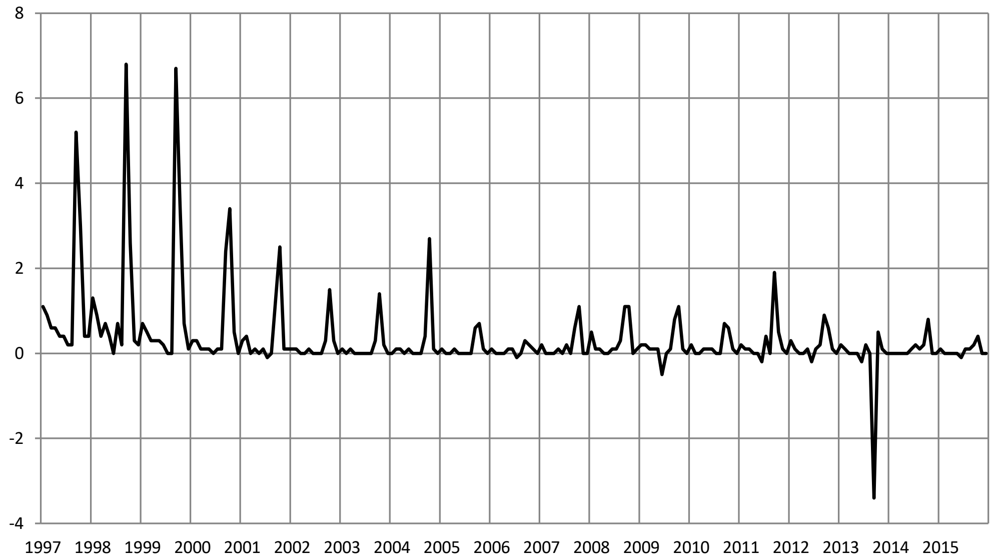


Fig. 1. Monthly rate of change for education COICOP (m-o-m percentage change, monthly data) from Jan. 1997 to Dec. 2015 (Poland).

		M1	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
	January	1	1	1	1	1	1	1	1	1
	February	1	1	1	1	1	1	1	1	1
	⋮									
	July	1	1	1	1	1	1	1	1	1
σ_t	August	1	1	1	1	1	1	1	1	1
	September	1	1	1	1	1	2	2	2	2
	October	1	1	1	1	1	2	3	2	3
	November	1	1	1	1	1	1	1	1	1
	December	1	1	1	1	1	1	1	1	1
	January	1	1	1	1	1	1	1	1	1
	February	1	1	1	1	1	1	1	1	1
	⋮									
	July	1	1	1	1	1	1	1	1	1
γ_t	August	1	1	1	1	1	1	1	1	1
	September	1	2	2	1	1	1	1	2	2
	October	1	2	3	1	1	1	1	2	3
	November	1	1	1	1	1	1	1	1	1
	December	1	1	1	1	1	1	1	1	1
	January	1	1	1	1	1	1	1	1	1
	February	1	1	1	1	1	1	1	1	1
	⋮									
	July	1	1	1	1	1	1	1	1	1
α_t	August	1	1	1	1	1	1	1	1	1
	September	1	1	1	2	2	1	1	2	2
	October	1	1	1	2	3	1	1	2	3
	November	1	1	1	1	1	1	1	1	1
	December	1	1	1	1	1	1	1	1	1

Table 1. Models parameters characteristics (under consideration).

We divide our sample into two parts: a training and a forecasting period. For the estimation we use 13-year rolling window. We start the estimation using data set up to Dec. 2009. After each new predictive distribution evaluation we add next observation and delete last one. Therefore, the length of the sample is constant over time (n=168). Finally, we collect sixty predictive distributions for 12 month ahead. For nowcasting the average logarithmic score was calculated (see Fig. 2). To compare logarithmic score for different models we use the test proposed by Amisano and Giacomini (2007). At significance level 5% in group of first five models we cannot reject null hypothesis that the logarithmic scores are different (comparison by pairs). The same conclusion concerns group of models M6-M9. Finally, if we compare any model from group M1-M5 with any model from group M6-M9 then we reject null hypothesis assuming equal logarithmic score, against alternative hypothesis that chosen model from group M6-M9 has higher logarithmic score. Summing up, the models with time-varying periodic in time variance of white noise improve forecasting performance over alternative models with constant variance of error term.

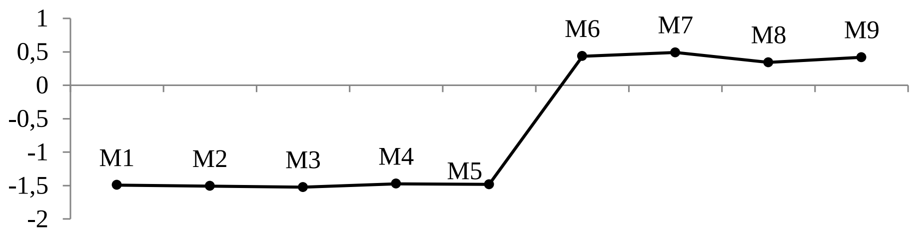


Fig. 2. Average logarithmic score for different considered models based on sixty predictive distributions (nowcasting).

Conclusions

In this paper we show the exact form for the linear innovations model with time-varying periodic in time parameters. It is Periodic Autoregressive Integrated Moving Average form with equality restrictions on parameters. Specific models ware considered with level, trend and seasonal pattern. The real data example was considered with monthly time series concerning price in education in Poland. We consider problem of forecasting performance based on logarithmic score rule. Our main findings from real data example is that in such case models with periodic in time variance of error term outperforms the considered models with constant variance of error terms. In addition, from statistical point of view (at significant level 5%) allowing seasonality only in smoothing parameters in level (α_t) and seasonal pattern (γ_t) doesn't produce better forecasting preference. Note that alternative model specifications should be considered and compared. Also the test proposed by Amisano and Giacomini (2007) should be adjust to the periodic case.

Acknowledgements

Publication was financed from the funds granted to the Faculty of Finance and Law at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Amisano G. & Giacomini. R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, 25, 177-190.

Berument, M. H. & Sahin, A. (2010). Seasonality in inflation volatility: Evidence from Turkey. *Journal of Applied Economics*, 13(1), 39–65.

Burridge, P., & Taylor, A.M. (2001). On regression-based tests for seasonal unit roots in the presence of periodic heteroscedasticity. *Journal of Econometrics*, 104, 91–117.

Doshi, A., & Frank, J., & Thavaneswaran, A. (2011). Seasonal volatility models. *Journal of Statistical Theory and Applications*, 10(1), 1–10.

Franses, P. H. & Paap, R. (2004). *Periodic Time Series Models*. Oxford University Press, New York.

Gladyshev, E. G. (1961). Periodically correlated random sequance. *Sov. Math.*, 2, 385–388.

Hannan, E. J. (1955). A test for singularities in Sydney rainfall. *Australian Journal of Physics*, 8, 289–279.

Hurd, H. L., & Miamee, A. G. (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice*. Wiley, Hoboken, New Jersey.

Hyndman, R. J., & Koehler, A. B., & Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Springer, Berlin.

Koopman, S. J., & Ooms, M. (2008). Forecasting daily time series using periodic unobserved components time series models. *Computational Statistics & Data Analysis*, 51, 885–903.

Lenart, Ł. (2017). Examination of Seasonal Volatility in HICP for Baltic Region Countries: Non-parametric Test versus Forecasting Experiment. *Central European Journal of Economic Modelling and Econometrics*, 9(1), 29–67.

Novales, A. & Fruto, R. F. (1997). Forecasting with periodic models, a comparison with time invariant coefficient models. *International Journal of Forecasting*, 13, 393–405.

Osborn, D. R. (1991). The implications of periodically varying coefficients for seasonal time-series processes. *J. Econometrics*, 48, 373–384.

Pagano, M. (1978). On periodic and multiple autoregression. *The Annals of Statistic*, 6(6), 1310–1317.

Proietti, T. (2004). Seasonal specific structural time series. *Studies in Nonlinear Dynamics & Econometrics*, 8(2).

Trimbur, T. M., & Bell, W. R. (2012). Seasonal Heteroskedasticity in Time Series Data: Modeling, Estimation, and Testing, In: W. R. Bell & S. H. Holan & T. S. McElroy (Eds.), *Economic Time Seires*, CRC Press.

Wyłomańska, A. (2008). Spectral measures of PARMA sequences. *Journal of Time Series Analysis*, 29(1), 1–13.

Business cycle analysis with short time series: a stochastic versus a non-stochastic approach

Łukasz Lenart[1], Błażej Mazur[2]

Abstract

The idea of modelling of business cycle fluctuations using an autoregressive model with complex roots in its characteristic polynomial is well established in econometric analysis. It can be shown that such an approach induces a spectral density function with concentration of mass around a value (a frequency) related to a complex root. The resulting stochastic process is often referred to as a stochastic cycle model. However, there exists an alternative approach where business cycles are generated by time-varying mean which is driven by a deterministic function (a deterministic cycle approach). It is often claimed that the deterministic cycle is not sufficient for description of actual economic fluctuations due to its inflexibility. However, it is not clear that it is true for short series of data available for e.g. the emerging economies. In order to deal with the problem we consider a general model that encompasses both mechanisms (that of stochastic cycle and that of deterministic cycle). We illustrate applicability of the model analysing Polish data on GDP growth rates.

Keywords: *business cycles, time-series analysis, Flexible Fourier Form, empirical macroeconomics*

JEL Classification: C22, E32

1 Introduction

A concept of stochastic cycle, used as a tool for investigation of business cycle fluctuations, is considered in the literature (see for instance: Harvey and Trimbur, 2003; Trimbur, 2006; Koopman and Shephard, 2015; Pelagatti, 2016). For such a stationary univariate process its autocovariance function decays to zero with some "pseudo-period". It is often assumed that the cyclical component is a stationary ARMA process with complex conjugate roots in the characteristic polynomial. An extension to a multivariate model with cyclical fluctuations was considered in Azevedo et al. (2006), see also a trivariate example in Harvey et al. (2007). Koopman and Azevedo (2008) consider a multivariate model with stationary multiple cyclical process with common frequency at each coordinate.

On the other hand, one might consider its deterministic counterpart using a tool similar to the Flexible Fourier Form; see Gallant (1981), though the approach is less popular in the applied macro literature. For example Harvey (2004) considers y_t of the form:

[1] Cracow University of Economics, Department of Mathematics, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: lukasz.lenart@uek.krakow.pl.

[2] Corresponding author: Cracow University of Economics, Department of Econometrics and Operations Research, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: blazej.mazur@uek.krakow.pl.

$$y_t = a\cos(\lambda_c t) + b\sin(\lambda_c t) + \varepsilon_t, \tag{1}$$

where λ_c is a frequency related to a cycle with period length of $2\pi/\lambda_c$; a and b are parameters and $\{\varepsilon_t\}$ represents a white noise process. The right-hand side of (1) can be interpreted as a restricted Flexible Fourier Form consisting of just one component. In general the deterministic model (1) is considered to be insufficient for description of actual business cycle fluctuations due to its inflexibility. However, it might be useful for analysis of short time-series that are available for emerging economies. To verify that we consider a model that encompasses the two concepts mentioned above. Note that a more general form of model (1) with multiple frequencies was considered in a nonparametric framework in Lenart and Pipień (2013a), Lenart and Pipień (2013b); Lenart and Pipień (2015) consider the idea of a deterministic cycle in examining the empirical properties of credit and equity cycle.

2 Stochastic cycles within an AR(p) model

Consider a general and zero-mean autoregressive process of order p, denoted by AR(p):

$$\Psi(L)y_t = \varepsilon_t, \tag{2}$$

with polynomial $\Psi(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_P L^P$ and a white noise process $\{\varepsilon_t\}$ with variance σ_ε^2. Let $\alpha_1, \alpha_2, ..., \alpha_{p_1}$ be a set of distinct values taken by the real roots, with multiplicities denoted by $k_1, k_2, ..., k_{p_1}$. By $\beta_1, \beta_2, ..., \beta_{p_2}$ we denote the vector of all (different) values of complex roots with multiplicities denoted by $l_1, l_2, ..., l_{p_2}$. Hence,

$$\Psi(L) = \left(\prod_{i=1}^{p_1}(1 - L/\alpha_i)^{p_i}\right)\left(\prod_{i=1}^{p_2}[(1 - L/\beta_i)(1 - L/\overline{\beta}_i)]^{l_i}\right).$$

The spectral density function of the process (defined on the interval (0, π)) is of the form:

$$f(\lambda) = \sigma_\varepsilon^2 (2\pi)^{-1}\left(\prod_{i=1}^{p_1}|1 - e^{-i\lambda}/\alpha_i|^{2p_i}\right)^{-1}\left(\prod_{i=1}^{p_2}[(1 - e^{-i\lambda}/\beta_i)(1 - e^{i\lambda}/\overline{\beta}_i)]^{l_i}\right)^{-1}. \tag{3}$$

Cyclical behaviour of the above processes is linked with a strong concentration of the spectral density in a neighbourhood of some frequency λ_0 (see for example Pelagatti, 2016). This frequency is one of the main quantities characterizing properties of cyclical fluctuations in such model.

Recall from the basic literature (see for example Brockwell and Davis, 2002) that for a simple AR(1) model with $\theta_1 < 0$, the spectral density function $f(\lambda)$ is an increasing function on (0, π), while for $\theta_1 > 0$, $f(\lambda)$ is decreasing function on (0, π). For an AR(2) model with two real roots, $f(\lambda)$ has local extremes at points: 0, π and might have one minimum in the interval $(0,\pi)$, while in the case of two complex (conjugate) roots $f(\lambda)$ reaches one maximum in the interval $(0,\pi)$ (see Brockwell and Davis, 2002). Therefore, in order to obtain a spectral density function with mass concentration in neighbourhood of some frequency one might consider autoregressive models with complex (conjugate) roots. This idea is well known in the literature, see Harvey and Trimbur (2003), Trimbur (2006), Koopman and Shephard (2015) and many others.

Let us consider the following AR(2) model with complex roots:

$$\Psi_z(L)y_t = \varepsilon_t \tag{4}$$

where $\Psi_z(L) = (1-zL)(1-\bar{z}L)$ and $z \in \mathrm{Z}$. Using a polar form $z = \rho e^{i\omega}$ we obtain:

$$\Psi_z(L) = \Psi_{(\rho,\omega)}(L) = (1-2Re[z]L + |z|^2 L^2) = (1-2\rho\cos(\omega)L + \rho^2 L^2). \tag{5}$$

with associated spectral density function (defined on the interval $(0,\pi)$) of the form:

$$f_y(\lambda) = \sigma_\varepsilon^2 (2\pi)^{-1}(1-2\rho\cos(\lambda+\omega)+\rho^2)^{-1}(1-2\rho\cos(\lambda-\omega)+\rho^2)^{-1}. \tag{6}$$

This is a basic model used in empirical modelling of cyclical fluctuations – it was examined in the first part of the last century by Samuelson (1939). Some extension to the AR(3) model with lag polynomial $(1-\alpha L)(1-zL)(1-\bar{z}L)$, where α is real number and z is a complex number was examined by Geweke (1988, 2016).

3 Cyclical properties of AR(*p*) model with multiple complex roots

The spectral density function (5) reaches a maximum at $\lambda_0 = \arccos\left((1+\rho^2)\cos(\omega)/(2\rho)\right)$, see Pelagatti (2016), p. 185 for similar computations. We propose the following reparametrization: $\omega = \arccos(2\rho\cos(\tilde{\omega})/(1+\rho^2))$. Consequently, the spectral density function has the form:

$$f_y(\lambda) = \sigma_\varepsilon^2 (2\pi)^{-1}\left(1-2\rho\cos\left(\lambda + g(\rho,\tilde{\omega})\right)+\rho^2\right)^{-1}\left(1-2\rho\cos\left(\lambda - g(\rho,\tilde{\omega})\right)+\rho^2\right)^{-1}, \tag{7}$$

with $g(\rho,\tilde{\omega}) = \arccos(2\rho\cos(\tilde{\omega})/(1+\rho^2))$. The spectral density function $f_y(\lambda)$ has a maximum at $\lambda_0 = \tilde{\omega}$. Note that another univariate cyclical process considered in the empirical economic literature, an ARMA(2,1) process with complex roots in the AR part, has spectral density function that differs from (6). The most important common properties of these two cyclical processes is that both spectral densities have maxima at known points. Note that for (6), $\lim_{|\rho|\to 1} f_y(\tilde{\omega}) = \infty$ which means that ρ determines the concentration of the spectral mass in the neighbourhood of frequency $\tilde{\omega}$.

In order to concentrate the spectral mass in the neighbourhood of the frequency of interest (with fixed ρ), Trimbur (2006) considers so-called *n*-th order cycle. In this paper we introduce an approach that is somewhat simpler. Our concept follows from purely mathematical properties of the spectral density function of an AR(*p*) process. Consider the general form:

$$\Psi^n_{(\rho,\tilde{\omega})}(L) y_{t,n} = \varepsilon_t. \tag{8}$$

Assuming $|\rho| < 1$, the above model is a stationary and inventible AR(2*n*) process. The spectral density function has the following form:

$$f_{y_n}(\lambda) = \sigma^2_\varepsilon (2\pi)^{-1} \left(1 - 2\rho\cos\left(\lambda + g\left(\rho, \tilde{\omega}\right)\right) + \rho^2\right)^{-n} \left(1 - 2\rho\cos\left(\lambda - g\left(\rho, \tilde{\omega}\right)\right) + \rho^2\right)^{-n}, \tag{9}$$

and it has a maximum at $\lambda_0 = \tilde{\omega}$, which means that the mass is concentrated in the neighbourhood of frequency $\tilde{\omega}$. Note that for $n \to \infty$ this maximum tends to infinity. Moreover, higher values of *n* imply stronger concentration around the maximum. The above concept is used here to introduce a stochastic cycle with known properties, controlled directly by model parameters.

4 A final model with stochastic and deterministic cycle

Finally we consider an AR(*p*) model with a Flexible Fourier-type time-varying mean μ_t:

$$\left(1 - \phi_1 L\right)\left(1 - \phi_2 L\right) \Psi^n_{(\rho,\tilde{\omega})}(L)(y_t - \mu_t) = \varepsilon_t, \tag{10}$$

where $\varepsilon_t \sim iiN\left(0, \sigma^2_\varepsilon\right)$ and

$$\mu_t = \delta_0 + a\sin\left(t\zeta\right) + b\cos\left(t\zeta\right), \tag{11}$$

and the latter corresponds to an almost periodic function with one frequency. We assume: $\delta_0, a, b \in R$, $\zeta, \tilde{\omega} \in (0, \pi)$, $-1 < \phi_i < 1$ for i = 1,2, $0 < \rho < 1$. The function μ_t reflects the deterministic cycle (1), while the polynomial $\Psi^n_{(\rho,\tilde{\omega})}(L)$ is related to the idea of stochastic cycle. The part $\left(1 - \phi_1 L\right)\left(1 - \phi_2 L\right)$ is related to the aspects of the dynamics of y_t that are not directly connected with its cyclical properties. Note that the spectral density function corresponding to the element $\left(1 - \phi_1 L\right)\left(1 - \phi_2 L\right)$ has a mass concentration around frequency 0 or π or around both. Parameters of the autoregressive part of the models are functions of $\tilde{\omega}$, ρ, ϕ_1 and ϕ_2.

The model is estimated using Bayesian techniques. As to the prior specification, we assume independent uniform priors for ζ, $\tilde{\omega}$, ρ, ϕ_1 and ϕ_2. As to δ_0, a, b, the priors are independent Student-*t*, and precision of the error process is a priori distributed as Gamma[3]. The estimation is undertaken using MCMC techniques (a random-walk Metropolis-Hastings algorithm). The inference is relatively easy as we consider only one frequency in the deterministic part. Bayesian estimation of deterministic cycle models with many frequencies is more complicated, as discussed by Lenart and Mazur (2016), see also Lenart et al. (2016).

5 Real data example

For the sake of illustration of the above concepts we analyse a series of quarterly growth rates of Polish GDP ([%], y-o-y, seasonally adjusted, 1999Q1-2015Q4, 68 observations)[4].

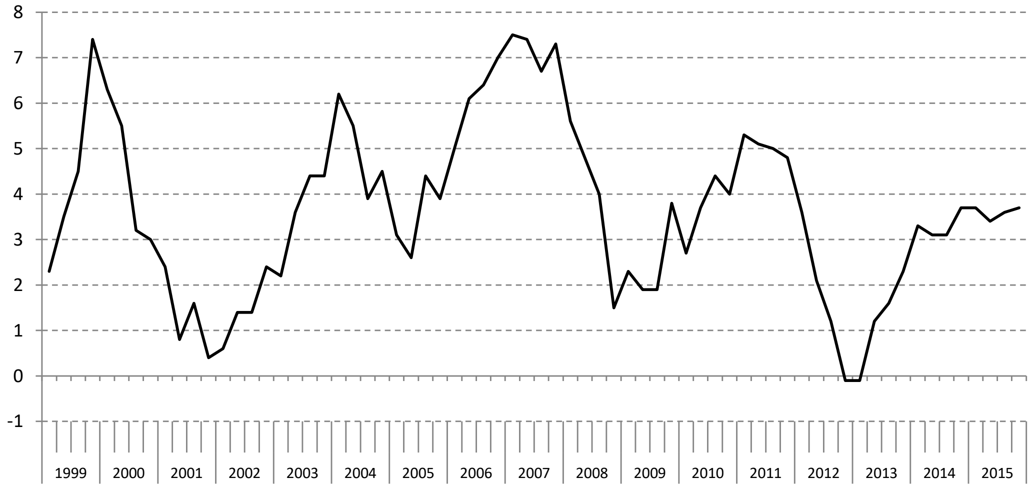


Fig. 1. GDP (y-o-y percentage change, quarterly data) from 1999Q1 to 2015Q4 (Poland).

In what follows we consider a sequence of models, beginning with a simple model with stochastic cycle only. It is then augmented as the deterministic cyclical component μ_t is introduced. Moreover, additional autoregressive parameters ϕ_1 and ϕ_2 (generating non-oscillatory behaviour) are also considered. For the additional autoregressive parameters we consider two cases with different inequality restrictions. The restrictions rule out negative values of parameters. In the general case there are two parameters taking values in the frequency domain that control the cyclical properties of the model: ζ for the deterministic

[3] Details of the prior specification and the estimation procedure are not elaborated here due to the space constraints. However, the details are available from the authors upon request.
[4] The data source is CEIC database accessed in March 2016.

cycle and $\tilde{\omega}$ for the stochastic cycle. Moreover, we assume that n = 2; A summary of specifications under consideration is given in Table 1.

Case no., *p*	**Stochastic cycle**	**Deterministic cycle**	**Real roots**	**Complex roots**
1, p = 6	Yes	No	$-1 < \phi_i < 1$	2
2, p = 6	Yes	No	$0 < \phi_i < 1$	2
3, p = 4	Yes	Yes	No	2
4, p = 6	Yes	Yes	$0 < \phi_i < 1$	2
5, p = 6	Yes	Yes	$-1 < \phi_i < 1$	2

Table 1. Models under consideration.

Posterior means for model components representing stochastic and deterministic cycles are reported in Fig.2. Shapes of marginal posterior distributions for selected model parameters are given in Fig. 3 (for all the cases in the figure priors are uniform). Moreover, Table 2 includes basic characteristics of posterior distributions for the standard deviation of the error term. Here we do not make an attempt to conduct a formal comparison of the nested cases. This is mostly because short macroeconomic series are not very informative so results of the formal comparison are seriously affected by prior assumptions.

Analysis of Fig. 2 suggests that the dynamic behavior of Polish GDP (in terms of the growth cycle) includes at least two different components. The first one represents regular fluctuations with period of approximately four years. It is captured by the stochastic part in the cases 1, 2 and 4 and by the deterministic part in cases 3 and 5. The other component is less regular and represents much longer cycles (lower frequencies). It is captured by the deterministic mechanism in case 4 and by the stochastic part in cases 3 and 5. The first component seems to be the dominant one.

It is interesting to ask which of the two components mentioned above is more likely to correspond to the stochastic cyclical mechanism. Interestingly, it seems to depend on the non-complex roots of the characteristic polynomial. If the roots are unrestricted and the deterministic component is present in the model, it takes over the fluctuations with period of approximately four years. However, if negative values of the parameters are ruled out (as in cases 2 and 4), the fluctuations are pinned down by the stochastic mechanism, while the deterministic part represents cycles characterized by longer period.

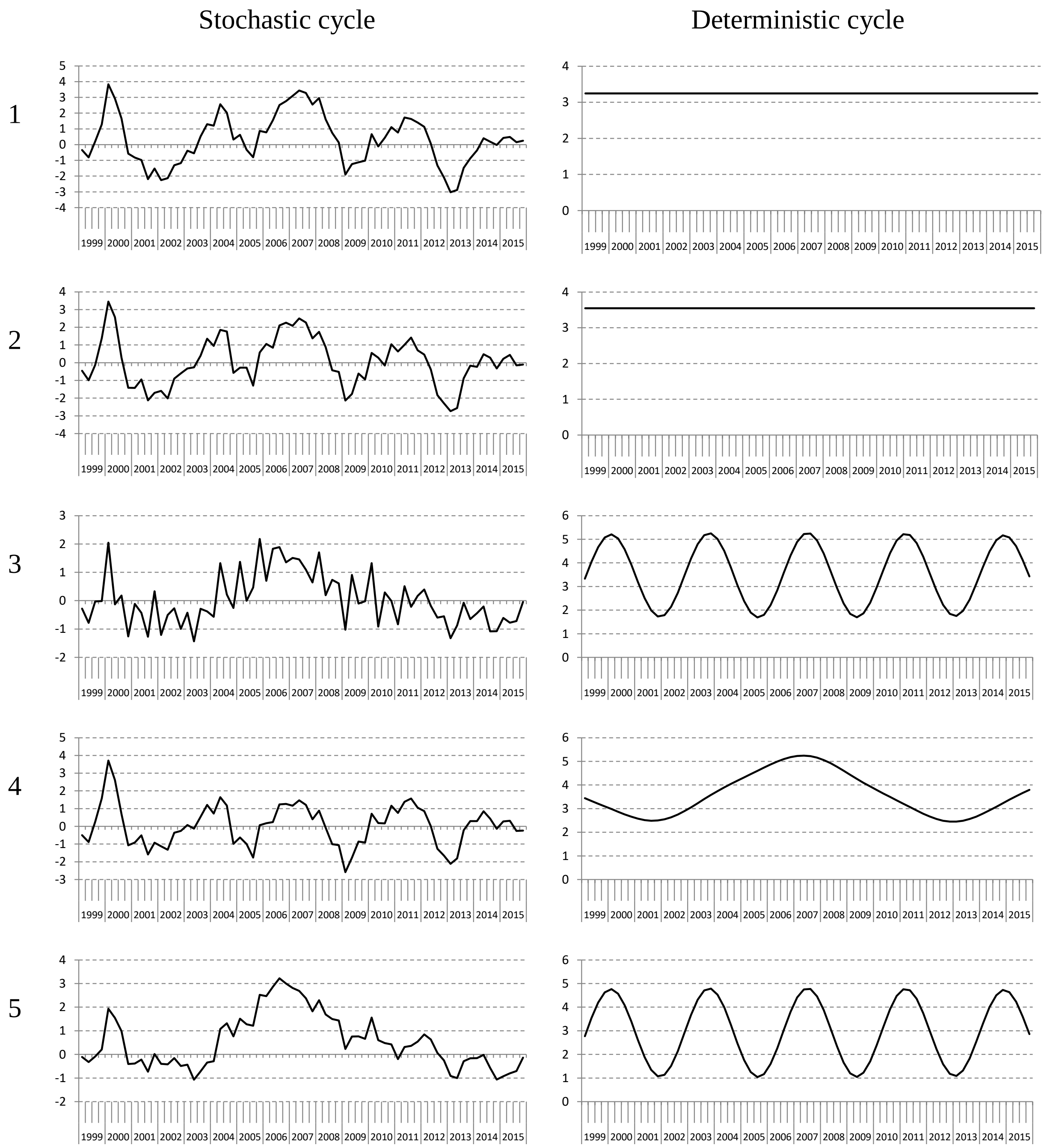


Fig. 2. Posterior means for (conditional) stochastic and deterministic components.

Similar effect is visible in Fig. 3, where higher values of ρ, corresponding to stochastic cycles with larger "amplitudes" can be seen in cases 2 and 4. The parameter $\tilde{\omega}$ is never separated from its lower bound. The posterior distribution of ζ in case 4 is bimodal, though the dominant mode represents fluctuations with longer period.

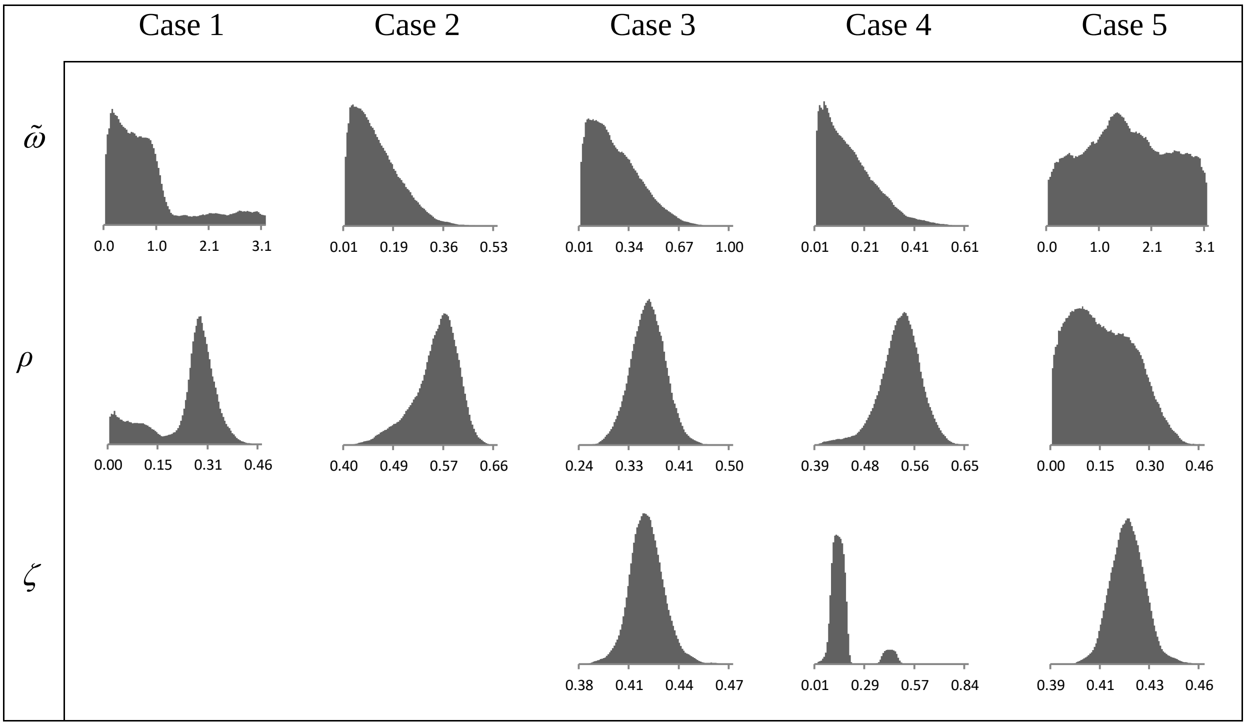


Fig. 3. Marginal posterior distributions for selected model parameters.

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**
mean	0.983	1.234	1.131	1.007	0.762
standard deviation	0.184	0.235	0.219	0.185	0.146

Table 2. Characteristics of posterior distribution for square root of σ_{ε}^{2}.

Analysis of Table 2 suggests that the most general specification (case 5) is characterized by the lowest standard deviation of the error term. This informally confirms empirical relevance of the general model considered here.

In general the model proposed here leads to empirical conclusions that are in line with other research results as to Polish business cycle fluctuations. A summary of empirical results (based on the industrial production index) is given by Lenart et al. (2016). In particular it seems that for the series at hand both the stochastic and the deterministic mechanisms are empirically relevant. Moreover, the decomposition of the GDP dynamics seems to be sensitive to various aspects of model assumptions.

Conclusions

As to theoretical contribution of the paper, we develop a univariate time-series model that encompasses two approaches to modelling of macroeconomic cyclical fluctuations that are present in the applied macro literature. In one approach the cycle in mean is approximated using a deterministic concept that can be interpreted in terms of the Flexible Fourier Form. Within the other approach cycles are driven by complex roots of the characteristic polynomial of a stochastic AR(p) process. We impose restrictions on the AR part of the process in order to ensure that the spectral density function of the autoregressive part has maximum at a given point in the frequency domain. Additional equality restrictions are imposed in order to ensure concentration of the mass around that value. We also introduce parameters corresponding to real roots. The resulting model has two distinct sources of cyclical fluctuations with two (potentially different) frequencies, which provides an advantage compared to models that are most often used in the applied work (often allowing for just one frequency).

As to the empirical part, we illustrate properties of the model analysing Polish data on GDP dynamics. The purpose is to check whether the deterministic cycle is relevant for the countries for which only short series are available (like the emerging economies). The empirical results we obtain are in line with other analyses of Polish business cycle fluctuations. Moreover, both the stochastic and the deterministic model components seem to be empirically important. Verification of the out-of-sample performance (or predictive consequences) of various nested cases is left for further research.

Acknowledgements

This research was supported by the National Science Centre (NCN) research grant (decision DEC-2013/09/B/HS4/01945).

References

Azevedo, J. V., Koopman, S. J., & Rua, A. (2006). Tracking the business cycle of the euro area: a multivariate model-based band-pass filter. *Journal of Business & Economic Statistics*, 24(3), 278–290.

Brockwell, P. J., & Davis, R.A. (2002). *Time Series: Theory and Methods*. Springer-Verlag, New York.

Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: the flexible Fourier Form. *Journal of Econometrics*, 15, 211–245.

Geweke, J. (1988). The Secular and Cyclical Behavior of Real GDP in 19 OECD Countries. *Journal of Business & Economic Statistics*, 6(4), 479–486.

Geweke, J. (2016). Sequentially Adaptive Bayesian Learning for a Nonlinear Model of the Secular and Cyclical Behavior of US Real GDP. *Econometrics*, 4(10), 1–23.

Harvey, A. C. (2004). Tests for cycles. In: A. C. Harvey & S. J. Koopman & N. Shephard (Eds.), *State space and unobserved component models*, CUP.

Harvey, A. C., Trimbur, T. M., & Van Dijk, H. K. (2007). Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics*, 140, 61-649.

Harvey, A. C., & Trimbur, T.M. (2003). General model-based filters for extracting cycles and trends in economic time series. *The Review of Economics and Statistics*, 85(2), 244–255.

Koopman S. J., & Azevedo, J. V. (2008). Measuring synchronization and convergence of business cycles for the euro area, UK and US. *Oxford Bulletin of Economics and Statistics*, 70(1), 23–51.

Koopman, S. J., & Shephard, N. (2015). *Unobserved Components and Time Series Econometrics*. Oxford University Press, Oxford.

Lenart, Ł., & Mazur, B. (2016). On Bayesian Inference for Almost Periodic in Mean Autoregressive Models. *Statistical Review* / *Przegląd Statystyczny*, 63(3), 255-271.

Lenart, Ł., Mazur, B., & Pipień, M. (2016). Statistical Analysis of Business Cycle Fluctuations in Poland Before and After the Crisis. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(4), 769-783.

Lenart, Ł., & Pipień, M. (2013a). Almost periodically correlated time series in business fluctuations analysis. *Acta Physica Polonica A*, 123(3), 567–583.

Lenart, Ł., & Pipień, M. (2013b). Seasonality revisited - statistical testing for almost periodically correlated processes. *Central European Journal of Economic Modelling and Econometrics*, 5, 85–102.

Lenart, Ł., & Pipień, M. (2015). Empirical Properties of the Credit and Equity Cycle within Almost Periodically Correlated Stochastic Processes - the case of Poland, UK and USA. *Central European Journal of Economic modelling and Econometrics*, 7, 168–186.

Pelagatti, M. M. (2016). *Time Series Modelling with Unobserved Components*. Taylor & Francis Group, Boca Raton.

Samuelson, P. A. (1939). Interactions between the multiplier analysis and the principle of acceleration. *Review of Economics and Statistics*, 21, 75–78.

Trimbur, T. M. (2006). Properties of higher order stochastic cycles. *Journal of Time Series Analysis*, 27, 1–17.

The evaluation of convergence between content description of academic textbooks in the form of textual notes and UDC expressions

Paweł Lula[1], Urszula Cieraszewska[2], Monika Hamerska[3]

Abstract

In the era of dynamic development of e-commerce solutions product descriptions available online have crucial importance for customers. Therefore, product characterisation should reflect its core features. In the paper the problem of the convergence between the content of available online product descriptions and their significant features is discussed. The analysis presented here was concentrated on the evaluation of the adequacy between prepared in the Polish language information about academic textbooks published by online bookstores and main topics presented in them.

In the research the Universal Decimal Classification (UDC) system was used for representation of essential content of every monograph. Whereas vector space model was used for representation of online available book descriptions. Using these two methods of content description cluster analysis of monographs was performed. Assuming that a dendrogram generating by a cluster method reflects the content-based relationships between monographs, the similarity analysis of obtained dendrograms can be treated as a tool of convergence evaluation between information content of monograph descriptions having two different form: UDC expressions and textual descriptions. Generally, the research outputs allow to assess the convergence between the content of published online textual description of academic textbooks and their core elements represented by UDC codes.

Keywords: *text mining, universal decimal classification (UDC) system, similarity of textual documents, similarity of cluster analysis results*

JEL Classification: C630, C88

1 Introduction

The analysis of convergence between information content of textual description of academic monographs and their bibliographic characteristics having the form of UDC expressions is the main goal of the research. First section of the paper presents Universal Decimal Classification (UDC) scheme and the algorithm of similarity calculation between two UDC expressions. The problem of expressing the similarity of textual documents is presented in the third section. Next, in the fourth section, the Fowlkes-Mallows index as a tool for similarity analysis between dendrograms is briefly presented. The fifth section of the paper presents the

[1] Corresponding author: Cracow University of Economics, Department of Computational Systems, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: pawel.lula@uek.krakow.pl.

[2] Cracow University of Economics, Main Library of the Cracow University of Economics, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: cieraszu@uek.krakow.pl.

[3] Cracow University of Economics, Science and Knowledge Transfer Department, 27 Rakowicka St., 31-510 Cracow, Poland, e-mail: hamerskm@uek.krakow.pl.

results of empirical research concerning the analysis of convergence between textual descriptions of academic textbooks offered by Polish online bookstores and their content represented by UDC descriptions. Conclusion presents main findings of the research.

2 Similarity of UDC classes and UDC expressions

Universal Decimal Classification (UDC) is a scheme of classification which is used for description of library resources (McIlwaine, 1997). It is worth to underline that the UDC system can be treated as a model of the whole knowledge and its structure (Rafferty, 2001). The knowledge is represented by the hierarchical structure (the knowledge tree) composed of nodes corresponding to fields or subfields of the knowledge. Every node in the knowledge tree is described by a UDC class identifier. The UDC class can be treated as an essential element of the knowledge. Every class has an identifier which have form of a sequence of digits which reflects the position of a given class in the knowledge tree. The knowledge is divided into fields represented by main classes indicated by one-digit identifiers (from 0 to 9). Descendants of every main class (which represent subfields of fields described by the main classes) have two-digit identifiers (the first digit represents the ancestor class and the second is a consecutive number of a descendant). This schema is used also for lower levels of the knowledge tree. For clarity a dot symbol is put between every group of three digits.

Every publication is characterized by an UDC expression which is composed of one or more UDC class identifiers and auxiliary symbols. Auxiliary symbols describe form, language, place of issue of the publication or define particulars of UDC classes (e.g. related to time or localization).

The problem of similarity calculation between UDC expressions can be solved provided that the measure of similarity between individual classes is defined. It causes that an algorithm for similarity calculation between UDC expressions should be preceded by a short description of the problem of similarity between UDC classes.

It can be assumed that an identifier of a given UDC class is represented by a sequence of digits:

$$c_i = n_1 n_2 n_3 \ldots n_N \,. \tag{1}$$

To calculate similarity between two classes: c_i and c_j first the identifier of the class which is the lowest common ancestor (LCA) of c_i and c_j is calculated. It is consisted of the leftmost digits which are common for the identifiers of both classes. Next the formula for similarity calculation proposed by Lin can be used (Lin, 1998):

$$sim(c_i, c_j) = \frac{2 \times \ln(p(c_{LCA}))}{\ln(p(c_i)) + \ln(p(c_j))} \tag{2}$$

where $p(c_i)$ is a probability of occurring of the symbol c_i.

In order to estimate the probability of occurrence of identifiers of UDC classes the analysis of the catalogue of the National Library of Poland was performed. The value of probability was calculated using the formula:

$$p(c_i) = \frac{n(c_i)}{N} \tag{3}$$

where $n(c_i)$ is a number of occurrence of the symbol c_i and N is a number of all UDC codes in the part of the catalogue which was taken into account during calculation.

The procedure of similarity estimation between UDC expressions is composed of two steps. First, identifiers of all UDC classes from a given expression are extracted. Next, a similarity coefficient is calculated.

Having two UDC expressions: $expr_1$ and $expr_2$ two sets can be defined: a set S_1 containing class identifiers extracted from $expr_1$:

$$S_1 = \{c_1, c_2, ..., c_N\} \tag{4}$$

and a set S_2 with identifiers extracted from $expr_2$:

$$S_2 = \{c_1, c_2, ..., c_M\}. \tag{5}$$

Then the similarity calculation between expressions $expr_1$ and $expr_2$ can be performed with the use of the formula (6) (Lula et al., 2014):

$$sim(expr_1, expr_2) = \frac{\sum_{i=1}^{N} \min_j (sim(c_i, c_j)) + \sum_{j=1}^{M} \min_i (sim(c_i, c_j))}{N + M}. \tag{6}$$

The formula (6) allows to build a similarity matrix for a given set of UDC expressions. Because all similarity coefficients are normalized to the range $[0;1]$ it possible to convert them into measures of distances subtracting them from value one. Matrix of distances calculated with the use of the algorithm presented above allows to perform cluster analysis of monographs based on their UDC codes.

3 Similarity of textual descriptions of monographs

Let’s assume, that a set of book descriptions prepared in the Polish language is available. To perform a comparison analysis of information content of descriptions, a similarity matrix can be calculated. The calculation of similarity matrix between descriptions consists of several stages (Manning and Schütze, 1999; Nasukawa and Nagano, 2001; Tuchowski et al., 2011):

1. Converting all texts into plain text format using UTF-8 coding system,
2. Performing a lemmatization process. During this step of analysis for every word its dictionary form (lemma) is identified. For lemmatization of documents in Polish several software packages are available (standalone applications or libraries which can be linked to other pieces of software). In the research describe here, Morfologik[4] package was used.
3. Creating a document-term matrix with rows representing documents and columns corresponding to words which appear at least in 2 and not more than in 6 documents with elements calculated as:

$$w_{ij} = f_{ij} \times \log\left(\frac{N}{n_j}\right) \tag{7}$$

   where f_{ij} indicates how many times a word j occurs in a document i, n_j is a number of documents containing a word j and N is a number of all documents.
4. Calculation the distance matrix for documents represented by rows of document-term matrix using Euclidean formula.

The matrix of distances constitutes the basis for cluster analysis of textual description of monographs.

4 Comparison of hierarchical clustering results with the Fowlkes-Mallows schema

The Fowlkes-Mallows schema for comparison of clustering results is presented in (Halkidi et al., 2001). Having results of a hierarchical clustering method for a set of N objects we can obtain the division of objects into k groups (where $k = 2,..., N-1$).

To compare the results of two clusterings Fowlkes and Mallows propose to calculate similarity index between divisions into the same number of groups obtained in both clustering processes. Taking into account two divisions of objects into k groups the Fowlkes-Mallows similarity index can be defined as:

$$FM(k) = \sqrt{\frac{T}{T+F_1} \times \frac{T}{T+F_2}} \tag{8}$$

[4] http://morfologik.blogspot.com/.

where:

T - number of objects that fall into the same groups in two studied divisions,

F_1 - number of objects that fall into the same group in the first division and simultaneously fall into different groups in the second division,

F_2 - number of objects that fall into different groups in the first division and simultaneously fall into the same group in the second division.

The value of $FM(k)$ is normalized to the range $[0;1]$ and its higher value indicates a greater similarity of two studied divisions. The similarity between two clusterings can be shown by analysing and plotting $FM(k)$ versus k for all $k = 2,..., N-1$.

5 Empirical analysis of convergence between UDC codes and textual description of academic textbooks

The set of arbitrarily chosen academic monographs considered in the empirical research was composed of 18 books related to the field of social sciences. The list presented below contains main information (author, title and UDC description) about every publication. Numbers assigned to consecutive positions on the list will serve as book identifiers.

1. Mruk H, Rutkowski I., *Strategia produktu*, 339.138:658.1/.5::66/69
2. Mazurek-Łopacińska K., *Badania marketingowe. teoria i praktyka*, 339.138(075.8)
3. Florek L., *Prawo pracy*, 349.2(438)(075.8)
4. Osińska M., *Ekonometria finansowa*, 330.43:336:51](075.8)
5. Waltoś S., *Proces karny. Zarys systemu*, 343.13(438)(075.8)
6. Dmowski A., et al., *Podstawy finansów i bankowości*, 336(075.8)
7. Mróz T., Stec M., *Prawo gospodarcze prywatne*, 346:347.44:347.7](438)(075.8)
8. Czarny E., *Mikroekonomia*, 330.101.542(075.8)
9. Flejterski S. et al., *Współczesna ekonomika usług*, 338.46(075.8)
10. Kończak G., Trzpiot G., *Metody statystyczne z wykorzystaniem programów komputerowych*, 311:004.42](075.8)
11. Budnikowski A., *Międzynarodowe stosunki gospodarcze*, 339.9(075.8)
12. Altkorn J., et al., *Podstawy marketingu*, 339.138(075.8)
13. Cziomer E., Zyblikiewicz L., *Zarys współczesnych stosunków międzynarodowych*, 327(4)"19"(075.8):[327.51:355.3:061.1A/Z](100-622)(075.8)
14. Witkowska D., *Podstawy ekonometrii i teorii prognozowania*, 330.43:338.27](075.8)
15. Osiatyński J., *Finanse publiczne. Ekonomia i polityka*, 336.1(075.8)

16. Błaszczuk D., *Wstęp do prognozowania i symulacji*, 338.27:330.43:519.876.5](075.8)

17. Ratajczak M., *Współczesne teorie ekonomiczne*, 330.83(075.8)

18. Florek M., *Podstawy marketingu terytorialnego*, 339.138:332.1](075.8)

In the first stage of analysis the similarity matrix between UDC descriptions was calculated (using the algorithm presented in the section 2). Probabilities of occurrence for UDC classes were estimated on the grounds of the part of the catalogues of the National Library of Poland which store the description of library resources in the field of social sciences (represented by the main table number 3 in the UDC system). First UDC descriptions were analysed with the uses of the UDC parser prepared in the Python programming language. Analysis of UDC class identifiers allowed to estimate similarities between them (formula (2)). Next UDC expressions of the monographs described in the Table 1 were analysed. Using the formula (6) the matrix of similarities between monographs was calculated (Fig. 1).

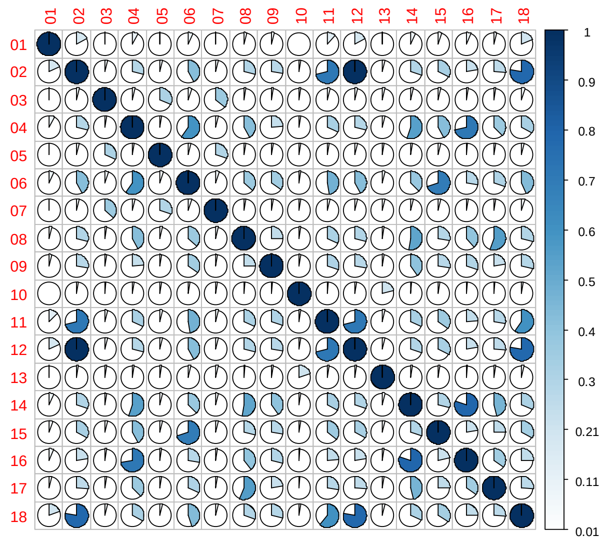


Fig. 1. Similarities between the monographs. Calculations based on UDC descriptions.

Next, the Ward method was used to perform cluster analysis of the monographs. The dendrogram is presented as Fig. 2.

After analysis of UDC codes textual descriptions of the monographs were retrieved from Polish online bookstores. Calculation of distances between textual descriptions of monographs was performed in the way presented in the section 3. The matrix of distances was normalized (by dividing all values by maximum distance) and transforming into the similarity matrix (by subtracting normalized values from 1) which is presented in the Fig. 3.

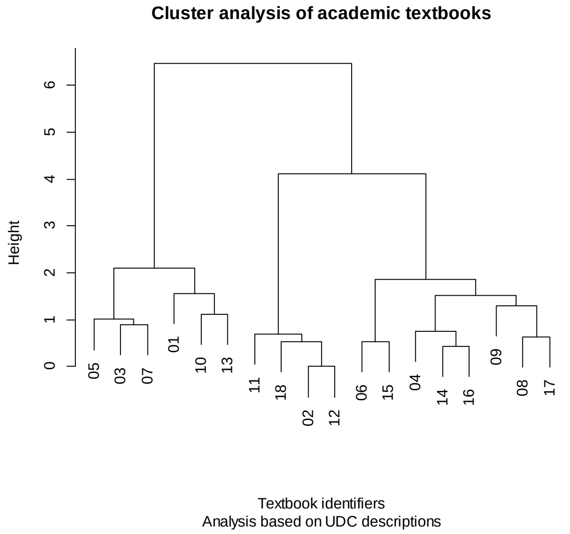


Fig. 2. Cluster analysis of the monographs. Calculations based on UDC descriptions.

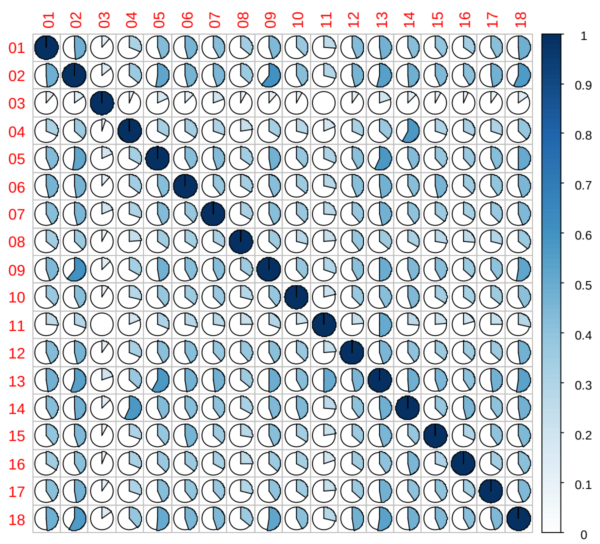


Fig. 3. Similarities between the monographs. Calculations based on textual descriptions.

Fig. 4. shows the results of cluster analysis of the set of monographs performed on the ground of their textual descriptions obtained from online stories.

The evaluation of the convergence between content description of academic textbooks in the form of textual notes and UDC expressions can be performed on the base of values of the Fowlkes-Mallows indexes presented in the Fig. 5.

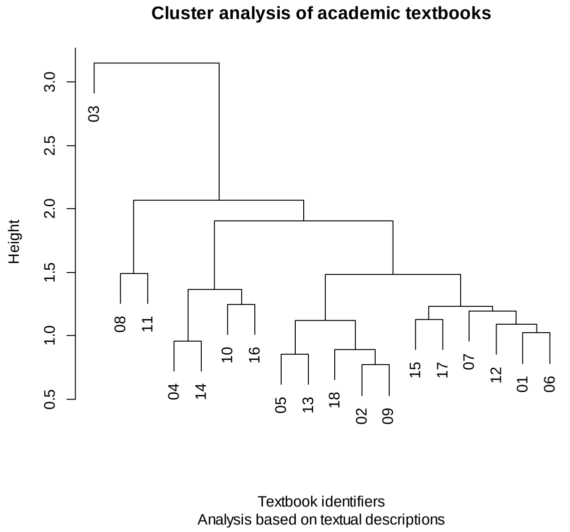


Fig. 4. Cluster analysis of monographs. Calculations based on textual descriptions.

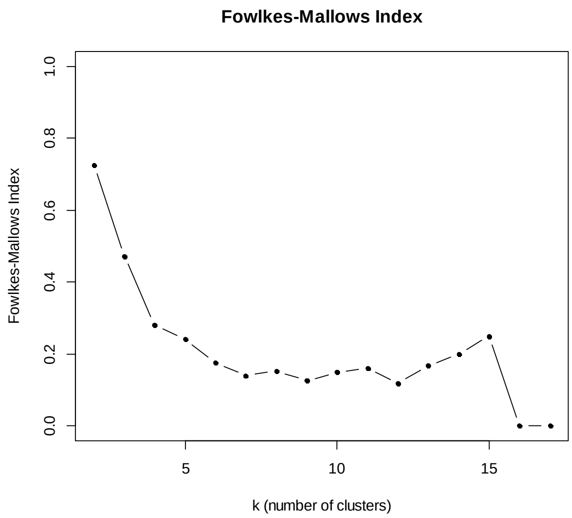


Fig. 5. Comparison of the results of monograph classification based on UDC codes and textual descriptions.

The values of Fowlkes-Mallows indexes show that the convergence between monograph content described with the help of UDC codes and monograph descriptions published by online bookshops is rather low. It is worth to notice that only comparisons of the monographs into two groups allows to obtain Fowlkes-Mallows index higher than 0.5.

Conclusions

The results obtained during analysis show that published online textual descriptions of academic textbooks do not reflect their content defined by UDC codes. Simultaneously it seems that presented in the paper the method of similarity calculation between UDC expressions was verified positively. Further investigations will be focused on theoretical and practical issues of automatic analysis of UDC description of library resources.

Acknowledgements

This research was financed by the statutory fund of the Faculty of Management, Cracow University of Economics.

References

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2-3), 107-145.

Lin, D. (1998). An Information-Theoretic Definition of Similarity. *ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning*, 296-304

Lula, P., Tuchowski, J., & Wójcik, K. (2014). Similarity between Compound Objects and its Application in Recruitment Process. In: D'Amico, A., Moschella, G., (eds.) *Enterprise in Hardship Economics, Managerial and Juridical Perspectives* (9-25). Ariccia: Aracne editrice.

Manning, C. D., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press.

McIlwaine, I., (1997). The Universal Decimal Classification: Some factors concerning its origins, development, and influence, *Journal of the American Society for Information Science*, 48(4), 331-339.

Nasukawa, T., Nagano, T., (2001). Text analysis and knowledge mining system, *IBM Systems Journal*, 40(4), 967-984.

Rafferty, P. (2001). The representation of knowledge in library classification schemes. *Knowledge Organization*, 28(4), 180-191.

Tuchowski, J., Wójcik, K., Lula, P., & Paliwoda-Pękosz, G. (2011). OBCAS - An Ontology-Based Cluster Analysis System. *Research in Systems Analysis and Design: Models and Methods*, 93, 106-112.

Relative Index of Enterprise Innovation Activity for Polish provinces

Małgorzata Markowska[1], Danuta Strahl[2]

Abstract

The aim of the paper is to propose the Relative Index of Enterprise Innovation Activity for Polish provinces. Calculations are based on data published in "Innovation activity of enterprises in 2013-2015" (Polish Central Statistical Office), which comes from the shorter version of Community Innovation Survey. Partial indexes are calculated for investments in innovation, innovation activities, public support, cooperation in innovations and economic results, separately for industry and services, and finally the global index is presented. Indexes are calculated with classical approach assuming normalization of variables into [0;1] interval and averaging them. A new iterative method which finds the best objects one by one starting from top of the ranking is also used. Variables are subject to three types of weights, one based on their hierarchical structure and one established by panel experts.

Key words: *innovations, composite index, Polish provinces*

JEL Classification: C19, O31, R11.

1 Introduction

Problems connected with innovativeness are considered on country economic level – macro scale, regional level – mezzo scale, and smaller spatial units (e.g. cities) – micro scale (Audretsch, 1998). This classification is based on territory, which just by its spatial existence does not determine the innovativeness which is understood as „the process of translating an idea or invention into a good or service that creates value or for which customers will pay. To be called an innovation, an idea must be replicable at an economical cost and must satisfy a specific need. Innovation involves deliberate application of information, imagination and initiative in deriving greater or different values from resources, and includes all processes by which new ideas are generated and converted into useful products. In business, innovation often results when ideas are applied by the company in order to further satisfy the needs and expectations of the customers"[3]. Creative and inventive personnel decide if new product, service, process, marketing or organizational solution is born on a given territory (Löfsten, 2014). The idea, the exchange of thoughts and experience, the contact network, and

[1] Wrocław University of Economics, Department of Regional Economy, Komandorska 118/120, 53-345 Wrocław, Poland, e-mail: malgorzata.markowska@ue.wroc.pl.

[2] Wrocław University of Economics, Department of Regional Economy, Komandorska 118/120, 53-345 Wrocław, Poland, e-mail: danuta.strahl@ue.wroc.pl.

[3] http://www.businessdictionary.com/definition/innovation.html.

innovation environment are needed for the creative destruction, as innovations were called by P. Drucker (1984), (Branstetter and Sakakibara, 2002).

In literature, we can find factors (Fukugawa, 2006), barriers (Hadjimanolis, 1999; Piattier, 1984), sources (Riggs and von Hippel, 1994) and determinants of innovations (Mohr, 1969; Romjin and Albaladejo, 2002). To measure this phenomenon we need methods (OECD, 2005), and indexes (Tadeu and Silva, 2014), which are being changed in time. Different set of measures were developed for the whole economy (Lundstorm et al., 2008), and different for innovativeness of firms (Kirner et al., 2008). There are descriptive and mathematical models trying to illustrate such complicated phenomenon as innovativeness. Rankings for countries and regions are developed based on sets of variables which measure innovation processes.

2 Method

The aim of the paper is to propose the Relative Index of Enterprise Innovation Activity for Polish Provinces. All variables used, are subject to transformation into [0;1] interval with the following formula:

$$x_i^* = \frac{x_i - \min_i \{x_i\}}{\max_i \{x_i\} - \min_i \{x_i\}}$$

All our variables are stimulants ("the bigger, the better"), so one transformation formula is enough. The proposed index is *relative* because variables are transformed using minimal and maximal values observed in the set of 16 Polish provinces. The value of a composite index shows the innovation level in relation to all other provinces, and not to the „objective" reference points. The classical composite index based on m variables, is a weighted (with w_j weights) arithmetic average of transformed variables:

$$RI_i = \frac{s \sum_{j=1}^{m} w_j x_{ij}^*}{\sum_{j=1}^{m} w_i}.$$

It takes values from [0;s] interval, while s is usually assumed as 1, 100, or 1000. The classical composite index is non-robust against outliers or very skewed distributions of individual variables. Outliers used as reference points introduce some kind of unwanted weighting system. Sokołowski and Markowska (2017) proposed a new iterative method for ranking multivariate object and calculating composite index. In the first step the best object is found, and got the rank number one. Then it is eliminated from the set searched for object number two, and so on, assigning just one rank at a time. Reference points (minimum and maximum) are selected from the set of objects not ranked yet, so we are looking for the best object out of those which are left to be ranked. The composite index is then calculated as:

$$RII_{(k)} = I_{(1)}^{1} \prod_{l=2}^{k} \frac{I_{(l)}^{l-1}}{I_{(l-1)}^{l-1}}$$

where:

$RII_{(k)}$ – relative iterative index for *k-th* ordered object,

l – rank (iteration number),

I^l – composite index obtained in l-step of the procedure.

We can see that the composite index for the best object is the same for classical and iterative procedures. Next values of *RII* are smaller by the ratio of local (for a given step) index values calculated for the latest step, when two consecutive objects where both considered.

3 Variables

Variables measuring innovations have been taken from the Central Statistical Office publication *"Innovation activity of enterprises in 2013-2015"* (Innovation, 2016), which comes from the shorter version of Community Innovation Survey. Partial indexes are calculated for:

- – I – investments in innovation (0.23),
- – S – public financial support (0.09)
- – A – innovation activities (0.24),
- – C – cooperation in innovations (0.19),
- – R – economic results (0.26),

separately for industry and services, and finally the global index is presented. There are three systems of weights. The first one comes from the hierarchical structure of variables. We have three levels of them and generally each lower level has the weight which is half of the upper level, and this weights are divided among homogeneous variables on the same level. The second weighting system for five innovation aspects mentioned above (weights are in brackets) is based on weights assigned individually by 10 experts. Finally, indexes for industry and services are joined with equal weights for global relative innovation index. The list of variables which is given in Table 1, is the same for both industry and services, as well as the first weighting system presented in the last column. Percentage of innovation active enterprises in industry is the lowest in Świętokrzyskie (14.5), and the highest in Opolskie (23.1). This range in services is between 6.6 (Warmińsko-Mazurskie) and 13.6 (Zachodniopomorskie).

	Variable	Industry		Services		Weight
		Range	Poland	Range	Poland	
I	Enterprises investing in innovations	10.9-16.4	14.0	4.6-11.1	7.4	1.00
I	Average innovation investments per enterprise in 1000's PLN	356-1362	968	19-1425	479	1.00
S	Enterprises receiving public financial support	3.6-7.5	4.9	0.6-4.1	2.1	1.00
S	Enterprises receiving public financial support from domestic institutions	1.6-4.9	2.7	0.4-3.5	1.4	0.50
S	Enterprises receiving public financial support from local authorities	0.4-2.1	1.0	0.0-2.3	0.8	0.25
S	Enterprises receiving public financial support from central authorities	1.1-3.0	1.9	0.2-3.5	0.8	0.25
S	Enterprises receiving public financial support from EU	2.5-5.9	3.7	0.3-3.9	1.5	0.50
S	Enterprises receiving public financial support from Horizon 2020	0.0-1.3	0.4	0.0-1.1	0.2	0.25
A	Innovation active enterprises	14.5-23.1	18.9	6.6-13.6	10.6	1.00
A	E.w.i.* innovations	13.7-21.5	17.6	6.3-13.0	9.8	1.00
A	E.w.i. new products	9.6-14.4	11.8	2.8-7.6	4.8	0.50
A	E.w.i. new processes	6.8-13.8	9.9	0.3-5.0	2.3	0.17
A	E.w.i new logistics	1.8-4.7	3.2	0.7-3.9	2.7	0.17
A	E.w.i supporting activities	4.0-7.5	5.9	2.3-8.8	5.4	0.17
A	E.w.i organisational or marketing innovations	8.2-16.3	11.4	2.7-18.1	10.7	1.00
A	E.w.i organisational innovations	5.3-10.5	8.1	2.3-15.2	8.1	0.50
A	E.w.i new business practices for organisational procedures	3.6-8.2	6.1	0.9-7.8	4.0	0.17
A	E.w.i. new methods in work responsibilities	2.8-7.9	5.0	1.6-9.3	5.5	0.17
A	E.w.i. new methods in external relations	1.9-4.7	3.1	0.5-7.2	3.5	0.17
A	E.w.i. marketing innovations	4.7-12.5	7.1	1.6-11.5	6.6	0.50
A	E.w.i. new packaging	2.6-9.0	4.2	0.4-3.9	2.2	0.13

A	E.w.i. new media and promotion	1.7-6.8	3.8	1.1-7.9	4.2	0.13
A	E.w.i. new product placement and sales	1.1-3.7	2.1	0.4-4.8	2.6	0.13
A	E.w.i. new pricing	1.4-5.4	2.7	0.8-6.1	3.2	0.13

* – *Enterprises which introduced*

Table 1. List of variables – part 1.

	Variable	Industry		Services		Weight
		Range	Poland	Range	Poland	
C	Enterprises cooperating	3.7-8.4	5.5	1.0-4.8	2.6	1.00
C	Enterprises cooperating for receiving access to intellectual property	0.1-0.3	0.2	0.2-1.3	0.6	0.17
C	Enterprises benefiting from free intellectual property	0.7-2.7	1.4	1.0-3.5	2.4	0.17
C	Enterprises using innovations protected by exclusive rights	1.6-3.7	2.3	0.5-4.7	2.4	0.17
R	Revenues from products to the market	3.6-18.3	9.5	0.3-5.1	3.0	0.50
R	Revenues from products new to the firm	1.1-13.1	5.5	0.0-2.9	1.7	0.50
R	Enterprises with applications for trademarks in Poland	0.4-6.5	3.0	0.7-5.4	3.2	1.00
R	Enterprises with applications for industrial designs in Poland	0.5-5.3	1.3	0.2-1.2	0.4	1.00
R	Enterprises with applications for utility models in Poland	0.2-3.3	1.0	0.0-1.4	0.3	1.00
R	Enterprises with applications for patents in Poland	1.2-4.6	2.3	0.1-1.9	0.7	1.00
R	Enterprises planning to apply for foreign patents	0.1-2.8	0.6	0.1-1.5	0.3	0.50
R	Enterprises with Polish patent applications resulted from internal R&D activities	1.1-3.7	1.6	0.2-1.9	0.6	0.50
R	Enterprises which obtained patents in Poland	1.0-4.9	2.0	0.2-2.7	1.4	1.00
R	Enterprises which made application for foreign patent	0.3-3.4	1.0	0.1-2.6	0.5	1.00

R	Enterprises which obtained foreign patents	0.1-3.0	0.7	0.0-1.2	0.3	1.00

Table 2. List of variables – part 2.

4 Results

For all indexes a value of s=1000 has been used. Indexes for industry and services are both based on 39 variables each, subject to weights for individual variable weights (Table 1 and 2) and expert weights. Global index comes as simple average from industry and services indexes. Values of indexes are given in Table 3 and ranks in Table 4.

Province	Classical Index			Iterative Index		
	Industry	Services	Global	Industry	Services	Global
Dolnośląskie	537	636	587	503	636	570
Kujawsko-pomorskie	258	217	238	233	239	236
Lubelskie	324	240	282	262	217	240
Lubuskie	256	187	221	200	222	211
Łódzkie	236	440	338	214	389	302
Małopolskie	490	386	438	488	399	444
Mazowieckie	467	730	598	468	730	599
Opolskie	503	183	343	481	186	334
Podkarpackie	620	542	581	582	520	551
Podlaskie	639	215	427	639	157	398
Pomorskie	318	311	314	324	346	335
Śląskie	591	263	427	591	298	445
Świętokrzyskie	108	246	177	92	199	146
Warmińsko-mazurskie	287	108	197	253	96	175
Wielkopolskie	358	186	272	365	227	296
Zachodniopomorskie	379	381	380	370	323	347

Table 3. Values of Relative Index of Enterprise Innovation Activity (RIEIA).

Values of classical and iterative indexes are of course very highly correlated – r=0.992 for industry, r=0.981 for services, and r=0.988 for global index. On Fig. 1 we can see that three provinces are the best in enterprise innovation activity. They are Mazowieckie, Dolnośląskie and Podkarpackie. Podlaskie which came first in industry was just 15th in services. Their leading positions are the same in classical and iterative rankings.

Correlations between two methods are high, but the changes in rankings are important. 7 out of 16 provinces changed their position in iterative methods comparing to classical one (See Fig. 2). Pomorskie went up two positions, Śląskie and Wielkopolskie by one position, while four provinces went down just one position: Lubelskie, Łódzkie, Małopolskie, and Opolskie.

Province	Classical Index			Iterative Index		
	Industry	Services	Global	Industry	Services	Global
Dolnośląskie	4	2	2	4	2	2
Kujawsko-pomorskie	13	11	13	13	9	13
Lubelskie	10	10	11	11	12	12
Lubuskie	14	13	14	15	11	14
Łódzkie	15	4	9	14	5	10
Małopolskie	6	5	4	5	4	5
Mazowieckie	7	1	1	7	1	1
Opolskie	5	15	8	6	14	9
Podkarpackie	2	3	3	3	3	3
Podlaskie	1	12	6	1	15	6
Pomorskie	11	7	10	10	6	8
Śląskie	3	8	5	2	8	4
Świętokrzyskie	16	9	16	16	13	16
Warmińsko-mazurskie	12	16	15	12	16	15
Wielkopolskie	9	14	12	9	10	11
Zachodniopomorskie	8	6	7	8	7	7

Table 4. Ranks for provinces.

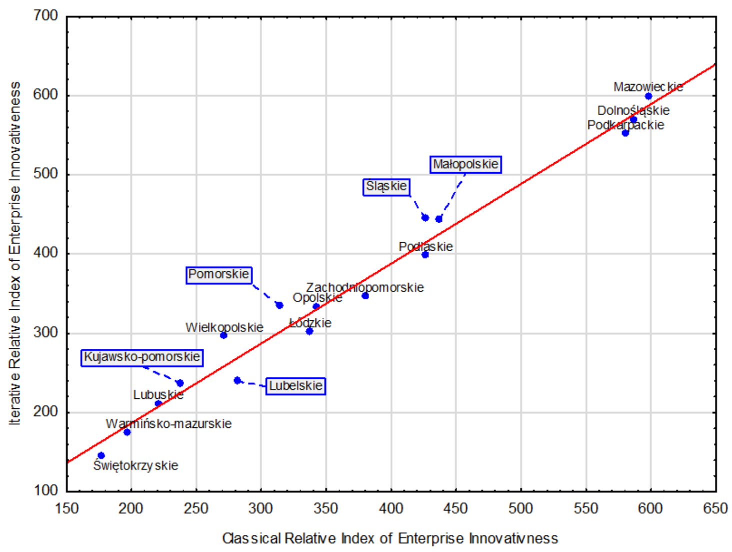


Fig. 1. Scatter diagram for provinces based on classical and iterative innovation indexes.

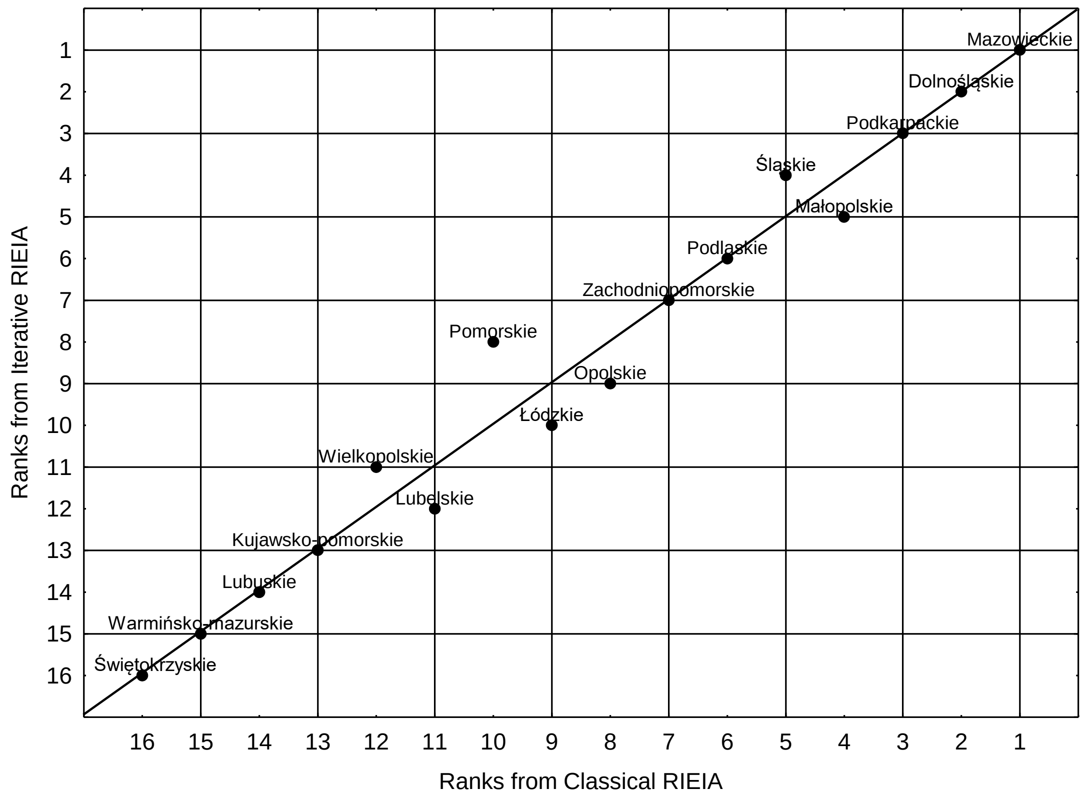


Fig. 2. Ranks based on classical and iterative innovation indexes.

What we found interesting is, a lack of correlation between industry and services. Fig. 3 illustrate the situation for iterative index. For classical one, the correlation is also nonsignificant (r=0.182; p=0.1815).

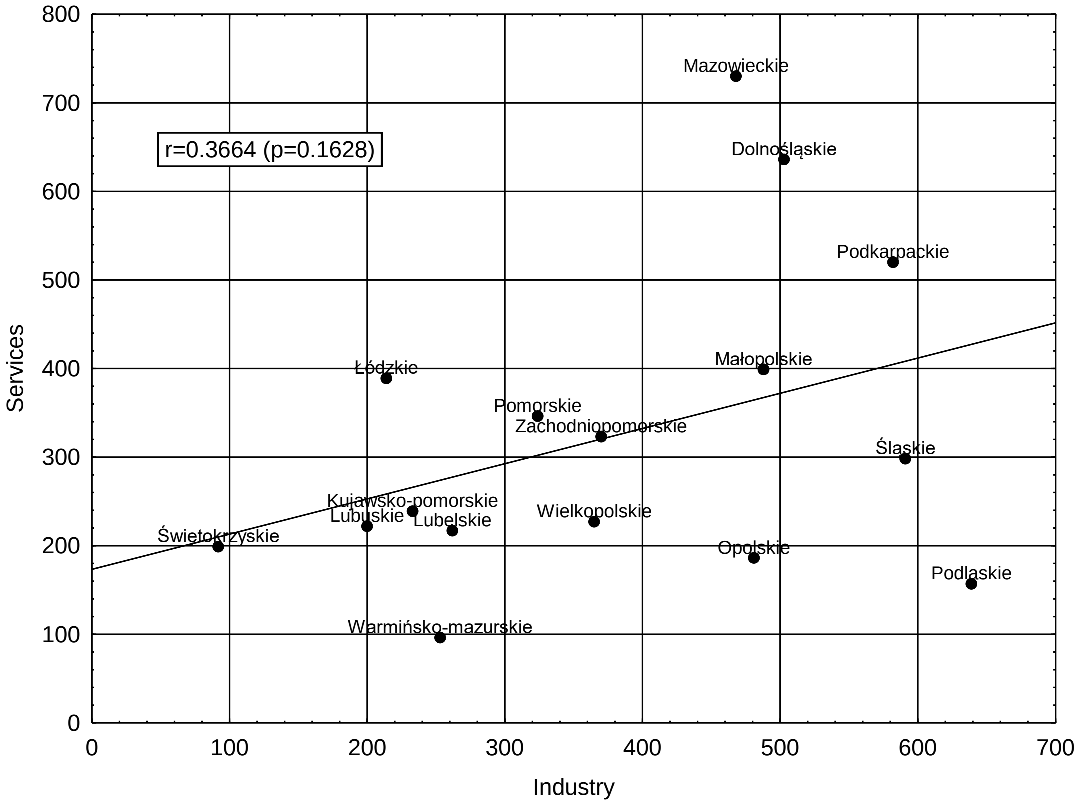


Fig. 3. Correlation of iterative indexes.

Conclusions

The main purpose of this paper was to introduce Relative Index of Enterprise Innovation Activity for Polish Provinces. The list of variables is always a matter for discussion, but the one we proposed is everything what could be extracted from mentioned Central Statistical Office publication on regional (NUTS 2) level. Iterative method seems to be the good choice since it is robust against outlying objects. It is interesting to find that innovation processes in industry and services look as uncorrelated.

Acknowledgements

The paper is financed by the National Science Centre: 2015/17/B/HS4/01021

References

Audretsch, D. (1998). Agglomeration and the Location of Innovative Activity, *Oxford Review of Economic Policy*, *14* (2), 18-29.

Branstetter, L. & Sakakibara, M. (2002), When do Research Consortia Work Well and Why? Evidence from Japanese Panel Data. *American Economic Review*, *92*, 143-159.

Drucker, P. F. (1984). Innovation and Entrepreneurship, Harper Collins, New York.

Fukugawa, N. (2006). Determining Factors in Innovation of Small Firm Networks: A case of Cross Industry Groups in Japan. *Small Business Economics, 27(2),* 181-193.

Godin, B. (2005). The Linear Model of Innovation: The Historical Construction of an Analytical Framework, *Science, Technology, and Human Values,* 31 (6), November 2006, 639-667.

Hadjimanolis, A., (1999). Barriers to innovation for SMEs in a small less developed country (Cyprus). *Technovation*, *19*, 561-570.

Innovation activity of enterprises in 2013-2015. (2016). Central Statistical Office, Szczecin.

Kirner, E., Som, O., Armbruster, H., Lay, G & Zhou, W. (2008). Measuring organisational innovation – concepts, indicators and outcomes. Fraunhofer Institute for Systems and Innovation Research (ISI) - 6 CP Workshop.

Löfsten, H. (2014). Product innovation processes and the trade-off between product innovation performance and business performance. *European Journal of Innovation Management, 17(1)*, 61-84.

Lundstrom, A., Almerud, M. & Stevenson, L. (2008). Entrepreneurship and Innovation Policies: Analysing Measures in European Countries, Swedish Foundation for Small Business Research, Stockholm

Mohr, L. B. (1969). Determinants of Innovation in Organizations, *American Political Science Review*, 63(1), 111-126.

OECD & Eurostat. (2005). Oslo Manual, Guidelines for Collecting and Interpreting Innovation Data, the Measurement of Scientific and Technological Activities, 3rd ed., A Joint Publication of OECD and Eurostat, Paris.

Piattier, A. (1984). Barriers to innovation. Frances Printer. London and Dover NH.

Riggs, W. & von Hippel, E. (1994). Incentives to innovate and the sources of innovation: the case of scientific instruments, *Research Policy, 23*, 459-469.

Romjin, H. & Albaladejo, M. (2002). Determinants of Innovation Capability in Small Electronics and Software Firms in Southeast England, *Research Policy, 31*, 1053-1067.

Sokołowski, A. & Markowska, M. (2017). Iterative Method for Linear Ordering of Multidimentional Objects, *Przegląd statystyczny* (in print).

Tadeu, H. & Silva, J. (2014). Management Indicators and Measurement of Innovation: Review of the Literature. *Management, 3(10)*, 52-58.

Economies of scope or specialisation in Polish dairy farms – an application of a new local measure

Jerzy Marzec[1], Jacek Osiewalski[2], Andrzej Pisulewski[3]

Abstract

Traditional measures of economies of scope, based on the definition introduced by Baumol et al. (1982), have serious shortcomings. However, when different products are expressed in the same units (e.g., using constant, reference prices), we can define a useful class of indices by considering the 1% increase in aggregate production through either the appropriate increase of only one product or the 1% increase of all products. The ratio of costs in these two situations leads to the coefficient of the effect of production structure change, which is our local counterpart of traditional measures of economies of scope or specialization. It has been illustrated using panel data and Bayesian inference in a two-product translog cost frontier model for Polish dairy farms.

Keywords: *multiproduct cost function, stochastic frontier model, Bayesian inference, agricultural economics*

JEL Classification: D22, C23, C51, Q12

1 Cost effects of production structure in multiproduct settings

Microeconomic (frontier) cost functions have been widely used in empirical analyses of costs of production. One of important areas is the analysis of economies of scope or specialisation of producers, with theoretical foundations presented by Panzar and Willig (1981). This area was particularly important in the banking sector in 1980s and 1990s, due to interest in measuring economic consequences of mergers and acquisitions; see Kim (1986), Berger et al. (1987), Lawrence (1989), Dietsch (1993), Hughes and Mester (1993), Mester (1993), Muldur and Sassenou (1993), Zardokoohi and Kolari (1994), Marzec and Osiewalski (2001).

Let $Q=(Q_1,\ldots,Q_G)\in R^G{}_+$ be the vector of quantities of G products, $C(Q_1,\ldots,Q_G;x,\delta)$ - the cost function, x - the vector of factor prices (and the quantities of fixed inputs in the case of short-run analysis) and δ - the vector of parameters. Economies of scope are present when

$$C(Q_1,\ldots,Q_G;\cdot)<C(Q_1,0,\ldots,0;\cdot)+C(0,Q_2,0,\ldots,0;\cdot)+\ldots+C(0,\ldots,0,Q_G;\cdot). \quad (1)$$

[1] Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: marzecj@uek.krakow.pl.

[2] Corresponding author; Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: eeosiewa@cyfronet.pl.

[3] University of Agriculture in Kraków, Al. Adama Mickiewicza 21, 30-001 Kraków, Poland, e-mail: a.pisulewski@ur.krakow.pl.

that is when the cost function is sub-additive, so that the cost of producing $Q=(Q_1,\ldots,Q_G)$ in one unit is smaller than the sum of costs of producing each product in a separate, specialised unit. If the inequality goes in the opposite direction, there are positive effects of specialisation.

Baumol et al. (1982) proved that, in the case of any twice differentiable multiproduct cost function, a sufficient (but not necessary) condition for economies of scope is

$$\frac{\partial^2 C(Q;x,\delta)}{\partial Q_g \partial Q_j} < 0 \qquad \text{for } g \neq j; \quad g, j = 1,\ldots,G. \tag{2}$$

The inequality in (2) means that the marginal cost of each individual product is a decreasing function of any other product. That is, slightly increasing any other product $(Q_1,\ldots,Q_{g-1}, Q_{g+1},\ldots,Q_G)$ results in a smaller value of the marginal cost of product g.

Baumol et al. (1982) proposed the coefficient that follows directly from the definition in (1):

$$SC = \frac{\sum_{g=1}^{G} C_g(Q_g;x,\delta) - C(Q;x,\delta)}{C(Q;x,\delta)}, \tag{3}$$

where $C_g(Q_g;\cdot) = C(0,\ldots,0,Q_g,0,\ldots,0;\cdot)$ is the cost of producing Q_g and no other products; $|SC|\cdot 100\%$ represents the percentage cost reduction (if $SC>0$) or increase (if $SC<0$) due to joint production (as compared to producing each product separately, in a specialised unit). This coefficient was extended to a group of products, see Kim (1986).

However, the definition of economies of scope and the corresponding coefficient SC can be impractical in empirical studies as $C_g(Q_g;\cdot)$ is based on zero levels of all other products, i.e. on values usually not met in the data. This may mean extrapolating the cost function far outside the region of reasonable approximation. Moreover, the popular Cobb-Douglas and translog forms of the cost function are defined only for strictly positive arguments. These considerations led to more practical economies of scope coefficients. In the case of only two products Zardkoohi and Kolari (1994) proposed:

$$\begin{aligned}
&SCP = \frac{\Delta C_1 + \Delta C_2 - \Delta C_{1,2}}{\Delta C_{1,2}}, \text{ where:} \\
&\Delta C_1 = C(Q_1^{\min} + \Delta Q_1, Q_2^{\min}; x,\delta) - C(Q_1^{\min}, Q_2^{\min}; x,\delta) \\
&\Delta C_2 = C(Q_1^{\min}, Q_2^{\min} + \Delta Q_2; x,\delta) - C(Q_1^{\min}, Q_2^{\min}; x,\delta) \\
&\Delta C_{1,2} = C(Q_1^{\min} + \Delta Q_1, Q_2^{\min} + \Delta Q_2; x,\delta) - C(Q_1^{\min}, Q_2^{\min}; x,\delta);
\end{aligned} \tag{4}$$

$Q_1^{\min}$ and $Q_2^{\min}$ are the smallest production levels observed in the data, ΔQ_1 and ΔQ_2 denote the differences between the smallest and average production levels, $\Delta C_{1,2}$ is the difference in the cost of producing both products on their average and minimum levels, and ΔC_1 is an additional cost of producing ΔQ_1 extra units above the minimum level of the first product (keeping the minimum level of the second product). While *SCP* is well defined for any functional form of the cost function, it is based on the increase of production from the minimum to the average level in the dataset. This may mean changes that are too large and do not represent production possibilities of any real production unit. Other, slightly modified, indices of the so-called *within-sample economies of scope* were proposed by Mester (1993) and Hughes and Mester (1993).

In order to avoid problems with using too large increases in production, Marzec and Osiewalski (2001) proposed a truly local measure of economies of scope or specialisation. The crucial assumption is that we can express the levels of all products in the same units. In practice, this will often mean that we have to resort to monetary units by using some constant prices of *G* products. Then the aggregate production $Q^A = Q_1 + \ldots + Q_G$ can be calculated and the proposed coefficient of the *effect of production structure change* (*EPSC*) is:

$$EPSC_g(r) = \frac{C(Q_1, \ldots, Q_g + r \cdot Q^A, \ldots, Q_G; x, \delta)}{C(Q_1 \cdot (1+r), \ldots, Q_g \cdot (1+r), \ldots, Q_G \cdot (1+r); x, \delta)} \text{ for } r \in (-u_g, +\infty), \tag{5}$$

where $u_g = Q_g/Q^A$ and r denotes the assumed change in the scale of production. $EPSC_g(r) < 1$ means that the cost of increasing the level of product g (only) by $r \cdot Q^A$ units is smaller than the cost of simultaneously increasing each product by $r \cdot Q_h$ units ($h=1,\ldots,G$). Since the latter corresponds to increasing scale of production without changing its structure, $EPSC_g(r) < 1$ indicates that the change of the share of product g in the aggregate product from u_g to $(u_g+r)/(1+r)$ leads to the cost reduction by $(1-EPSC_g)\times 100\%$.

2 Short-run translog frontier cost function for Polish dairy farms

The Bayesian frontier cost model presented in this study is estimated using balanced panel data from 846 Polish dairy farms observed over the period 2004-2011 (8 years); the data come from the Farm Accountancy Data Network (FADN). The construction of the variables is based on other studies (on dairy farms) in which FADN data were used (Maietta, 2000; Frahan et al., 2011).

In the present study we consider a variable cost function with two output categories: production of milk (Q_1, including milk products) and other production (Q_2, i.e. total crop production, livestock output and livestock products except milk) and five input categories.

Both products are expressed as the deflated total net farm revenues from sales excluding the value of feed, seeds and plants produced on the farm.

Variable inputs consist of the following: capital (buildings and machinery, *K*), materials and services (*M*), utilised agricultural area (*A*), herd of dairy cows (*Z*). We assume that labour (*L*) is a fixed factor; hence it means that it was not subject to optimization. The cost of capital is measured by the sum of financial expenses – in particular, the costs of repairs and maintenance of capital equipment, interest paid, and annual depreciation on buildings and machinery. The category "materials and services" aggregates mainly expenses on fertilisers, pesticides, seeds, feeds, fuel, energy and veterinary services. The utilised agricultural area is the value of owned and rented land. The herd of dairy cows corresponds to the yearly average value of dairy cows. The price of capital is obtained by dividing the cost of capital by the value of capital. The price of intermediate inputs is constructed by aggregating Laspayres indices of the components weighted by farm-specific cost shares. The resulting series are farm specific due to differences in input composition. The price of area at the farm level is measured by the rental rates for farm land provided by FADN. When a farm uses only own area, the price is calculated as the average rental rate (at the same year) over the farms belonging to the same region. The price of herd was calculated using FADN data; it was obtained by dividing the value of dairy cows by their number. Labour input is measured as the total labour (family and hired) expressed in hours. The variable cost (*VC*) is computed as the sum of all costs associated with four variable inputs used in the production process.

In this study the cost frontier is formulated by a short-run cost function which relates the observed variable cost to output quantities, prices of variable inputs and quantities of fixed inputs, allowing for inefficiency and random noise. The stochastic cost frontier model based on panel data is defined as

$$y_{it} = h_t(x_{it}, \beta) + v_{it} + z_{it}, \qquad (i = 1,\ldots, N;\ t = 1,\ldots, T), \tag{6}$$

where y_{it} is the natural logarithm of the observed cost for unit i at time t (i=1,..., N; t=1,…,T); x_{it} is a row vector of exogenous variables; h_t is a known parametric functional form with β as a (k×1) vector of parameters, v_{it} and z_{it} are random terms, one symmetric about zero and the other non-negative. In the case of a cost frontier, the inefficiency term z_{it} captures the overall cost inefficiency, reflecting cost increases due to both technical and allocative inefficiency of farm i at time t. Here the inefficiency term may reflect not only a farm specific effect but also a time-varying component. Cost efficiency score is calculated as $r_{it}=\exp(-z_{it})$, which is an easily interpretable quantity in the interval (0; 1].

Empirical analyses require a particular functional form of the cost function. In (6) we use the translog specification, since it is a local second-order Taylor series approximation of any sufficiently smooth "true" cost function. The estimated translog frontier should be a cost function, so linear homogeneity in input prices is imposed by normalising both the cost and input prices by one input price (e.g., by w_Z). Consequently, for y_{it} defined as the log cost minus $\ln(w_{Z,it})$, the functional form of the variable cost model is:

$$\begin{aligned} & h_t(x_{it};\beta) = \beta_0 + \sum_{h=1}^{6}\beta_h x_{it,h} + \sum_{h=1}^{6}\sum_{j\geq h}^{6}\beta_{hj}x_{it,h}x_{it,j} + \\ & + \beta_{Q1}\cdot t\cdot \ln Q_{it,1} + \beta_{Q2}\cdot t\cdot \ln Q_{it,2} + \beta_{trend,1}\cdot t + \beta_{trend,2}\cdot t^2, \\ & x_{it} = \left[\ln\frac{w_{K,it}}{w_{Z,it}} \quad \ln\frac{w_{M,it}}{w_{Z,it}} \quad \ln\frac{w_{A,it}}{w_{Z,it}} \quad \ln L_{it} \quad \ln Q_{1,it} \quad \ln Q_{2,it}\right], \end{aligned} \tag{7}$$

where w_h is the price of the variable input h ($h\in\{K, M, A, Z\}$). Additionally, the trend variable t has been introduced in order to capture the influence of technical progress; it allows to model the variability of returns to scale (RTS) over the time period considered.

To define our statistical model, we make the usual assumption that the symmetric error terms v_{it} are independent and normally distributed with the same unknown variance, i.e., they are $iid\ N(0,\sigma_v^2)$. This study employs the standard Bayesian normal-exponential stochastic frontier model with Varying Efficiency Distribution (VED) specification proposed by Koop et al. (1997). Subsequently, the inefficiency terms vary over time and production units, but not freely. Namely, z_{it} are independent and follow the exponential distributions with means (and standard deviations) λ_{it} that depend on exogenous variables $s_{it,j}$ (j=1,...,m), i.e. $\ln\lambda_{it} = -\sum_{j=1}^{m} s_{it,j}\cdot\ln\phi_j$ where ϕ_j>0 are additional unknown parameters; see also Osiewalski and Steel (1998). The important special case when $m = 1$ and $s_{it,1} = 1$ is called the Common Efficiency Distribution (CED) specification.

In our VED specification we use nine dummy variables to describe λ_{it}; they explain possible systematic differences in efficiency levels due to some farm characteristics. The binary exogenous variables are: farm size measured by land area (small, large), the economic size and the amount of dairy cows; type of specialization (one when milk production is the main source of farm income, zero otherwise), information on whether the farmer has received less favoured areas subsidies or investment subsidies. Other two factors that could potentially influence farm efficiency are whether the farmer rents land or uses hired labour.

The statistical modelling and inference is based on the Bayesian Stochastic Frontier Analysis, proposed by van den Broeck et al. (1994) and Koop et al. (1997), which is now regarded as being relatively standard. The complexity of the stochastic frontier model requires advanced numerical methods to describe the posterior distribution. As Koop et al. (1997) and Osiewalski and Steel (1998) showed, Gibbs sampling, a relatively simple Markov Chain Monte Carlo algorithm, is an efficient tool for generating samples from the posterior distribution.

3 Results on cost frontier and cost efficiency

In Table 1 we present the posterior means and standard deviations for main characteristics of the cost function. The individual posterior means of cost elasticities with respect to all factor prices and two outputs have the expected sign for almost every farm and every period; this holds particularly strongly for the outputs and the prices of capital, materials and livestock.

Positive posterior means of the cost elasticity with respect to the fixed factor (labour) suggest that farms are far from long-run cost minimisation; about 66% of the estimated elasticities are (slightly) positive. A typical Polish dairy farm is characterised by increasing short-run returns to scale; the *RTS* coefficient is approximately 1.2. Only 6% of farms operate at decreasing short-run returns to scale and the vast majority of them are the largest producers in the sample.

Cost elasticity w.r.t.:	**Posterior Mean**	**Posterior Std. dev.**	**Prior Mean**	**Prior Std. dev.**	**Positive sign (% of the sample)**
price of capital	0.141	0.009	0.25	0.5	98%
price of materials	0.617	0.016	0.25	0.5	100%
price of area	0.064	0.010	0.25	0.5	84%
price of livestock	0.177	0.014	0.25	0.5	99%
production of milk (Q_1)	0.644	0.005	0.5	0.5	100%
other production (Q_2)	0.211	0.004	0.5	0.5	98%
labour (L)	0.026	0.013	-0.1	0.5	66%

Table 1. Posterior moments for elasticities of cost with respect to outputs, labour and input prices (for a typical farm, with average values of logs of explanatory variables).

Note that, in Table 1, the posterior standard deviations of elasticities are much smaller than the prior standard deviations. This means that our prior distribution, which imposes microeconomic regularity in a very weak fashion, has not distorted the evidence coming from the likelihood function, based on quite informative data.

Regarding cost efficiency, the average posterior mean of r_{it} is 0.94, while the average posterior standard deviation is 0.05. This implies that Polish dairy farms could have decreased variable cost by about 6% on average. Most of the farms (91%) were relatively efficient, with efficiency above 0.9. The minimum estimate of cost efficiency (among all the observations in our sample) is 0.5. The detailed results are not reported here due to space constraints.

4 Effects of production structure change in Polish dairy farms

The main goal of this study is to obtain results regarding cost effects of (small) changes in the product structure. The estimates of the *EPSC* coefficient (5) are presented in Figures 1 and 2. For illustrative purposes, this measure was calculated for three firms of different size and production structure. We assume in our interpretations below that the change in the scale of production (r) belongs to the interval (0; 0.1].

The large farm with decreasing returns to scale (Figure 1) is characterized by positive cost effect of the change in production structure if it increases the share of Q_1 (milk), although it is already high. The down sloping curve in Figure 1 shows for r = 0.02 and $EPSC_1$=0.9988 that the change of the share of milk in the aggregate product from 89.1% to 89.2% leads to the cost reduction by 0.12% in comparison to the increase of production scale without changing its structure. The upward sloping curve shows that the cost of increasing the level of Q_2 (other production) by 43000 PLN (0.02×2144000) is by 0.82% ($EPSC_2$ = 1.0082) greater than the cost of increasing the aggregate product by the same amount without changing the shares of Q_1 and Q_2. Alternatively, in the case of diminishing total production, reducing only Q_2 is advisable because it results in a smaller cost than decreasing scale of production without changing its structure.

The average size farm (Figure 2) has approximately 41% share of Q_1 in aggregate product and operates under increasing returns to scale. This medium farm should prefer to increase the share of Q_2 in the aggregate product, because it leads to the cost reduction as compared to the situation when this unit expands the scale of production without changing its structure. An increase of the share of Q_1 is not desirable; a rise in milk production corresponding to r = 0.01 leads to the cost higher by approximately 0.04% ($EPSC_1$ = 1.004) than in the case of proportional growth of Q_1 and Q_2. Finally, for the small farm in Figure 3 we observe positive cost effects of specialisation in each product, so there are two ways to reduce cost.

This study has illustrated that *EPSC*, defined in (5), is a precise and useful measure of cost effects of changes in production structure.

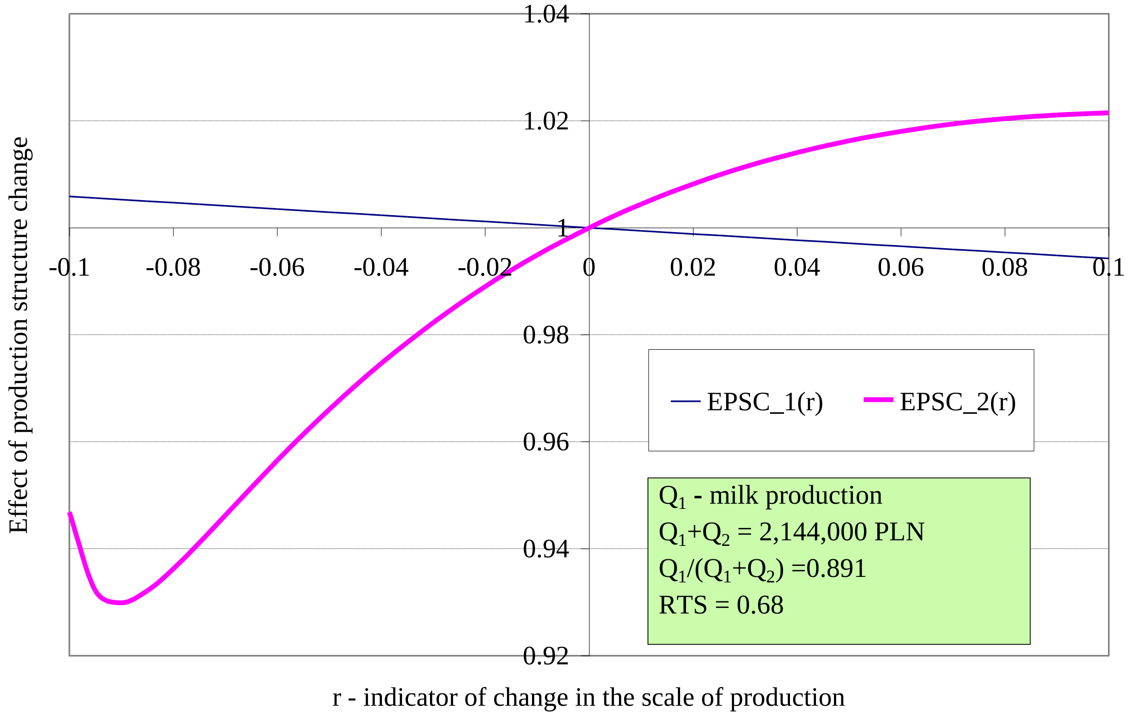


Fig. 1. *EPSC* for a large farm.

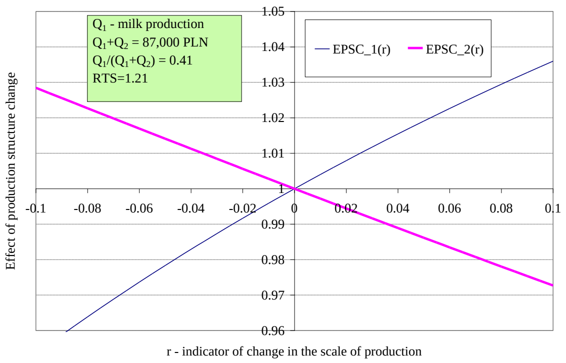


Fig. 2. *EPSC* for a medium farm.

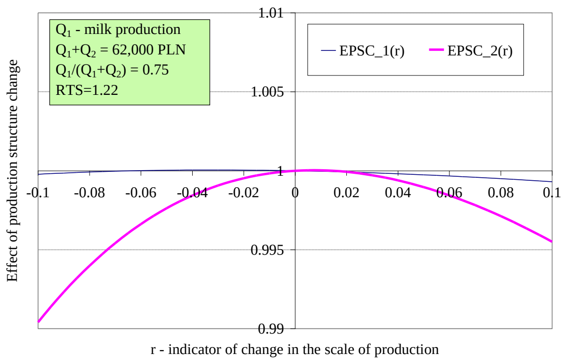


Fig. 3. *EPSC* for a small farm.

Acknowledgements

The first two authors acknowledge support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential. The third author acknowledges the financial support from the National Science Centre, Poland, based on decision no. DEC-2013/09/N/HS4/03833.

References

Baumol, W.J., Panzar J.C., & Willig, R.D. (1982). *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace Jovanovich: San Diego.

Berger, A.N.D., Hanweck, G., & Humphrey, D. (1987). Competitive Viability in Banking. Scale, Scope, and Mix Economies. *Journal of Monetary Economics*, 20(3), 501-520.

Broeck van den, J.V, Koop, G., Osiewalski, J., & Steel, M.F. (1994). Stochastic Frontier Models: A Bayesian Perspective. *Journal of Econometrics*, 61(2), 273-303.

Dietsch, M. (1993), Economies of Scale and Scope in French Commercial Banking Industry. *Journal of Productivity Analysis*, 4(1-2), 35-50.

Frahan, B. H., Baudry, A., Blander, R. D., Polome, P., & Howitt, R. (2011). Dairy Farms Without Quotas in Belgium: Estimation and Simulation With a Flexible Cost Function. *European Review of Agricultural Economics*, 38(4), 469-495.

Hughes, J., & Mester, L.J. (1993). A Quality and Risk-adjusted Cost Function for Banks: Evidence on the „too-big-to-fail" Doctrine. *Journal of Productivity Analysis*, 4(3), 293-315.

Kim, Y. H. (1986), Economics of Scale and Economies of Scope in Multiproduct Financial Institutions: Further Evidence from Credit Unions. *Journal of Money, Credit, and Banking*, 18(2), 220-226.

Koop, G., Osiewalski J., & Steel, M.F. (1997). Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers. *Journal of Econometrics*, 76(1-2), 77-105.

Lawrence, C. (1989). Banking Cost, Generalized Functional Forms, and Estimation of Economies of Scale and Scope. *Journal of Money, Credit, and Banking*, 21(3), 368-379.

Maietta O.W. (2000). The Decomposition of Cost Inefficiency Into Technical and Allocative Components With Panel Data of Italian Dairy Farms. *European Review of Agricultural Economics*, 27(4), 473-495.

Marzec, J., Osiewalski, J. (2001). Funkcja kosztów dla oddziałów banku – mierniki korzyści specjalizacji. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 895, 155-165.

Mester, L.J. (1993). Efficiency in the Savings and Loan Industry. *Journal of Banking and Finance*, 17, 267-286.

Muldur, U., & Sassenou, M. (1993). Economies of Scale and Scope in French Banking and Savings Institutions. *Journal of Productivity Analysis*, 4(1-2), 51-72.

Murray, J.D., & White, R.W. (1983). Economies of Scale and Economies of Scope in Multiproduct Financial Institutions: A Study of British Columbia Credit Unions. *Journal of Finance*, 38(3), 887-902.

Osiewalski, J., & J., Steel, M.F. (1998). Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models. *Journal of Productivity Analysis*, 10(1), 103-117.

Panzar, J.C., Willig, R.D. (1981). Economies of Scope. *American Economic Review*, 71, 268-272.

Zardkoohi, A., & Kolari, J. (1994). Branch office economies of scale and scope: evidence from savings banks in Finland. *Journal of Banking & Finance*, 18(3), 421-432.

Density forecasting performance of alternative GARCH and DCS models for daily financial returns

Błażej Mazur[1]

Abstract

The objective of the paper is to verify predictive performance (in terms of density forecasts) of some of recently proposed models for daily financial returns. We focus on Beta-Gen-*t*-EGARCH specification proposed by Harvey and Lange (2015,2016), which uses a generalized-*t* conditional distribution. We conduct an empirical comparison of out-of-sample predictive performance against some well-established univariate alternatives. We use daily S&P500 logarithmic returns and investigate sequences of 1525 density forecasts for horizons of 1 to 6 trading days ahead (3rd of January, 2011 till 30th of December, 2016) and the training sample begins on 3rd of January, 2006. We obtain the density forecasts using Bayesian methods. A general Beta-Gen-*t*-EGARCH model dominates other specifications in terms of criteria for density forecast comparison like LPS and CRPS. Moreover, it provides best estimates of Value-at-Risk. In general the results indicate the importance of asymmetric modelling of the news impact curve and of heavy-tailed conditional distribution, which is not unexpected; the results also suggest importance of tail asymmetry which is not addressed here.

Keywords: *empirical finance, Value-at-Risk, generalized t distribution, dynamic conditional score, EGARCH, density forecasting, Bayesian inference*

JEL Classification: G17, C58

1 Introduction

There exists a growing body of methods and models available for empirical financial analyses focusing on daily asset returns dynamics. An issue that has been receiving a lot of attention is that of risk evaluation. The related problem of time-series modelling is that of density forecasting. Crucial issues under consideration are those of adequate quantification of uncertainty and estimation of probability of extreme events. At the same time, the volume of available data seems to be growing very rapidly. Risk evaluation often requires costly computational methods and for real-time analysis it is important to balance advancement of computations against feasibility.

Here we focus on recent developments in the field of GARCH models that are popular in applied finance also because of their computational convenience. Well-established tools include extensions of the basic GARCH specification like EGARCH, GJR-GARCH among others. Moreover, simple conditional normality is often generalized to capture features of the

[1] Cracow University of Economics, Department of Econometrics and Operations Research, Rakowicka 27, 31-510 Kraków, Poland, e-mail: blazej.mazur@uek.krakow.pl.

conditional distribution like its heavy tails. A popular choice is that of *t* distribution, though some authors suggest the GED distribution instead.

An interesting class of dynamic models is based on the idea of Dynamic Conditional Score approach which was proposed by Harvey (2013). A related strand of research develops models under the label of Generalized Autoregressive Score (GAS), see Creal et al. (2011, 2013). The general idea of the approach is that features of the conditional sampling distribution are updated in a way that takes into account an autoregressive update and a score of the conditional sampling distribution. This leads to a class of dynamic models that is similar to GARCH models in the spirit – the models are not based on latent processes, though are likely to provide gains in terms of estimation feasibility.

In particular we focus on Beta-Gen-t-EGARCH model, see Harvey and Lange, (2015, 2016) and Harvey and Sucarrat (2014). The model not only makes use of the novel idea of dynamic updating. It also features an interesting (and flexible) form of the conditional sampling distribution, being a generalized *t* distribution. The purpose of the paper is to verify whether the model features lead to benefits in the applied work, in particular whether out-of-sample density predictive performance is competitive compared to some well-established specifications. An important point emphasized here is that – contrary to the work of Harvey and his co-authors – the paper makes use of Bayesian inference methods. This is because it is difficult to find any other modelling approach that allows for estimation uncertainty to be taken into account in a coherent way when constructing density forecasts.

The rest of the paper is organized as follows. Firstly, the types parametric of conditional distributions under consideration are reviewed. Secondly, basic GARCH models are recalled and Beta-Gen-*t*-EGARCH specification is introduced. Thirdly, some remarks on Bayesian model specification, estimation and prediction are considered. Fourthly, an empirical analysis (using daily S&P500 returns) is conducted, with out-of-sample predictive experiments on expanding sub-samples. The general predictive performance is examined using scoring rules that are appropriate for evaluation of density forecasts, namely the Log-Predictive Score (LPS) and Continuous Ranked Probability Score (CRPS) – see Gneiting and Raftery (2007). Moreover, adequacy of tail forecasts is examined in detail, as quality of Value-at-Risk estimates is discussed.

2 A generalized *t* distribution

There exists a number of generalizations of *t*-distribution that are of interest. Here we focus on a special case with the probability density function of the following form:

$$f(y)=\frac{1}{\sigma}K(\nu,\gamma)\left(1+\frac{1}{\nu}\left|\frac{y-\mu}{\sigma}\right|^{\gamma}\right)^{-(\nu+1)/\gamma}, \tag{1}$$

with location parameter μ, scale parameter σ and two shape parameters ν and γ (see Theodossiou, 1998, Harvey and Lange, 2016). The normalizing constant is given by:

$$K(\nu,\gamma)=\frac{\gamma}{2\nu^{1/\gamma}}\frac{1}{B(\nu/\gamma,1/\gamma)}. \tag{2}$$

The distribution nests a number of interesting special cases. Firstly, as $\gamma=2$ it becomes a Student *t*-distribution with γ degrees of freedom. Moreover, as $\gamma\to\infty$, the limiting case is the GED i.e. the Generalized Error Distribution, also known as the Exponential Power Distribution (EPD). Hence the generalized *t*-distribution also encompasses the Gaussian and the Laplace cases - see the discussion in Harvey and Lange (2016) and references therein for derivations of the limiting cases. The distribution seems appealing for financial applications since its tail behavior depends on two parameters (namely ν and γ). Though γ controls also the shape of the distribution around the mode, it seems reasonable to verify whether such a general distribution can be used to improve performance of density forecasts and risk evaluation.

3 GARCH and DCS models under consideration

Consider the following dynamic specification for daily financial logarithmic returns y_t:

$$y_t=\delta_t+u_t \tag{3}$$

with

$$u_t=\rho u_{t-1}+\varepsilon_t \tag{4}$$

and

$$\varepsilon_t=\sqrt{h_t}z_t. \tag{5}$$

A standard AR(1)-GARCH(1,1) specification assumes that z_t represents a sequence of independent and identically distributed zero-mean and unit-variance random variables. Moreover, $\delta_t=\delta$ and:

$$h_t=\alpha_0+\alpha_1\varepsilon_{t-1}^2+\beta_1 h_{t-1}. \tag{6}$$

In the GJR GARCH specification of Glosten et al. (1993) the parameter α_1 takes different values depending on the sign of ε_{t-1}:

$$h_t=\alpha_0+\alpha_1^-\varepsilon_{t-1}^2 I_{\varepsilon_{t-1}<0}+\alpha_1^+\varepsilon_{t-1}^2 I_{\varepsilon_{t-1}>0}+\beta_1 h_{t-1}. \tag{7}$$

This allows for asymmetry of the news impact curve (in such a model it is possible that negative “surprises” lead to increased conditional volatility while positive surprises do not). The GJR effect is very often found empirically important. Another option is to consider so-called In-Mean effect that aims to capture the leverage effect:

$$\delta_t = \delta + \eta\sqrt{h_t} \ . \tag{8}$$

Parametric restrictions are of course necessary to ensure that the conditional variance h_t is positive; further restrictions ensure stationarity of the model.

The simple GARCH-type models have certain disadvantages; an effort to deal with some of them has led to a development of the EGARCH model, where log-volatility is updated instead of volatility, see Nelson (1991). Here we are going to follow Harvey and Lange (2015, 2016) who derive similar model from the idea of Dynamic Conditional Score approach. It can be used to construct the following dynamic updating mechanism. Assume that $\lambda = \ln\sigma$ i.e. λ denotes logarithm of scale of the conditional distribution of error terms. Let:

$$\lambda_t = \omega(1-\phi) + \phi\lambda_{t-1} + \kappa w_{t-1} . \tag{9}$$

where w_{t-1} denotes score of the conditional distribution wrt. λ at *t*-1 i.e. a derivative of log-(conditional) density at the actual outturn. Harvey and Lange (2016) consider also an asymmetric specification:

$$\lambda_t = \omega(1-\phi) + \phi\lambda_{t-1} + \kappa w_{t-1} + \kappa^* w^*_{t-1} . \tag{10}$$

with:

$$w^*_t = \mathrm{sgn}(\delta_t - y_t)(w_t + 1), \tag{11}$$

where ω, ϕ, κ and κ^* are model parameters. Similar approach can be taken for updating of conditional location parameter. Moreover, the expressions for the score functions (depending on the class of the conditional distribution) as well as the relationship between location and log-scale score is discussed in Harvey (2013) and Harvey and Lange (2016).

The above formulas define model classes considered in the paper. The first group of “traditional” models is that of AR-GARCH class. For the model class we assume that the conditional distribution is Student *t* with ν degrees of freedom. The models under consideration include a limiting case of conditional normality with $\nu \to \infty$. The volatility equation is either of GARCH (1,1) form or of GJR (1,1) form.

The other class of models that includes the “novel” specifications assume that the conditional distribution is of the generalized *t* form. Moreover, the conditional location and conditional scale is updated according to the DCS idea. Below a convention for labelling is used according to which the above EGARCH model is denoted by S/L-DCS(1,1)-Gt-AL-AS, where S/L-DCS(1,1) denotes DCS-updating of conditional scale (S) and location (L), Gt denotes the conditional distribution (being generalized *t* given by (1)), and AL and AS indicate that both scale and location equation feature asymmetric form of the type (10). Special (or limiting) cases as to the conditional distribution are labelled GED, t or N for GED, Student *t* and Gaussian distribution.

4 Bayesian model specification, estimation and prediction

Bayesian estimation of the model class outlined above requires certain prior assumptions as to model parameters. Here the assumptions are somewhat simple, since the number of observations is rather large, as it is usually the case with financial returns. Hence we assume that even if the prior structure is not fully justified from the theoretical point of view, it is not going to affect the predictive performance in a serious way, as any irrelevant prior information should be updated by the information coming from the training sample.

Priors used in the paper are independent and proper, the latter is important as it is well known that the use of improper priors for parameters controlling tail thickness might cause theoretical problems as to the existence of the proper posterior with finite moments. Throughout the paper all the autoregressive parameters feature flat priors do not exceeding the (-1,1) interval, flat priors are also assumed for α_1 and β_1. The score parameters κ are assumed to be a priori *t*-distributed. For γ we assume a prior that is restricted to arbitrarily chosen range (1.05,6.05) covering empirically relevant cases and is practically uniform over the range. Prior for ν assumes that the parameter is restricted to the range (2.01,150) since it ensures that the conditional variance exists. The upper bound is used for numerical convenience only and the restriction has very limited empirical consequences.

The inference is conducted using MCMC techniques. Firstly, the posterior distribution is evaluated using a Metropolis-Hastings algorithm with a random walk proposal. Based on the results an Independent version of the Metropolis-Hastings algorithm is constructed. This is crucial for the predictive exercise conducted here. As all the models have likelihood that is fully sequential, if one can construct a proposal that is valid for all the subsamples within the expanding sample exercise, evaluation of the likelihood for all the samples is equivalent to its evaluation for the longest subsample. Hence the same proposal can be used for all the instances of the model (though the acceptance has to be addressed separately, so that each subsample has the associated MCMC chain). Within such a setup it is possible to undertake a sequential forecasting exercise of that complexity in a practical timespan.

5 Empirical model comparison: analysis of daily S&P500 returns 2006-2016

For the sake of the empirical comparison we analyse daily S&P500 logarithmic returns (based on closing prices), depicted in Figure 1. We use a sequence of expanding samples, the first observation corresponding to the 3rd of January, 2006. The shortest sample ends on 31st of December, 2010 (1259 observations) and it covers the period of the global financial crisis, with large negative returns of 15th of October and 1st of December 2008. The longest sample ends on 30th of December, 2016 (2751 observations).

Model label	LPS	CRPS	std. dev.	$q_{0.95}$-$q_{0.05}$	IQR	$q_{0.6}$-$q_{0.4}$
AR(1)-N-GARCH(1,1)	-1873.59	0.486	0.92	3.04	1.24	0.47
AR(1)-t-GARCH(1,1)	-1834.82	0.485	0.95	2.95	1.07	0.39
AR(1)-t-IM-GARCH(1,1)	-1832.24	0.484	0.95	2.95	1.07	0.39
AR(1)-N-GJR(1,1)	-1833.31	0.482	0.92	3.01	1.23	0.46
AR(1)-t-GJR(1,1)	-1803.54	0.481	0.94	2.95	1.08	0.40
AR(1)-t-IM-GJR(1,1)	-1796.75	0.480	0.94	2.96	1.08	0.40
S/L-DCS(1,1)-Gt-AL-AS	-1795.70	0.480	0.94	3.09	1.06	0.37
S/L-DCS(1,1)-Gt-AS	-1796.42	0.480	0.95	3.08	1.06	0.37
S/L-DCS(1,1)-GED-AS	-1800.23	0.480	0.95	3.12	1.05	0.36
S/L-DCS(1,1)-t-AS	-1798.89	0.482	0.94	3.00	1.11	0.41
S/L-DCS(1,1)-Gt	-1837.52	0.485	0.93	3.05	1.03	0.36
S/L-DCS(1,1)-t	-1843.27	0.487	0.93	2.95	1.08	0.40

Table 1. Overall predictive performance for $h = 1$; LPS is cumulated, CRPS is averaged, std. dev. denotes average standard deviation of the predictive distribution, the last three columns depict averaged distance between respective quantiles of predictive distributions.

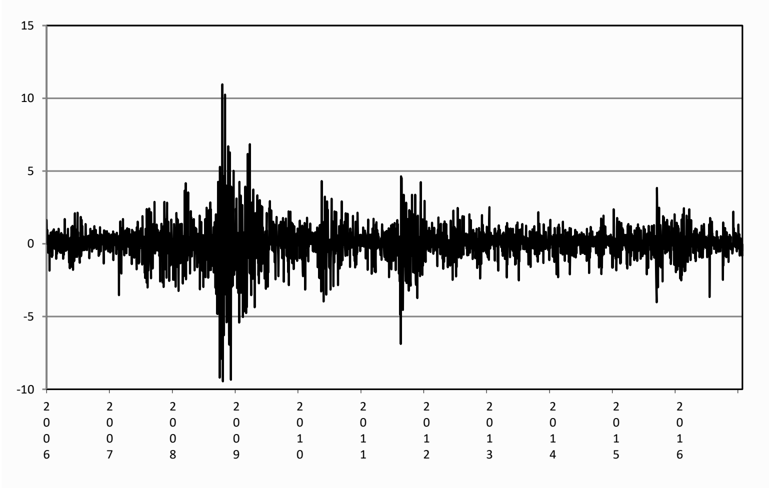


Fig. 1. S&P500 daily logarithmic returns, 2006-2016.

Model label	LPS	CRPS	std. dev.	$q_{0.95}$-$q_{0.05}$	IQR	$q_{0.6}$-$q_{0.4}$
AR(1)-N-GARCH(1,1)	-1887.41	0.487	0.93	3.06	1.25	0.47
AR(1)-t-GARCH(1,1)	-1844.5	0.487	0.96	2.98	1.07	0.39
AR(1)-t-IM-GARCH(1,1)	-1841.27	0.485	0.96	2.98	1.07	0.39
AR(1)-N-GJR(1,1)	-1843.15	0.484	0.92	3.03	1.23	0.46
AR(1)-t-GJR(1,1)	-1815.43	0.483	0.95	2.97	1.08	0.40
AR(1)-t-IM-GJR(1,1)	-1813.43	0.482	0.94	2.96	1.08	0.40
S/L-DCS(1,1)-Gt-AL-AS	-1811.49	0.482	0.95	3.1	1.06	0.37
S/L-DCS(1,1)-Gt-AS	-1814.38	0.482	0.95	3.09	1.05	0.37
S/L-DCS(1,1)-GED-AS	-1817.22	0.483	0.95	3.12	1.05	0.36
S/L-DCS(1,1)-t-AS	-1816.8	0.483	0.95	3.01	1.11	0.41
S/L-DCS(1,1)-Gt	-1854.29	0.486	0.94	3.08	1.03	0.36
S/L-DCS(1,1)-t	-1852.54	0.49	0.94	2.97	1.09	0.40

Table 2. Overall predictive performance for h = 2, LPS is cumulated, CRPS is averaged, std. dev. denotes average standard deviation of the predictive distribution, the last three columns depict averaged distance between respective quantiles of predictive distributions.

The data from the first two weeks of January 2017 are used for forecast evaluation, as we consider prediction horizons ranging from 1 to 6 trading days ahead. Hence the *ex-post* evaluation of predictive performance is based on sequences of 1510 realized forecasts for each horizon. All the models are reestimated with each new observation added. Basic results of the comparison are given in Table 1 and Table 2, for one-day-ahead and two-days-ahead forecasts, respectively. The results are similar for horizons up to 6 days ahead.

Model label	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
AR(1)-N-GARCH(1,1)	2.03	3.02	4	4.66	5.51	6.36	7.08	8.2	8.85	9.57
AR(1)-t-GARCH(1,1)	1.25	2.62	4	5.38	6.3	7.61	8.72	9.77	10.69	11.87
AR(1)-t-IM-GARCH(1,1)	1.25	2.62	3.93	5.11	6.3	7.87	8.59	9.77	10.69	12.07
AR(1)-N-GJR(1,1)	1.77	3.02	3.54	4.39	5.31	6.36	6.89	7.74	8.26	8.85
AR(1)-t-GJR(1,1)	1.11	2.75	3.8	5.05	6.16	7.34	8.2	8.98	9.77	10.75
AR(1)-t-IM-GJR(1,1)	1.11	2.75	3.8	4.85	6.16	7.28	8.13	8.98	9.9	11.02
S/L-DCS(1,1)-Gt-AL-AS	1.05	2.3	3.34	4.13	5.18	6.56	7.48	8.72	9.38	10.03
S/L-DCS(1,1)-Gt-AS	1.05	2.36	3.34	4.26	5.44	6.62	7.48	8.79	9.57	10.03
S/L-DCS(1,1)-GED-AS	1.05	2.16	3.21	4.13	5.38	6.43	7.41	8.72	9.51	10.03
S/L-DCS(1,1)-t-AS	1.31	2.75	3.67	4.79	6.03	6.89	7.8	8.72	9.31	10.43
S/L-DCS(1,1)-Gt	1.44	2.56	3.74	4.98	6.43	7.34	8.72	9.77	10.62	11.48
S/L-DCS(1,1)-t	1.64	2.82	4.26	5.7	7.08	7.87	9.11	9.9	10.82	11.87

Table 3. Lower tail predictive performance (h = 1) – empirical fraction of realized forecasts that exceed quantile of order α = 1%,…,10% of the predictive distribution.

Values of LPS (computed with natural log, the higher the better) and CRPS (the lower the better, as it can be perceived as a generalization of *ex-post* absolute error) shown in Table 1 and Table 2 suggests that the model that dominates the comparison is the Beta-Gen-t-EGARCH with asymmetry in scale and location (S/L-DCS(1,1)-Gt-AL-AS), though it is followed by the AR(1)-GJR-GARCH model with conditional *t*-distribution and In-Mean effect. Asymmetry in the volatility equation (introduced via κ^* or α_1^+ parameters) seems to be the most important factor for goodness-of-fit measured by the LPS. Differences in the average predictive standard deviation or average length of predictive intervals (constructed from quantiles) convey some information as to differences in shape of the predictive distribution between the models implied by the differences in assumptions concerning the type of the conditional distribution. VaR estimates perform best for S/L-DCS(1,1)-Gt-AL-AS, followed by the model with GED distribution. Interestingly for h = 1 N-GARCH model performs badly for $0.01 < \alpha < 0.05$ and has good performance for $0.05 < \alpha < 0.1$.

Conclusions

The objective of the paper is to verify predictive performance (in terms of density forecasts) of some of recently proposed models for daily financial returns. We focus on Beta-Gen-*t*-EGARCH specification of Harvey and Lange (2015, 2016). We conduct an empirical comparison of out-of-sample predictive performance of the model against some of alternatives that are well-established in empirical finance, like GJR-*t*-GARCH model. As the point of interest here is on density forecasting performance (in particular risk evaluation) it is important that estimation uncertainty problem is addressed. This is even more important as the model under consideration uses a complicated conditional sampling distribution, namely a Generalized *t*-distribution. The distribution nests a number of interesting special cases such as *t* distribution and GED distribution. Quantification of the estimation uncertainty for shape (in particular tail) parameters might be challenging though it has to be taken into account for the sake of density forecasting. This is why we make use of Bayesian inference methods which is an interesting contribution.

The empirical comparison conducted here uses daily S&P500 logarithmic returns data. We investigate sequences of 1525 density forecasts. The results confirm that the new specification is an interesting tool of empirical financial analyses, as Beta-Gen-t-EGARCH (with asymmetry in log-scale equation) dominates other specifications in terms of criteria for density forecast comparison like LPS or CRPS in all the horizons under consideration (though the difference against the best 'traditional' model is not large). Moreover, Beta-Gen-t-EGARCH provides best VaR estimates during the verification period. However, the gain over the AR(1)-t-GJR-In-Mean model is rather moderate. In general the results indicate the importance of asymmetric modelling of the news impact curve and of heavy-tailed conditional distribution, which is not unexpected. Moreover, the results suggest that tail asymmetry might be an important feature of such models. This is because all the models underestimate risk in the lower tail and overestimate risk in the upper tail (according to VaR results) which is a suggestion for further research.

Acknowledgements

The research was supported by the funds granted to the Faculty of Management, Cracow University of Economics within the framework of the subsidy for the maintenance of research potential.

References

Creal, D., Koopman, S. J. & Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics*, 29, 552-563.

Creal, D., Koopman, S. J. & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28, 777-795.

Glosten, L., Jagannathan, R. & Runkle, D. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779-1801.

Gneiting, T., & Raftery, A. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477).

Harvey, A.C. (2013). Dynamic models for Volatility and Heavy Tails: with applications to financial and economic time series. Econometric Society Monograps, Cambridge University Press.

Harvey, A.C., & Lange, J.-R. (2016). Volatility Modelling with a Generalized *t*-distribution. *Journal of Time Series Analysis*, 38, 175-190.

Harvey, A.C., & Lange, J.-R. (2015). Modelling the Interactions between Volatility and Returns. *Cambridge Working Papers in Economics 1518*.

Harvey, A.C. & Sucarrat, G. (2014). EGARCH Models with Fat Tails, Skewness and Leverage. *Computational Statistics and Data Analysis*, 26, 320–338.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, 347-370.

Theodossiou, P. (1998). Financial data and the skewed generalized *t* distribution, *Management Science*, 44, 1650-1661.

A bivariate model of the number of children and the age at first birth

Jacek Osiewalski[1], Beata Osiewalska[2]

Abstract

We formulate a joint statistical model of two important demographic variables: (i) the number of children born by a given woman and (ii) her age at the birth of her first child. The proposed specification is based on the so-called ZIP-CP model of bivariate Poisson-type regression that enables to easily examine dependence between two count variables. In our specification the number of children is a ZIP-type variable (in the hurdle model version), while the conditional distribution of the age at first childbirth given the number of children is a Poisson distribution either left-truncated (when a woman has not had any child) or right-truncated (if a woman gave birth to at least one child). The expected values of the underlying Poisson distributions as well as the relation between both variables are functions of the age of a woman and some socio-economic explanatory variables.

Keywords and Phrases: *count data models, bivariate Poisson regression models, Bayesian inference, fertility, fertility forecasting*

JEL Classification: J13, C35, C51

1 Introduction

Identifying socio-economic factors determining fertility (as well as its forecasting) is a crucial issue of current demographic research. The very low level of fertility observed nowadays in many European countries might be caused by the combination of both the postponement of childbearing to older age (the tempo effect) and the tendency to have smaller families (the quantum effect) (Bongaarts and Feeney, 1998; Sobotka, 2003). Examining which of these two effects plays a major role in the reproductive behaviour of a contemporary woman is necessary to effectively forecast her completed fertility (Lee, 1981; Kohler et al., 2001).

It seems that the completed family size and the age at first birth are usually negatively related, meaning that women who enter motherhood at late ages have fewer children (Trussell and Menken, 1978; Kohler et al., 2001). There are many biological and social reasons for such negative correlation (Leridon and Slama, 2008; Schmidt et al., 2012), nevertheless the dependence may change with contextual factors and socio-economic characteristics of a woman (Neels and De Wachter 2010; Berrington et al., 2015). Thus, it is important to jointly model the two basic variables: age at entry into motherhood and the number of

[1] Corresponding author: Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: eeosiewa@cyf-kr.edu.pl.

[2] Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: beata.osiewalska@uek.krakow.pl.

children – first, in order to test their dependence and its possible changing character; second, to make statistical inferences more efficient when the tested dependence is present.

In this paper we propose a joint bivariate statistical model of the number of children born by a given woman and her age at the birth of her first child. The first variable can take only non-negative integer values, while the second can be treated either as a continuous variable or as a count variable. Here we assume that age is measured in full years (above some threshold, e.g. 15 years old), so we jointly model two count variables. By adopting such an approach we can take advantage of recent specifications proposed in the statistical literature.

Modelling univariate count data by means of Poisson type regression models is nowadays a routine approach, and several specifications have been proposed for the bivariate case; see e.g. Cameron and Trivedi (1998, 2005). In this paper we modify the so-called ZIP-CP (*zero inflated Poisson – conditional Poisson*) model, analysed by Osiewalski (2012) and Osiewalski and Marzec (2016), which is the generalised version of the P-CP (*Poisson – conditional Poisson*) specification, proposed by Berkhout and Plug (2004). We replace the regular Poisson conditional part by a more appropriate distribution of the age at first childbirth given the number of children. This conditional distribution is a Poisson distribution, either left-truncated (when a woman has not had any child) or right-truncated (if a woman gave birth to at least one child). We obtain the likelihood function for our non-standard bivariate Poisson-type model, formulate important parametric hypotheses and consider forecasting issues. In order to conduct exact small-sample inference, we propose the Bayesian approach equipped with MCMC simulation tools. A preliminary empirical illustration is based on the *Generations and Gender Survey* (GGS) data for Poland.

In the next section we present the probabilistic foundations of our model, i.e. the discrete bivariate distribution used to jointly describe two demographic variables. Section 3 is devoted to our statistical model, the form of the likelihood function and the Bayesian analysis. Section 4 contains a preliminary empirical example.

2 Foundations of the new statistical model

We consider the joint distribution of two random variables (Y_1, Y_2) that can take non-negative integer values. In the bivariate P-CP distribution analysed by Berkhout and Plug (2004) the probability distribution of (Y_1, Y_2) is as follows:

$$\Pr\{Y_1 = i, Y_2 = j\} = \Pr\{Y_1 = i\} \Pr\{Y_2 = j \mid Y_1 = i\} = g(i)\, h(j,i)\,, \tag{1}$$

where $i, j \in N \cup \{0\}$, the marginal distribution of Y_1 is Poisson with mean and variance λ_1, and the conditional distribution of Y_2 given Y_1 is Poisson with mean and variance $\lambda_2 \exp(\alpha Y_1)$, i.e.

$$g(i) = \exp(-\lambda_1)(\lambda_1)^i / i!, \quad h(j,i) = \exp[-\lambda_2 \exp(\alpha i)](\lambda_2)^j \exp(\alpha i j) / j!. \tag{2}$$

If $\alpha \neq 0$, then two count variables are stochastically dependent and the variance of Y_2 is greater than its expectation. The dependence between these variables leads to the inflated variance of Y_2, which is often observed in empirical count data. The Poisson distribution of Y_1 does not have this property. Also, the P-CP model puts restrictions on the dependence of two variables, as the sign of covariance between Y_1 and Y_2 depends only on the sign of α, and not on λ_1 or λ_2, which are described by explanatory variables in statistical applications. An appropriate generalisation was proposed by Osiewalski (2012); it allows for the dependence of the sign of covariance on λ_1. This more general class of distributions (called ZIP-CP) is characterized by the same conditional distribution of Y_2 given Y_1, $\Pr\{Y_2 = j \mid Y_1 = i\} = h(j,i)$, and by the ZIP-type distribution of Y_1, with zero treated separately:

$$\Pr^*\{Y_1 = i\} = g^*(i) = \begin{cases} \gamma & for\ i = 0, \\ \dfrac{1-\gamma}{1-g(0)}\, g(i) & for\ i \in N, \end{cases} \tag{3}$$

where γ belongs to the (0, 1) interval, and g and h are the functions as in (2). If $\gamma = g(0)$, then $\Pr^*\{Y_1 = i\} = g^*(i) = g(i) = \Pr\{Y_1 = i\}$; we have the P-CP case. If $\gamma > g(0)$, the distribution of Y_1 is of the ZIP type. However, the specification (3) is more general as it also allows $\gamma < g(0)$; it is known in the literature as the *hurdle model* (Cameron and Trivedi, 2005, p. 680) and is compared to the original ZIP model by Winkelman (2008). The *hurdle model* form of our ZIP type specification for Y_1 leads to a very simple statistical specification, making estimation – as well as testing of the standard Poisson case – relatively easy. The ZIP-CP distribution enables inflating variances of both count variables (although they are not symmetrically treated) and making their dependence more complex than in the P-CP case.

As yet we have not focused on the interpretation of Y_2 (the age at first birth); this variable cannot have the regular Poisson conditional distribution given Y_1. Let a be the actual age of a woman and let b denote the end of the reproductive age. We assume that the conditional distribution of Y_2 given Y_1 is Poisson, but either right-truncated at $\min\{a, b\}$ (if $Y_1>0$) or left-truncated at a (if $Y_1=0$). In the first case (a woman has already had at least one child) the age at first birth has to be between the beginning of the reproductive age and either woman's current age a or the end of the reproductive age b (whichever is smaller). In the latter case, when a woman has had no children ($Y_1=0$), the age at first birth has to be between a and b.

Thus $\Pr\{a \le Y_2 \le b \,|\, Y_1 = 0\}$ is the probability that a woman, which is childless at age *a*, will have a child (and obviously equals to zero if $a>b$), while $\Pr\{Y_2 > b \,|\, Y_1=0\}$ is the probability that she will remain childless (it equals to one if $a \ge b$). For further consideration, assume that age is counted in full years exceeding some threshold, e.g. in years over 15; that is, $b=34$ (as the reproductive age of a woman is [15, 49]) and at this threshold $a=0$ and $Y_2=0$.

While the interpretation of Y_1 is straightforward (the number of children ever born by a woman), the meaning of Y_2 is more subtle, as only its values up to b can represent the age at first birth. In the conditional distribution of Y_2 given $Y_1=0$, which is not truncated at b, the values above b serve to describe childlessness after the reproductive age, by attaching a positive probability to such situation. Summing up the assumptions we have already introduced, we propose the following joint distribution of Y_1 and Y_2:

$$\Pr^*\{Y_1 = i, Y_2 = j\} = \begin{cases} \gamma \dfrac{h(j,0)}{1-\sum_{l=0}^{a-1} h(l,0)}, & i=0 \wedge j \ge a, \\ \dfrac{1-\gamma}{1-g(0)} \dfrac{g(i)\, h(j,i)}{\sum_{l=0}^{\min\{a,b\}} h(l,i)}, & i>0 \wedge j \le \min\{a,b\}, \\ 0, & i>0 \wedge j > \min\{a,b\},\ i=0 \wedge j<a. \end{cases} \tag{4}$$

3 The statistical model, its likelihood function and Bayesian analysis

The statistical model proposed in this paper is designed to cope with cross-section micro-level data for women that may differ in terms of age and other characteristics. Consider K bivariate observations (Y_{1k}, Y_{2k}; k = 1,2,...,K), where Y_{1k} is the number of children born by the k-th woman and Y_{2k} denotes her age at first birth (in full years over 15). The pairs (Y_{1k}, Y_{2k}) are independent and have different distributions with the probability function of the general form (4), so our model amounts to the following parametric class of distributions:

$$\Pr^*\{Y_{1k} = i, Y_{2k} = j;\ \theta\} = \begin{cases} \gamma_k \dfrac{h_k(j,0)}{1-\sum_{l=0}^{a_k-1} h_k(l,0)}, & i=0 \wedge j \ge a_k, \\ \dfrac{1-\gamma_k}{1-g_k(0)} \dfrac{g_k(i)\, h_k(j,i)}{\sum_{l=0}^{\min\{a_k,b\}} h_k(l,i)}, & i>0 \wedge j \le \min\{a_k,b\}, \\ 0, & i>0 \wedge j > \min\{a_k,b\},\ i=0 \wedge j<a_k, \end{cases} \tag{5}$$

where a_k is the actual age of the k-th woman (in full years over 15),

$$g_k(i) = \exp(-\lambda_{1k})(\lambda_{1k})^i / i!, \quad \lambda_{1k} = \exp(x_k\beta_1), \quad \gamma_k = \exp(-e^{z_k\delta}\lambda_{1k}) = \exp(-\exp(z_k\delta + x_k\beta_1)),$$

$$h_k(j,i) = \exp[-\lambda_{2k}\exp(\alpha_k i)](\lambda_{2k})^j \exp(\alpha_k ij)/j!, \quad \lambda_{2k} = \exp(w_k\beta_2), \quad \alpha_k = s_k\beta_3.$$

In the above formulas x_k, z_k and w_k are row vectors of explanatory variables that determine the marginal probabilities of Y_{1k} and conditional probabilities of Y_{2k} given Y_{1k}, respectively, while s_k is a row vector of explanatory variables explaining possible differences in dependence between Y_{2k} and Y_{1k} for different groups of women. The age variable, a_k, seems an obvious explanatory variable, appearing in all four vectors – x_k, z_k, w_k and s_k. Obviously, the role of the explanatory variables depends on the column vectors of parameters β_1, β_2, β_3 and δ, grouped in θ, the vector of all parameters. In particular, stochastic independence between the number of children (Y_{1k}) and mother's age at first birth (Y_{2k}) is equivalent to β_3=0, while $\delta\neq0$ means that $\Pr^*\{Y_{1k}=0;\ \theta\}$ deviates from the value corresponding to the Poisson distribution with mean (and variance) λ_{1k}.

When specifying the likelihood function that corresponds to (5) we have to remember that the age at first birth (Y_{2k}) is not observed when the woman has not born a child (Y_{1k}=0). Thus, the likelihood function is the product of $K=K_0+K_1$ factors, where K_0 factors (of the form γ_k) correspond to the probability of zero in the marginal distribution $\Pr^*\{Y_{1k}=i;\ \theta\}$ and K_1 factors correspond to the joint probability (5) for $i > 0$ and $j \le \min\{a_k, b\}$. Denote the observed values of Y_{1k} and Y_{2k} as y_{1k} and y_{2k}, respectively; then the likelihood function takes the form

$$L(\theta; y) = \left[\prod_{k:\,y_{1k}=0} \gamma_k\right]\left[\prod_{k:\,y_{1k}>0} \frac{1-\gamma_k}{1-g_k(0)} \frac{g_k(y_{1k})h_k(y_{2k}, y_{1k})}{\sum_{l=0}^{\min\{a_k,b\}} h_k(l, y_{1k})}\right], \tag{6}$$

where y groups all the values y_{1k} and y_{2k}. The likelihood function (6) enables us to test many specific parametric hypotheses, but the most important from the theoretical point of view is the one stating that β_3=0. If β_3=0, Y_{1k} and Y_{2k} are independent random variables that lead to two separate models and likelihood functions: one built for $K=K_0+K_1$ values y_{1k} and involving β_1 and δ, the other built for K_1 values y_{2k} and involving β_2. Only for $\beta_3\neq0$ our joint bivariate model can lead to inferential gains and makes joint forecasting of both demographic variables more efficient than treating them separately.

Our inference on the parameters and unobserved values of both demographic variables will follow the Bayesian statistical approach, where a probability measure (prior distribution) on the parameter space is defined. We assume prior independence among all parameters in θ and the standard normal prior $N(0, 1)$ for each individual parameter. Zero prior expectations mean that the simplest model (with no ZIP effect, no dependence and no explanatory variables) gets the highest prior chance, but unitary standard deviations ensure significant prior chances for specifications being far from the simplest one. It seems that such simple joint prior distribution brings little initial information and guarantees easy Monte Carlo simulations from the posterior distribution. Obviously, the sensitivity of inferences with respect to the form of the prior distribution is an empirical question, to be answered with the data at hand. Following Bayesian statistical paradigm makes our inference not only exact (small-sample), despite a non-standard form of the likelihood (6), but also coherent and intuitive.

The statistical analysis based on our model (5) can serve different purposes. First, using the posterior density $p(\theta \mid y) \propto p(\theta) L(\theta; y)$, where $p(\theta)$ denotes the prior density, we can test basic hypotheses and point at the explanatory variables that are most important in determining the number of children and the age at first birth. The simplest way to test hypotheses is the so-called Lindley-type approach, see e.g. Osiewalski and Marzec (2016).

Second, our model can serve different forecasting purposes. For a woman outside the dataset, but with given characteristics represented by the row vectors x_f, z_f, w_f and s_f, the predictive probability that $Y_{1f} = i$ and $Y_{2f} = j$ is obtained by averaging (5), interpreted as the conditional probability given θ, with $p(\theta \mid y)$ as the weight function:

$$\Pr^*\{Y_{1f} = i, Y_{2f} = j \mid y\} = \int_\Theta \Pr^*\{Y_{1f} = i, Y_{2f} = j \mid \theta\} p(\theta \mid y)\, d\theta . \tag{7}$$

In order to examine the explanatory power of our model and to infer on fertility of women with certain characteristics, we compute the predictive probabilities: marginal for $Y_{1f} = i$ and conditional for $Y_{2f} = j$ given $Y_{1f} > 0$:

$$\Pr^*\{Y_{1f} = i \mid y\} = \int_\Theta \Pr^*\{Y_{1f} = i \mid \theta\} p(\theta \mid y)\, d\theta , \tag{8}$$

$$\Pr^*\{Y_{2f} = j \mid y, Y_{1f} > 0\} = \sum_{i>0} \Pr^*\{Y_{1f} = i, Y_{2f} = j \mid y\} \Big/ \sum_{i>0} \Pr^*\{Y_{1f} = i \mid y\} ; \tag{9}$$

they can be compared to observed frequencies.

Third, interesting forecasts can be made for women from the dataset. In this case we can consider one of the complimentary predictive probabilities: that a woman, which is childless at age a, will have a child or that she will remain childless. However, this is left for future research.

In order to simulate samples from the joint posterior distribution of θ, the vector of model parameters, and to approximate the integrals above, we will use the Metropolis-Hastings sequential chain, one of the Markov Chain Monte Carlo (MCMC) simulation techniques.

4 A preliminary empirical example

Our empirical example is based on the first wave GGS data for Poland. The survey was conducted in 2011 and includes respondents between 18 and 79 years old. The only purpose of our preliminary analysis is to check whether the proposed model can describe small data sufficiently well. Thus we have selected only women at the age of 33, who have completed tertiary education, are married and live in urban areas. Our final sample consists of only 52 women, of which 7 are childless (13,5%). The half of the women have one child (48,1%), one-third have two children (34,6%) and the rest (3,8%) have already three children. The most common age at first birth is 28 (22,2%). Almost half of the women gave birth after that age (46,7%), and one-third have the first child before the age of 28 (31,1%). At this stage we do not include any explanatory variables (only intercepts), thus the final model includes four scalar parameters β_1, β_2, β_3 and δ.

The basic characteristics of marginal posterior distributions (means, standard deviations, 0.05 and 0.95 quantiles) are presented in Table 1. All the distributions, besides the marginal posterior distribution of β_1, are separated from zero. In particular, it is *a posteriori* almost certain that β_3 is negative. This confirms the negative dependence between the number of children and the age at first birth; it also proves the necessity to jointly model these two variables. In addition, the parameter δ is positive (with very high posterior probability), thus the ZIP effect is present and childlessness seems more frequent than the standard Poisson distribution may capture.

$\boldsymbol{\theta_i}$	$\boldsymbol{E(\theta_i \vert y)}$	$\boldsymbol{D(\theta_i \vert y)}$	$\boldsymbol{q_{0.05}(\theta_i \vert y)}$	$\boldsymbol{q_{0.95}(\theta_i \vert y)}$
β_1	-0.144	0.212	-0.499	0.198
β_2	3.054	0.166	2.786	3.327
β_3	-0.256	0.099	-0.419	-0.095
δ	0.814	0.285	0.355	1.290

Table 1. Characteristics of marginal posterior distributions.

The marginal predictive distribution of the number of children, given by (8), is compared to the frequencies observed in the data and presented in Figure 1 (left-hand side). The model efficiently represents the data and properly predicts the number of children of a given woman with chosen characteristics. Although at this stage it underestimates the probability of having two children, we believe that the accuracy will be improved through enlarging the sample size (by considering women of different age and other characteristics) and, therefore, including explanatory variables. The conditional predictive distribution (9) of the age at first birth and the frequencies observed in the data (for 45 women with at least one child) are shown in Figure 1 (right-hand side). As for the very limited sample, the model performs exceptionally well and accurately represents the data.

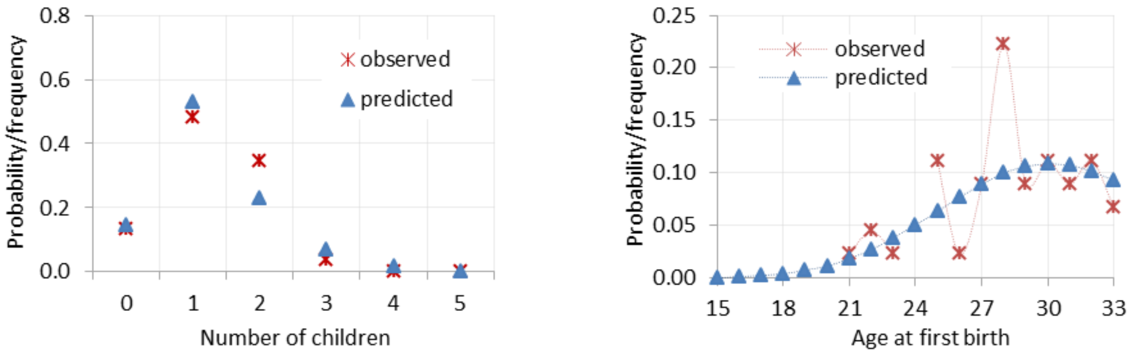


Fig. 1. The marginal predictive distribution of the number of children and the conditional predictive distribution of the age at first birth versus frequencies observed in the data.

To conclude, using our modified ZIP-CP model to jointly analyse the number of children and the age at first birth seems to be well justified by both the dependence between the two variables and the overrepresentation of zero (childlessness). The model also provides with reasonable predictions and thus serves a promising tool to infer about the fertility of a woman (on the basis of much larger and more informative samples).

Acknowledgement

The authors acknowledge support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Berkhout, P., & Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica*, 58, 349–364.

Berrington, A., Stone, J., & Beaujouan, E. (2015). Educational differences in timing and quantum of childbearing in Britain: a study of cohorts born 1940-1969. *Demographic Research*, 33(26), 733.

Bongaarts, J., & Feeney, G. (1998). On the quantum and tempo of fertility. *Population and development review*, 24(2), 271-291.

Cameron, A.C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press, New York.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and application*. Cambridge University Press, New York.

Kohler, H. P., Skytthe, A., & Christensen, K. (2001). *The age at first birth and completed fertility reconsidered: findings from a sample of identical twins*. MPIDR Working Paper 2001-006. Max Planck Institute for Demographic Research, Rostock.

Lee, R. D. (1981). A model for forecasting fertility from birth-expectations data. In: Hendershot, G. E., & Placek, P. J. (eds.), *Predicting Fertility: Demographic Studies of Birth Expectations*, Lexington, MA: Heath, 75-99.

Leridon, H., & Slama, R. (2008). The impact of a decline in fecundity and of pregnancy postponement on final number of children and demand for assisted reproduction technology. *Human Reproduction*, 23(6), 1312-1319.

Neels, K., & De Wachter, D. (2010). Postponement and recuperation of Belgian fertility: how are they related to rising female educational attainment?. *Vienna Yearbook of Population Research*, 8, 77-106.

Osiewalski, J. (2012). *Dwuwymiarowy rozkład ZIP-CP i jego momenty w analizie zależności między zmiennymi licznikowymi*. In: *Spotkania z królową nauk (Księga jubileuszowa dedykowana Profesorowi Edwardowi Smadze)*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków, 147–154.

Osiewalski, J. & Marzec, J. (2016). Joint modelling of two count variables when one of them can be degenerate. Technical report, Cracow University of Economics.

Schmidt, L., Sobotka, T., Bentzen, J. G., & Andersen, A. N. (2012). Demographic and medical consequences of the postponement of parenthood. *Human reproduction update*, 18(1), 29-43.

Sobotka, T. (2003). Tempo-quantum and period-cohort interplay in fertility changes in Europe: Evidence from the Czech Republic, Italy, the Netherlands and Sweden. *Demographic Research*, 8(6), 151-214.

Trussell, J., & Menken, J. (1978). Early childbearing and subsequent fertility. *Family Planning Perspectives*, 10(4), 209-218.

Winkelman, R. (2008). *Econometric analysis of count data*. Springer-Verlag, Berlin Heidelberg.

Determinants of the renewable energy development in the EU countries. A 20-year perspective

Monika Papież[1], Sławomir Śmiech[2], Katarzyna Frodyma[3]

Abstract

The objective of the paper is to identify factors shaping energy policy in the EU countries in the middle of 1990s. This objective is achieved in two stages. During the first one, using PCA, we describe the distribution of energy sources in 26 EU countries in 1995. We find four main principal components distinguishing countries in terms of energy source consumption in 1995, and then, during the second stage, use them as potential determinants of renewable energy development. We also consider several factors related to energy security, environmental concerns, economy and politics.

Applying two methods of variable selection, namely, the best subset regression and the lasso method, we demonstrate that the present (in 2014) share of RES in the energy mix significantly depends on the condition of the EU countries in the middle of 1990s. Our study reveals that the distribution of energy sources in 1995 is the main determinant of renewable energy development. Countries without their own fossil fuel sources are the ones with the greatest development of renewable energy. Other factors affecting RE development include: GDP per capita, the Shannon–Wiener index (SWI) reflecting the concentration of energy supply, and the cost of the consumption of energy obtained from fossil fuels in relation to GDP.

Keywords: *Renewable energy, European Union, LASSO, PCA*

JEL Classification: C31, C38, N74, Q2

1 Introduction

Fossil fuels lay the foundation of energy balance in the European Union member countries. Their share in the total primary energy supply (hereafter TPES) in 2014 amounted to 34.4% for oil, 21.4% for natural gas and 16.7% for coal. Nuclear energy constituted 14.1% of the TPES, while renewable energy (hereafter RE) – 12.5%. In 2014 the TPES in the EU member countries equalled to 1606 Mtoe, and net import constituted 54.8% of the TPES and increased in comparison with 1995, when it constituted 44%. A growing dependence of the EU on imported energy, diminishing deposits of its own resources, and the necessity to provide energy at acceptable prices lead to the increase of the significance of the factors connected

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: papiezm@uek.krakow.pl.

[2] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: smiechs@uek.krakow.pl.

[3] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: frodymak@uek.krakow.pl.

with energy security and energy policy. The EU dependence on import of energy sources contributes to the growing interest in renewable energy sources (hereafter RES), which is reflected in the introduction of relevant directives in the area of energy policy. Specifying realistic energy policy targets requires thorough knowledge of this area, including knowledge of factors influencing RE development. One might expect abundant subject literature addressing this important issue, however, only several papers directly investigate the determinants of such development (e.g. Marques et al., 2010; Marques and Fuinhas, 2011; Aguirre and Ibikunle, 2014; Cadoret and Padovano, 2016; Lucas et al., 2016, Frodyma, 2017).

The aim of the paper is to identify the main determinants of RE development in the EU countries, with emphasis on how the distribution of energy sources in the middle of 1990s influences RES in 2014. The year 1995 is used as a reference point because it was the year in which the EU initiated legal procedures aimed at promoting RE development and in which the European Commission published *Green Paper* (1995) delineating the European Union energy policy and listing three basis targets connected with gas and electricity monopolies.

The share of RE in the TPES is selected as the dependent variable. This particular variable is chosen because, as stated by Aguirre and Ibikunle, 2014, firstly, policy targets are focused on achieving a certain share of energy from renewable sources in the energy mix, and, secondly, it is expected that RE will successively displace more-polluting energy sources in the energy mix.

In our study it is assumed that RE development is a long-lasting process, and this is why we decide to replace a panel approach commonly used in other studies (see: Marques et al., 2010; Marques and Fuinhas, 2011; Aguirre and Ibikunle, 2014; Cadoret and Padovano, 2016; Lucas et al., 2016), with a cross-sectional regression and a reliable method of variable selection to investigate to what extent the determinants from 1995 influence RE development in 2014. This approach allows us to analyse the impact of a set of determinants connected with the environment, security of supply, economy, and politics.

The first stage of our study focuses on the distribution of energy sources, which seems a key factor in RE development. Several factors contribute to its significant role. Firstly, it is inextricably connected with national energy security, as most EU countries have to import energy sources (mainly oil and gas) from countries outside the EU, and only several of them have sufficient domestic energy sources. Secondly, the share of energy sources in the energy mix in a given country exerts a direct influence on the natural environment, since using e.g. coal is connected with high emissions of pollutants, while using solar energy is not. As opposed to earlier studies (Marques et al., 2010; Marques and Fuinhas, 2011; Lucas et al., 2016; and Cadoret and Padovano, 2016; Frodyma, 2015), which consider only the share of particular energy sources in the overall energy consumption, in our study we use PCA.

During the second stage, we analyse a relatively large set of potential variables, which includes the ones taken into consideration by policy makers when they make decisions regarding the distribution of energy sources, such as energy security, energy self-sufficiency, environmental costs, international treaties and commitments, political costs of potential changes in the structure of energy production, or the political power of interest groups, including miners. In order to distinguish the key determinants of RE development, we use two statistical methods: the best subset regression (hereafter BSR) and the lasso method (Tibshirani, 1996).

2 Methodology

In order to find the determinants of RE development, we analyse 15 (some highly correlated[4]) variables which describe various aspects of European countries in 1995 (outside the area of energy policy) and could be related to the share of RE in the TPES in 2014. The analysis is conducted with the use of standard linear regression framework. The model is given by:

$$Y = X\beta + \varepsilon, \tag{1}$$

where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ is a vector of i.i.d. random variables with mean zero and variance σ^2. Y is a $n \times 1$ vector of response, $X = (X_1, ..., X_p)$ is the $n \times p$ matrix of predictors, and β is a *p*-dimensional vector of model parameters.

There are two approaches to limiting the number of variables in the regression model.

The first one is known as the best subset regression (BSR) model. In this approach, each possible subset of predictors of size *k* is considered, where $k \in \{1,2,.., p\}$, and *p* is the total number of variables. Each model is estimated by least squares and then compared to other models by applying a specific criterion (usually adjusted R^2, Akaike information criterion (AIC) or Bayesian information criterion (BIC)[5]). The best subset of variables (and the best model) is the one which maximizes the criterion used. However, this method is time consuming since the total number of combinations of *p* variables is given by: $2^p - 1$. The BSR is, as Efron et al. (2004, p.409) claim, "overly greedy, impulsively eliminating covariates which are correlated with" other covariates.

To avoid this problem, Tibshirani (1996) proposes the lasso (the least absolute shrinkage selection operator) method, which, in order to estimate regression parameters, minimizes the

[4] Detailed results of correlations are available from the authors upon request.
[5] BIC generally places a heavier penalty on models with many variables, and hence results in the selection of more parsimonious models.

sum of squares of residuals with a constraint for parameters:

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \| Y - X\beta \|_2^2 + \lambda \| \beta \|_1, \tag{2}$$

where $\| \cdot \|_1$ is the L_1 norm (the sum of absolute values of the vector's entries), $\lambda \geq 0$ is a tuning parameter which controls the amount of shrinkage applied to the estimates, and $\| \cdot \|_2$ is the standard L_2 norm. For $\lambda = 0$, Eq. (1) is the standard Ordinary Least Squares (without any regularizations). For large λ, $\hat{\beta}$ shrinks to 0, which results in the empty model. Due to the form of the penalty in the lasso (L_1 norm), for moderate λ there are only several non-zero estimates (among all possible choices), thus, the method is useful in the variable selection problem and is our second methodological approach. In 2002 Efron et al. (2002) proposed the LARS (least angle regression) algorithm, which provides an efficient way of using the lasso method and connects the lasso with forward stagewise regression. That is why this method is considered to be one of the most effective ways of solving the variable selection problem in regression applications.

3 Data description and preliminary statistics

We conduct our empirical analysis with the data describing the share of RES in TPES in a sample of 26 European Union member countries. The analysis covers the period between 1995 and 2014, and the data are obtained from the European Commission websites[6]. The study does not include Cyprus and Malta due to the fact that in the analysed period the share of RES in TPES in these countries is almost zero. The first dataset describes the distribution of energy sources. The variables analysed reflect the share of energy sources in TPES and include: solid fuels (including hard coal) (COAL), crude oil and petroleum products (OIL), natural gas (GAS), nuclear energy (NUCL) and renewable energy sources (RES). The second stage of our study is devoted to the search for the determinants of the share of RES. On the basis of literature, we have compiled a list of potential determinants whose role in renewable energy development is verified in our study. Additionally, a set of determinants is expanded by principal components describing the distribution of energy sources in 1995 obtained in the second stage of the study. Following Marques et al. (2010), Cadoret and Padovano (2016) and Lucas et al. 2016, we have chosen a set of potential predictors and divided them into four categories: environmental variables (carbon intensity of energy use - CITPES; the CO_2

[6] Energy datasheets: EU-28 countries (https://ec.europa.eu/energy/en/data-analysis/country, accessed on 30.10.2016 r.

emissions per capita - CICAP), security of supply variables (net energy import in relation to the TPES - DEP; energy self-sufficiency rate - ESS; energy mix concentration - SWI; energy mix diversification - HHI; diversification of energy sources in electricity generation - HHIE; net energy import to GDP - NDEPGDP; net energy import per capita - NDEPCAP; net energy import in relation to GDP – VNTDEPGDP), economic variables (GDP per capita - GDPCAP; energy consumption per capita – EICAP; energy consumption per GDP - EIGDP; consumption of energy obtained from fossil fuels in relation to GDP – VEIFFGDP), and political variables (a dummy variable to identify the EU countries in the year 1995)[7].

4 Results and discussion

The first stage of our study focuses on the analysis of the distribution of energy sources in 1995 in the EU countries and reveals that the first principal component is related to coal and oil variables, which means that it distinguishes countries which either use or do not use coal or crude oil as their main energy sources. Consequently, these countries can be described as countries using either dirty or clean energy. Similarly, the fourth principal component[8] represents the European countries which primarily use clean energy as their main energy source and limited amounts of other energy sources. Crude oil is positively and nuclear energy is negatively correlated with the second principal component, i.e. the countries which use oil as their main energy source and do not use nuclear energy are described by the second principal component. The third principal component is related to all variables, but the factor loading is negative only for natural gas, which means that this component distinguishes countries using natural gas as their main energy source[9].

The second stage of our analysis is devoted to finding determinants of RE development. All variables described in Section 3 are assumed to be potentially important. At first, we present the results of BSR models.

The final BSR model includes seven variables. The detailed results obtained for this model are presented in Table 1. All variables included in the model are significant. Two variables, PC1 and energy consumption per capita (EICAP), have negative parameters and are highly related to high coal consumptions. Thus, it seems obvious that for the countries with a large share of coal consumption in the energy mix RE development is not a priority. It

[7] Detailed information regarding variable definitions and descriptive statistics are available from the authors upon request.
[8] The fifth principal component represents only 1.2% of total variance, thus, is not taken into account.
[9] Detailed data on principal components are available from the authors upon request.

means that countries with high energy consumption per capita in 1995 are not interested in investing in RES. Our results are confirmed by Aguirre and Ibikunle (2014), but not confirmed by Marques et al. (2010), and Lucas et al. (2016), who report opposite relations.

Two other determinants in the model reflect the distribution of energy sources in 1995, namely PC3 and PC4. Both have positive parameters in the final BSR. Countries with large values of these two principal components have either a low share of natural gas consumption in the energy mix (PC3) or a high share of RES (PC4). Countries with such a profile of energy consumption reveal a considerable development of their RES. It is not difficult for countries with significant shares of RES in 1995 to continue their development because they already have relevant legal regulations, technological facilities or conductive geographical and climatic conditions.

However, there is a convincing interpretation of a negative impact of high share of natural gas on RE development. Countries which use natural gas as their main energy source (the Netherlands, the United Kingdom, Romania) are the largest natural gas producers in the UE, and, as such, are not dependent on import of energy sources. Since they can satisfy their basic energy needs with energy they produce themselves, they are not affected by political risks resulting from other energy producers suspending their fossil fuel supplies. Thus, as they already are energy self-sufficient, they are not interested in developing RE. What is more, natural gas is a relatively clean energy source, so, even if a country reduces its consumption, it will not translate into a substantial reduction of CO_2 emissions (which happens when coal consumption is reduced). A positive parameter for GDP per capita (GDPCAP) means that the richest countries are more prone to invest in RE: they can afford expensive RE technologies and support subsidies for promoting and regulating RE. A positive impact of income on the promotion of RE is also found by Marques et al. (2010). Energy security parameter SWI, which describes concentration of energy supply, is also positive. The more diversified energy sources a country has, the more it is interested in developing RE. This means that the relatively even distribution of energy sources in the energy mix motivates the EU countries to promote the development of RE, and this relation is confirmed by Lucas et al. (2016). Finally, the cost of the consumption of energy obtained from fossil fuels in relation to GDP (VEIFFGDP) parameter is positive. Higher prices may make RE more economically viable, thereby encouraging countries to invest in RE.

	Coef	Stand. Error	t-stat	p-value
Intercept	0.05750	0.0556	1.035	0.3146
PC1	-0.13190	0.0457	-2.885	0.0098 ***
PC3	0.30430	0.0615	4.949	0.0001 ***
PC4	0.96170	0.0922	10.428	0.0000 ***
GDPCAP	0.00520	0.0016	3.169	0.0053 ***
EICAP	-0.00004	0.0000	-3.500	0.0026 ***
SWI	0.08880	0.0418	2.126	0.0476 **
VEIFFGDP	0.90800	0.4056	2.239	0.0380 **

Note: Residual standard error: 0.03599 on 18 degrees of freedom, Multiple R^2: 0.9012, Adjusted R-squared: 0.8628; F-statistic: 23.45 on 7 and 18 DF, p-value: 0.0000;
***, ** Represent significance at the 1% and 5% levels respectively

Table 1. The results of the final best subset regression model.

In order to find crucial predictors, the lasso method is used next, and the variables of interest are ordered according to their importance as predictors for RE development. PC4 turns out to be the most influential variable (see Table 2), as it enters the model first (M1). The estimates of the coefficient are positive and remain positive when the tuning parameter increases (for models M2 – M6 including from two to six variables[10]). The second most influential variable, PC3, has positive coefficients, which again remain positive for other models. The next three variables: carbon intensity of energy use (CITPES), PC1 and the CO_2 emissions per capita (CICAP) are negatively related to RE development. All these variables are linked with high concentration of coal in the energy mix. This means that these determinants, apart from having a negative impact on the environment, also have a negative impact on RE development. It is a rather surprising effect, as it might be reasonably expected that high pollutant activity will act as a powerful motivator for investing in RE. Our results are confirmed by studies conducted by Marques et al. (2010), Marques and Fuinhas (2011), and Lucas et al. (2016). The next two variables which characterise security of energy supplies included in the model represent diversification of energy sources: HHIE and HHI. As the parameters of these variables are negative, the results confirm that the countries with less diversified energy sources are not interested in RE development. The volume of net energy import per capita (NDEPCAP) and the SWI are related to the distribution of energy sources.

[10] The table contains selected models, all results are available from the authors upon request.

The parameters of NDEPCAP are negative, while the parameters of SWI are positive, which again confirms the unwillingness of countries with high concentration of energy sources to invest in RE development. Finally, it is worth noticing that the results obtained with the use of BSR and the lasso methods overlap to some extent. Most variables which are included in the BSR when BIC criterion is applied are the ones which enter the regression model first after the application of the lasso approach or are strongly correlated with these variables. The most significant determinants include three variables describing the distribution of energy sources: PC4, PC3, PC1. In case of the best subset method, the final model includes SWI, while the lasso approach includes HHI, which is strongly correlated with SWI, and HHIE, which carries similar information. The direction of the impact of variables obtained in both methods is the same for the most significant variables. High values of PC3 and PC4 stimulate RE development. Large values of PC1 and remaining variables related to large coal consumption are obstacles for RE development, while high diversification of energy sources (small values of HHI, HHIE and a large value of SWI) encourage RE development.

	M1	**M2**	**M3**	**M4**	**M5**	**M6**
PC4	0.436	0.489	0.549	0.560	0.620	0.663
PC3		0.040	0.095	0.104	0.173	0.216
CITPES			-0.00008	-0.000008	-0.00008	-0.00009
PC1				-0.00511	-0.04965	-0.06306
CICAP					-0.000002	-0.000002
HHIE						-0.02459

Table 2. The estimates of the most influential variables indicated by lasso shrinkage.

Conclusions

Our analysis reveals that RE development in the EU member countries is relatively diverse. In the analysed period all EU countries increase their shares of RES in the energy mix, however, the increase is uneven, and the shares of particular RES in particular countries are not the same. The objective of the study is to identify factors determining energy policy in the EU member countries in the middle of 1990s.

This objective is achieved in two stages. During the first one we use PCA to identify the distribution of five main energy sources in 1995. We find four main principal components distinguishing countries in terms of energy sources consumption in 1995. These principal components allow us to indicate groups of countries with particular energy mixes: the ones which either use both coal and crude oil as their main energy sources (i.e. dirty energy) use or clean energy, or the ones which use mainly RES or other energy sources. Principal components also describe the contrast between the countries which use nuclear energy and do not use crude oil, the countries which use mainly crude oil and do not use nuclear energy, and the countries with natural gas as their main energy source with little share of other sources. Principal components are used during the next stage of our study as potential determinants of RE development.

During the second stage we identify the potential determinants of RE development, which include factors related to energy security, environmental concerns, economy and politics. We use two methods of variable selection to identify the main determinants of RE development, and both of them yield similar results, which can be summarised in the following way.

Firstly, the results indicate that the present (in 2014) share of RES in the energy mix significantly depends (high R^2 obtained for regression) on the set of circumstances in the EU member countries in the middle of 1990s. Secondly, the distribution of five main energy sources in the energy mix in 1995 is a key factor affecting RE development.

The most general conclusion of our study states that countries with the lowest shares of RE are the countries with relatively high energy self-sufficiency, i.e. countries with high shares of their domestic fossil fuel resources in the energy mix. There are two groups of such countries: countries with coal resources (Poland, Czech Republic, Bulgaria, Estonia) and countries with natural gas resources (the Netherlands, the United Kingdom, Romania, Hungary), which are not threatened with other countries' suspension of the export of energy supplies. Consequently, they do not need to develop their RE sector to the extent the countries without their own energy sources do if they want to minimize their dependence on energy import. Moreover, countries with their own energy sources have a well-developed mining industry, which generates employment, thus a sudden transformation from their own fossil fuel market to RES would entail huge changes in the labour market, which might prove risky for both: the country's policy and its economy.

Not only do we demonstrate that the distribution of energy sources is the main determinant of RE development, but we also identify other factors conductive to the increase of the share of RES in the energy mix: GDP per capita, the Shannon–Weiner index (SWI), concentration of the energy supply, and the cost of the consumption of energy obtained from fossil fuels in relation to GDP. Energy consumption per capita hinders RE development. Our study reveals that decisions regarding energy policy are of pragmatic nature and tend to accommodate the needs and requirements of the local environment (the country's energy security and care of its labour market) rather than universal values connected with climate protection.

Acknowledgments

This paper is supported by funds from the National Science Centre (NCN, Poland) through grant No 2015/17/B/HS4/02685.

References

Aguirre, M., & Ibikunle, G. (2014). Determinants of renewable energy growth: A global sample analysis. *Energy Policy, 69*, 374-384.

Cadoret, I., & Padovano, F. (2016). The political drivers of renewable energies policies. *Energy Economics, 56*, 261-269.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499.

EU Commission. (1995). Green Paper: For a European Union Energy Policy. European Commission, COM(94) 659, Brussels 1995.

Frodyma, K. (2015). Rynek energii ze źródeł odnawialnych w Unii Europejskiej–analiza statystyczna. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 941, 63-78.

Frodyma, K. (2017). The analysis of changes in the distribution of renewable energy consumption in the EU countries. In M. Papież & S. Śmiech (Eds.). *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow: Foundation of the Cracow University of Economics, 280-289.

Furnival, G. M., & Wilson, R. W. (2000). Regressions by Leaps and Bounds. *Technometrics, 42*(1), 69-79.

Lucas, J. N., Francés, G. E., & González, E. S. (2016). Energy security and renewable energy deployment in the EU: Liaisons Dangereuses or Virtuous Circle? *Renewable and Sustainable Energy Reviews, 62*, 1032-1046.

Marques, A. C., & Fuinhas, J. A. (2011). Drivers promoting renewable energy: A dynamic panel approach. *Renewable and Sustainable Energy Reviews, 15*(3), 1601-1608.

Marques, A. C., Fuinhas, J. A., & Manso, J. P. (2010). Motivations driving renewable energy in European countries: A panel data approach. *Energy Policy,38*(11), 6877-6885.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

The impact of development of the renewable energy sector in the EU on the energy – growth nexus

Monika Papież[1], Sławomir Śmiech[2], Katarzyna Frodyma[3]

Abstract

The aim of the study is to assess the impact of development of the renewable energy sector on the relations between renewable and non-renewable energy consumption and economic growth in the EU countries. First, two groups of EU member countries with different levels of development of their renewable energy sectors are identified. Then, the dynamic common correlated effect estimator proposed by Chudik and Pesaran (2015) is used in order to estimate the error correction models in both groups of countries as well as in all EU countries. The results reveal that the level of development of the renewable energy sector influences energy-growth nexus. The findings on long-term output elasticities suggest that both renewable and non-renewable energy plays a significant role in economic growth in countries with a relatively high level of development of the renewable energy sector. The results obtained for a short-term perspective indicate bidirectional causal relations in this group. In the other group only non-renewable energy consumption influences economic growth.

Keywords: *renewable energy, economic growth, EU countries, dynamic common correlated estimator*

JEL Classification: C33, C38, Q43

1 Introduction

During the last two decades European Union countries experienced a significant change in the distribution of their energy consumption. The total primary energy supply (TPES) from renewable energy sources (RES) rose from 84.6 Mtoe in 1995 to 210.0 Mtoe in 2015, with an average annual growth rate of 9.5 per cent. At the same time, TPES from non-renewable energy sources (including coal, oil, gas and nuclear sources) decreased from 1583 Mtoe in 1995 to 1402 Mtoe in 2015. Over this period, the share of renewable energy sources in TPES grew from 5.1 per cent to 12.9 per cent. The largest part of renewable energy sources in TPSE came from biofuels and waste, which accounted for about 64.4 per cent of the renewable energy supply in 2015.

Although the share of "new" renewables in EU countries increases rapidly, they still constitute only a small portion of overall energy consumption (in 2015 the share of wind

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: papiezm@uek.krakow.pl.

[2] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland; e-mail: smiechs@uek.krakow.pl.

[3] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: frodymak@uek.krakow.pl.

energy in TPES amounted to 1.6%, and the share of solar energy in TPES to 0.8%). What is more, EU countries differ in their development of "new" renewable energy sources. The data provided by EurObserv'ER's (2015) barometers indicate three areas of importance of renewable energy development in economy, i.e. the share of employment, investments and turnover in renewable energy in the total economy, which makes it possible to distinguish countries with reference to their share of the renewable energy sector in economy in the last several years. Obviously, the larger the share of renewable energy in economy is, the more significant its interrelations with economy should be. The question of interest is, however, whether the renewable energy sector is large enough to exert a noticeable impact on economy. Our study could be placed among numerous papers (see for e.g. Ozturk, 2010; Smyth and Narayan, 2015; Śmiech and Papież, 2014; Tiba and Omri, 2017) which focus on energy consumption-growth nexus. However, as we distinguish renewable and non-renewable energy consumption (like Bhattacharya, 2016, Apergis and Payne, 2012), the results obtained in our study give a more detailed picture of the relations than the results obtained in other papers.

The objective of the paper is to assess the impact of development of the renewable energy sector in economy on the relationship between renewable and non-renewable energy consumption and economic growth in the EU countries. In order to examine this impact, first, two groups of countries with a similar level of development of the renewable energy sector are identified. Next, the modern panel econometric tools are applied to uncover the energy (renewable and non-renewable) – growth nexus within the groups. The analysis is based on the annual panel data in EU member countries from the period 1995–2014.

The study is divided into two stages. The aim of the first one is to identify countries with a similar level of development of the renewable energy sector. In order to measure the significance of this sector in national economies, three variables - employment, turnover and investment - are used, and the *k*-means method is used to identify groups of similar countries.

During the second stage, dynamic error correction models are applied to investigate causality between (renewable and non-renewable) energy consumption and economic growth, and Cobb-Douglas production function is used as a theoretical framework. The analysis is conducted within the groups obtained in the first step, as well as for the whole sample. This stage consists of several steps, each of which is based on econometric methodology allowing for cross-sectional dependence.

The paper contributes to the existing literature in three main aspects.

First, our analysis is one of the first attempts directed at describing the relationship between renewable and non-renewable energy consumption and economic growth in

EU countries. Second, we notice that the share of the renewable energy sector in the EU economy is still limited, but, at the same time, considerable differences between countries can be noticed. Thus, the empirical strategy we employ takes into account a potential impact of the level of development of the renewable energy sector on the relations we analyse. Third, we apply a modern panel econometric approach including the dynamic common correlated effect estimator proposed by Chudik and Pesaran (2015), which allows for commonly observed cross-sectional dependence and potential heterogeneity in a panel of countries.

2 Methodology

To test for the presence of such cross-sectional dependence in our data, we apply Pesaran's (2004) cross-sectional dependence (CD) test, with the null hypothesis claiming no cross-sectional dependence. Next, the panel unit roots tests are applied. We investigate the possible non-stationarity property of the data by applying second-generation panel CIPS unit root test recently proposed by Pesaran (2007), which allows for cross-sectional dependence. The second generation cointegration tests developed by Westerlund (2007) are used, as they take into account cross-sectional dependencies. The long-term relationship between the renewable and non-renewable energy consumption and economic growth is concluded from the Cobb-Douglas production function, as it is convenient in the field. In order to estimate the equations, the Common Correlated Effects estimator (henceforth CCE) proposed by Pesaran (2006), which is robust to the presence of cross-sectional dependence, is used. During the final step of the analysis, the error correction model is estimated with the use of Chudik and Pesaran (2015) dynamic common correlated effect estimator with heterogeneous coefficients.

3 Data

The assessment of the impact of development of the renewable energy sector in economy on the relationship between renewable and non-renewable energy consumption and economic growth is based on the annual panel data. The analysis covers the period 1995–2015 and takes into account 26 European Union member countries (Cyprus and Malta are excluded).

The dataset used to determine the level of development of the renewable energy sector in economy includes three variables. The first variable is the share of employment in renewable energy technologies in total employment (from 15 to 64 years) in each European Union country (EMP). The second variable denotes the share of combined turnover of renewable energy sectors in GDP (in current prices) in each European Union country (TURN). The third variable describes the share of asset finance (investment) in the newly built capacity for all renewable energy sectors in GDP (in current prices) in each European Union country (INVEST). The data on employment, turnover and asset finance are obtained from the EurObserv'ER barometers (2011-2015). Taking into account access to data (the data have been recorded since 2010), we use all available records from the period between 2010 and 2014. In order to establish the engagement of particular countries in the renewable industry (economy), we calculate average values from the period 2010-2014.

The second stage of our study is devoted to the analysis of the relations between renewable and non-renewable energy consumption and economic growth. In order to study these relations, we use real gross domestic product per capita (variable: GDP) in constant 2010 U.S. dollars obtained from the World Development Indicators (World Bank, 2016). Energy consumption from non-renewable energy sources (variable: NREC) and energy consumption from renewable energy sources (variable: REC) are represented by energy use in kg of oil equivalent per capita. These data are obtained from the European Commission websites[4]. As the analysis is conducted using the Cobb-Douglas production function, a set of variables is extended to include two more: real gross capital formation per capita (K) in constant 2010 U.S. dollars as a proxy of capital and labour force participation rate (% of total population aged 15+) (L). Both variables come from the World Development Indicators (World Bank, 2016). All variables are in natural logarithms.

4 Empirical results and discussion

The first stage of the analysis is aimed at finding groups of countries with a similar level of development of their renewable energy sectors. The groups are identified by comparing three variables: the employment in renewable energy sector in total employment (EMP), the share of turnover of renewable energy sector in GDP (TURN), and the share of investments in renewable energy sector in GDP (INVEST). All variables are initially standardized. To visualize the arrangement of countries relevant to these three variables, principal component analysis is applied, and its results are reported in Table 1. The first principal component (PC1) represents the common impact of all three variables. The larger the part of economy related to renewable energy (i.e. more employment in renewable energy sector, larger turnover, more investment in RES), the smaller the value of the PC1. The PC2 represents the contrast between employment and investment in the renewable energy sector. Both PCs represent

[4] Energy datasheets: EU-28 countries (https://ec.europa.eu/energy/en/data-analysis/country, accessed on 28.02.2017 r.

about 82 per cent of total variance. The PC3 shows the contrast between turnover and other two variables (EMP, INVEST).

	PC1	PC2	PC3
EMP	-0.599	-0.536	0.595
TURN	-0.718	-	-0.695
INVEST	-0.354	0.844	0.404
Cumulative variance	0.460	0.817	1.000

Table 1. The results of principal component analysis.

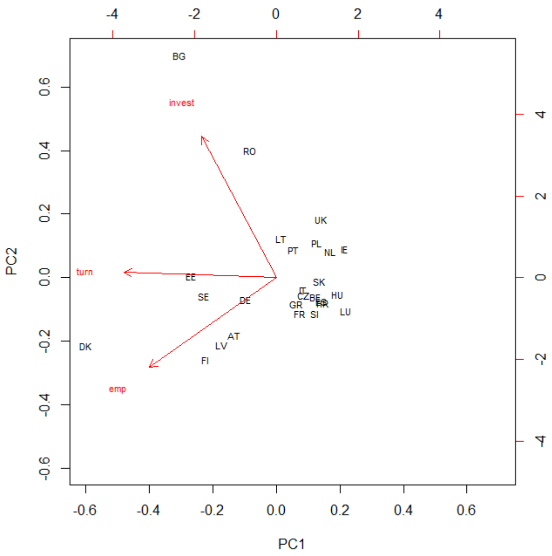


Fig.1. Biplot for variables representing countries' involvement in renewable energy.

The biplot presented in Fig.1 demonstrates the distribution of countries related to the first two principal components. There is a homogeneous group of countries with a positive value of the first principal component, and moderate values of the second principal component. The group consists of countries which, in general, have a low share of the renewable sector in economy. In the remaining countries at least one of the variables is larger than average. Thus, these countries could be seen as the ones which develop their renewable energy sectors most noticeably. Finally, the *k*-means method is employed, and the division (into two groups) observed in the biplot is confirmed.

The first group (GROUP 1) consists of Austria, Bulgaria, Denmark, Germany, Estonia, Finland, Latvia, Romania, and Sweden. The second group (GROUP 2) includes the remaining countries.

The analysis of the relations between non-renewable and renewable energy consumption and economic growth is conducted for both groups (GROUP 1 and GROUP 2) as well as for all European countries (ALL). In order to formally examine the presence of cross-sectional dependence, Pesaran's (2004) cross-sectional dependence (CD) test is used. The CD test statistics are highly significant, so our conclusion is that each series exhibits cross-sectional dependence[5]. The results of the CIPS unit root test proposed by Pesaran (2007) confirm that at least three of the time series are non-stationary, and the Westerlund (2007) cointegration test confirms the existence of the long-term relationship between economic growth and non-renewable and renewable energy consumption[6]. Next, the cointegrating vector is estimated, and the results of the estimation of the long-term relationship are reported in Table 2, which presents the results obtained for CCE MG method proposed by Pesaran (2006).

All variables are expressed in natural logarithms, and the estimated coefficients can be interpreted as long-term elasticity. The empirical findings for all European countries (ALL) (see Table 2) reveal that long-term elasticities of non-renewable energy and renewable energy consumption are equal to 0.124 and 0.039, respectively. Similarly, in countries from the first group (GROUP 1), which display a relatively high level of development of the renewable energy sector, long-term elasticities of non-renewable energy and renewable energy consumption are equal to 0.086 and 0.062, respectively. The coefficients are significant at the 5 per cent level, and the signs of these four variables are positive and follow our expectations for both energy proxies. In turn, the results obtained for the countries from the second group (GROUP 2) (see Table 2), which are characterized by very low employment, turnover and investment in the renewable energy sector, indicate that a one-per cent increase in non-renewable energy consumption increases output growth by 0.117 percentage points.

During the final step of the analysis we estimate a panel vector error correction model using dynamic CCE estimator (Chudik and Pesaran, 2015) to investigate Granger causality between the variables. Table 3 demonstrates F-statistics and t-statistics, which are useful in the short- and long-term Granger causality analyses for all European countries (ALL), countries from the first group (GROUP 1), and countries from the second group (GROUP 2),

[5] Detailed results of the CD test are available from the authors upon request.
[6] Detailed results of Pesaran (2007) as well as Westerlund's panel cointegration tests (2007) are available from the authors upon request.

respectively. For all specifications the results reveal that the estimated coefficients of the error correction terms are all statistically significant (with one exception, see Table 3) and negative (as expected). These results indicate that all variables used in this study respond to deviations from the long-term equilibrium.

Statistics	ALL	GROUP 1	GROUP 2
NREC	**0.124*** **	**0.086*** **	**0.117** **
REC	**0.039** **	**0.062** **	**0.022**
K	0.195***	0.183***	0.223***
L	0.190	0.281	-0.103
const	2.796	1.731	0.383
CD Statistic	1.92*	-1.95*	-1.50
CIPS test for residuals CADF(0)	-12.655***	-9.085***	-11.342***

Notes: The Pesaran (2006) CCE MG is estimated using the Stata "xtdcce2" command. ***, ** indicate statistical significance at 1, and 5 per cent level of significance, respectively.

Table 2. Panel Cointegration Estimation Results. (Cointegration regression).

The results reported in Table 3 indicate the lack of causality between renewable and non-renewable energy consumption and economic growth in all European countries. Thus, they confirm the neutrality hypothesis, which states that renewable and non-renewable energy consumption and economic growth are independent and do not affect each other in the short-run. Similar results for renewable energy consumption and economic growth are obtained by Menegaki (2011).

However, the results reveal bidirectional causality between the economic growth and renewable energy consumption in the countries belonging to the first group (GROUP 1), which means that their renewable energy consumption and economic growth are mutually dependent. Additionally, we demonstrate that development of the renewable energy sector, i.e. employment growth and the increase in turnover and investment in this sector, have an impact on economic growth. Thus, the results confirm the neutrality hypothesis between non-renewable energy consumption and economic growth.

The results obtained in the countries from the second group (GROUP 2) confirm the feedback hypothesis between non-renewable energy consumption and economic growth only at the 10 per cent level of significance. It means that economies in these countries can be called 'non-renewable energy dependent', and that non-renewable energy consumption can play an important role in their economic growth and vice versa. However, we find no causality between renewable energy consumption and economic growth, which confirms the neutrality hypothesis.

Dependent variable	Source of causation (independent variables)			
	Short term - F-statistics			Long term - t-stat
	ΔGDP	**ΔNREC**	**ΔREC**	ECT_{t-1}
ALL				
ΔGDP	-	1.67	1.75	-4.98***
ΔNREC	0.90	-	0.53	-4.86***
ΔREC	1.58	0.83	-	-3.33***
GROUP 1				
ΔGDP	-	0.67	**6.11****	-3.54***
ΔNREC	1.58	-	2.54	-5.59***
ΔREC	**7.25*****	0.51	-	-4.44***
GROUP 2				
ΔGDP	-	2.75	0.34	-1.04
ΔGDP[a]	-	**3.63***	0.02	-
ΔNREC	**3.12***	-	2.92*	-2.66***
ΔREC	1.24	0.26	-	-3.42***

Notes: Regression with 1 lag estimated by the panel DCCE (Chudik and Pesaran, 2015) using the Stata “xtdcce2” command (see: Ditzen, 2016). ***, **, * indicate statistical significance at 1, 5 and 10 per cent level, respectively. a As the coefficient of ECT in ΔGDP equation is insignificant, as robustness check we present the result of Eq. 3a when error-correction term ECT is excluded.

Table 3. Panel Granger causality test results.

Conclusions

The objective of the study is to assess the impact of the level of development of the renewable energy sector in the EU on the relationship between renewable and non-renewable energy consumption and economic growth. Our findings can be summarised in four points.

First, we find that Austria, Bulgaria, Denmark, Germany, Estonia, Finland, Latvia, Romania, and Sweden are the countries with the relatively well-developed renewable energy sectors. In comparison to other countries, they have both larger turnover and larger investment stock in the renewable energy sector.

Second, the findings on long-term output elasticities suggest that, along with traditional inputs, such as capital and labour force, both renewable and non-renewable sources play a significant role in the process of economic development in countries with a relatively high level of development of the renewable energy sector. In countries with a relatively low level of development of the renewable energy sector only non-renewable sources play a significant role in their economic development.

Third, the feedback hypothesis is confirmed for the group of countries with a relatively high level of development of the renewable energy sector, which means that renewable energy consumption and economic growth are mutually dependent in countries belonging to this group.

Forth, in the second group of countries, which are characterized by a low level of development of the renewable energy sector in economy, the relationship between renewable energy consumption and economic growth is not confirmed. The feedback hypothesis for non-renewable energy consumption and economic growth is confirmed only for this group. It means that economies in these countries can be called ‘non-renewable energy dependent’, and exclusively non-renewable energy consumption plays an important role in their economic growth.

Acknowledgments

This paper was supported by funds from the National Science Centre (NCN), Poland through grant No 2015/17/B/HS4/02685.

References

Apergis, N., Payne, J. E., 2012. Renewable and non-renewable energy consumption-growth nexus: Evidence from a panel error correction model. *Energy Economics*, *34*(3), 733-738.

Bhattacharya, M., Paramati, S. R., Ozturk, I., Bhattacharya, S., 2016. The effect of renewable energy consumption on economic growth: Evidence from top 38 countries. *Applied Energy*, *162*, 733-741.

Chudik, A., Pesaran, M. H., 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, *188*(2), 393-420.

Ditzen, J., 2016. xtdcce: Estimating Dynamic Common Correlated Effects in Stata. SEEC Discussion Paper Series No 8, July 2016 Available at http: //seec.hw.ac.uk

Menegaki AN., 2011. Growth and renewable energy in Europe: a random effect model with evidence for neutrality hypothesis. *Energy Econ 33,* 257–63.

Ozturk, I., 2010. A literature survey on energy–growth nexus. *Energy policy*, *38*(1), 340-349.

Pesaran, M.H., 2004. General diagnostic tests for cross section dependence in panels. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=572504, access: 3.02.2017.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74, 967–1012.

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22 (2), 265–312.

Smyth, R., Narayan, P. K., 2015. Applied econometrics and implications for energy economics research. *Energy Economics*, *50*, 351-358.

Śmiech, S., Papież, M., 2014. Energy consumption and economic growth in the light of meeting the targets of energy policy in the EU: The bootstrap panel Granger causality approach. *Energy Policy*, *71*, 118-129.

Tiba, S., Omri, A., 2016. Literature survey on the relationships between energy, environment and economic growth. *Renewable and Sustainable Energy Reviews*. 69, 1129-1146.

Westerlund, J., 2007. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics*, *69*(6), 709–748.

World Bank, 2016. World Development Indicators. Accessed at: http://www.worldbank.org/data/onlinedatabases/onlinedatabases.html.

Prediction of company bankruptcy in the context of changes in the economic situation

Barbara Pawełek[1]

Abstract

The papers on the prediction of company bankruptcy emphasise that there is no uniformity of a research set, including time uniformity of financial data. The economic conditions of business activities change over time. Therefore, the researchers have made attempts to study the impact of changes in the economic environment on the results of the prediction of company bankruptcy. The studies assume that the level of bankruptcy threat depends not only on the assessment of the financial condition of a given company, but it may also change depending on the economic situation in a given country or industry. The purpose of the work is to compare the predictive accuracy of selected bankruptcy prediction models built on the basis of financial data derived from various years, with the predictive accuracy of a corresponding aggregate model consisting of models built on the basis of data referring to individual years. The analysis uses the financial data of companies operating in the industrial processing sector in Poland in the years 2005-2008. The research covered selected data classification methods such as: a classification tree, *k*-nearest neighbours algorithm, support vector machine, a neural network, random forests, bagging, boosting, naive Bayes, logistic regression and linear discriminant analysis. The predictive accuracy of the constructed models was assessed on the basis of a test set. The following measures were used: sensitivity and specificity. The calculations were performed in the R programme.

Keywords: *company bankruptcy, economic situation, prediction, predictive accuracy*

JEL Classification: C380, C520, G330

1 Introduction

Corporate bankruptcy is a phenomenon characteristic for the market economy. Due to the social and economic impact of company bankruptcy, effective prediction of company bankruptcy risk is of constant interest for politicians, business practitioners and researchers. The issues related to corporate bankruptcy, including the possibility of predicting bankruptcy risk, have been widely discussed in the economic literature. Works on the methodology for forecasting company bankruptcy are being continued. The works comprise both developing methods applied traditionally in this research area (e.g. Hauser and Booth, 2011) and proposing the application of new methods, including Data Mining (e.g. Pawełek and Grochowina, 2016; Virág and Nyitrai, 2014).

[1] Corresponding author: Cracow University of Economics, Department of Statistics, 27 Rakowicka Street, 31-510 Kraków, Poland, e-mail: barbara.pawelek@uek.krakow.pl.

In papers on corporate bankruptcy prediction special attention is paid to, among other things, the problem of the homogeneity of the research set, including time-related homogeneity of financial data (e.g. Pawełek and Pociecha, 2012). Predictions of corporate bankruptcy risk are made using methods based primarily on data derived from financial statements of enterprises. If it is not possible to compile a large enough set of data for a single year then the database is built using financial data from several years. Economic conditions in which companies operate change over time. Therefore, attempts are made to study the impact of changes in the economic environment on the outcomes of corporate bankruptcy predictions, and the paper is part of this research trend. In such research it is assumed that the degree of corporate bankruptcy risk depends not only on the assessment of the financial standing of a given business but may also change depending on the economic situation in a given country or sector of economy. The literature on the subject offers proposals that take into account changes of the economic situation in bankruptcy prediction models. One proposal involves the addition of dummy variables that identify the analysed years to the set of the model explanatory variables (e.g. Pawełek et al., 2016). Another proposal points to the relevance of the incorporation of selected macroeconomic ratios in the set of explanatory variables of the model used to forecast bankruptcy (e.g. De Leonardis and Rocci, 2014).

The presented empirical research assumes that the economic situation in a given country influences the financial situation of companies. On the other hand, the evaluation of the financial situation of companies is based on the values of financial ratios. Changes of financial ratios are caused partially by changes in the economic environment of companies, including changes in the economic situation. Thus, it has been assumed that the information concerning the year of the financial statements is the link between the financial ratios and the economic situation. Partial models were built on the basis of financial data for individual years and then combined to form an aggregate model. This was intended to take account of the changes in the economic situation in the process of predicting company bankruptcy. Then, the predictive accuracy of the aggregate model was compared with the predictive accuracy of the model developed on the basis of financial data from various years. It aimed at checking whether the time-related heterogeneity of financial data, being the basis for the development of contemplated corporate bankruptcy prediction models, affects their predictive accuracy.

The aim of the paper is to present the results of comparing the predictive accuracy of selected company bankruptcy prediction models based on financial data from various years with the predictive accuracy of the corresponding aggregate model composed of models built on the data concerning individual years. The added value of the paper lies in the presentation and comparison of outcomes obtained using selected data classification methods to predict bankruptcy in the case of companies operating in the industrial processing sector in Poland on the basis of the financial data from 2005-2008[2], where such data were analysed in aggregate and for particular years. In addition, it describes the results of the verification of the hypothesis stating that values of a selected predictive accuracy measure for a given method obtained as a result of the application of two research approaches to the problem of time-related heterogeneity of financial data are derived from populations with the same averages. The calculations were made in R software using mainly ‘rminer’ (Cortez, 2015), ‘stats’ (R Core Team, 2015) and ‘HH’ packages (Heiberger, 2016).

2 Data and research procedure

The empirical research was carried out on the basis of the set of 5,920 companies operating in the industrial processing sector in Poland. Among the analysed objects there were 123 companies which the court had declared bankrupt in 2007-2010 (bankrupt companies). The financial data were related to the years 2005-2008 i.e. they were derived from financial statements published two years prior to the declaration of bankruptcy. There were 5,797 ‘healthy’ enterprises in the database i.e. companies continuing business in 2007-2010 (non-bankrupt companies). When selecting ‘healthy’ companies for the data set, the authors were guided, among other things, by their main area of business and the availability of financial data for the years 2005-2008. The data were downloaded from the Emerging Markets Information Service (https://www.emis.com/pl). Company bankruptcy risk was predicted two years in advance.

In view of the aim of the paper, the input data set was analysed in aggregate and for particular years. Four sets were separated which contained financial data from particular years in the period 2005-2008. The data from 2005 were related to 1,203 companies including 16 bankrupt companies; from 2006, to 1,368 companies including 17 bankrupt companies; from 2007, to 1,663 companies including 63 bankrupt companies; and from 2008, to 1,686 companies including 27 bankrupt companies.

On the basis of each of the four separated data sets concerning particular years, balanced research sets were created[3]. The sets comprised, respectively, 32 companies for 2005 (S_{05} set),

[2] The period between 2005 and 2008 includes the beginning of the worldwide financial crisis (e.g. Pawełek at al., 2016).

[3] The research presented in this paper is pilot research; similar deliberations will be carried out for non-balanced sets.

34 companies for 2006 (S_{06} set), 126 companies for 2007 (S_{07} set) and 54 companies for 2008 (S_{08} set). The selection of 'healthy' companies for the research sets was made at random[4]. Each of the four balanced research sets was divided at random (30 times each) into the training part (⅔ of all companies) and the test part (⅓ of all companies)[5]. The fifth balanced research set was obtained as a result of combining the four balanced sets concerning particular years from the period 2005-2008. This set included 246 companies and contained financial data from the years 2005-2008 ($S_{05\text{-}08}$ set). The training and test parts for this set were created by combining the training and test parts of the four sets related to particular years. The procedure used to compile the research sets and to divide them into training and test parts aimed at ensuring the comparability of the outcomes obtained on the basis of the $S_{05\text{-}08}$ set and after aggregating the outcomes obtained on the basis of the S_{05}, S_{06}, S_{07} and S_{08} sets.

The research entailed a dummy variable which equalled '1' in the case of companies that had been declared bankrupt in 2007-2010 and '0' for 'healthy' companies. In addition, the research involved an analysis of 32 financial ratios divided into groups: liquidity ratios (4 variables), liability ratios (10 variables), profitability ratios (7 variables) and operating efficiency ratios (11 variables).

The research comprised ten data classification methods i.e. a classification tree (rpart), *k*-nearest neighbours algorithm (knn), support vector machine (svm), a neural network (mlp), random forests (randomForest), bagging, boosting, naive Bayes (naiveBayes), logistic regression (lr) and linear discriminant analysis (lda)[6]. The methods chosen for the analysis include both traditional methods of bankruptcy prediction (e.g. a classification tree, logistic regression) and methods that are becoming increasingly popular in corporate bankruptcy risk predictions (e.g. support vector machine, random forests). Calculations were made using selected functions of 'rminer' package from R environment. In functions performing the contemplated methods, the default settings of this package were left unchanged[7].

[4] In further research, random selection of 'healthy' companies for the research sets will be repeated several times in order to assess the generality of conclusions formulated on the basis of the findings of the pilot research.

[5] The proportions for the division of the set into the training and test parts and for the number of division repetitions were assumed arbitrarily.

[6] Abbreviated names of the methods that were used in the tables are given in parentheses.

[7] Such a solution, due to the aim of the research, was considered to be justified. At the adopted level of the generality of deliberations there is no need to introduce changes in function parameters. In further analyses being the continuation of the research and dedicated to a particular data classification method, it is worth considering changes of the values of some parameters of functions to values dedicated to the analysed phenomenon of corporate bankruptcy.

Based on the outcomes obtained for each of the contemplated methods, the companies were classified into two groups: companies at risk of bankruptcy in two years (‘B’ group) and companies not at risk of bankruptcy in two years (‘NB’ group).

The evaluation of the predictive accuracy of the analysed methods was based on the values of the following classification accuracy measures for companies from the test set:

- sensitivity (the percentage of bankrupt companies classified in the ‘B’ group);
- specificity (the percentage of ‘healthy’ companies classified in the ‘NB’ group).

The hypothesis stating that values of a selected measure of the classification accuracy of a given method, calculated on the basis of the test sample created from $S_{05\text{-}08}$ set or generated after aggregating the outcomes obtained on the basis of test samples from S_{05}, S_{06}, S_{07} and S_{08} sets originating from populations with the same averages, was verified using univariate ANOVA. In order to verify whether the assumptions regarding the normality of the distribution of a variable in groups and the equality of variances for a variable in groups were observed, the Shapiro-Wilk test (Shapiro and Wilk, 1965) and the Brown-Forsythe test (Brown and Forsythe, 1974) were applied. In the case of non-compliance with such assumptions, the Kruskal-Wallis test (Kruskal and Wallis, 1952) was used.

3 Results of the empirical research

As a result of the calculations, 30 values of sensitivity and specificity measures were separately obtained for each of the ten contemplated methods. These values indicate the classification accuracy of objects from the test samples for a given method. The mean values of sensitivity and specificity measures will become the basis for the evaluation of the predictive accuracy of contemplated methods applied on $S_{05\text{-}08}$ set or S_{05}, $S_{06,}$ S_{07} and S_{08} sets.

At the first stage of the research the authors checked whether the values of the selected measure of the classification accuracy of a given method, calculated on the basis of the test sample created from $S_{05\text{-}08}$ set or generated after aggregating the outcomes obtained on the basis of test samples from S_{05} S_{06}, S_{07} and S_{08} sets, originate from populations with the same averages. The rejection of the null hypothesis (at the adopted significance level) stating the equality of averages will support the conclusions regarding the advantage (in terms of the adopted criterion) of one of the two approaches used in company bankruptcy predictions in the case of using financial data from several years.

As such, the authors checked whether the assumptions of the parametric univariate ANOVA concerning the normality of the distribution of a variable in groups and the equality of variances for a variable in groups were met. The Shapiro-Wilk test was applied in the case of each of the ten contemplated methods and each of the two analysed classification accuracy measures for objects from the test sample. Table 1 presents *p*-value in the Shapiro-Wilk test for the group of bankrupt companies (the sensitivity measure) and Table 2 – for the group of non-bankrupt companies (the specificity measure). At the significance level of 0.10, the obtained outcomes indicate the failure to meet the assumption about the normality of the distribution of both measures only in a few cases. For the sensitivity measure the null hypothesis (α = 0.10) is rejected for the classification tree and the naive Bayes method in the case of based on the $S_{05\text{-}08}$ set. For the specificity measure this is the case for the *k*-nearest neighbours algorithm and the naive Bayes method in the case of based on the S_{05} S_{06}, S_{07} and S_{08} sets.

Method	**Shapiro-Wilk test**		**Brown-Forsythe test**	**ANOVA**	**Kruskal-Wallis test**
	$S_{05\text{-}08}$	S_{05}–S_{08}			
rpart	0.0124	0.3322	0.2986	0.3581	0.2531
knn	0.2718	0.2734	0.5172	0.3609	0.3191
svm	0.7772	0.5448	0.4893	0.0132	0.0189
mlp	0.3657	0.7462	0.1544	0.1833	0.2150
randomForest	0.3439	0.6207	0.6501	0.7757	0.7945
bagging	0.4851	0.2156	0.8516	0.0576	0.0549
boosting	0.5835	0.5603	0.7305	0.7668	0.8001
naiveBayes	0.0000	0.1124	0.0216	0.0000	0.0000
lr	0.4112	0.6241	0.4728	0.0020	0.0015
lda	0.1102	0.4836	0.8773	0.0129	0.0063

Table 1. *p*-value in the Shapiro-Wilk test, the Brown-Forsythe test, ANOVA and the Kruskal-Wallis test for the sensitivity measure.

The fulfilment of the assumption about the equality of variances for a variable in the groups was verified using the Brown-Forsythe test. Table 1 presents *p*-value in the Brown-Forsythe test for the sensitivity measure and Table 2 – for the specificity measure. Test results (α = 0.10) indicate the non-satisfaction of the assumption about the equality of variances for measures only for the naive Bayes method.

Due to the fact that in the majority of analysed cases the assumptions regarding the normality of the distribution of a variable in groups and the equality of variances for a variable in groups were met, the parametric univariate ANOVA was used. The outcomes for the sensitivity measure are shown in Table 1 and for the specificity measure – in Table 2. The

Kruskal-Wallis test was also applied to be able, in cases when at least one of the assumptions of parametric ANOVA is not fulfilled, to comment (at the adopted significance level) on the equality of the averages (Tables 1 and 2).

Method	**Shapiro-Wilk test**		**Brown-Forsythe test**	**ANOVA**	**Kruskal-Wallis test**
	$S_{05\text{-}08}$	S_{05}–S_{08}			
rpart	0.1344	0.4679	0.7015	0.8501	0.8121
knn	0.8908	0.0007	0.2969	0.3754	0.6329
svm	0.2537	0.1528	0.2931	0.0034	0.0083
mlp	0.4768	0.2663	0.6945	0.9382	0.9941
randomForest	0.9415	0.2943	0.3226	0.6525	0.6232
bagging	0.3515	0.2226	0.4718	0.5709	0.5313
boosting	0.1102	0.4101	0.3335	0.4760	0.7321
naiveBayes	0.1136	0.0853	0.0017	0.0000	0.0000
lr	0.3329	0.4121	0.2993	0.0640	0.0468
lda	0.4341	0.7826	0.1157	0.0000	0.0000

Table 2. *p*-value in the Shapiro-Wilk test, the Brown-Forsythe test, ANOVA and the Kruskal-Wallis test for the specificity measure.

On the basis of the outcomes of the tests ($\alpha = 0.10$), only in a few cases did the null hypothesis have to be rejected in favour of the alternative hypothesis stating that values of a given measure of the classification accuracy, calculated on the basis of the test sample created from $S_{05\text{-}08}$ set or generated after aggregating the outcomes obtained on the basis of test samples from S_{05}, S_{06}, S_{07} and S_{08} sets, originate from populations with different averages. Such was the case for the sensitivity measure in the support vector machine, bagging, the naive Bayes method, the logistic regression and the linear discriminant analysis. It means that only for these five methods is it possible to identify a research approach supporting a statistically significant improvement of the classification accuracy of bankrupt companies from the test set. For the specificity measure there were four cases found in which, at the significance level of 0.10, the null hypothesis had to be rejected. This was the case for the support vector machine, the naive Bayes method, the logistic regression and the linear discriminant analysis. Hence, in the case of these four methods, one can point to the relevance of one of the two analysed research approaches for the improvement of the classification accuracy of non-bankrupt companies from the test set.

At the second stage of the research, values of basic descriptive measures for the sensitivity and specificity measures on the test set were calculated for each of the ten contemplated data classification methods. Table 3 below presents only the values of the arithmetic mean and standard deviation.

Method	Sensitivity measure				Specificity measure			
	Mean		Stand. deviation		Mean		Stand. deviation	
	$S_{05\text{-}08}$	S_{05}–S_{08}	$S_{05\text{-}08}$	S_{05}–S_{08}	$S_{05\text{-}08}$	S_{05}–S_{08}	$S_{05\text{-}08}$	S_{05}–S_{08}
rpart	0.698	0.671	0.119	0.105	0.663	0.668	0.104	0.095
knn	0.656	0.674	0.081	0.069	0.688	0.671	0.077	0.071
svm	0.695**	0.635	0.085	0.097	0.684***	0.618	0.087	0.080
mlp	0.642	0.670	0.088	0.070	0.641	0.639	0.086	0.075
randomForest	0.711	0.705	0.079	0.075	0.708	0.716	0.075	0.063
bagging	0.742*	0.704	0.075	0.078	0.715	0.725	0.064	0.068
boosting	0.701	0.695	0.075	0.073	0.676	0.691	0.086	0.071
naiveBayes	0.262	0.479***	0.098	0.134**	0.934***	0.765	0.039	0.108***
lr	0.673***	0.607	0.085	0.074	0.664*	0.628	0.069	0.081
lda	0.658**	0.617	0.066	0.056	0.717***	0.597	0.065	0.086

In ANOVA: * – p-value < 0.10, ** – p-value < 0.05, *** – p-value < 0.01.

In the Brown-Forsythe test: * – p-value < 0.10, ** – p-value < 0.05, *** – p-value < 0.01.

Table 3. Mean and standard deviation values for the sensitivity and specificity measures.

On the basis of the values contained in Table 3, one can conclude that only in the case of three out of ten contemplated methods did the inclusion of the S_{05}, S_{06}, S_{07} and S_{08} sets in the analysis result in an increase in the arithmetic mean of the sensitivity measure against the mean value derived on the basis of the $S_{05\text{-}08}$ set. These were the following methods: the k-nearest neighbours algorithm, the neural network and the naive Bayes method. However, the outcomes of the tests applied in the first phase of the research make it possible to recognise the advantage of the approach based on the S_{05}, S_{06}, S_{07} and S_{08} sets as statistically significant (α = 0.10) only in the case of the naive Bayes method. On the other hand, the superiority of the approach based on the $S_{05\text{-}08}$ set can be recognised (α = 0.10) in the case of the support vector machine, bagging, the logistic regression and the linear discriminant analysis. The remaining differences between the arithmetic mean values should be approached with caution as the outcomes of the tests (α = 0.10) did not provide grounds for rejecting the null hypothesis stating the equality of averages in populations.

In the case of the specificity measure, basing the analysis on the $S_{05\text{-}08}$ set for the six contemplated methods resulted in a higher arithmetic mean than in the case of basing it on the S_{05}, S_{06}, S_{07} and S_{08} sets. However, the test outcomes make it possible to recognise as statistically significant (α = 0.10) the advantage of the approach based on the $S_{05\text{-}08}$ set only in the case of four methods, namely the support vector machine, the naive Bayes method, the logistic regression and the linear discriminant analysis.

Conclusions

Based on the results of the research, one can conclude that the predictive accuracy of certain methods of company bankruptcy prediction may an increase as a result of applying the research approach in which the fact that the data are from various years is ignored, as compared with the approach taking into account the time-related heterogeneity of financial data. Such a situation occurred in the case of the support vector machine, the logistic regression and the linear discriminant analysis – both in the group of bankrupt companies and in the group of non-bankrupt companies. On the other hand, for the naive Bayes method the aforementioned regularity was observed only in the group of non-bankrupt companies. The advantage of the approach based on the data concerning individual years over the solution based on data originating from various years was revealed in the case of the naive Bayes method, but only in the group of bankrupt companies. However, for half of the analysed corporate bankruptcy prediction methods, at the significance level of 0.10, the advantage (in terms of the adopted criterion) of one of the two analysed approaches to the problem of the time-related heterogeneity of financial data has not been proven. Thus, research on the impact of changes in the economic situation on company bankruptcy prediction results should be continued.

In further research the authors intend primarily to repeat the analysis for unbalanced sets, apply *v*-fold cross-validation method and expand the set of measures of the predictive accuracy of the analysed methods.

Acknowledgements

Publication was financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Brown, M. B., & Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, *69(346)*, 364–367.

Cortez, P. (2015). *rminer: Data Mining Classification and Regression Methods*. R package version 1.4.1. URL: http://CRAN.R-project.org/package=rminer.

De Leonardis, D., & Rocci, R. (2014). Default Risk Analysis via a Discrete-time Cure Rate Model. *Applied Stochastic Models in Business and Industry*, *30(5)*, 529–543.

Hauser, R. P., & Booth, D. (2011). Predicting Bankruptcy with Robust Logistic Regression. *Journal of Data Science*, *9(4)*, 565–584.

Heiberger, R. M. (2016). *HH: Statistical Analysis and Data Display: Heiberger and Holland*. R package version 3.1-31. URL: http://CRAN.R-project.org/package=HH.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47(260)*, 583–621.

Pawełek, B., & Grochowina, D. (2016). The random subspace method in the prediction of the risk of bankruptcy of companies in Poland. In: Malina, A., Węgrzyn, R. (eds.), *Knowledge – Economy – Society. Challenges and Development of Modern Finance and Information Technology in Changing Market Conditions*, Foundation of the Cracow University of Economics, Kraków, 25–34.

Pawełek, B., & Pociecha, J. (2012). General SEM Model in Researching Corporate Bankruptcy and Business Cycles. In: Pociecha, J., Decker, R. (eds.), *Data Analysis Methods and its Applications*, C.H. Beck, Warsaw, 215–231.

Pawełek, B., Pociecha, J., & Baryła, M. (2016). *Dynamic Aspects of Bankruptcy Prediction Logit Model for Manufacturing Firms in Poland*. In: Wilhelm, A. F. X., Kestler, H. A. (eds.), *Analysis of Large and Complex Data*, Studies in Classification, Data Analysis, and Knowledge Organization, Switzerland: Springer, 369–382.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52(3/4)*, 591–611.

Virág, M., & Nyitrai, T. (2014). The application of ensemble methods in forecasting bankruptcy. *Financial and Economic Review*, *13(4)*, 178–193.

Forecasting the bankruptcy of companies: the study on the usefulness of the random subspaces and random forests methods

Barbara Pawełek[1], Dorota Grochowina[2]

Abstract

The subject matter of the paper falls into the mainstream of research on an important economic issue, which is the bankruptcy of companies. The issues related to the bankruptcy of companies have been widely discussed in the economic literature. Works are being continued on the methodology of forecasting the bankruptcy of companies, including the use of Data Mining methods for this purpose. The purpose of the paper is to present the results of empirical studies on the usefulness of the random subspaces and random forests methods for forecasting the bankruptcy of companies in Poland. Both of the aforementioned methods belong to the multi-model approach to data analysis. The random subspaces method relies on the random selection of variables, while the random forests methods uses both the random selection of variables, as well as the random selection of observations. The analysis was performed for balanced sets of companies. The usefulness of the random subspaces and random forests methods for forecasting the bankruptcy of companies was assessed based on the values of the classification accuracy measures for companies from the test part of the input set of objects. The basis for building the set of objects was a set of companies active in the industrial processing sector in Poland in the years 2013-2014. The financial data were taken from the Emerging Markets Information Service website. The computations were done in the R program.

Keywords: *bankruptcy, forecasting, random subspaces method, random forests method, classification accuracy*
JEL Classification: C380, C520, G330

1 Introduction

Establishing new enterprises and closure of the business by some of the existing companies are typical phenomenons for the market economy. Among enterprises, both finishing their activity in the market and continuing their business, entities can highlighted that are subject to court's declaration of bankruptcy by liquidation or bankruptcy with a possibility of concluding an agreement (restructuring proceedings). These companies are colloquially known as the bankrupts. The phenomenon of bankruptcy of enterprises is an important economic issue. Practitioners and researchers are interested in forecasting the bankruptcy risk of companies due to social and economical issues that emerge as a result of such a phenomenon. Issues related to forecasting the bankruptcy of enterprises have been widely

[1] Corresponding author: Cracow University of Economics, Department of Statistics, 27 Rakowicka Street, 31-510 Kraków, Poland, e-mail: barbara.pawelek@uek.krakow.pl.
[2] Warsaw School of Economics, Collegium of Economic Analysis, 6/7 Madalińskiego Street, 02-513 Warszawa, Poland, e-mail: Dorota.Grochowina@doktorant.sgh.waw.pl.

discussed in the economic literature (e.g. Pociecha et al., 2014). Works on the methodology of forecasting the bankruptcy of companies are being continued, including the use of the Data Mining methods (e.g. Min, 2016a; Min, 2016b; Panov and Džeroski, 2007; Pawełek and Grochowina, 2016; Virág and Nyitrai, 2014).

The aim of this paper is to present the results of empirical studies on the usefulness of the random subspaces and random forests methods for forecasting the bankruptcy of enterprises in Poland. The added value of the paper consists in the presentation of results obtained using the selected Data Mining methods for forecasting the bankruptcy of companies operating in the industrial processing sector in Poland, along with the verification of the hypothesis that the values of the classification accuracy measures for the particular considered methods, calculated on the basis of the testing part, come from the populations that have the same average values. The computations were done in the R program, first of all using the packages 'randomForest' (Liaw and Wiener, 2002), 'rpart' (Therneau et al., 2015), 'PMCMR' (Pohlert, 2014).

2 Data and the research procedure

The analysis used a set containing 7223 companies operating in the industrial processing sector in Poland. Among the considered objects were 42 companies that were subject to court's declaration of bankruptcy in 2014 or 2015 (bankrupts, B). The financial data related to the years 2013 and 2014, i.e. they were obtained from the financial statements which were published a year before the bankruptcy. The database also contained 7181 "healthy" companies, i.e. those continuing business activity in 2014 and 2015 (non-bankrupts, NB). The selection of "healthy" companies for the database was based, inter alia, on their core business activity and the availability of financial data for 2013 and 2014. The data were taken from the Emerging Markets Information Service website (https://www.emis.com/pl). The companies' bankruptcy risk was forecast with one year in advance.

The input dataset was used to generate balanced research sets that contained 84 companies (42 B – ½ of total and 42 NB – ½ of total)[3]. The "healthy" companies were included in the research sets through the random selection, repeated 30 times. Each of the thirty research sets was randomly divided (100 times) into a training part (⅔ of total) and a testing part (⅓ of total)[4]. In the analysis, the dependent variable had the zero-one characteristic and received

[3] The considerations presented in this paper are a pilot project. Similar considerations will be made for the unbalanced sets of objects.
[4] The arbitrary proportions and the number of repetitions were adopted for data set division into the training part and testing part.

category "1" in the case of companies that declared bankruptcy in 2014 or 2015, and category "0" for "healthy" companies. 16 financial ratios were used as the independent variables, divided into the following groups: liquidity ratios (3 variables), liability ratios (4 variables), profitability ratios (5 variables), productivity ratios (4 variables).

The methods of random subspaces (Ho, 1998) and random forests in the Forest-RI version (Breiman, 2001), based on the classification tree CART (Breiman et al., 1984)[5] were used in the analysis. Both of the aforementioned methods belong to the ensemble approach to data analysis. The ensemble approach consists in aggregating M base models D_1, ..., D_M into one aggregated model D^*, in order to improve the forecasting accuracy (Gatnar, 2008). The random subspaces method relies on the random selection of variables that are used to the base models, while the random forests method uses both the random selection of the variables and the random selection of observations. The analysis was carried out for one hundred base models (M = 100). The classification of companies into two groups: entities without bankruptcy risk in a year (group "NB") and entities with bankruptcy risk in a year (group "B")[6] was carried out based on each of one hundred base models created for each of three thousand training sets. The aggregation of one hundred base models, i.e. the combination of the forecasting results based on the base models, was performed according to the majority voting method, and resulted in 3000 aggregated models (Gatnar, 2008).

The evaluation of the usefulness of the random subspaces and random forests methods for forecasting the bankruptcy of companies in Poland was performed on the basis of the values of the following classification accuracy measures calculated for companies included in the testing set:

- total error (percentage of companies incorrectly classified into "B" or "NB" groups);
- error type I (percentage of the bankrupts classified into "NB" group);
- error type II (percentage of "healthy" companies classified into "B" group).

The classification accuracy of the random subspaces and random forests methods was compared with the classification accuracy of the single classification tree CART, which is one of the popular methods for forecasting the bankruptcy. 3000 single classification trees were created, one for each of the analysed training sets. The input set of the independent variables included all the considered financial indicators.

[5] Default settings of the given package were left during calculations using `rpart` and `randomForest` functions. Changing some parameters may help to achieve better results than those presented in this paper.
[6] 30 research sets divided 100 times into the training part and the testing part.

Using univariate ANOVA, it was planned to verify the hypothesis that the values of the classification accuracy measures for the particular methods, calculated on the basis of the testing sets, came from populations that had the same average values. The Shapiro-Wilk (Shapiro and Wilk, 1965), Levene and Brown-Forsythe (Brown and Forsythe, 1974) tests were used in order to check if the ANOVA assumptions regarding the normality of the variable distribution in the sets and the equality of the variable variance in the sets, were met. Due the fact that the ANOVA assumptions had not been met, it was decided to use a non-parametrical variance analysis. The Kruskal-Wallis test (Kruskal and Wallis, 1952), followed by the post-hoc Conover test (Conover and Iman, 1981) and post-hoc Dunn test (Dunn, 1964), were used, whereas, the Bonferroni multiple testing correction was considered in the post-hoc tests.

A common feature of the random subspace and random forests methods is the reduction of the variables space dimension. The calculations were performed for all possible variants of the reduction of the set of variables, i.e. the number of variables ranging from 1 to 15. Researchers involved in the ensemble approach to the data analysis often point to the usefulness of a formula derived from the information theory for the determination of the dimension of the reduced space (e.g. Amit and Geman, 1997; Breiman, 2001). According to this approach, the number of variables should be equal to the integer number from the interval $[\log_2 K, \log_2 K+1)$, where K is the number of input variables (Cover and Thomas, 2006). In the research carried out by the authors, $K=16$, therefore the results for four variables are presented.

3 Results of empirical research

In the first phase of the research, on the basis of each of the thirty research sets and each of one hundred training sets, the bankruptcy of the companies was projected with one year in advance, using the single classification tree, random subspace and random forests methods. Afterwards, the classification errors were calculated (total, type I and type II) for the companies included in the respective testing sets. It resulted in 27 000 values of errors (30 research sets $\times$ 100 divisions into training and testing parts $\times$ 3 Data Mining methods $\times$ 3 error types). The obtained results provided the basis for analysis performed in the subsequent phases.

The second phase of the analysis consisted in checking, whether the values of the classification accuracy measures for the considered methods, calculated on the basis of the test sample, came from the populations that had the same average values. For this purpose, it was checked whether the ANOVA assumptions regarding the normality of the variable distribution in the sets and the equality of the variable variance in the sets, had been met.

The Shapiro-Wilk test was used for each of the thirty research sets and each of the three considered methods, based on the error values (respectively: total, type I and type II), which occurred during the classification of objects belonging to the one hundred test samples analysed. Table 1 (columns 2-4) contains numbers indicating how many times *p*-value in the Shapiro-Wilk test was lower than 0.05 for 30 research sets. The results show that the assumption of the normality of distribution of the total, type I and type II errors, which define the classification effectiveness of the considered methods (as measured based on the test sample) in most of the analysed cases at a significance level of 0.05, has not been met.

Group	**Shapiro-Wilk test**			**Levene test**	**Brown-Forsythe test**	**Kruskal-Wallis test**
	Tree	**RS**	**RF**			
Total	20	25	30	11	11	28
Bankrupts	29	29	30	20	20	28
Non-bankrupts	24	30	30	26	26	27

Symbols: Tree – single classification tree, RS – random subspace method, RF – random forests method.

Table 1. The number of research sets (max = 30) for which the *p*-value was lower than 0.05 respectively in the Shapiro-Wilk (columns 2-4), Levene (column 5), Brown-Forsythe (column 6), Kruskal-Wallis (column 7) tests.

Then the assumption about the equality of the variables variances in the groups was verified. The Brown-Forsythe and Levene tests were used for this purpose for each of the thirty research sets. Table 1 contains the numbers indicating how many times *p*-value in the Levene (column 5) and Brown-Forsythe (column 6) tests was lower than 0.05 for 30 research sets. The test results at a significance level of 0.05 show that the assumption of the equality of the variance of the type I and type II errors that define the classification accuracy of the considered methods (as measured based on the test set) have not been met in most of the analysed cases. In the group including bankrupts and non-bankrupts, the test results at a significance level of 0.05, in most cases show that there are no grounds to reject the null hypothesis that assumes the equality of the variance of type I and type II errors that define the classification effectiveness of the considered methods (as measured based on the test set).

In the third phase of the analysis, as the ANOVA assumptions had not been met by the considered variables in the majority of the analysed cases, it was decided to use a non-parametrical variance analysis. The Kruskal-Wallis test was employed. Table 1 contains numbers indicating how many times *p*-value in the Kruskal-Wallis test (column 7) was lower than 0.05 for 30 research sets. In most cases, at a significance level of 0.05, the null hypothesis was rejected in favour of an alternative hypothesis that stated that for at least two out of three considered methods, the values of the classification accuracy measures, as calculated based on the test set, come from populations that had different average values.

In the fourth phase of the analysis, the post-hoc Conover and Dunn tests with the Bonferroni correction for the multiple testing were used to determine which sets of the results did not come from the populations that had the same average values. Table 2 contains figures indicating how many times *p*-value in the post-hoc Conover (columns 2-4) and post-hoc Dunn (columns 5-7) tests was lower than 0.05 for 30 research sets.

Group	**post-hoc Conover test**			**post-hoc Dunn test**		
	Tree – RS	**Tree – RF**	**RF – RS**	**Tree – RS**	**Tree – RF**	**RF – RS**
Total	26	24	21	26	24	18
Bankrupts	22	22	14	22	22	14
Non-bankrupts	22	19	15	22	18	13

Symbols: Tree – single classification tree, RS – random subspace method, RF – random forests method.

Table 2. The number of research sets (max = 30) for which the *p*-value was lower than 0.05 respectively in the post-hoc Conover (columns 2-4), post-hoc Dunn (columns 5-7) tests.

The results in table 2 show that in most cases, at a significance level of 0.05, it should be considered that the values of the classification accuracy measures, as calculated based on the test set, for the single tree and the random subspace method and for the single tree and the random forests method come from the populations having different average values. The same conclusion can be made for the results for the random subspace method and the random forests method in the group of bankrupts and non-bankrupts. In the case of separate sets of bankrupts and non-bankrupts, the results are not conclusive. At the significance level of 0.05, the number of cases in which it was necessary to reject the null hypothesis and the number of cases in which there were no grounds to reject the null hypothesis were similar.

In the last, fifth phase of the analysis it was verified, which of the considered methods was characterized by higher classification accuracy, as measured based on the test set. The arithmetic average the errors (respectively: total, type I and type II) in the classification of the objects included in the one hundred test sets analysed, was calculated for each of the thirty research sets and each of the three considered methods. Then, the methods were compared in terms of the obtained result. Table 3 (columns 2-4) contains numbers indicating how many times the difference between the average value of the errors occurred for the thirty analysed research sets.

Group	**Arithmetic average**			**Ranks average (post-hoc Conover test)**		
	Tree > RS	**Tree > RF**	**RF > RS**	**Tree > RS**	**Tree > RF**	**RF > RS**
Total	30	30	17	30 (26)	29 (24)	17 (14)
Bankrupts	29	29	18	29 (22)	28 (22)	17 (9)
Non-bankrupts	28	26	18	28 (22)	25 (19)	18 (7)

Symbols: Tree – single classification tree, RS – random subspace method, RF – random forests method.

Table 3. The number of the research sets (max = 30) for which the specific difference between the arithmetic average values occurs (columns 2-4) and the number of the sets for which the specific difference between the average rank values occurs, and additionally for which the *p*-value in the post-hoc Conover test was lower than 0.05 (values in brackets) (columns 5-7).

The analysis of the results presented in table 3 (columns 2 and 3) leads to conclude that the use of the random subspace and random forests methods in forecasting the bankruptcy of enterprises in the industrial processing sector in Poland, in majority of cases resulted in an increase of the classification accuracy of the test set objects, compared to the single classification tree. Comparison of the average error values calculated for the random subspace and random forests methods (table 3, column 4), shows the advantage of the random subspace method. However, considering fact that 30 pairs of values were compared, the results should be approached with some reservation.

The Kruskal-Wallis, post-hoc Conover and post-hoc Dunn tests are based on the ranks and their arithmetic averages. The calculating of the average value of the ranks means an operation unacceptable for the measurements on the ordinal scale and involves a strengthening of the measurement scale. However, in order to determine which of the considered methods was characterized by higher classification accuracy, as measured based on the test set, in combination with examination of significance of the differences between the pairs of the average values of the populations, it was decided to calculate the average rank values and to compare the methods in terms of obtained results. Table 3 (columns 5-7) contains numbers indicating how many times the difference between the average rank values occurred for the thirty research sets analysed. The numbers in parentheses indicate how many times the specific differences between the average rank values are accompanied by the *p*-value of the post-hoc Conover test smaller than 0.05. The discussions considered only the post-hoc Conover test due to the similarity between the results of this test and the results of the post-hoc Dunn test.

The obtained results strengthen the conclusion concerning the improvement of the classification accuracy of the objects from the test set as a result of using the random subspace and random forests methods, compared to the single classification tree. In addition, the believe about the need for exercising caution when attempting to indicate a more effective method between the random subspace and random forests methods has been confirmed.

Conclusions

On the basis of the results of the research carried out it can be found that the use of the random subspaces and random forests methods in forecasting the bankruptcy of companies helps to improve the classification accuracy of test set objects compared to the single classification tree.

The results do not provide the grounds for distinguishing any of the analysed methods (i.e. either the random subspaces method or the random forests method), as the one the application of which more strongly favours the improvement in the classification accuracy of the test set objects. This conclusion concerns in particular a situation where sets of bankrupts and non-bankrupts are considered separately. Therefore, no justification has been found for using the more complex method, which is the random forests method, compared to the random subspaces method,. This problem requires further analysis, including e.g. a change of the parameters in the `rpart` and `randomForest` functions from the default ones to those dedicated to the phenomenon under study.

The above conclusions have been formulated based on the results of analyses carried out on the basis of actual data. However, it should be borne in mind that every empirical study has specific limitations. The presented study was focused on companies active in the industrial processing sector in Poland. The financial data related to the years 2013 and 2014, and the forecast was performed one year in advance. The research set was balanced. The training part included ⅔ of the companies of the research set and the test part included the rest of the objects, i.e. ⅓ of all companies.

In further studies, the authors intend to repeat the analysis for unbalanced sets. It is also planned to include the Forest-RC type of the random forests method and the *v*-times cross validation in the analysis.

Acknowledgements

Publication was financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation, 9*, 1545–1588.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC Press, Boca Raton.

Brown, M. B., & Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association, 69(346)*, 364–367.

Conover, W. J., & Iman, R. L. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician*, *35(3)*, 124–129.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory*. Second Edition, A Wiley-Interscience publication, A John Wiley & Sons, Inc., Hoboken, New Jersey.

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6(3)*, 241–252.

Gatnar, E. (2008). *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji [The multi-model approach to discrimination and regression problems]*. PWN Scientific Publishers, Warsaw.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8)*, 832- 844.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47(260)*, 583–621.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2(3)*, 18–22.

Min, S.-H. (2016a). A genetic algorithm-based heterogeneous random subspace ensemble model for bankruptcy prediction. *International Journal of Applied Engineering Research*, *11(4)*, 2927–2931.

Min, S.-H. (2016b). Integrating instance selection and bagging ensemble using a genetic algorithm. *International Journal of Applied Engineering Research*, *11(7)*, 5060–5066.

Panov, P., & Džeroski, S. (2007). Combining Bagging and Random Subspaces to Create Better Ensembles. In: Berthold, M. R., Shawe-Taylor, J., & Lavrač, N. (eds.), *IDA 2007, LNCS 4723,* Springer-Verlag, Berlin Heidelberg, 118–129.

Pawełek, B., & Grochowina, D. (2016). The random subspace method in the prediction of the risk of bankruptcy of companies in Poland. In: Malina, A., Węgrzyn, R. (eds.), *Knowledge – Economy – Society. Challenges and Development of Modern Finance and Information Technology in Changing Market Conditions*, Foundation of the Cracow University of Economics, Cracow, 25–34.

Pociecha, J. (ed.), Pawełek, B., Baryła, M., & Augustyn, S. (2014). *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej [Statistical methods of forecasting the bankruptcy in changing economic conditions*]. Foundation of the Cracow University of Economics, Cracow.

Pohlert, T. (2014). The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). *R package*, URL: http://CRAN.R-project.org/package=PMCMR.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52(3/4),* 591–611.

Therneau, T., Atkinson, B., & Ripley, B. (2015). rpart: Recursive Partitioning and Regression Trees. *R package version 4.1-10,* URL: http://CRAN.R-project.org/package=rpart.

Virág, M., & Nyitrai, T. (2014). The application of ensemble methods in forecasting bankruptcy. *Financial and Economic Review*, *13(4)*, 178–193.

Economic development of Polish voivodeships in the years 2010-2014. Application of taxonomic measure of development with entropy weights

Michał Bernard Pietrzak [1], Adam P. Balcerzak [2]

Abstract

Implementing policy and forming socioeconomic conditions that support sustainable and equable growth of regions is currently an important objective both at European and national level. Regional development policy is supported by significant resources from European Union funds. As a result, a constant monitoring of the development process at regional level with application of quantitative methods is an important scientific and practical task. Thus, the aim of the article is to assess the level of economic development in Poland at the voivodeships level (NUTS 2). Economic development is considered here as a multiple-criteria phenomenon. In the case of multiple-criteria analysis a common dilemma is attributed to the problem of applying appropriate weights for variables used in the research. Therefore, in order to provide a rating of voivodeships a taxonomic measure of development with entropy weights was applied here. The research was conducted for the years 2010-2014. It was based on the data provided by Central Statistical Office of Poland. The results of the analysis confirm that in spite of a progress obtained by all voivodeships significant disparities between them are still present.

Keywords: *entropy weighs, multiple-criteria decision analysis, taxonomic measure of development, regional development*

JEL Classification: O18, P25, C38

1 Introduction

The main empirical objective of the article is to assess the level of economic development in Poland at the voivodeships level (NUTS 2) in the years 2010-2014. Currently it is commonly accepted that a single measure of economic development or welfare – especially the most commonly used one such as GDP per capita – provides oversimplified information on the subject. The phenomenon should be considered as a complex and multivariate problem. As a result in the research a multiple-criteria decision analysis tools are applied. For this purpose a taxonomic measure of development based on the method proposed by Hellwig will be used, where for proposing weights for given aspects entropy values are applied.

In the research the following hypothesis is given: In Poland one can see the process of improvement in the level of economic development at regional level.

[1] Corresponding author: Nicolaus Copernicus University/Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: michal.pietrzak@umk.pl.
[2] Nicolaus Copernicus University/Department of Economics, e-mail: adam.balcerzak@umk.pl.

2 Taxonomic measure of development: methodology

The problem of economic development must be considered as a multivariate phenomenon that can be characterised with many economic aspects (Pietrzak et al., 2013; Balcerzak, 2016a, 2016b; Balcerzak and Pietrzak, 2016; Jantoń-Drozdowska and Majewska, 2016; Łyszczarz, 2016; Małkowska and Głuszak, 2016; Pietrzak, 2016). In order to model and measure such phenomenon one can use taxonomic measure of development (*TMD*) that was originally proposed by Hellwig (see also: Balcerzak, 2016b; Balcerzak and Pietrzak, 2016).

In the case of multiple-criteria decision analysis problems a researcher usually faces a dilemma concerning the potential differences in the importance of given aspects (specific variables) and thus differences in their influence on the analyzed phenomenon. In order to solve that problem a set of different weights for each variable can be used, where normalized variables are multiplied by weights in the process of obtaining taxonomic measure of development (Balcerzak and Pietrzak, 2016c; Żelazny and Pietrucha, 2017). In the case of current research the objective weights based on the entropy values are used (see: Wang and Lee, 2009). In the procedure the definition of Shannon entropy prosed and developed in the context of information theory is applied.

The procedure of assessing *TMD* based on Hellwig method with application of entropy weights is given in the following steps (Balcerzak and Pietrzak, 2016):

1. A selection of a set of object O_i (i=1,2,..., m) and diagnostic variables X_j (j=1,2,.., n) for a given economic phenomenon.
2. Assessing a set of entropy weights w_j for variables X_j based on the entropy value (see: Wang, Lee, 2009). In the case of the current research the proposed procedure of assessing entropy weights with application of Shannon entropy is extended by the authors with the time dimension. In the first step the entropy value e_{jt} is assessed (equation 1).

$$e_{jt} = -\frac{1}{\ln(m)} \sum_{i=1}^{m} p_{ijt} \ln(p_{ijt}) \quad , \quad p_{ijt} = \frac{x_{ijt}}{\sum_{i=1}^{m} x_{ijt}} \tag{1}$$

In the next step, based on the obtained values, entropy weights w_j with the equation 2 are assessed, where the sum of entropy weights w_j is equal to 1.

$$w_{jt} = \frac{1 - e_{jt}}{n - \sum_{j=1}^{n} e_{jt}} \tag{2}$$

3. A normalization of diagnostic variables[3], which enables to obtain a set of normalised diagnostic variables Z_j.
4. Including entropy weights w_j for every normalised diagnostic variable Z_j with equation 3.

$$y_{ijt} = z_{ijt} w_{jt} \tag{3}$$

5. An assessment of a pattern of development x_{ojt} for every diagnostic variable Z_j with equation 4. In the case of dynamic analysis, the arithmetic average values, the values of standard deviation applied for standardisation, and the values of pattern of development are set as constant for the whole analysed period. It is a condition for obtaining a compatibility of the objects in different points in time t.

$$y_{0jt} = \max_{it} y_{ijt} \tag{4}$$

6. For every object O_i an assessment of distance to the pattern of development in a given point in time with equation 5.

$$d_{i0t} = \sqrt{\sum_{j=1}^{p} (y_{ijt} - y_{0jt})^2} \tag{5}$$

7. The value of taxonomic measure of development TMD_{it}, which describes the level of development of analyzed phenomenon for every object O_i in time t can be given with equation 6

$$TMD_{it} = 1 - \frac{d_{i0t}}{d_{0t}}, \tag{6}$$

where $d_{0t} = \bar{d}_{0t} + 2s_{dt}$, and $\bar{d}_{0t}$, s_{dt} are given with formula (7).

$$\bar{d}_{ot} = \frac{1}{n}\sum_{i=1}^{n} d_{i0t}, \quad s_{ot} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (d_{i0t} - \bar{d}_{ot})^2} \tag{7}$$

The values of TMD_{it} are on the scale of 0-1, where the high values of TMD_{it} indicate high level of development of a given phenomenon. In recent years an interesting direction of development of the concept of taxonomic measure of development relates to the problem of taking into account spatial interdependence in the design of the measure, which can be found in the works of Pietrzak (2016).

[3] In the current resarch a standarisation of variables based on the arithmetic average value and standard deviation was conducted.

3 Assessment of socio-economic development at regional level in Poland

Based on the aim of the article the values of taxonomic measure of development were assessed for the years 2010 and 2014. The set of diagnostic variables used in the research is given in table 1. The values of the variables were provided by Central Statistical Office of Poland and are available in the service: http://wskaznikizrp.stat.gov.pl/. The set of diagnostic variables can be considered as incomplete. The selection of the variables is based on the conducted literature review devoted to the determinants of growth and economic welfare (Hadaś-Dyduch, 2015; Kordalska and Olczyk, 2016; Kondratiuk-Nierodzińska, 2016; Ciburiene, 2016; Shuaibu and Oladayo, 2016; Kryk, 2016; Zemtsov et al., 2016). However, the authors are aware that the choice of variables given in table 1 can be considered as arbitrary.

Based on the procedure described in previous section the empirical research was started with an assessment of the weights, where the entropy values were used. The results are given in table 2. In the case of all the variables related to different aspects of economic development the weights values close to 0,25 were obtained. As a result it can be said that the influence of all the variables on the TMD_{it} is generally at the same level. What is more the changes in the values of weights between the years 2010 and 2014 are also quite small. Additionally, one can see that there is also a tendency to equate the values of the weights, which can indicate that all the variables tend to have the same importance for economic development in the whole analyzed period.

Economic development	
Area 1 (EO_1) – Economy	
X_1 – Gross domestic product per capita	stimulant
X_2 – Investments outlays per capita	stimulant
Area 2 (EO_2) – Zatrudnienie	
X_3 – Employment rate by age	stimulant
Area 3 (EO_3) – Innowacyjność	
X_4 – Expenditure on R&D activity in relation to GDP	stimulant

Table 1. Diagnostic variables.

Weights (year)			
2010			
$\mathbf{w_1}$	$\mathbf{w_2}$	$\mathbf{w_3}$	$\mathbf{w_4}$
0.248	0.254	0.262	0.236
2014			
$\mathbf{w_1}$	$\mathbf{w_2}$	$\mathbf{w_3}$	$\mathbf{w_4}$
0.251	0.251	0.253	0.245

Table 2. Weights based on the entropy values.

Next the values of TMDit for economic development for the years 2010 and 2014 were assessed. Based on the obtained values of TMDit a ranking of voivodeships for both years was given. Additionally, the voivodeships were grouped into four relatively homogenous subsets, where the voivodeships characterised with the highest level of economic development was classified in the class 4 and the once with its lowest level were grouped in the class 1. For this purpose a natural breaks method was applied. The results are given in table 3 and figure 1.

In the years 2010-2014 the level of economic development in the case of most of the voivodeships was improved, which can be seen in the changes in their grouping. Additionally, an increase in the value of TMDit was obtained. In the year 2010 in the first class grouping the voivodeships with the lowest level of economic development one could find five voivodeships (lubelskie, opolskie, warmińsko-mazurskie, podlaskie, zachodniopomorskie), whereas in the year 2014 only two voivodeships (świętokrzyskie and warmińsko-mazurskie) could be found here. A negative example could be seen in the case of świętkorzystkie voivodeship, which in the first year of analysis was classified in the 2 group, whereas in the year 2014, it was found in the 1 class – characterised with the lowest level of development.

Economic development							
2010				2014			
Voivodeships	TMR	Rank	Class	Voivodeships	TMR	Rank	Class
mazowieckie	0.721	1	4	mazowieckie	1.000	1	4
wielkopolskie	0.417	2	3	wielkopolskie	0.507	2	3
pomorskie	0.347	3	3	dolnośląskie	0.498	3	3
łódzkie	0.344	4	3	pomorskie	0.483	4	3
dolnośląskie	0.333	5	3	łódzkie	0.478	5	3
małopolskie	0.315	6	3	małopolskie	0.434	6	3
śląskie	0.298	7	2	śląskie	0.41	7	3
lubuskie	0.294	8	2	podlaskie	0.339	8	2
świętokrzyskie	0.266	9	2	lubelskie	0.301	9	2
kujawsko-pomorskie	0.254	10	2	opolskie	0.296	10	2
podkarpackie	0.226	11	2	zachodniopomorskie	0.293	11	2
lubelskie	0.205	12	1	kujawsko-pomorskie	0.287	12	2
opolskie	0.192	13	1	podkarpackie	0.279	13	2
Warmińsko-mazurskie	0.169	14	1	lubuskie	0.262	14	2
podlaskie	0.164	15	1	świętokrzyskie	0.202	15	1
zachodniopomorskie	0.146	16	1	Warmińsko-mazurskie	0.166	16	1

Table 3. Ranking and grouping of voivodeships based on the level of economic development.

Mazowieckie voivodeship is the one with the highest level of economic development. Both in 2010 and 2014 it forms individually the 4 class, which is characterised with the highest level of development.

In the 3 class with high level of economic development in the year 2010 one could see dolnośląskie, wielkopolskie, pomorskie, łódzkie and małopolskie voivodeships, whereas in the year 2014 one could additionally find śląskie voivodeship in that group. In the case of the class with an average level of development in the year 2014 one could see zachodniopomorskie, podlaskie, lubelskie, opolskie, kujawsko-pomorskie, podkarpackie and lubuskie voivodeships. Additionally, in the analyzed period a promotion of lubelskie, opolskie, zachodniopomorskie and podlaskie voivodeships form the 1 to 2 class was recorded.

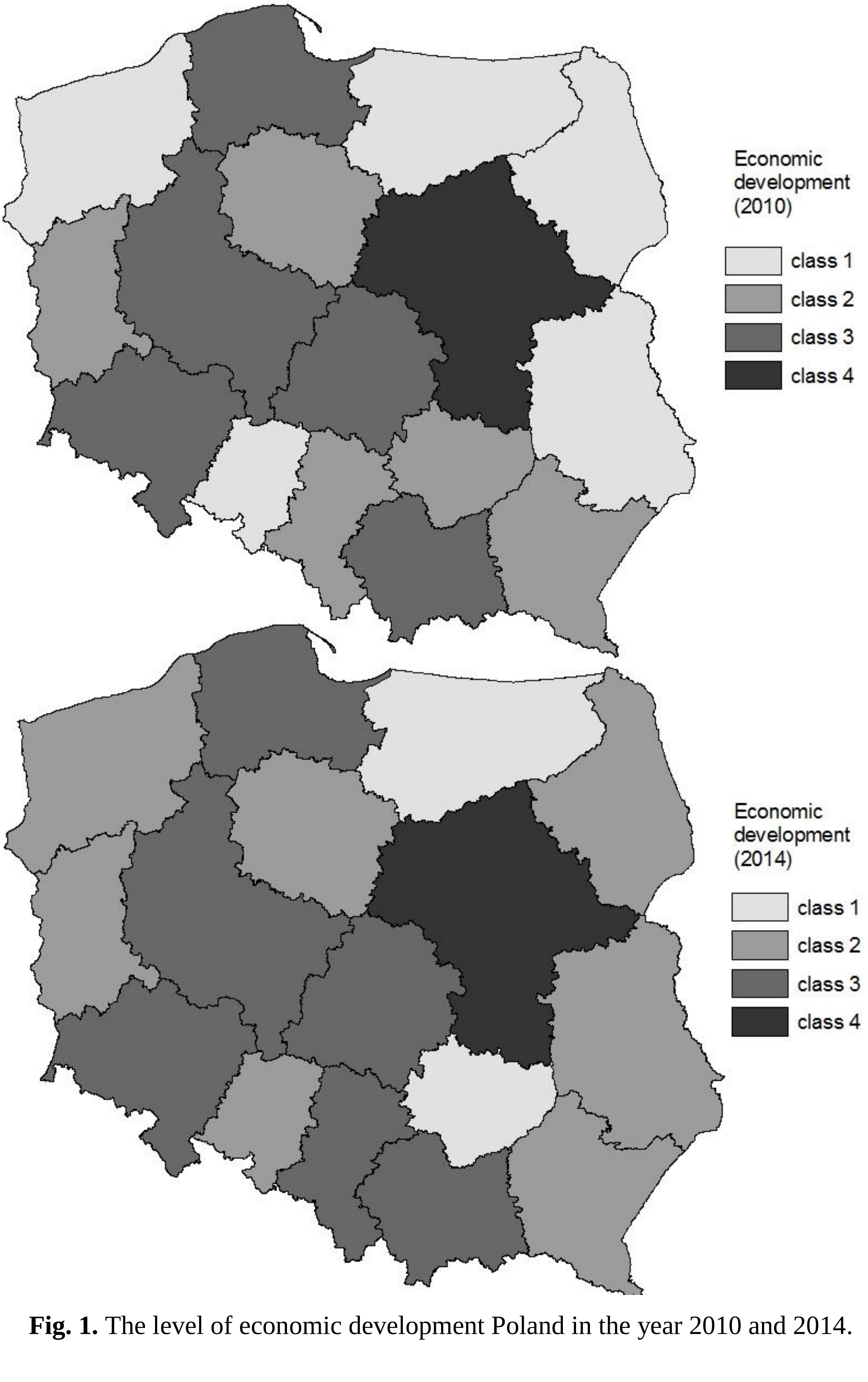


Fig. 1. The level of economic development Poland in the year 2010 and 2014.

Conclusions

The aim of the article was to assess the level of economic development in Poland at the voivodeships level in the years 2010-2014. In the analysis the phenomenon of regional economic development was considered as a multivariate problem. Thus, a taxonomic measure of development with the weights based on entropy values were applied.

Based on the obtained result, it can be said that the empirical hypothesis of the research given as follow – In Poland one can see the process of improvements in the level of economic development at regional level – was not rejected.

References

Balcerzak, A. P. (2016a). Fiscal Burden in the European Union Countries. *Economic Annals XXI*, 161(9-10), 4-6.

Balcerzak, A. P. (2016b). Technological Potential of European Economy. Proposition of Measurement with Application of Multiple Criteria Decision Analysis. *Montenegrin Journal of Economics*, 12(3), 7-17.

Balcerzak, A. P., & Pietrzak, M.B. (2016). Quality of Institutions for Knowledge-based Economy within New Institutional Economics Framework. Multiple Criteria Decision Analysis for European Countries in the Years 2000–2013. *Economics & Sociology*, 9(4), 66-81.

Ciburiene, J. (2016). The Evaluation of Economic Development Index of Poland and Lithuania in the Context of the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 437-449.

Hadaś-Dyduch, M. (2015). Polish macroeconomic indicators correlated-prediction with indicators of selected countries. In M. Papież & S. Śmiech (Eds.). *Proceedings of the 9th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena, Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 68-76.

Jantoń-Drozdowska, E., & Majewska, M. (2016). Investment Attractiveness of Central and Eastern European Countries in the Light of New Locational Advantages Development. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(1), 97-119.

Kondratiuk-Nierodzińska, M. (2016). New Knowledge Generation Capabilities and Economic Performance of Polish Regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy,* 11(3), 451-471.

Kordalska, A., & Olczyk, M. (2016). Global Competitiveness and Economic Growth: A One-Way or Two-Way Relationship?. *Equilibrium. Quarterly Journal of Economics and Economic Policy,* 11(1), 121-142.

Kryk, B. (2016). Accomplishment of the European Union Lifelong Learning Objectives in Poland. *Oeconomia Copernicana*, 7(3), 389-404.

Łyszczarz, B. (2016). Public-private Mix and Performance of Health Care Systems in CEE and CIS Countries. *Oeconomia Copernicana*, 7(2), 169-185.

Małkowska, A., & Głuszak, M. (2016). Pro-investment Local Policies in the Area of Real Estate Economics - Similarities and Differences in the Strategies Used by Communes. *Oeconomia Copernicana*, 7(2), 269-283.

Pietrzak, M. B. (2016). The Problem of the Inclusion of Spatial Dependence Within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), 69-86.

Pietrzak M. B, Wilk J., & Siekaniec M. (2013). The Impact of Metropolitan Areas on Internal Migrations in Poland. The Case of Southern Regions. In M. Papież & S. Śmiech (Eds.). *Proceedings of the 7th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow: Foundation of the Cracow University of Economics, 124-132.

Shuaibu, M., & Oladayo, P. T. (2016). Determinants of Human Capital Development in Africa: A Panel Data Analysis. *Oeconomia Copernicana*, 7(4), 523-549.

Wang, T. C., & Lee, H. D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications*, 36(5), 8980-8985.

Zemtsov, S. P., Pavlov, P. N., & Sorokina, A. V. (2016). Specifics of Cluster Policy in Russia. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 499-536.

Żelazny R., & Pietrucha J. (2017). Measuring Innovation and Institution: the Creative Economy Index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1).

The use of random fields in the Modifiable Areal Unit Problem

Michał Bernard Pietrzak[1], Bartosz Ziemkiewicz[2]

Abstract

The focus of the research will be on the modifiable areal unit problem (MAUP) within which two aspects will be considered: the scale problem and the aggregation problem. In the article we consider the use of random fields theory for the needs of the "Scale Problem" issue. The Scale Problem is defined as a volatility of the results of analysis as a result of a change in the aggregation scale. In the case of the scale problem empirical studies should be conducted with application of simulations. Within the simulation analysis the realisations of random fields referred to irregular regions will be generated. First, the internal structure of spatial processes will be analysed. Next, we consider the theoretical foundations for random fields relative to irregular regions. The accepted properties of random fields will be based on the characteristics established for economic phenomena. The outcome of the task will be the development of a procedure for generating the vector of random fields with specified properties. Procedure for generating random fields will be used to simulations within the scale problem too. The research is funded by National Science Centre, Poland under the research project no. 2015/17/B/HS4/01004.

Keywords: *spatial econometrics, Scale Problem, random fields, Modifiable Areal Unit Problem, simulations*

JEL Classification: C10, C15, C21

1 Introduction

The article deals with the problem of Modifiable Areal Unit Problem (MAUP) which is found to be significant in the area of economic spatial analysis (see: Arbia, 1989). The MAUP issue concerns the possibility of obtaining different results due to changes in the level of aggregation (see: Pietrzak, 2014). The main purpose of this work is to consider the Scale Problem, which is one of the aspects of the issue of MAUP. Analysis of the Scale Problem will be conducted based on the example of the assessment of socio-economic development of which one aspect can be expressed by means of number of entities of the national economy per capita. Examining the number of entities of the national economy makes only one aspect of the complex phenomenon of socio-economic development. Various aspects of socio-economic development both at the regional and national levels have been examined in many works (see: Ciburiene, 2016; Łyszczarz, 2016; Balcerzak, 2016a, 2016b; Jantoń-Drozdowska

[1] Corresponding author: Nicolaus Copernicus University, Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: pietrzak@umk.pl.

[2] Nicolaus Copernicus University, Department of Probability Theory and Stochastic Analysis, ul. Chopina 12/18, 87-100 Toruń, Poland, e-mail: bartek@mat.umk.pl.

and Majewska, 2016; Małkowska and Głuszak, 2016; Pietrzak and Balcerzak, 2016; Żelazny and Pietrucha, 2017).

Research conducted in this work enabled us to consider the scale problem based on the example of spatial development of the number of business entities. Empirical analysis of the properties of this phenomenon allowed the spatial trend in its internal structure to be identified. Then, a simulation analysis was performed with the assumption of identified empirical properties. It turned out that the simulation analysis performed at various levels of aggregation led to the obtainment of similar parameter estimates of the spatial trend and a different correlation structure for simulated processes. This allowed the evaluation of selected elements of the internal structure of spatial processes.

2 Analysis of the internal structure of selected spatial processes

In the performed analysis emphasis was laid on distinguishing processes expressed in the absolute quantities from those expressed in relative quantities. This results from to the fact that spatial studies ought to be predominantly based on the analysis of processes expressed in relative quantities referred to certain values characterizing the selected region (area, number of residents). This ensures the comparability of data and the correctness of the results obtained. In case of spatial economic analysis of business entities, final conclusions should be based on a process expressed in relative quantities, *i.e.*, on the number of business entities referred to the number of residents. This process combines two processes expressed in the absolute quantities, the number of business entities and the number of residents. The processes adopted in the study are: X1 - population in 2016, X2 - number of entities of the national economy in 2016 and Y - number of entities of the national economy per capita in 2016. In addition, it is assumed that data used in the context of spatial economic analyses may be treated as a realization of a two-dimensional random field X(u1,u2), where u1,u2 denote the coordinates on the plane (see: Arbia, 1989). The two-dimensional random field defined in such a way will be further referred to in the work as a spatial process.

The scale problem should be examined only for a composition of territorial units forming Quasi Composition of Regions. Therefore, the next step in the study of the Scale Problem should consist in determining Quasi Composition of Regions for the examined number of business entities in Poland. The concept of Quasi Composition of Regions was introduced by Pietrzak (2014a). This concept refers to a composition of regions which is formed of individual compositions of territorial units for subsequent levels of aggregation. Individual compositions of territorial units need to be selected in such a way that research conducted based on them should provide valid conclusions. In this study, Quasi Composition of Regions will be limited to two individual regions from NUTS 4 and NUTS 3 territorial units.

Since economic analysis will be performed based on selected spatial processes, it is very important to examine their internal structure (see: Pietrzak 2014). Therefore, after determining Quasi Composition of Regions, another step consisted in studying the internal structure of selected spatial processes X1, X2, Y. Studying the internal structure of spatial processes means providing a correct description of their properties. The following components of the internal structure can be distinguished: the component related to unsystematic heterogeneity, the component related to systematic heterogeneity, and the component of the structure where the spatial process is homogeneous (see: Pietrzak 2014a).

The analysis focused on the study of systematic heterogeneity in the form of parameters estimation of the model of spatial trend[3].For each process X1, X2, Y we estimated a linear spatial trend model determined by the following equation

$$\mathbf{Y} = \alpha_0 + \alpha_1 \mathbf{U}_1 + \alpha_2 \mathbf{U}_2 + \boldsymbol{\varepsilon} . \qquad (1)$$

where $\mathbf{Y}$ is the vector of spatial process, $\mathbf{U}_1$, $\mathbf{U}_2$ are vectors of geographic coordinates, $\boldsymbol{\varepsilon}$ is the vector of spatial noise, α_0, α_1, α_2 are parameters.

It is expected that a spatial linear trend in the process of the number of business entities per capita will occur and the phenomenon is expected to become more intense in the east-west directions. The cause of the occurrence of a linear spatial trend specified in such a way results from a higher level of socio-economic development of western parts of Poland (see: Pietrzak et al., 2013; Hadaś-Dyduch, 2016; Kondratiuk-Nierodzińska; 2016; Czaplak, 2016; Murawska, 2016). Therefore, for the selected processes we estimated the spatial model parameters of a linear trend following formula 1.The results obtained are shown in Table 1. The identification of the spatial line trend was made only for the Y process - the number of entities of the national economy per capita. The parameter with the variable referring to longitude coordinates proved to be statistically significant. The negative parameter estimate indicates a spatial increase in the number business entities per capita in Poland, as long as we move to the west.

After the estimation of linear spatial trend models, we estimated the value of Pearson's correlation coefficients between the processes, where for the Y process a further spatial line

[3] We deliberately omitted here the internal structure in the form of spatial autocorrelation (spatial homogeneity) to focus in the article solely on the analysis of heterogeneity in the form of a systematic spatial trend. The problems related to spatial autocorrelation were raised in the works of Pietrzak et al., (2014), Pietrzak (2016).

trend was identified. The performed analysis of the results led to the conclusion that there are strong positive correlations among all processes. At the NUTS 4 aggregation level the highest level of correlation dependence occurs between the processes X1 and X2 and reaches level of 0.971. For the pairs of the processes (Y, X1) and (Y, X2) the dependence reaches similar levels of 0.481 and 0.421 respectively. The correlation coefficients change during the transition to a higher level of aggregation. At the NUTS 3 aggregation level the correlative dependence between the processes of X1 and X2 decreased to the level of 0.889. In turn, the correlative dependence between the processes of Y and X2 remained practically unchanged and reaches level of 0.433, and between the processes Y and X1 it rose to 0.741.

Process (Aggregation level)							
Process Y - Aggregation level NUTS 4				**Process Y - Aggregation level NUTS 3**			
Parameter	Estimate	p-value	R^2	Parameter	Estimate	p-value	R^2
α_1	-0.661	~0.000	0.186	α_1	-0.748	0.001	0.245
α_2	0.143	0.072	-	α_2	0.187	0.318	-
Process X1 - Aggregation level NUTS 4				**Process X1 - Aggregation level NUTS 3**			
Parameter	Estimate	p-value	R^2	Parameter	Estimate	p-value	R^2
α_1	-9.113	0.8116	0.002	α_1	112.248	0.473	0.031
α_2	-64.551	0.103	-	α_2	-166.083	0.308	-
Process X2 - Aggregation level NUTS 4				**Process X2 - Aggregation level NUTS 3**			
Parameter	Estimate	p-value	R^2	Parameter	Estimate	p-value	R^2
α_1	-530.752	0.421	0.007	α_1	-1606.984	0.604	0.005
α_2	-298.757	0.662	-	α_2	109.743	0.972	-

Table 1. The results of the estimation of model parameters of a linear spatial trend.

Based on the empirical analysis of the properties of processes we can conclude that the linear trend did not change during the transition to a higher level of aggregation. Also in the case of correlations dependence, there were no significant changes during the transition from the NUTS4 level to the NUTS3 level. The next phase of the study should consist in performing simulation analysis and checking whether similar results will be obtained. The simulated processes should have similar properties to those set for the empirical processes. Simulation analysis should allow us to find out whether between the simulated processes at the aggregate NUTS 4 level and NUTS 3 level there are any differences in a form of a spatial linear trend as well as in the structure of correlations dependence.

Therefore, for the purposes of simulation analysis we assumed for spatial processes an adequate correlation structure, for the pairs of the processes (Y, X1) – 0.39 ,(Y, X2) – 0.71 (X1,X2) – 0.91. We also assumed adequate parameters of the spatial trend in the case of process Y, where α_0=10, α_1=-0.4, α_2=0.1. In addition, it was assumed that between the processes there is a relation defined by the following equation

$$Y = \frac{X2}{X1}. \tag{2}$$

The relationship defined by the formula (2) causes a problem in the simulation of the random fields vector with given properties (see Arbia, 1989), hence the assumptions in correlation structure and parameters of the spatial trend are different from the properties of empirical spatial processes. Accordingly, the simulation of the random field vector was performed in the following steps. In the first step, a simulation was performed of two random fields Y and X1 with the assumed correlation dependence at the level of 0.39. Next the spatial trend values were added to the received results of the process Y. Then, based on the spatial processes Y and X1,the values of the process X2 were determined according to the formula 2. In the last step a correction of the correlation structure was made.

As a result of the simulation, we obtained simulated values of the random fields vector, where every individual component is a spatial process related to the composition of NUTS 4 spatial units. Thus we obtained a starting set of spatial processes at the aggregation NUTS 4 level with specified properties. This was followed by aggregation of simulated spatial processes so that the aggregated processes could relate to the composition of NUTS 3 territorial units. Aggregation of spatial processes was carried out by an appropriate summation of the spatial processes X1 and X2. We calculated the sum of the values of the process from respective regions at the NUTS 4 level that make up a selected region at the NUTS 3 level. Then for the aggregated processes X1 and X2 we determined their quotient and thus we received the values of the spatial process Y referred to the NUTS 3 composition. Therefore, aggregation was performed only for the processes X1 and X2. However, the values of the process Y at the NUTS 3 level were obtained based on the formula 2.

As a result of the simulation, we received a thousand simulated values of the random fields vector at the aggregate NUTS 4 level. In turn, the aggregation performed allowed to obtain a thousand of simulated values of the random fields vector at the aggregate NUTS 3 level. The possessed simulated values of variables at the two levels of aggregation, i.e., NUTS 4 and NUTS 3, allowed us to estimate model parameters of the linear spatial trend, and then to determine the value of the correlation dependence between the processes devoid of spatial trend. In this way two resultant sets were obtained - a set of estimations of correlation coefficients and a set of parameter estimations of the spatial model. Based on the two sets, the mean values and standard deviations were determined (see: Table 2 and Table 3).

Correlation (aggregation level)			
Correlation between processes (aggregation level NUTS 4)			
Statistics	(Y, X1)	(Y,X2)	(X1,X2)
Mean	0.393	0.713	0.912
Standard deviation	0.002	0.002	0.002
Correlation between processes (aggregation level NUTS 3)			
Statistics	(Y, X1)	(Y,X2)	(X1,X2)
Mean	0.138	0.337	0.972
Standard deviation	0.102	0.094	0.004

Table 2. The results of the estimation of the correlation structure on simulated data.

The results obtained allowed us to draw up the following conclusions. For all spatial processes there were no changes in the nature of the linear spatial trend due to the aggregation process. According to the assumption made, in case of the processes X1 and X2, a spatial trend was not present at the NUTS 4 level and after aggregation it also proved to be statistically insignificant at the NUTS 3 level. However, in the case of the process Y, the presence of a spatial trend at the aggregation NUTS 4 level was assumed. As a result of the estimation of the trend model parameters for the process Y at the aggregation NUTS 3 level, we obtained similar evaluation parameters to the aggregation at the NUTS 4 level. Based on the simulation performed, it can be concluded that the results of parameters estimation of a linear model of the spatial trend do not change depending on the choice of the aggregation level. Obviously, a condition must be fulfilled that the adopted compositions of territorial units belong to Quasi Composition of Regions. It should be further noted that the higher the aggregation level, the better fit of the spatial trend model to empirical data (see Table 3).

The simulation analysis conducted also allowed us to conclude that the correlation structure of spatial processes changed as a result of the aggregation process. There was decline in the value of the correlation dependence between the pairs of processes (Y, X1) and (Y, X2) and an increase in the level of the correlation dependence between the pairs of processes (X1, X1) (see Table 2). In case of the analysis of empirical processes, the spatial correlation structure of these processes did not change. It is possible that this is due to the presence of other properties, including spatial autocorrelation. Therefore, taking into account spatial autocorrelation in simulations should be the subject of further study.

Statistics	**α_0**	**p-value**	**α_1**	**p-value**	**α_2**	**p-value**	**R^2**
Process Y (Aggregation level NUTS 4)							
Mean	10.470	~0.000	-0.424	~0.000	0.099	0.129	0.216
Standard deviation	0.370	~0.000	0.039	~0.000	0.047	0.201	0.027
Process Y (Aggregation level NUTS 3)							
Mean	10.528	~0.000	-0.413	~0.000	0.101	0.202	0.501
Standard deviation	0.441	~0.000	0.055	~0.000	0.053	0.263	0.095
Process X1 (Aggregation level NUTS 4)							
Mean	1116.218	~0.000	-25.210	0.071	2.910	0.506	0.017
Standard deviation	75.284	~0.000	9.522	0.154	11.286	0.313	0.013
Process X1 (Aggregation level NUTS 3)							
Mean	4787.751	0.003	-23.327	0.826	21.784	0.303	0.009
Standard deviation	4371.663	0.003	557.875	0.129	6.453	0.131	0.012
Process X2 (Aggregation level NUTS 4)							
Mean	11707.71	~0.000	-643.677	~0.000	116.828	0.402	0.086
Standard deviation	872.7975	~0.000	110.782	~0.000	119.756	0.299	0.025
Process X2 (Aggregation level NUTS 3)							
Mean	53283.251	0.001	-2480.641	0.204	239.923	0.231	0.039
Standard deviation	5040.059	0.001	643.462	0.121	696.878	0.126	0.022

Table 3. Estimation of the spatial linear trend model parameters based on simulated data.

Conclusions

The subject of the article concerned the scale problem whose presence may lead to different results obtained from spatial economic analysis. The scale problem was analysed based on an empirical example of the formation of number of entities of the national economy per capita in Poland. For the needs of this research, we established Quasi Composition of Regions which consisted of two single areas of territorial units, the NUTS 4 and NUTS 3 systems. The performed analysis of the empirical properties of processes at both levels of aggregation allowed the identification of their internal structure. The existence of a spatial trend for the number of entities of the national economy per capita was established, and after taking into account this fact the correlation structure was determined for selected spatial processes.

Then a simulation analysis was made where the simulated spatial processes displayed similar properties to the empirical processes examined. It turned out that the simulation analysis performed at various levels of aggregation led to the obtainment of similar parameter evaluations of the spatial trend. This means that the scale problem does not significantly influence the nature of the spatial trend when affected by an aggregation process. Simulation analysis also allowed the identification of changes in the correlation structure of the processes resulting from aggregation. The comparison of the results obtained for empirical processes and results gained from the simulation analysis indicate that the lack of changes in the correlation structure for empirical processes may result from the presence of spatial autocorrelation. Therefore, the obtained results indicate the need to broaden the scope of research into the scale problem by the inclusion of the issue of spatial autocorrelation.

Acknowledgements

The research is funded by National Science Centre, Poland under the research project no. 2015/17/B/HS4/01004.

References

Arbia, G. (1989), Spatial *Data Configuration in Statistical Analysis of Regional Economics and Related Problems*, Kluwer Academic Publisher, Dordrecht.

Balcerzak, A. P. (2016a). Technological Potential of European Economy. Proposition of Measurement with Application of Multiple Criteria Decision Analysis. *Montenegrin Journal of Economics*, 12(3), 7-17. DOI: 10.14254/1800-5845.2016/12-3/1.

Balcerzak, A. P. (2016b). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.

Ciburiene, J. (2016). The Evaluation of Economic Development Index of Poland and Lithuania in the Context of the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 437-449.

Czaplak, J. (2016). Analysis of the Effectiveness of Public Services Provision by Local Governments in the Years 2007–2013. *Oeconomia Copernicana*, 7(2), 223-243.

Hadaś-Dyduch M. (2016). Econometric-wavelet prediction in the aspect of space-time. In: Papież, M., Śmiech, S. (eds.), *Proceedings of the 10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*,.

Conference Proceedings, Cracow: Foundation of the Cracow University of Economics, 45-52.

Jantoń-Drozdowska, E., & Majewska, M. (2016). Investment Attractiveness of Central and Eastern European Countries in the Light of New Locational Advantages Development. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(1), 97-119.

Kondratiuk-Nierodzińska M. (2016). New knowledge generation capabilities and economic performance of polish regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*,11(3), 451-471. DOI: http://dx.doi.org/10.12775/EQUIL.2016.021

Łyszczarz, B. (2016). Public-private Mix and Performance of Health Care Systems in CEE and CIS Countries. *Oeconomia Copernicana*, 7(2), 169-185.

Małkowska, A., & Głuszak, M. (2016). Pro-investment Local Policies in the Area of Real Estate Economics - Similarities and Differences in the Strategies Used by Communes. *Oeconomia Copernicana*, 7(2), 269-283.

Murawska A. (2016). Differences in unemployment among persons in a special situation on the labour market on the example of polish provinces. *Oeconomia Copernicana*, 7(3), 375-388. DOI: http://dx.doi.org/10.12775/OeC.2016.022

Pietrzak M. B, Wilk J., & Siekaniec M. (2013). The Impact of Metropolitan Areas on Internal Migrations in Poland. The Case of Southern Regions. In: Papież, M., Śmiech, S. (eds.), *Proceedings of the 7th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Cracow: Foundation of the Cracow University of Economics, 124-132.

Pietrzak M. B. (2014a). Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale Problem. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 9(2), 111-132.

Pietrzak M.B., Wilk J.,. Kossowski T., & Bivand R. (2014). The Identification of Spatial Dependence in the Analysis of Regional Economic Development - Join-count Test Application. In: Papież, M., Śmiech, S. (eds.), Proceedings of the 8th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Cracow: Foundation of the Cracow University of Economics, 135-144.

Pietrzak, M. B. (2016). The Problem of the Inclusion of Spatial Dependence Within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), 69-86.

Pietrzak, M. B., & Balcerzak, A. P. (2016). Assessment of Socio-Economic Sustainability in New European Union Members States in the years 2004-2012. In: Papież, M., Śmiech, S. (eds.), *The 10th Professor Aleksander Zelias International Conference on Modelling and*

Forecasting of Socio-Economic Phenomena. Conference Proceedings, Cracow: Foundation of the Cracow University of Economics, 120-129.

Żelazny R., & Pietrucha J. (2017). Measuring Innovation and Institution: the Creative Economy Index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1).

Risk measurement for goals realization in a household financial plan

Radosław Pietrzyk[1], Krzysztof Piontek[2], Paweł Rokita[3]

Abstract

This article presents a concept of household financial plan integrated risk measures with short-term and long-term approach to financial plan risk mitigation combined in one procedure. Short-term and long-term risk measures, based on household default probability, are introduced. Then, two approaches to their application in financial plan management process are proposed.

Keywords: *household, financial planning, risk measurement*

JEL Classification: D91, E21, J26, D10

1 Introduction

When constructing a financial plan for a household, it is very important to be able to incorporate risk, related to realization of its financial goals, into the model. The risk should be part of the plan choice criteria. The plan choice process should also be suited to natural psychological perspective of decision makers and it should be understandable for them.

People tend to think of their financial situation in rather a short-term perspective (Ballinger et al. 2003; Carbone and Hey, 2004; Carbone and Infante, 2012), whereas they need a life-long plan indeed. Amongst other reasons, the last is necessary for successful accomplishment of retirement goal. An approach that allowed to accommodate these perspectives would be thus welcomed. Moreover, it may be worth consideration to analyse risk aversion towards short-term and long-term threats separately. Treating risk aversion as one general characteristic of an investor, regardless the decision horizon, may be misleading.

Whereas there is a lot of discussion about intertemporal choice in respect of consumption in the existing literature, the solutions proposed there do not address the potential conflict between short-term risk minimization and minimization of the whole-plan risk, understood as the threat that the household will not be able to fully realize its life-long financial plan.

The work takes up the discussion about augmentation and modification of the known concepts in the area of household financial planning, to overcome the problem that has been

[1] Wroclaw University of Economics, Department of Financial Investments and Risk Management, Komandorska 118/120, 53-345 Wroclaw, Poland, e-mail: Radoslaw.Pietrzyk@ue.wroc.pl.

[2] Wroclaw University of Economics, Department of Financial Investments and Risk Management, Komandorska 118/120, 53-345 Wroclaw, Poland, e-mail: Krzysztof.Piontek@ue.wroc.pl.

[3] Wroclaw University of Economics, Department of Financial Investments and Risk Management, Komandorska 118/120, 53-345 Wroclaw, Poland, e-mail: Pawel.Rokita@ue.wroc.pl.

pointed out here. A proposal of risk measurement approach is presented. It combines risk of short- and long-term goals realization. Also such plan optimization constraint is proposed that household's preferences towards short- and long-term risk mitigation are incorporated.

2 Household financial planning model used

The discussion about household financial plan risk is based here on a conceptual framework for financial plan models, proposed by Jajuga et al. (2015). The discussion is, however, more general and not limited to the assumptions of this model only.

The framework is cash-flow based, discrete time, with a discrete space of scenarios. A scenario variable is a random vector with a discrete joint distribution. Elements of the random vector are risk factors. Each realization of the scenario variable is a scenario. A two-person household is assumed. The two persons are referred to as main household members. Though they are the only decision makers, cases with more persons (like children) are also covered by the model. Any number of financial goals is possible, with the retirement goal treated as a distinguished one – the one that must be set. For most households, this goal is unachievable without long-term planning (relatively big value and lack of post-financing possibility).

The model was developed in stages. The first, basic variant, called here the baseline model, assumes only two goals: retirement and bequest (if the household members show a bequest motive). Only two risk factors are taken into account there. These are dates of death of the main household members. The baseline model assumes only retirement investment.

The goal function of the optimization procedure is called value function of the household (Pietrzyk and Rokita, 2015b, 2016b). It is based on expected discounted utility of consumption and expected discounted utility of residual wealth. Its general idea builds on the classical concept by Yaari (1965), but it is suited to the discrete-time, discrete-space-of-states, two-person household model with a bequest motive. The household value function is maximized under two constraints: minimum consumption and the requirement that the pre-planned life annuity purchase expense is covered, fully and on time. The only decision variables in the financial plan optimization procedure are: consumption rate (of the household taken as a whole) in the first year of the plan and the proportion in which retirement investment contribution of the household is divided between systematic investment plans of each person.

The baseline model is constructed so that its augmentation is relatively easy. It may be extended to cover more financial goals, risk factors, as well as more ways of goal financing.

The general aim of the household is to realize all financial goals. The goal function of the optimization procedure is an augmented version of the one from the baseline model (Jajuga et al., 2015). Like in the baseline model, it is called value function of the household. Also new decision variables are introduced there. For example, the proportion in which each post-financeable goal is pre-financed (own contribution) and post-financed (debt) may be optimized. The set of boundary conditions consists now of: the minimum consumption constraint and the subset of constraints that all pre-planned cash outflows, that will result from accomplishment of the financial goals, must be covered fully and on time.

Originally, no bankruptcy level constraint or maximum indebtedness constraint was used in the Jajuga et al. (2015) model. If the result of optimization had shown an unacceptably high probability of bankruptcy, the decision makers would have had to review their goals, and then they would have started a new optimization procedure with a less ambitious goals set. Thus, risk was a criterion of plan choice in general, but not of the automated optimization procedure.

3 The role of integrated risk measures in financial plan management

One of key problems in the search for a life-long financial plan for a household is management of its risk. The easiest, but not recommended, approach is to analyse its different types separately. Risk steering methods that are suited to their specific risk types may be locally efficient, but life-long financial planning requires a more integrated approach.

The general scheme of risk management is more or less similar as for any other risky situation. That is, the main elements of *risk management process* are:

1) definition of risk management aims,
2) identification of risk types and risk factors,
3) risk measurement,
4) risk steering,
5) risk control.

The last (risk control) refers both to verification of risk steering effects and to overall control of all steps of the process.

In addition to the very risk management process, it is necessary to be aware of the role it plays in a more general process of financial plan management. The risk should be understood as the threat that the whole financial plan fails to be successfully accomplished. It may happen due to any of the possible reasons. This is an argument for integrated risk management.

In the proposal by Jajuga et al. (2015), integrated risk measurement is a part of household financial plan management process at the stage after plan optimization. Risk measures are used there to check if the plan that has already been optimized under a given set of constraints bears acceptable level of risk (where one of the optimization constraints is the requirement that the whole bunch of household financial goals should be realized). If the optimal plan is unacceptable, then the household members need to revise their goals and try to find a new optimal solution with the new set of boundary conditions. The reason for which the plan choice procedure is not fully automated is that no algorithm can autonomously change goals of the household. The concept of cyclical process of financial plan management and the position of risk measurement in this process is illustrated in **Fig. 1**.

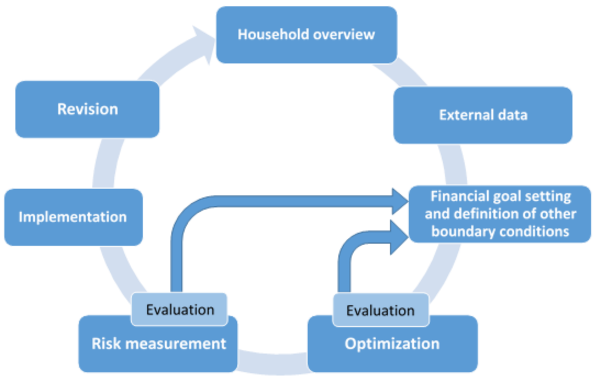


Fig. 1. Household financial plan management – a cyclical dynamic process.
Source: Jajuga, Feldman, Pietrzyk and Rokita (2015).

Nevertheless, some of the risk measures proposed there, as well as the new ones that are introduced in this article, allow to better integrate risk measurement with the plan management process. It would be close to an ideal situation if maximization of value function led automatically to an optimal trade-off between maximization of some measure of success and minimization of integrated risk measure. This would make the plan optimization problem analogous to portfolio optimization in the classical Markowitz world. At the present stage of model development, no such elegant proposal has been produced yet, nor the one that is proposed here fulfils this postulate. It is, however, possible to rebuild the optimization procedure so that a maximum acceptable level of risk is imposed there as a constraint.

Thanks to this solution, one may use the risk measure at an earlier step of financial plan management process, namely – on the stage of plan optimization. The risk management process may be then described like in the **Fig. 2**.

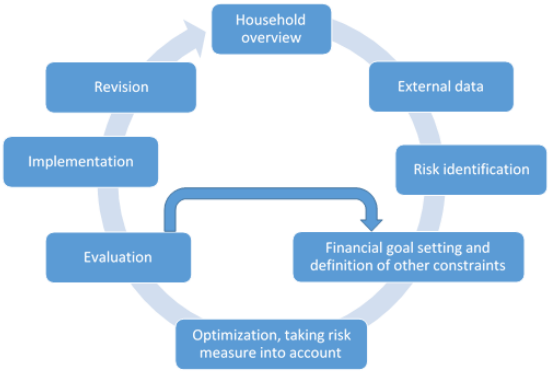


Fig. 2. Household financial plan management with risk incorporated into plan optimization procedure – a cyclical dynamic process.

Now, we need a measure of risk that may be conveniently used in the construction of a boundary condition. It must be also integrated and it should reflect threats to realization of the whole financial plan throughout the whole life of the household. By the term *integrated measure of risk* it is understood a measure that incorporates information on all types of risk and all risk factors that are taken into account in a given financial plan optimization model.

For the needs of this work, it is proposed to use measures based on household default probability (Jajuga et al., 2015; Pietrzyk and Rokita, 2016a). Section 3 describes this measure in more details, whereas section 4 discusses short-term and long-term measures based on it.

4 Household Default Probability (*HDP*)

Household default probability – a measure of risk proposed for the first time by Pietrzyk and Rokita (2015a) – is the probability that cumulated net cash flow falls below a critical shortfall level, called *default threshold*. In the model with a discrete number of scenarios and with discrete time, calculation of this probability consists in:

1) Analysing of each scenarios to check if cumulated net cash flow falls below the default threshold, at any moment of this scenario, since the start of the plan until the date when, under this scenario, the longer-surviving household member dies.
2) Marking, as a default scenario, each one in which cumulated net cash flow intersects the default threshold at least once.
3) Calculating probabilities of the default scenarios.
4) Summing probabilities of all default scenarios to obtain the probability that the household, having implemented this plan, defaults (*household default probability*).

A default threshold is determined at the plan starting moment (t_0) – for example, credit worthiness at t_0. Then, the default threshold changes with financial situation of the household.

Let us denote the multivariate random variable "scenario" with the symbol **Z** and some particular value of it – as $\mathbf{Z}^*$. Generally speaking, scenario is here a vector of risk factors that are taken into account in a given financial plan model. Generally speaking, scenario variable may be formally described as: $\mathbf{Z} = \left[X_1, X_2, ..., X_n\right]$, where the variables X_i are all random variables which are treated as risk factors under the assumptions of the model.

For example, in the baseline model, the scenario variable is bivariate and it consists only of dates of death. If we call one of the two main household members Person 1 and the second – Person 2, and denote the moment of death of Person 1 as D_1 and the moment of death of Person 2 as D_2, then the scenario variable is defined as: $\mathbf{Z} = \left[D_1, D_2\right]$. One particular scenario (a value or a realization of the random variable **Z**) may be then denoted as: $\mathbf{Z}^* = \left[D_1^*, D_2^*\right]$.

Definition of a default scenarios, that is – the ones under which cumulated shortfall exceeds the default threshold, may be formally described like in the formula (1):

$$\mathbf{T}^* = \mathbf{Z}^* : \underset{t=1,..,T_B^*}{\exists} CSp_t^{(\mathbf{Z}^*)} < DTh_t^{(\mathbf{Z}^*)}, \tag{1}$$

where: $CSp_t^{(\mathbf{Z}^*)}$ denotes cumulated net cash flow (cumulated surplus) at a moment t and under a scenario $\mathbf{Z}^*$, T_B^* is the moment of household end, under $\mathbf{Z}^*$, which is the moment of death of the household member who lives longer in this scenario: $T_B^* = \max\left\{D_1^*, D_2^*\right\}$, $DTh_t^{(\mathbf{Z}^*)}$ denotes the default threshold (its level as for the moment t, under $\mathbf{Z}^*$).

Probabilities of default are defined by the formula (2):

$$p_{\mathbf{T}_i^*} = \begin{cases} p_{\mathbf{Z}_i^*} & \text{if} \quad \underset{t=1,..,T_B^*}{\exists} CSp_t^{(\mathbf{Z}_i^*)} < DTh_t^{(\mathbf{Z}_i^*)} \\ 0 & \text{if} \quad \underset{t=1,..,T_B^*}{\forall} CSp_t^{(\mathbf{Z}_i^*)} \geq DTh_t^{(\mathbf{Z}_i^*)} \end{cases}, \tag{2}$$

where $p_{\mathbf{Z}_i^*}$ is probability of i-th scenario (i-th realization/value of the random variable **Z**).

Household default probability is a sum given in the formula (3)):

$$HDP = \sum_{i=1}^{n} p_{\mathbf{T}_i^*} . \tag{3}$$

The choice of *HDP* as the basic concept for the proposed risk measures is justified by the fact that it is relatively easy to impose constraints on it. A check of current value of the sums as defined in formula (3) may be performed in each iteration of the optimization procedure. It is possible to compare default probabilities for different plans. Thanks to this, a single default probability limit may be set for all suboptimal plans and the optimal one. For all of them, the level of default probability may be calculated and the constraint may be checked.

5 Short and long-term measures

The idea underlying short-term and long-term risk measures is that household members may have different preferences as to mitigation of risk in a short and long horizon. Having short-term and long-term measures defined, it is possible to add new constraints to the plan optimization procedure, which may allow household members to better express their preferences. Some acceptable levels of short-term and long-term risk may be set separately.

Let us now divide financial goals of the household into two groups:

- Group 1: *short term goals* –time until planned realization is shorter or equal to 1 year;
- Group 2: *long term goals* –longer time until planned realization (longer than 1 year).

Like for *HDP*, a default threshold is determined at the moment t_0 and it may change in time (e.g., different capability of financial shock absorbing in different life cycle phases).

For short term, scenarios in which cumulated shortfall intersects the default threshold in the first year are identified as short-term default scenarios – formally defined in the eq. (4):

$$\mathbf{T}^1 = \mathbf{Z}^* : CSp_1^{(\mathbf{Z}^*)} < DTh_1^{(\mathbf{Z}^*)} . \tag{4}$$

Probabilities of short-term default scenarios are defined by the formula (5):

$$p_{\mathbf{T}_i^1} = \begin{cases} p_{\mathbf{Z}_i^*} & \text{if} \quad CSp_1^{(\mathbf{Z}_i^*)} < DTh_1^{(\mathbf{Z}_i^*)} \\ 0 & \text{if} \quad CSp_1^{(\mathbf{Z}_i^*)} \geq DTh_1^{(\mathbf{Z}_i^*)} \end{cases} . \tag{5}$$

Then, probabilities of short-term default scenarios are summed (see the formula (6)). In this way, probability that any of the short-term default scenario realizes is obtained:

$$HDP_1 = \sum_{i=1}^{n} p_{\mathbf{T}_i^1} . \tag{6}$$

A long-term default scenario is a scenario in which no default is encountered in the first year, but there is incurred a default during whichever of the next periods – formula (7):

$$\mathbf{T}^2 = \mathbf{Z}^* : CSp_1^{(\mathbf{Z}^*)} \geq DTh_1^{(\mathbf{Z}^*)} \wedge \underset{t=2,..,T_B^*}{\exists} CSp_t^{(\mathbf{Z}_i^*)} < DTh_t^{(\mathbf{Z}_i^*)} . \tag{7}$$

Formula (8) gives a definition of a long-term default scenario probability:

$$p_{\mathbf{T}_i^2} = \begin{cases} p_{\mathbf{Z}_i^*} & \text{if} \quad CSp_1^{(\mathbf{Z}_i^*)} \geq DTh_2^{(\mathbf{Z}_i^*)} \wedge \underset{t=2,..,T_B^*}{\exists} CSp_t^{(\mathbf{Z}_i^*)} < DTh_t^{(\mathbf{Z}_i^*)} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

and the sum (9) is the long-term household default probability.

$$HDP_2 = \sum_{i=1}^{n} p_{\mathbf{T}_i^2} \, . \tag{9}$$

6 Application of short-term and long-term risk measures in plan optimization

Following a common-sense assumption that people would rather control risk of a short-term goal realization first, and then start to think about long-term ones, and being aware of the fact that, after all, a whole-life financial plan is necessary to be able to accomplish all financial goals, we propose some general workflow patterns. These are also ways in which the proposed risk measures may be implemented in the financial plan management process. They are based on the premise that household members may prefer to be able to express their short-term risk aversion separately from long-term risk aversion.

Let us consider two approaches. The first is very similar to the one by Jajuga et al. (2015). It does not involve any risk measures into financial plan optimization procedure. The only difference is that the household uses short-term and long-term measures to evaluate the result of optimization, besides the whole-plan risk measure.

Approach 1:

1. An optimal plan is found by an automated optimization formula that does not incorporate any risk measure.
2. After optimization, three risk measures are calculated for the optimal plan: *HDP*, *HDP*1 and *HDP*2.
3. On the basis of the three risk measures, the household makes decision if the plan may be accepted or some revision of financial goals and other boundary conditions is needed (**Fig. 1**). If the household chooses the revision variant, optimization is repeated after the revision. Since the revision means rethinking and resetting financial goals, a new plan is constructed indeed. The new optimization means, thus, searching for a new plan.

Approach 2

In this approach, some limits are imposed on default probabilities and these limits are used as constraints at the stage of plan optimization (**Fig. 2**). In this way, the measures become parts of the optimization procedure. Two variants of the approach 2 are discussed beneath.

Variant A (of the Approach 2)

In the first one, a weighted average of short-term and long-term default probability is calculated, where the weighting multiplier reflects household’s preference of short-term risk mitigation versus long-term risk mitigation. The weighted sum is presented in the formula (10) and the parameter ξ denotes the preference to secure short-term goal realization. To that, the whole-plan default probability is calculated. Household members declare maximum acceptable values of the two probabilities – the weighted sum of the short-term and long-term default probabilities and the whole-plan household default probability (*HDP*). These two upper limits are used as constraints in the plan optimization procedure.

The new boundary conditions to be added are (formulas (10) and (11)):

$$\xi HDP_1 + (1-\xi) HDP_2 \le p^*, \tag{10}$$

$$HDP \le p^{**}. \tag{11}$$

Where upper limits of default probabilities (p^* and p^{**}) are declared by the household.

Variant B (of the Approach 2)

In the second variant, the household imposes two upper limits: one on the shorty-term household default probability and the second – on the long-term household default probability. The plan is optimized under the constraint that the two limits are held.

Beneath, short summary of the both variants is presented.

The new constraint to be added to the set of constraints is (formula (12)):

$$\left(HDP_1 \le p^{***}\right) \wedge \left(HDP_2 \le p^{****}\right), \tag{12}$$

where upper limits of default probabilities (p^{***} and p^{****}) are declared by the household.

It is important to point out that the Approach 1 gives different solutions than the Approach 2. In the first one, if the optimization result is rejected by the household due to high risk, financial goals are revised and optimization is performed with a new set of goals. In the second one, risk is measured within the optimization procedure and risk limits are treated as its constraints. If an optimal solution is feasible, it must be in compliance with pre-declared risk limits. Financial goal revision is, thus, unnecessary, unless no optimal solution exists.

7 Summary

All measures of risk discussed here are integrated, in the meaning that they incorporate information of all types of risk which are included in a given financial plan model. Moreover, measures proposed in section 4 allow to distinguish between short-term and long-term risk.

The proposed risk measures are based on household default probability (*HDP*), because it is relatively easy to impose a constraint on it.

The constraints on short-term and long-term risk are a way of reflecting household’s aversion against risk in a short and long perspective. Moreover, setting of these constraints incorporates integrated risk measurement directly into plan optimization procedure. There have been no solutions proposed so far that would combine integrated risk measures with plan optimization, taking, moreover, short-term and long-term risk aversion into account.

References

Ballinger, T. P., Palumbo, M. G., & Wilcox, N. T. (2003). Precautionary saving and social learning across generations: an experiment. *The Economic Journal*, *113*(490), 920–947.

Carbone, E., & Hey, J. D. (2004). The effect of unemployment on consumption: an experimental analysis. *The Economic Journal*, *114*(497 (July)), 660–683.

Carbone, E., & Infante, G. (2012). *The Effect of a Short Planning Horizon on Intertemporal Consumption Choices* (LABSI Working Paper No. 43/2012).

Jajuga, K., Feldman, Ł., Pietrzyk, R., & Rokita, P. (2015). *Integrated Risk Model in Household Life Cycle*. Publishing House of Wrocław University of Economics.

Pietrzyk, R., & Rokita, P. (2015a). Integrated measure of risk in a cumulated-surplus-based financial plan optimization model for a 2-person household. In *33rd International Conference Mathematical Methods in Economics. Conference Proceedings* (pp. 647–652). Cheb: University of West Bohemia, Plzen.

Pietrzyk, R., & Rokita, P. (2015b). On a concept of household financial plan optimization model. *Research Papers of Wrocław University of Economics*, (381), 299–319.

Pietrzyk, R., & Rokita, P. (2016a). A decent measure of risk for a household life-long financial plan – postulates and properties. In *34th International Conference Mathematical Methods in Economics. Conference Proceedings* (pp. 687–692). Liberec: Technical University of Liberec.

Pietrzyk, R., & Rokita, P. (2016b). Facilitating Household Financial Plan Optimization by Adjusting Time Range of Analysis to Life-Length Risk Aversion. In A. Wilhelm & H. A. Kestler (Eds.), *Analysis of Large and Complex Data* (pp. 357–367). Cham: Springer International Publishing.

Yaari, M. E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies*, *32*(2), 137–150.

Comparative analysis of the total factor of productivity changes for banks of Visegrad group for the period 2009-2013

Anatoliy Pilyavskyy[1], William Aaronson[2], Liběna Černohorská[3]

Abstract

The article studies the total factor of productivity (TFP) changes differences between banks of the Visegrad group (V4) ~~(~~Czech Republic, Poland, Hungary and Slovak) during the period 2009-2013 .

The analysis of TFP changes has been done to determine the productivity changes of the selected banking sector in Visegrad group countries during and after the financial crisis.

We found that ~~(~~TFP~~)~~ changes across all countries were relatively stable in 3 of the 4 observation periods. Nevertheless, there was a substantial decline in TFP in 2011-12. Examination of the trends for each of the countries showed that Hungary overly influenced the sample mean. The TFP remained stable during this period for all Poland and the Czech Republic, declined slightly for Slovakia, but declined precipitously for Hungary in 2011-12.

Keywords: *Performance of Banks, Visegrad Group, Technical Efficiency, Total Factor of Productivity Changes*

JEL Classification: G34, M12

1 Introduction

In 2007-2008 a financial crisis struck the global economy. A number of large banks in the USA and the European Union required government bailouts. In the better cases, profit declined by tens of percentage points or showed actual losses. This was not, however, the case for Czech, Slovak and Polish banks. Banks from these countries survived the financial crisis without the need for government intervention and, in most cases, even achieved distinct profit. One of the reasons for these excellent – and, in Europe, unique – results is considered to be the fact that only a few years had gone by since a costly bailout of the banks by the governments. The result of government intervention was that the banks had not been able to accumulate poor quality assets. At that time, the governments were required to spend hundreds of millions of dollars to save the largest banks. Subsequently, foreign financial groups privatized these banks. As of now, foreign

[1] Lviv University of Trade and Economics, Dept of Mathematics and Quantitative Methods. 10 Tuhan Baranovsky Str, Lviv, 79011, Ukraine, e-mail: apiliavs@mail.lviv.ua.

[2] Temple University, Dept of Health Serv Admin & Policy, 521 Ritter Annex 1301 Cecil B Moore Ave, Philadelphia, PA 19122, USA, e-mail: william.aaronson@temple.edu.

[3] Univesity of Pardubice, Institute of Economic Sciences, 95 Studentská Str., Pardubice 2, 532 10, Czech Republic, e-mail: Libena.Cernohorska@upce.cz.

entities have acquired nearly all Czech, Slovak and Poland banks. The banking sectors in these three countries have been characterized by unprecedented stability and have shown very healthy profits, despite the global financial and economic crisis of 2007 and 2008 (Teplý et al., 2010). The competitive ability of transition economies within the global financial markets became apparent.

Of the Visegrad Group (VG) banks, only the Hungarian bank sector remains unprofitable after 2010 mainly due to the implementation of a bank tax (Dec & Maiukiewicz, 2011). However, the government of Hungary tried to resolve the situation through use of public funds. In 2013 a tax on financial transactions was imposed. All of this could have led to decreasing support from local branches from foreign parent banks in Hungary (The Economist, 2013).

Economic analyses are used as a foundation for decision making by bank management. At the same time, economic analyses are used extensively by government bodies that regulate and oversee the financial markets. They are also used when adopting adequate measures for preserving the stability of the banking sector (Vodová, 2013; Cernohorska and Cernohorsky, 2014).

As in previous studies (Pilyavskyy & Matsiv, 2009, 2010) we use Data Envelopment Analysis (DEA) (Charnes et al.,1978; Banker et al., 1984**)** for estimation of the Malmquist index for analysis of productivity changes (Malmquist, 1953; Fare et al., 1991, Fare et al., 1992). We applied this method to the study of banks of the Visegrad group from the beginning of 2009 till the end of 2013.In this paper we focus methods exclusively on the derivation and application of a Malquist type index, Total Factor of Productivity (TFP) to analyse banking productivity changes. This current study narrows the methodological focus compared to the methods we recently used in studying the Visegrad banking system. In previous studies we decomposed the index in order to discover potential sources of increasing TFP. This allows us to focus on the influence of exogenous change on TFP.

There are a number of studies of bank efficiency in Visegrad countries that employ Data Envelopment Analysis (DEA). This region provides an excellent opportunity to study efficiency. Each published study provides a different perspective on bank efficiency in the Visegrad region. Due to the importance of banking sector efficiency to macroeconomic stability and strong competitive pressure in this sector, a substantial research was done to measure efficiency of banking institutions in developed countries and to benchmark them (Zimková, 2014). As to studies which cover individual banking systems in Visegrad countries, Stavárek & Řepková

(2013) analysed the Czech banking sector and its efficiency over the period of 2000 to 2009. They founded that the average efficiency in the Czech commercial banks in the period 2001 – 2010 remained nearly unchanged during the periodof estimation. Řepková (2014) applied DEA window analysis on the data of the Czech commercial banks and to examine the efficiency of the Czech banking sector during the period 2003–2012. The paper employed an extended DEA approach, specifically DEA window analysis for the efficiency assessment of commercial banks in the Czech Republic. The group of large Czech banks were lower efficient than other banks in the banking industry. The reasons of the inefficiency of the group of large banks were the excess of deposits in balance sheet and inappropriate size of operation. Wozniewska (2008) examined the efficiency of the Polish banking sector over the period of 2000 to 2007. Palečková (2015) examined the efficiency of the banking sectors in Visegrad countries during the period 2009–2013.The results show that average efficiency was slightly decreasing within the period 2010–2011. But significant decrease in efficiency in 2012, it was probably as a result of financial crisis. Then average efficiency increased in 2013. This finding confirms results of Anayiotos et al. (2010) who presented that banking efficiency decreased during the crisis period.

Stochastic Frontier Analysis (SFA) is also frequently used for the analysis of banks performance. The disadvantage of this method is that the analytic model must be exactly defined. On the other hand, parametric methods allow for random error in the estimation process, while nonparametric methods do not. There is no agreement in the literature as to which of the methods is preferable (Holod & Lewis, 2011). Casu & Molyneux (2000) compare parametric and non-parametric estimates of productivity change in European banking between 1994 and 2000. They find that the competing methodologies do not yield markedly different results in terms of identifying the main components of productivity growth.

The paper is organized as follows. In section 2 DEA method for the calculation of Malmquist index is considered. In section 3 the data and model that we made use of for the calculations are presented. In section 4 the main results of the research are discussed. In section 5 we make conclusions.

2 Method

Let us consider N banks, each of them uses n inputs for producing m outputs. Then, let $x_i \in \Re_+^n$ and $y_i \in \Re_+^m$ denote input and output vectors for the $i-\text{th}$ bank. We consider each bank in two periods of time $t=0$ and $t=1$. Then a production technology transforming inputs into outputs can be presented in the form of the following set S^t:

$$S^t = \{(x^t, y^t) \mid x^t \text{can produce } y^t\}.$$

A set of outputs $P^t(x^t)$ then is defined as:

$$P^t(x^t) = \{y^t \mid (x^t, y^t) \in S^t\}.$$

Note that the set S^t can represent a certain production technology only when it meets certain properties (for more details see Fare & Primont, 1995).

For the analysis of the productivity changes of banks we use Malmquist-type Index and the output distance function offered by Shephard, 1970.

Shephard's output distance function $D_i^t(x^t, y^t)$ (Shephard, 1970) for bank i is defined on the output set $P^t(x^t)$ as:

$$D_i^t(x^t, y^t) = \inf\left\{\theta \mid \theta > 0, \frac{y_i^t}{\theta} \in P^t(x^t)\right\}. \quad (1)$$

In practice the output distance function $D_i^t(x^t, y^t)$ for bank i can be calculated solving the following linear programming (LP) problem (Charnes et al., 1979); Banker et al., 1984),:

$$\left[D_i^t(x_i^t, y_i^t)\right]^{-1} = \max\{\varphi_i \mid -\varphi_i y_i^t + Y^t\lambda \geq 0, \lambda \geq 0, \vec{1}\lambda = 1, \lambda \geq 0\}. \quad (2)$$

Note, that solving of the LP problem (2) makes it possible to receive a value of parameter that measures bank's efficiency, if a technology is characterized by variable return to scale (VRS). But in the case it is characterized by constant return to scale (CRS), the LP problem (2) must be solved without the constraint: $\vec{1}\lambda = 1$.

If there are data about activity of a bank for two periods of time $t=0$ and $t=1$, outputs distance function for bank i in the period $t=0$ can be defined with respect to the technology of the period $t=1$:

$$D_i^1(x_i^0, y_i^0) = \inf\left\{\theta \mid \theta > 0, \frac{y_i^0}{\theta} \in P^1(x^1)\right\}.$$

Distance function $D_i^0(x_i^1, y_i^1)$ is built analogically.

Building of such functions (solving the 4 appropriate linear programming (LP) problem (2)) allows us to use Malmquist's idea (Malmquist, 1953)] for analysis of banking productivity.

In the article we use the following Malmquist-type index (Total Factor of Productivity (TFP))

$$TFP^{0,1} = \left(\frac{D^0(x^1, y^1)}{D^0(x^0, y^0)} \bullet \frac{D^1(x^1, y^1)}{D^1(x^0, y^0)} \right)^{\frac{1}{2}}, \tag{3}$$

that was suggested by Fare and colleauges (Fare et al., 1991, 1992). A value of the index (3) greater than 1 indicates increasing of productivity, a value less than 1 indicates decreasing.

3 Data and Model

3.1 Data

This study is based on annual data published in the database, Bankscope. The data extracted for use in estimating the models included net loans, total securities, fixed assets, deposits and short term funding. We included number of employees derived from annual reports of banks included in the sample. We used activity data from Czech, Slovak, Polish and Hungarian commercial banks for each year in the period 2009-2013. Our data set contained 205 observations. We chose ten banks from the Czech Republic, eight banks from Hungary, eleven from Poland and twelve from Slovak.

Several banks (four banks from the Czech Republic, fifteen banks from Hungary, seventeen from Poland and five banks from Slovak) were removed from the data set as they were not purely commercial banks. We also excluded banks that had missing data for our key variables. Once these banks were excluded, we were left with 205 observations. Therefore, we are confident that we included comparable decision making units which is a fundamental requirement of DEA.

3.2 Model

The two methods that are most frequently used are *production and intermediation* approaches. Using production approach banks are considered to be “producers” of services for debtors and investors. For the first time this approach was suggested in (Benston, 1965). In intermediation approach banks are considered financial intermediaries between debtors and investors. This approach was used in one of the early research studies of bank efficiency (Colwell & Davis,1992).

Specification of inputs and outputs is one of the major problems for measurement of productivity changes. To determine inputs and outputs, we made use of an assets approach (Sealey & Lindley, 1977) that treats banks as classical intermediators between depositors and borrowers.

We have determined that three inputs (personnel, physical capital, purchased funds) and two outputs: net loans, total securities best model efficiency. All of the financial data are presented in USD. Physical capital can be measured by the book value of fixed assets. Purchased funds consist of loanable funds that include all the kinds of bank deposits and short term funding and securities emitted by bank. Net loans of a bank contain all the kinds of loans (either for legal entities or individuals) reduced on the sum of reserves. Total securities consist of public and private funds in other banks. Descriptive statistics of inputs and outputs is given in Table 1.

	Net Loans (th USD)	**Total Securities (th USD)**	**Fixed Assets (th USD)**	**Deposits & Short term funding (th USD)**	**Number of Employees**
Mean	7535614	3308057	149959.7	10095488	4057.532
Median	4617496	1255396	57590	5367297	2302
Standard Deviation	8903289	4682699	214246	11793878	5732.452
Minimum	36875	3579	42	89761	14
Maximum	46417473	23172151	992808	48691992	32811

Table 1. Descriptive statistics of inputs and outputs.

4 Results of the research

The TFP across all countries (see Table 2, the last column) was relatively stable in 3 of the 4 observation periods. However, there was a substantial decline in TFP in 2011-12 to 0.954). Examination of the trends for each of the countries showed that Hungary overly influenced the sample mean (see Table 2 for Hu). The TFP remained stable at about 1.000 during this period for all Poland and Czech Republic, declined slightly for Slovakia, but declined precipitously for Hungary in 2011-12 to 0.751. Let us notice also that the index of TFP change for Hungarian banks from 2012 to 2013 was the highest (1.212).

In order to understand this trend we examined the underlying variables to see if there was a root cause of the decline. Input variable trended similarly across the four countries. At the same time the output variable “net loans” increased for Polish, Slovak and Czech banks throughout the

4 years (see Table 3). However, the value for Hungarian banks declined each year (from 5151004.38 th. USD in 23009 to 3313570.50 th. USD in 2013). The value of "total securities" grew for Slovak and Czech banks, but declined for Hungarian banks between 2010 and 2012 (from 1835760.00 th. UDS in 2010 to 1479625.63 in 2011) (see Table 4).

We then asked why the TFP change would be anomalous for Hungary. Hungary had been a trailblazer among the Visegrad group resulting from decades of experience with economic reform beginning in the 1960s (Valentinyi, 2012). However, due to growing indebtedness Hungary's economic position was on the decline when the European banking crisis of 2008 hit. In 2010 a center-right party (Fidesz) was elected in Hungary (Than, 2012). The new government instituted important economic reforms that precipitated a financial crisis in 2011-12 (Valentinyi, 2012; Simon, 2012). The general economic decline coupled with the financial crisis in 2011-12 can be seen clearly in the declining value of the output variables compared to the other three countries studied. This largely explains the anomalous results we observed in terms of the TFP change in 2012-2013.

Years	Cz	Hu	Po	Sk	Total
2009/2010	0.994	1.095	1.016	1.191	1.074
2010/2011	1.012	1.038	1.081	1.058	1.049
2011/2012	1.036	**0.751**	0.986	1.016	**0.954**
2012/2013	1.049	1.212	1.012	1.017	1.059

Table 2. Comparative Total factor of Productivity Chang.

Years	Cz	Hu	Po	Sk
2009	8927011.40	5151004.38	11245522.82	2953844.17
2010	9448379.10	4704462.38	12440336.82	3157286.50
2011	9377627.20	4264459.88	12950572.64	3221320.42
2012	9572748.10	3597873.88	12802936.45	3429124.83
2013	10326889.30	3313570.50	14480369.91	3647279.42

Table 3. Net loans for countries during the study period.

Years	Cz	Hu	Po	Sk
2009	4539182.80	1673907.75	4146345.18	1206250.17
2010	5619595.00	1995547.13	4374193.55	1455116.25
2011	5853310.10	1835760.00	4646055.55	1455812.92
2012	5742468.20	1479625.63	4334709.55	1417588.58
2013	6150271.40	1901441.13	4537195.36	1597707.25

Table 4. Total Securities for countries during the study period.

Conclusions

The use of TFP to understand differences in productivity is proving to be an invaluable tool that can benefit banking policy and the actions of banks in response to changes in government policy. The Visegrad banking system presented a unique opportunity to demonstrate the value of TFP. Due to a set of circumstances that immunized Visegrad banks from a global banking crisis, we were able to study the impact on each of the Visegrad country's banking policies on productivity in the banking sector.

This type of analysis can be applied to study of productivity in banking systems in other countries. This study clearly shows the value of TFP in cross-national as well as intra-national studies of banking productivity and efficiency.

Acknowledgements

The paper has been created with the financial support of GACR No. 17-02509S, "Emerging financial risks during a global low interest rate environment".

References

Anayiotos, G., Toroyan, H., & Vamvakidis, A. (2010). The efficiency of emerging Europe's banking sector before and after the recent economic crisis. *Financial Theory and Practice*, *34*(3), 247-267. Bankscope database (2015). *Bureau van Dijk.*

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

Cernohorska, L., & Cernohorsky, J. (2014). The Criticism of Basel II Based on Credit Risk Modelling. In *SGEM2014 Conference on Political Sciences, Law, Finance, Economics and Tourism,* 721-728.

Charnes, A., Cooper, W., & Rhodes E. (1979). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, *3*(4), 339.

Colwell, R. J., & Davis, E. P. (1992). Output and productivity in banking. *The Scandinavian Journal of Economics*, S111-S129.

Dec, P., & Masiukiewicz, P. (2011). Bank Tax in the European Union. *Economics and Applied Informatics. 17*(2).

Fare R., Grosskopf S., Lindgren B., & Roos P. (1992). Productivity Changes in Swedish Pharmacies 1980-1989: A Non-Parametric Approach. *Journal of Productivity Analysis 3*, 85-101.

Fare R., & Primont D. (1995). Multi-Output Production and Duality: Theory and Application. *Kluwer Academic Publishers.*

Holod, D., & Lewis, H. F. (2011). Resolving the deposit dilemma: A new DEA bank efficiency model. *Journal of Banking & Finance*, *35*(11), 2801-2810.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadistica y de Investigacion Operativa*, *4*(2), 209-242.

Palečková, I. (2015). Banking Efficiency in Visegrad Countries: A Dynamic Data Envelopment Analysis. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *63*(6), 2085-2091.

Pilyavskyy, A., & Matsiv, Y. (2009). Performance of Banks in Ukraine (2005–2008). *Zeszyty Naukowe "Współczesne Problemy Zarządzania*, 7-20.

Pilyavskyy, A., & Matsiv, Y. (2010). An Analysis of the Efficiency and Productivity of Ukrainian Banks. *Studia i Prace Uniwersytetu Ekonomicznego w Krakowie*, (11), 91-106.

Řepková, I. (2014). Efficiency of the Czech banking sector employing the DEA window analysis approach. *Procedia Economics and Finance, 12*, 587-596.

Sealey, C. & Lindley, J. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance*, *32*(4), 1251-1266.

Shephard, R. (1970). Cost and Production Functions. *Princeton University Press, Princeton.*

Simon, Z. (2012) Hungarian Economy Slips Into Second Recession in 4 Years. http://www.bloomberg.com/news/articles/2012-08-14/hungarian-economy-slips-into-second-recession-in-4-years

Stavarek, D. (2003). Banking efficiency in Visegrad countries before joining the European Union. *Available at SSRN 671664*.

Stavárek, D., & Řepková, I. (2013). Efficiency in the Czech banking industry: A non-parametric approach. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *60*(2), 357-366.

Teplý, P., Stárová, H., & Černohorský, J. (2010). Value creation of European bank mergers and acquistitions in the 1998 – 2007 period. *Ekonomický časopis. 58*(5). 458-470.

Valentinyi, A. (2012) The Hungarian crisis. http://voxeu.org/article/hungarian-crisis. The material is based on "The Hungarian Crisis" in *The EEAG Report on the European Economy, CESifo, Munich 2012, pp. 115-130.*

Vodová, P. (2013). Liquid assets in banking what matters in the Visegrad countries. *E+M Economics and Management*, *16*(3), 113-129.

Wozniewska, G. 2008. Methods of measuring the efficiency of commercial banks: an example of Polish banks. *Ekonomika*, 84, 81-91.

Zimková, E. (2014). Window analysis of technical efficiency: case of the slovak banking system. In *17th Applications of Mathematics and Statistics in Economics, International Scientific Conference*, 294-300.

On comparing populations based on two sets of variables

Dominika Polko-Zając[1]

Abstract

In an economic and social studies it is often necessary to test the differences between the two sets of variables. Multidimensional comparisons allow researchers to a thorough analysis of the studied phenomenon. The article concerned the problem of comparing multidimensional populations using canonical correlation analysis. In order to identify differences between the analysed sets of variables permutation tests were used. These tests do not require additional assumptions about the form of the distribution in the population; are suitable for small sample sizes and are robust to outliers. The properties of these tests were characterized using a computer simulation in R program.

Keywords: *multidimensional data, canonical correlation, permutation tests, Monte Carlo study*

JEL Classification: C12, C15, C30

1 Introduction – testing differences between two correlation coefficients

Significance tests that verify equality of correlation coefficients are important in population studies. From two populations where tested variables have two–dimensional normal distributions of unknown correlation coefficients ρ_1 and ρ_2 samples of n_i elements $(n_i > 4)$ for $i = 1,2$ are taken. The null hypothesis which says that the correlation coefficients in both compared independent populations are equal

$$H_0 : \rho_1 = \rho_2 \tag{1}$$

against alternative hypothesis

$$H_1 : \rho_1 \neq \rho_2, \tag{2}$$

can be verified using test statistic in form of (Domański, 1990):

$$Z = (z_1 - z_2)\sqrt{\frac{(n_1 - 3)(n_2 - 3)}{n_1 + n_2 - 6}}, \tag{3}$$

where: $z_i = \frac{1}{2}\ln\frac{1 + r_i}{1 - r_i}$, for $i = 1,2$ and r_i are sample canonical correlation coefficient.

The statistic has asymptotic distribution $N(0,1)$ when the null hypothesis is assumed to be true. Hypothesis H_0 is rejected in favour of H_1 if $|Z| > z_\alpha$.

[1] Corresponding author: University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, Bogucicka 14, 40-226 Katowice, Poland, e-mail: dpolko@gmail.com.

2 Correlation coefficients in multidimensional analysis

Multidimensional methods are significant part of statistical methods. This stems mainly from the fact that in many areas of empirical research they relate to phenomena of complex, multidimensional structure. The Pearson's correlation coefficient measures relation between only two random variables *Y* and *X*

$$(Y, X) \longrightarrow \rho = \frac{Cov(Y, X)}{\sqrt{Var(Y)Var(X)}} \tag{4}$$

and is a quantity ranging from -1 to 1.

The multiple correlation coefficient measures, in turn, the relationship between a variable *Y*, and sets of *q* variables **X**=($X_1, X_2, \ldots, X_q$). In fact this is the maximal correlation coefficient between a variable *Y* and a linear combination of variables *X*

$$(Y, \mathbf{X}) \longrightarrow \rho = \max_{\mathbf{B}} \frac{Cov(Y, \mathbf{B}^{\mathbf{T}}\mathbf{X})}{\sqrt{Var(Y)Var(\mathbf{B}^{\mathbf{T}}\mathbf{X})}} \tag{5}$$

where: **B**=($B_1, B_2, \ldots, B_q$) and *n* – the number of observations of each variable.

The multiple correlation coefficient has values from the interval $\langle 0,1 \rangle$.

The canonical analysis is a generalization of the concept of the multiple correlation to the case of correlation between the two sets of random variables. The objective of the canonical analysis is to test the strength of the relationship between two sets of variables **Y**=($Y_1, Y_2, \ldots, Y_p$) and **X**=($X_1, X_2, \ldots, X_q$)

$$(\mathbf{Y}, \mathbf{X}) \longrightarrow \rho = \max_{\mathbf{A},\mathbf{B}} \frac{Cov(\mathbf{A}^{\mathbf{T}}\mathbf{Y}, \mathbf{B}^{\mathbf{T}}\mathbf{X})}{\sqrt{Var(\mathbf{A}^{\mathbf{T}}\mathbf{Y})Var(\mathbf{B}^{\mathbf{T}}\mathbf{X})}} \tag{6}$$

where: **A**=($A_1, A_2, \ldots, A_p$), **B**=($B_1, B_2, \ldots, B_q$), *n* – the number of observations of each variable.

The canonical correlation coefficient is a quantity ranging from 0 to 1.

3 Canonical correlation analysis

The basic concept of the canonical analysis was developed by Hotelling (1936). Hotelling introduced the idea of canonical variables and canonical correlation and studied the relationship between the sets of variables on the skills of fast comprehension reading, and fast executing arithmetic calculations.

In the canonical correlation analysis both the independent variable and the dependent variable are multidimensional. The population with *p*+*q*–dimensional characteristic is considered to study the dependency between *p*–coordinates of the characteristic which are considered as *p*–dimensional dependent variable and remaining *q*–coordinates of the characteristic which are considered as *q*–dimensional independent variable (Kosiorowski, 2008). The sample of *n* vectors of observations of the tested characteristic can be specified as follows

$$(\mathbf{Y_i}, \mathbf{X_i})^T = (Y_{i1}, Y_{i2}, \ldots, Y_{ip}, X_{i1}, X_{i2}, \ldots, X_{iq})^T, \ i = 1,2,\ldots,n. \tag{7}$$

The sample covariance matrix of $p+q$ variables is in the form of

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{8}$$

where: $\mathbf{S}_{11}$ – covariance matrix of *Y*'s of dimension $p \times p$,

$\mathbf{S_{12}} = \mathbf{S_{21}^T}$ – covariance matrix of *Y*'s and *X*'s of dimension $p \times q$,

$\mathbf{S}_{22}$ – covariance matrix of *X*'s of dimension $q \times q$.

The objective of the canonical analysis is to test the strength of the association. Coefficient vectors **A** and **B** are pursued such that linear combinations of dependent variables $\mathbf{U=A^TY}$ and linear combinations of independent variables $\mathbf{V=B^TX}$, called canonical variables (canonical variates) are maximally correlated. These combinations give an insight into the relationship between the two sets of variables.

In the first stage of the canonical analysis coefficient vectors of the first pair of canonical variables are sought. Coefficient vectors are selected to maximize the correlation between the first pair of canonical variables, that is, maximize the expression

$$r_1 = r_{u_1,v_1} = \frac{\left(\mathbf{A_1^T S_{12} B_1}\right)}{\left[\left(\mathbf{A_1^T S_{11} A_1}\right)\left(\mathbf{B_1^T S_{22} B_1}\right)\right]^{\frac{1}{2}}} \tag{9}$$

where r_{u_1,v_1} stands for canonical correlation coefficient.

Eigenvalues are obtained from equations

$$\left|\mathbf{S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}} - \lambda \mathbf{I}\right| = 0$$

$$\left|\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}} - \lambda \mathbf{I}\right| = 0.$$

Number of non–zero eigenvalues of equations is denoted as $k = \min(p,q)$. Once eigenvalue with greatest λ_1 value is found we search for coefficient vectors for the first pair of canonical variables. They are determined by solving the characteristic equations

$$\left(\mathbf{S}_{11}^{-1}\mathbf{S_{12} S_{22}^{-1}}\mathbf{S}_{21} - \lambda_1 \mathbf{I}\right)\mathbf{A}_1 = 0$$

$$\left(\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1}}\mathbf{S}_{12} - \lambda_1 \mathbf{I}\right)\mathbf{B}_1 = 0$$

where:

$\lambda_1 = r_1^2$ – eigenvalue of matrix $\mathbf{S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}}$ or $\mathbf{S_{22}^{-1} S_{21} S_{11}^{-1} S_{12}}$.

When we obtain the first canonical correlation vectors for *Y* and *X*; further canonical correlation vectors can be found in the uncorrelated directions to the previous ones in the same manner. A few pairs of canonical variates are expected to represent the original sets of variables to explain their relation and variabilities.

In general there are *k* canonical correlations $r_1, r_2, \ldots, r_k$ corresponding to the $k = \min(p,q)$ pairs of canonical variates $u_i = \mathbf{A_i^T Y}$ and $v_i = \mathbf{B_i^T X}$:

$$\begin{array}{llllll} r_1 & u_1 & = & \mathbf{A_1^T Y} & v_1 & = \mathbf{B_1^T X} \\ r_2 & u_2 & = & \mathbf{A_2^T Y} & v_2 & = \mathbf{B_2^T X} \\ \vdots & \vdots & & & \vdots & \\ r_k & u_k & = & \mathbf{A_k^T Y} & v_k & = \mathbf{B_k^T X} \end{array}$$

for each *i*=1,2,…,*k*, r_i is the sample correlation between u_i and v_i; that is $r_i = r_{u_i, v_1}$.

The pairs (u_i,v_i), *i*=1,2,…,*k*, provide the *k* dimensions of the relationship.

Canonical correlation analysis is a useful technique for simplifying the correlation structure between two sets of variables (Yamada and Sugiyama, 2006). The most important assumptions of classical canonical analysis are:

- multidimensional normality of distribution of variables in the population,
- suitable size of the sample, at least 20 times larger than number of variables,
- the lack of collinear variables,
- the lack of outliers.

4 Testing differences based on canonical correlation coefficients

Comparing populations based on multi–dimensional sets of variables the null hypothesis can be stated as follows: "in the underlying populations all corresponding canonical correlation coefficients are equal". The alternative hypothesis is formulated: "in the underlying populations corresponding canonical correlation coefficients are not equal for at least one pair". To test the null hypothesis the permutation test can be used. Formally these hypotheses can be written as follows

$$H_0 : \begin{pmatrix} \rho_{11} \\ \rho_{21} \\ \cdots \\ \rho_{m1} \end{pmatrix} = \begin{pmatrix} \rho_{12} \\ \rho_{22} \\ \cdots \\ \rho_{m2} \end{pmatrix} \tag{10}$$

and the alternative

$$H_1 : \rho_{i1} \neq \rho_{j2} \tag{11}$$

for some *i*, *j* = 1, 2, …, *m*, where *i*=*j*.

It is more interesting for researcher to employ directional alternative hypotheses. The one–sided alternative hypothesis can be stated as follows

$$H_1 : \rho_{i1} > \rho_{j2} \ \ or \ \ H_1 : \rho_{i1} < \rho_{j2} \ . \tag{12}$$

The most important is to test the significance of the difference between the first canonical correlations, which determines to the greatest extent the relationship between the two sets of variables. To test null hypothesis against alternative hypothesis we can use the following test statistic

$$T = r_{i1} - r_{j2} \tag{13}$$

where: r_{i1}, r_{j2} (for *i, j=1*) are first, sample canonical correlation coefficients.

Tests based on permutations of observations were introduced by R.A. Fisher in 1930's (Welch, 1990). Because of the need to perform complex calculations method was widely used only in recent decades, when computing capabilities of computers increased. The basic idea behind permutation methods is to generate a reference distribution by recalculating a statistic for many permutations of the data (Ernst, 2004). Permutation tests in general take a test statistic *T* used for a parametric test, or one derived intuitively (Baker, 1995). Currently, the problem of using the permutation tests in statistical analysis is popular among researchers. The most important references are Good (1994, 2005, 2006), Pesarin (2001), Basso et al. (2009), Pesarin and Salmaso (2010) and Kończak (2016).

These tests do not require additional assumptions about the form of the distribution in the population; are suitable for small sample sizes and are robust to outliers. The goal of the test is to verify hypothesis at certain level of significance to discover a correlation between data sets. After the value of the statistic T_0 had been calculated, *N* permutations of variables were performed and values T_i ($i = 1, 2, \ldots, N$) were determined. The decision concerning a verified hypothesis is made on the basis of *ASL* (*achieved significance level*) value (Efron and Tibshirani, 1993):

$$ASL = P_{H_0}\{|T| \geq |T_0|\} . \tag{14}$$

On the basis of the random variable of large size (it is recommended in most cases the number of permutation to be greater than 1000) taken from the set of all possible permutations of the data set the *ASL* is determined using formula (Kończak, 2016)

$$A\hat{S}L = \frac{card\{i : |T_i| \geq |T_0|\}}{N} . \tag{15}$$

Sample data

Y_1	Y_2	. .	Y_p	X_1	X_2	...	X_q
y_{111}	y_{121}	...	y_{1p1}	x_{111}	x_{121}	...	x_{1q1}
y_{211}	y_{221}	...	y_{2p1}	x_{211}	x_{221}	...	x_{2q1}
...	...	...	...	...	...	...	...
y_{n11}	y_{n21}	...	y_{np1}	x_{n11}	x_{n21}	...	x_{nq1}
y_{112}	y_{122}	...	y_{1p2}	x_{112}	x_{122}	...	x_{1q2}
y_{212}	y_{222}	...	y_{2p2}	x_{212}	x_{222}	...	x_{2q2}
...	...	...	...	...	...	...	...
y_{n12}	y_{n22}	...	y_{np2}	x_{n12}	x_{n22}	...	x_{nq2}

Subsequent permutations of variables

1

Y_1	Y_2	...	Y_p	X_1	X_2	...	X_q
y_{211}	y_{221}	...	y_{2p1}	x_{211}	x_{221}	...	x_{2q1}
y_{112}	y_{122}	...	y_{1p2}	x_{112}	x_{122}	...	x_{1q2}
...	...	...	...	...	...	...	...
y_{n11}	y_{n21}	...	y_{np1}	x_{n11}	x_{n21}	...	x_{nq1}
y_{512}	y_{522}	...	y_{5p2}	x_{512}	x_{522}	...	x_{5q2}
y_{212}	y_{222}	...	y_{2p2}	x_{212}	x_{222}	...	x_{2q2}
...	...	...	...	...	...	...	...
y_{811}	y_{821}	...	y_{8p1}	x_{811}	x_{821}	...	x_{8q1}

… ...

N

Y_1	Y_2	...	Y_p	X_1	X_2	...	X_q
y_{712}	y_{722}	...	y_{7p2}	x_{712}	x_{722}	...	x_{7q2}
y_{311}	y_{321}	...	y_{3p1}	x_{311}	x_{321}	...	x_{3q1}
...	...	...	...	...	...	...	...
y_{n12}	y_{n22}	...	y_{np2}	x_{n12}	x_{n22}	...	x_{nq2}
y_{511}	y_{521}	...	y_{5p1}	x_{511}	x_{521}	...	x_{5q1}
y_{312}	y_{322}	...	y_{3p2}	x_{312}	x_{322}	...	x_{3q2}
...	...	...	...	...	...	...	...
y_{111}	y_{121}	...	y_{1p1}	x_{111}	x_{121}	...	x_{1q1}

Fig. 1. The scheme of permutation variables (where x_{ijk}, y_{ijk} are respectively the *i*-th observation of the *j*-th variable for the *k*-th population).

The smaller the value of *ASL*, the stronger evidence against H_0. Formally we choose a significance level α and reject H_0 if *ASL* is less than α .

The steps in the permutation test conducted to determine the significance of the difference between two sets of variables are as follows:

1. Assume the level of significance α ;
2. Calculate the value of the statistics T_0 for the sample data;
3. Proceed permutations of data that destroy existing structure of variables (see. Fig. 1) and calculate test statistic values T_i for these permutations.
4. Create empirical distribution of T_i, where (i = 1, 2, …, N) and locate calculated value of T_0 on this distribution and estimate *ASL* value.

The proposed test procedures do not assume underlying distributions of **Y** and **X**. The simulation study was performed using R program (R Core Team, 2016). Package CCA with function cancor() is freely available from the Comprehensive R Archive Network (CRAN, http://CRAN.R-project.org/) (González et al., 2008).

5 Empirical example

To illustrate the possibilities of application the method for the analysis of economic data, the data provided by the Central Statistical Office of Poland (GUS) and Social Diagnosis study (Diagnoza Społeczna) were used. Three variables representing a subjective evaluation of life satisfaction of respondents, and three variables determining the level of socio–economic development were determined. In the Table 1 and Table 2 observations of two periods: year 2000 and year 2015 grouped by voivodships were presented.

Voivodship	Y_1	Y_2	Y_3	X_1	X_2	X_3
Dolnośląskie	11.39	13.41	28.38	8.90	21.30	102.90
Kujawsko-Pomorskie	14.69	13.38	29.18	8.20	17.80	89.60
Lubelskie	14.62	11.38	35.53	9.20	14.10	71.40
Lubuskie	17.17	14.54	26.38	8.30	20.70	89.40
Łódzkie	13.82	11.54	26.94	9.90	16.60	88.60
Małopolskie	10.35	8.37	37.25	10.60	11.70	89.70
Mazowieckie	15.57	15.65	31.43	11.70	13.10	152.80
Opolskie	15.32	16.33	36.24	8.40	15.40	83.40
Podkarpackie	14.80	10.49	33.08	8.10	15.90	72.70
Podlaskie	17.42	11.30	26.23	10.00	15.20	73.40
Pomorskie	16.20	15.48	34.26	10.30	16.70	98.90
Śląskie	17.02	13.03	29.27	6.60	17.50	106.20
Świętokrzyskie	10.83	10.40	30.50	8.10	15.70	77.90
Warmińsko-Mazurskie	15.14	14.81	24.80	8.00	23.60	77.50
Wielkopolskie	17.23	16.83	37.13	7.90	13.60	106.80
Zachodniopomorskie	13.98	15.41	29.14	9.50	19.10	99.00

Table 1. Data determining the subjective evaluation of life satisfaction and the level of socio–economic development in 2000.

Source: Social Diagnosis and Central Statistical Office of Poland (GUS).

Voivodship	Y_1	Y_2	Y_3	X_1	X_2	X_3
Dolnośląskie	34.32	22.22	43.43	24.20	7.00	111.50
Kujawsko-Pomorskie	30.13	24.62	40.54	18.80	8.00	81.60
Lubelskie	27.56	17.00	37.79	23.70	9.30	68.60
Lubuskie	35.14	25.68	42.19	19.50	6.30	83.50
Łódzkie	28.67	21.78	42.82	23.80	7.70	93.50
Małopolskie	33.27	23.48	45.31	24.90	7.20	90.10
Mazowieckie	30.49	23.94	41.90	33.50	6.40	159.40
Opolskie	39.70	29.97	47.05	21.90	6.50	80.80
Podkarpackie	25.29	17.44	41.44	22.00	11.70	70.70
Podlaskie	25.88	19.25	41.34	24.40	6.90	71.10
Pomorskie	39.13	32.46	53.41	24.30	6.60	95.90
Śląskie	35.17	26.77	45.94	22.80	7.20	104.10
Świętokrzyskie	30.93	20.74	43.36	21.80	10.10	72.40
Warmińsko-Mazurskie	29.57	21.13	29.39	19.70	9.40	71.00
Wielkopolskie	35.73	28.08	51.05	22.90	5.80	108.80
Zachodniopomorskie	36.67	22.25	42.91	21.60	7.50	84.90

Table 2. Data determining the subjective evaluation of life satisfaction and the level of socio–economic development in 2015.

Source: Social Diagnosis and Central Statistical Office of Poland (GUS).

First set of variables contains:

Y_1 – the percentage of people satisfied or very satisfied with the financial situation of their own family (in %),

Y_2 – the percentage of people satisfied or very satisfied with the prospects for the future (in %),

Y_3 – the percentage of people satisfied or very satisfied with their education (in %).

Second set contains following variables:

X_1 – percentage of population with university education (in %),

X_2 – unemployment rate (in %),

X_3 – gross domestic product per capita (Poland takes ratio = 100; in 2015 the estimated value).

To test null hypothesis (10) against alternative hypothesis $H_1 : \rho_{11} > \rho_{12}$ the permutation test was used. Significance level α = 0.05 was assumed and N=1000 permutations of variables were performed. As test statistic (13) was used. Empirical distribution of statistic was presented on Figure 2. *ASL* value calculated with empirical distribution of statistic

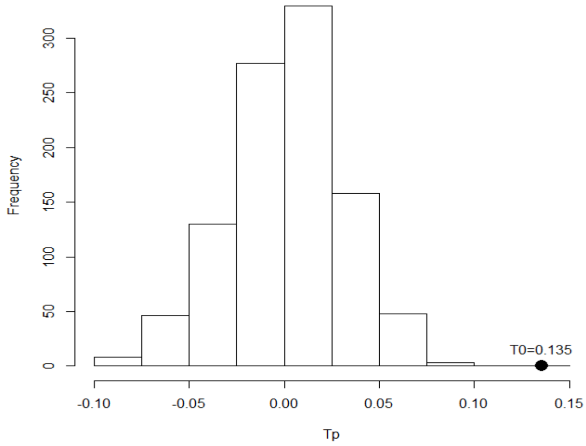


Fig. 2. Empirical distribution of statistic.

is lower than assumed significance level and amounts 0. Verified hypothesis H_0 should be rejected in favour of alternative hypothesis. There is the significance difference between the first canonical correlations so the dependency between socio-economic situation and satisfaction with life is stronger in 2000 than in 2015.

Conclusion

The limitation of commonly used classical statistical methods makes the simulation methods being used to an increasing extent in a variety of analyses for both quantitative and qualitative data. Many classical statistical methods not have their counterparts for multidimensional data. The paper presents a method for testing the differences between sets of variables. In order to identify differences between the canonical correlation coefficients permutation test was proposed. The advantage of the proposed method is that the method can be used even when the required assumptions (e.g.: on the distribution of variables in the population) are not fulfilled. The procedure using the permutation test is used to estimate the distribution of the test statistics. The proposed method is illustrated by an empirical example.

References

Baker, R. D. (1995). Two permutation tests of equality of variances. In: *Statistics and Computing*, 5, 289–296.

Basso, D., Pesarin, F., Salmaso, L., & Solari, A. (2009). *Permutation Tests for Stochastic Ordering and ANOVA*. Heidelberg: Springer Science + Business Media.

Domański, C. (1990). *Testy statystyczne*. Warszawa: Państwowe Wydawnictwo Ekonomiczne.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. In: *Statist. Sci.*, *19*(4), 676–685.

González, I., Déjean, S., Martin, P. G., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. In: *Journal of Statistical Software*, *23*(12), 1–14.

Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. New York: Springer Science Business Media.

Good, P. (2006). *Resampling Methods. A Practical Guide to Data Analysis*. Boston–Basel–Berlin: Birkhauser.

Good, P. I. (1994). *Permutation Tests: A Practical Guide for Testing Hypotheses*. New York: Springer–Verlag.

Hotelling, H. (1936). Relations between two sets of variates. In: *Biometrika*, 28, 321–377.

Kończak, G. (2016). *Testy permutacyjne. Teoria i zastosowania*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

Kosiorowski, D. (2008). *Wstęp do wielowymiarowej analizy statystycznej zjawisk ekonomicznych. Kurs z wykorzystaniem środowiska R*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Pesarin, F., & Salmaso, L. P. (2010). *Permutation Tests for Complex Data. Theory, Applications and Software*. Chichester: John Wiley & Sons, Ltd.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Welch, W. J. (1990). Construction of Permutation Tests. In: *Journal of the American Statistical Association. Theory and Methods, 85*(411), 693–698.

Yamada, T., & Sugiyama T. (2006). On the permutation test in canonical correlation analysis. In: *Computational Statistics & Data Analysis*, *50*(8), 2111–2123.

Expenditure rules in the context of a balanced budget

Agnieszka Przybylska-Mazur[1]

Abstract

Expenditure rules that are one of the types of fiscal rules are important in budgetary policies. They facilitate to maintain a stable budget in accordance with the adopted strategy in the medium and long term, they allow to coordinate the budgetary expenditure and they enable mitigate the effects of negative shocks associated with budgetary revenue. The aim of the article is the presentation of the method of determination of expenditure rules based on the optimal control model in the context of a balanced budget. In this paper we present the fiscal policy rule that is the solution of the Quadratic Linear Problem. This rule help develops the economy in accordance with the desired path and this rule minimize the deviation of inflation rate, GDP growth and unemployment rate from the desired values

Keywords: *expenditure rule, balanced budget, economic policy model, control theory*

JEL Classification: E62, C54, C61, H50

1 Introduction

Government actions that affect on the country's economic situation and its development are taken through the implementation of economic policy. The monetary and fiscal policies are important in the realization of economic policy in the short and medium term. Government making decisions about the fiscal policy may affect on the realization of sustainable economic development. Often used method of fiscal decision-making are decisions based on the rules. In this paper we determine the optimal fiscal policy rules based on optimal control theory model. This optimal fiscal rules are a solution of the economic policy model consists of the function criterion and the model of the economy. The model of fiscal policy, which we use, should lead to long-term stable economic growth. Thus, models of economic policy can be the basis for determination of the strategy which the effect is the achievement of the desired values of selected variables, such as inflation, output and unemployment in the future.

The other authors analyze the optimal fiscal policy. The optimal fiscal policy in a stochastic endogenous growth model with private and public capital is studied by Tamai (2016). Public investment, the rate of return, and optimal fiscal policy in a stochastically growing economy are analyzed by Tamai (2016). The optimal fiscal rules within a stochastic

[1] University of Economics in Katowice, Department of Mathematical and Statistical Methods in Economics, ul. 1 Maja 50, 40-287 Katowice, e-mail: agnieszka.przybylska-mazur@ue.katowice.pl.

model of Keynesian type are analyzed by Correani et al. (2014). About Taylor rule for fiscal policy write Kliem and Kriwoluzky (2014) and Kendrick and Amman (october 2011).

2 The importance of fiscal policy based on rules

The fiscal policy involves the government decisions on the size and structure of public expenditure, budget deficit and public debt. Fiscal sustainability, as an integral part of macroeconomic stability will strengthen the protection of the economy against various types of shocks. One of the ways of making decisions are the decisions based on rules. When we conduct the rules-based fiscal policy, it is strengthened the prudence and the objectivity in the realization of fiscal policy. The fiscal rules have a significant impact on the economy. One of the benefits is the creation of favorable conditions for increase of potential GDP growth. The decisions on the basis of the rules, including the expenditure rule, allow to coordinate the budgetary expenditure. These decisions allow to mitigate the effects of negative shocks related to budgetary revenue and they help maintain a stable budget in accordance with the adopted strategy in the medium and long term. If fiscal impulse is necessary for stimulation of economic business, the emergence of a budget deficit should be possible. However, it should be noted that the fiscal impulse cannot lead to a permanent increase of budget deficit and public debt.

Furthermore, it should also be noted that Article 5 of the Council Directive of the European Union (COUNCIL DIRECTIVE 2011/85/EU of 8 November 2011 on requirements for budgetary frameworks of the Member States) says that „Each Member State shall have in place numerical fiscal rules which are specific to it and which effectively promote compliance with its obligations deriving from the TFEU (ed. of the Treaty on the functioning of the European Union) in the area of budgetary policy over a multiannual horizon for the general government as a whole…" Whereas, the Article 7 of this Directive requires that „the annual budget legislation of the Member States shall reflect their country-specific numerical fiscal rules in force".

Currently in Poland it holds true the modification of stabilizing expenditure rule. Therefore, the basis for the analysis work is to determine the expenditure rule in the context of a balanced budget. The application of these rules allows the economy to develop according to the desired path.

In order to determine these rules, we applied the control theory and the selected form of dynamic model of fiscal and monetary policies.

3 Economic Policy Model

A typical economic policy model (EPM) in deterministic version consists of optimized function criterion (FC) and the model describing the economy (EM), and we can write the EPM in the general following form:

$$\text{(EPM)} \quad \begin{cases} (FC) \quad \underset{\textit{policy instrument}}{\textit{extremum}} \left\{ \sum_{\textit{time}} \text{discount factor} \left[\text{objective function} \begin{pmatrix} \text{policy objectives} \\ \text{state variables} \\ \text{policy instruments} \\ \text{target weights} \\ \dots \end{pmatrix} \right] \right\} \\ (EM) \; [\text{state variables}] = F([\text{state variables}], [\text{policy instruments}]) \end{cases} \quad (1)$$

We can consider the model of economic policy as a problem of dynamic optimization, and the solution of this problem may take the form of a formula defining the optimal relation between the instruments and state variables and policy objectives. We have the following:

$$\text{(PR)} \qquad [\text{policy instruments}] = G([\text{policy objectives}], [\text{state variables}]) \qquad (2)$$

The important advantage of the discrete dynamic programming method is the possibility of its application to the determination of optimal control processes in which we take into account a random shock.

Below presented optimization problem of economic policy, which we will use to determine the expenditure, is the quadratic linear problem. In this problem the criterion function is a quadratic function and as a constraint we take into account linear dynamic model including the additive shock (Kendrick, 2005).

The Quadratic Linear Problem can be formulated following: (Benigno & Woodford, 2012; Ellison; Przybylska-Mazur, 2016): for each $t = 1,2,...,T$ we determine the control vector U_t^* for which the function defined as:

$$G(X,U) = E_t\left(\sum_{t=0}^{T-1} \left(\left(X_t - X_t^o\right)^T V_t \left(X_t - X_t^o\right) + \left(U_t - U_t^0\right)^T Z_t \left(U_t - U_t^o\right)\right)\right) \qquad (3)$$

reaches a minimum with the constraint that is the linear dynamic model with additive shock. This model can be written in matrix form as follows (Kendrick and Amman, October 2011):

$$X_t = A \cdot X_{t-1} + B \cdot U_t + C \cdot \varepsilon_t \text{ for all } t = 1,2,...,T \qquad (4)$$

with the initial condition

$$X_0 = \tilde{X}_0 \qquad (5)$$

and with the restriction on an acceptable range of coordinates of control $U_{it} \in D_i$ for $t = 0,1,2,...,T-1$ and for each i,

where: ε_t - random shock , $\varepsilon_t \sim IID(0, I)$, T - planning horizon, X_t - vector of state variables at time *t*, U_t - control vector at time *t*, X_t^o - vector of desired values of the state variables at time *t*, U_t^o - vector of desired control values at time *t*, $\tilde{X}_0$ - given initial value of state vector, A - matrix of state vector coefficients, B - matrix of control vector coefficients, B is multiplier matrix of impact of control variables, C - variance and covariance matrix of random shocks ε_{t+1}. V_t are symmetric positive definite matrices of penalties of deviations of state variables from the desired values of state variables and Z_t are symmetric positive definite matrices of penalties of deviations of control variables from the desired values. By D we denote the following vector $D = \begin{bmatrix} D_1 & D_2 & \dots & D_p \end{bmatrix}^T$, p is the number of instruments of fiscal policy.

About the selected methods of choice of the optimal monetary policy transmission horizon write Przybylska-Mazur (2013).

For determination of the optimal solution of economic policy model we define the Bellman function for the end section of the discrete trajectory of state in the following form (Bellman, 2010, Bellman, Dreyfus, 2015):

$$S(X_t) = \min_{\substack{U_k \in D \\ k=t,t+1,\dots,T-1}} \sum_{k=t}^{T-1} \left(\left(X_k - X_k^o\right)^T V_k \left(X_k - X_k^o\right) + \left(U_k - U_k^o\right)^T Z_k \left(U_k - U_k^o\right) \right) \quad (6)$$

Since the relationship is true:

$$S(X_t) = \min_{U_t \in D} \left(\left(X_t - X_t^o\right)^T V_t \left(X_t - X_t^o\right) + \left(U_t - U_t^o\right)^T Z_t \left(U_t - U_t^o\right) + S(X_{t+1}) \right) \quad (7)$$

then recursive Bellman equation for discrete process control is of the following form

$$S(X_t) = \min_{U_t \in D} \left(\begin{array}{r} \left(X_t - X_t^o\right)^T V_t \left(X_t - X_t^o\right) + \left(U_t - U_t^o\right)^T Z_t \left(U_t - U_t^o\right) + \\ + S(A \cdot X_{t-1} + B \cdot U_t + C \cdot \varepsilon_t) \end{array} \right) \quad (8)$$

for $t = T-1, T-2, ..., 0$ with the final condition $S(X_T) = 0$.

This equation is the basis of discrete dynamic programming method, which reduces the determination of optimal controls sequence $\{U_t^*,\ t = 0,1,2,...,T-1\}$ to the determination of the individual controls U_t^* from the equation recursive Bellman.

For the determination of optimal controls U_t^* and the optimal values of the state vector X_t^* we will use the algorithm of the stochastic dynamic programming method for the optimal control problem with discrete time (the algorithm backward), which consists the following steps described below (Bar-Shalom, 1982; Kendrick, 1982):

1) We substitute $t = T-1$ and we solve the optimalization problem for the final stage of the process

$$S(X_{T-1}) = \min_{U_{T-1}\in W}\left(E_{T-1}\left(\begin{array}{l} \left(X_{T-1} - X_{T-1}^o\right)^T V_{T-1}\left(X_{T-1} - X_{T-1}^o\right) + \\ + \left(U_{T-1} - U_{T-1}^o\right)^T Z_{T-1}\left(U_{T-1} - U_{T-1}^o\right)^T \end{array}\right)\right) =$$

$$= \min_{U_{T-1}\in W}\left(\begin{array}{l} \int\limits_{-\infty}^{+\infty} P(\varepsilon_{T-1})(\left(X_{T-1} - X_{T-1}^o\right)^T V_{T-1}\left(X_{T-1} - X_{T-1}^o\right) + \\ + \left(U_{T-1} - U_{T-1}^o\right)^T Z_{T-1}\left(U_{T-1} - U_{T-1}^o\right)^T + \\ + S_{T-1}\left(A\cdot X_{T-2} + B\cdot U_{T-1} + C\cdot \varepsilon_{T-1}\right))d(\varepsilon_{T-1}) \end{array}\right) \tag{9}$$

determining the optimal control for the last stage of the process U_{T-1}^* as the function of the initial state of this stage X_{T-1}, thus $U_{T-1}^* = U^*(X_{T-1})$.

2) Using the function Bellman $S(X_{t+1})$ at $t-$th stage of the process, we solve the optimalization problem of this stage resulting from the recursive Bellman equation:

$$S(X_t) = \min_{U_t\in W}\left(E_t\left(\begin{array}{l} \left(X_t - X_t^o\right)^T V_t\left(X_t - X_t^o\right) + \\ + \left(U_t - U_t^o\right)^T Z_t\left(U_t - U_t^o\right)^T + S(f(X_t, U_t, \varepsilon_t)) \end{array}\right)\right) =$$

$$= \min_{U_t\in W}\left(\begin{array}{l} \int\limits_{-\infty}^{+\infty} P(\varepsilon_t)(\left(X_t - X_t^o\right)^T V_t\left(X_t - X_t^o\right) + \\ + \left(U_t - U_t^o\right)^T Z_t\left(U_t - U_t^o\right)^T + S(f(X_t, U_t, \varepsilon_t))d(\varepsilon_t) \end{array}\right) \tag{10}$$

determining the optimal control of $t-$th stage of the process $U_t^* = U^*(X_t)$ as the function of the initial state X_t of this stage.

3) After reaching the initial stage for $t = 0$ we calculate the value of the optimal control $U_0^* = U^*(X_0)$ for this stage using the initial condition $X(0) = \breve{X}_0$.

4) Next we calculate the optimal sequence of controls on the basis of the relationships $U_t^* = U^*(X_t^*)$, where: $X_t^* = f(X_{t-1}^*, U_{t-1}^*, \varepsilon_{t-1})$, for $t = 1,2,...,T-1$.

Since the inflation rate, GDP growth and the unemployment rate are the main variables taken into account in the assumptions of the budget law we take into account these three variables in the article. Therefore $X_t = [\pi_t \quad Y_t \quad N_t]^T$, where π_t - inflation rate, Y_t - GDP growth and N_t - unemployment rate. Considering in analyze only fiscal policy we take into account as the control variables: W_t - budget expenditure and P_t - budget revenue. Thus $U_t = [W_t \quad P_t]^T$.

Moreover, as vectors of desired values of the state variables and control variables we take $X_t^0 = \left[\pi_t^0 \quad Y_t^0 \quad N_t^0\right]^T$ and $U_t^0 = \left[W_t^0 \quad P_t^0\right]^T$, where π_t^0 - the inflation target, Y_t^0 - the potential GDP, N_t^0- the natural rate of unemployment, W_t^o - the desired level of budget expenditure planned in the assumptions of the budget law generating a budget deficit no higher than 3% of GDP, P_t^0 - the desired level of budget revenues not less than assumed in the draft Budget Act. In addition, we assume the constant values of weight matrices $V_t = V = \begin{bmatrix} \lambda_\pi & 0 & 0 \\ 0 & \lambda_Y & 0 \\ 0 & 0 & \lambda_N \end{bmatrix}$, $Z_t = Z = \begin{bmatrix} \lambda_W & 0 \\ 0 & \lambda_P \end{bmatrix}$ for each t.

4 The empirical analysis

For the calculation of optimal values of control variables – the instruments of fiscal policy we use the annual inflation rate data (corresponding period of the previous year = 100), the GDP growth (annual data) and the annual unemployment rates (data published by Central Statistical Office, source: www.stat.gov.pl) and also the budget expenditures (annual data) and the budget revenues (annual data) (data published by the Ministry of Finance). For analysis we take into account the data for the Poland from the period 2004 to 2016. We take the planning horizon T =3. As desired values of the state variables we take: the inflation target equals to 2.5%, the potential GDP and natural unemployment rate determined on the basis of the Hodrick – Prescott filter and the general government deficit equals to 3% of GDP from the convergence criteria. However the desired values of control vector are: the budget expenditures and the budget revenues planned in the draft Budget Act. As the forecasts of potential GDP growth, and the forecasts of unemployment rate for 2017 and 2018 we take the values published in the publication "Long-term Financial Plan for the years 2016-2019". The desirable budget revenue and desirable budget expenditure for the 2017 are taken from the draft Budget Act contained in the Budget Act for 2017. As desired budget revenue in 2018 is the value calculated on the basis of average value of improving tax collection contained in the publication "Long-term Financial Plan for the years 2016-2019." We assume in the analysis for each t the following constant matrices $V_t = V = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{6} \end{bmatrix}$, because we put the greatest attention on the achievement of the highest growth and we take into account price stability that is the main objective of the strategy of direct inflation targeting. We study the expenditure rule, thus we assume the following weight matrices of control variables $Z_t = Z = \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$ for each t.

The following table shows the optimal values of the control variables and optimal values of the state variables determined on the basis of the presented economic policy model.

The application of the optimal instrument of fiscal policy – the optimal budget expenditures and optimal budget revenues in the realization of fiscal policy allows to achieve the minimum deviation of state variables from the desired value of these variable, ie inflation from the inflation target, GDP growth from potential GDP growth and the unemployment rate from natural unemployment rate.

Based on this study we can conclude that the optimal values of control variables indicate to achieve the desired values of these variables. However, the practical application of obtained optimal instruments fiscal policy will lead to significant divergence between the optimal and desired values of state variables. This shows the necessary to change the desired values of the state variables in the years 2016-2018.

Horizon *t*	**Year**	**Optimal values of**	
0	2016	control variable	
		budget expenditure	368548.53
		budget revenue	313808.53
1	2017	state variable	
		inflation rate	2.18
		GDP growth	0.80
		uemployment rate	4.47
		control variable	
		budget expenditure	384773.50
		budget revenue	325428.00
2	2018	state variable	
		inflation rate	2.72
		GDP growth	-0.96
		uemployment rate	8.90
		control variable	
		budget expenditure	390773.50
		budget revenue	328258.53
3	2019	state variable	
		inflation rate	1.15
		GDP growth	-0.31
		uemployment rate	12.40

Table 1. Optimal control values and optimal state variables.

Conclusion

In the article we determined the fiscal policy rule. This rule is the solution of the Quadratic Linear Problem. We calculated the optimal values of fiscal instrument, that are the budget expenditures and the budget revenues. If the economy is regarded as a dynamic system with control, the application of the solution of Quadratic Linear Problem will help the economy develop in accordance with the desired path. The determined optimal fiscal policy rule has the positive impact on economy because it minimizes the deviation of inflation rate, GDP growth and unemployment rate from the desired values. In the simple proposed optimal fiscal rule, the budget expenditures and budget revenues depend on the inflation rate, the GDP growth and the unemployment rate. For the analysis ex ante we must use the forecasts of desired values of control variable and the forecasts of state variables, that can be determined different methods.

When we take into account the desired values of control vector and the desired values of state vector from the draft Budget Act and from the publication "Long-term Financial Plan for the years 2016-2019”, we didn’t received a satisfactory optimal values of the state variables. But for assumed desired values of the state and for desired values of control variables these optimal values allow to achieve the minimum deviation of state variables from the desired value of these variable, ie inflation from the inflation target, GDP growth from potential GDP growth and the unemployment rate from natural unemployment rate.

References

Bar-Shalom, Y. (1982). Stochastic control for economic models. *Journal of Economic Dynamics and Control,4*, 311-313.

Bellman, R. (2010). *Dynamic programming*. Princeton, N.J: Princeton University Press**.**

Bellman, R. E., & Dreyfus, S. E. (2015). *Applied dynamic programming*. Princeton University Press.

Benigno, P., & Woodford, M. (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory, 147*(1), 1-42.

Correani, L., Dio, F., & Patrì, S. (2014). Optimal choice of fiscal policy instruments in a stochastic IS–LM model. *Mathematical Social Sciences,71*, 30-42.

DIRECTIVES COUNCIL DIRECTIVE 2011/85/EU of 8 November 2011 on requirements for budgetary frameworks of the Member States, Official Journal of the European Union L 306/41.

Ellison, M. (n.d.). *Optimal Linear Quadratic Control*. Lecture. Retrieved from http://users.ox.ac.uk/~exet2581/recursive/lqg_mat.pdf.

Lecture notes in Recursive Methods for Macroeconomics.

Kendrick, D. (1982). Stochastic Control for Economic Models. *Journal of Economic Dynamic and Control, 1*(3), 311-313.

Kendrick, D. (2005). Stochastic control for economic models: past, present and the paths ahead. *Journal of Economic Dynamics and Control,29*(1-2), 3-30.

Kendrick, D., & Amman, H. (october 2011). *A Taylor Rule for Fiscal Policy*. Utrecht, Netherlands: Utrecht School of Economics, Tjalling C. Koopmans Research Institute. Discussion Paper Series.

Kliem, M., & Kriwoluzky, A. (2014). Toward a Taylor rule for fiscal policy. *Review of Economic Dynamics,17*(2), 294-302.

Przybylska-Mazur, A. (2016). Application of selected dynamic model to the analysis of the impact balanced budget rule on the economy. In *The 10th Professor Aleksander Zelias Internetional Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings / ed. Monika Papież, Sławomir Śmiech* (pp. 139-148). Cracow: Foundation of the Cracow University of Economics**.**

Przybylska-Mazur, A. (2013). Selected methods of choice of the optimal monetary policy transmission horizon. In M. Papież & S. Śmiech (Eds.), *Proceedings of the 7th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phrnomena* (pp. 133-140). Cracow: Foundation of the Cracow University of Economics.

Tamai, T. (2016). Public investment, the rate of return, and optimal fiscal policy in a stochastically growing economy. *Journal of Macroeconomics,49*, 1-17.

Symmetry of output effects of government expenditure and government revenue in Ukraine

Victor Shevchuk[1], Roman Kopych[2]

Abstract

This paper empirically analyzes the effects of fiscal policy in Ukraine, using the vector error correction model (VECM). For quarterly data of the 2001–2016 period, we find a robust positive impact of both government expenditure and revenue upon output in Ukraine. Otherwise, the fiscal policy transmission mechanism exhibits several standard features (e.g., as an increase in government expenditure after a positive shock to government revenue or widening of the budget deficit following an interest rate hike). Our results reflect the prediction of the Mankiw–Summers model that tax cuts could be restricted under (i) strong demand for money of consumption expenditure in comparison to the investment-based demand for money combined with (ii) significant interest rate elasticity of investments. The results suggest feasibility of revenue-based austerity policies in Ukraine, as higher tax rates and better tax collection may contribute to economic growth even in the short run. The findings also imply that the real exchange rate depreciation brings about a decline in output and a symmetrical decrease in either government revenue or government expenditure. Also, there is a rather strong inverse relationship between interest rate and output.

Keywords: *fiscal policy, interest rate, real exchange rate, the Mankiw–Summers model, Ukraine*

JEL Classification: C5, E1, E6, H6

1 Introduction

As of the beginning of 2017, fiscal austerity is still among top priorities of macroeconomic stabilization policies in Ukraine. However, it is not clear whether expenditure-based or revenue-based austerity measures should be implemented in a specific Ukrainian case. IMF experts admit that the particular mix of fiscal policy measures could depend on country-specific conditions, capacities and preferences (IMF 2015). To complicate matters even more, different theoretical models often offer opposite results, with sign and magnitude of the fiscal policy effects being dependent on assumptions regarding such structural features as the existence of nominal rigidities in the economy, the elasticity of the labour supply, the interest rate elasticity of investment, the degree of openness of the economy, the exchange-rate

[1] Corresponding author: Cracow University of Technology, Institute of Economy, Sociology and Philosophy, Warszawska 24, 31-155 Kraków, Poland, e-mail: vshevchuk@pk.edu.pl.

[2] Lviv National University named by Ivan Franko, Ukraine, Universytetska 1, 79-005 Lviv, Ukraine, e-mail: kopych_r@bigmir.net.

regime, the magnitude of the wealth effects, the role played by rational expectations or the interest rate and income elasticities of money demand (De Castro and de Cos, 2008).

Despite the fact that macroeconomic theories developed since the late 1980s have de-emphasized the role of money aggregates in the transmission of monetary policy, the demand for money remains relevant for macroeconomic discussions (Duca and VanHoose, 2004). One point of interest refers to the specification of money demand. As it is assumed by Mankiw and Summers (1986), disaggregated income explains money demand better, being consistent with economic theories and empirical studies. Consequently, it is shown that lower taxes are not necessarily expansionary with respect to the aggregate demand and output, while it is still the case that the government expenditure multiplier is positive. Using a modified demand equation within the theoretical framework of IS─LM model, the authors show that tax cut can be contractionary to the U.S. economy. It is worth noting that standard Keynesian models imply contractionary effects of higher taxes and government expenditure cuts, while the models of so-called non-Keynesian effects provide with positive output responses to both type of abovementioned fiscal consolidation measures.

Although empirical studies of the U.S. economy are overwhelmingly in support of a Keynesian assumption that expansionary fiscal shocks increase output (Ahtiala and Kanto, 2002; Blanchard and Perotti, 2002; De Castro and de Cos, 2008), there are numerous empirical studies for other countries that suggest that the fiscal consolidation programs do not incur costs in terms of output losses (Adam and Bevam, 2005; Afonso, 2010; Afonso et al., 2006; Ardagna, 2004; Giavazzi and Pagano, 1995; Giudice and Turrini, 2003). However, it is common for studies of expansionary fiscal consolidations that a favorable outcome is brought about by cutting government spending rather than by increasing taxes (Alesina and Ardagna, 2010; Alesina et al., 2015). On the other hand, Giavazzi et al. (2005) find that higher taxes could stimulate private consumption.

The aim of this study is to estimate the effects of fiscal policy in Ukraine. Section 2 reviews the Mankiw─Summers model. Data and statistical methodology are presented in Section 3. Estimates of the implied VECM are interpreted in Section 4, which is followed by a conclusion.

2 Theoretical framework

Conventional econometric models relate the demand for money to the level of GDP, serving as the scale variable determining the transactions demand for money balances. Referring to portfolio and transaction models of money demand as justification for the disaggregated money equation within the familiar IS─LM framework, Mankiw and Summers (1986) demonstrate that tax cuts can constrain the aggregate demand, holding that money supply is constant. The model presents as follows:

$$Y = C(Y - T, r) + I(Y, r) + G + CA(E, Y, Y^*), \tag{1}$$

$$C_Y,\, I_Y > 0,\, C_Y,\, I_Y > 0,\, CA_E, CA_{Y^*} > 0,\, CA_Y < 0\,,$$

$$M / P = L(C, I, G, r), \quad L_C > L_I > L_G > 0,\, L_r < 0, \tag{2}$$

$$CA(E, Y, Y^*) + k(r - r^*) = 0, \tag{3}$$

where Y and Y^* are domestic and foreign output, C is consumption, I is investment, r and r^* are domestic and foreign interest rate, G and T are government expenditure and government lump-sum taxes, respectively, CA is the current account, M is the money supply, P is the price level, E is the nominal exchange rate.

Equation (1) relates the aggregate demand to private consumption, investments, government expenditure, and price and income effects on foreign trade. Both consumption and investments are proportional to income and inversely related to interest rate. Similar contracted channel is provided by the relationship between income and imports. Aggregate demand is stimulated by exchange rate depreciation and higher income abroad. In Equation (2), the money supply in real terms is balanced with the demand for money, which is an increasing function of disaggregated income and a lower interest rate. For simplicity, there is no difference between nominal and real interest rates in specifications for the goods and money markets. Equation (3) defines the balance-of-payments (BOP) equilibrium. The current account balance is equated with the net capital inflows. It is assumed that capital flows are dependent on the differential interest rate. For the case of capital immobility ($k = 0$), the BOP equilibrium is achieved solely through the relative price adjustment. Under inefficiency of the relative price mechanism, a decline in income is necessary to improve the external balance through a decrease in demand for imports.

For a flexible exchange rate regime, a comparative static analysis yields fiscal policy multipliers as follows:

$$\frac{dY}{dG} = \frac{CA_q\left[(L_G - L_I)I_r + (L_G - L_C)C_r - L_r + kL_G\right]}{\Omega}, \tag{4}$$

$$\frac{dY}{dT} = -\frac{CA_q C_Y\left[(L_C - L_I)I_r - L_r + kL_C\right]}{\Omega}, \tag{5}$$

where $\Omega = -CA_q\left[(1 - C_Y - I_Y)(L_r + L_C C_r + L_I I_r) + (C_r + I_r)(L_C C_Y + L_I I_Y) + k(L_C C_Y + L_I I_Y)\right]$.

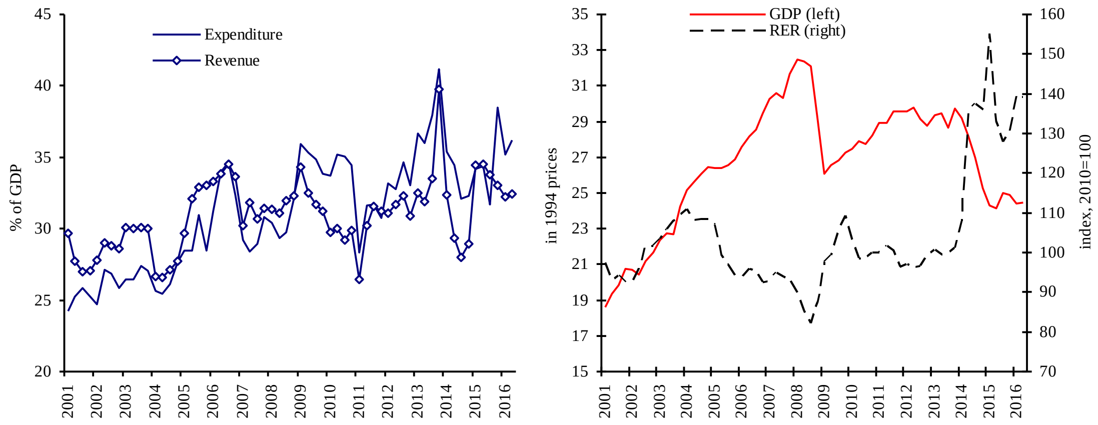
Regardless of the capital mobility, the determinant Ω is unambiguously negative under the standard assumptions that $C_Y, I_Y > 0$, $C_r, I_r < 0$, $L_r < 0$, and $C_Y + I_Y < 1$.

For a closed economy ($k = 0$), the multipliers reduce to those obtained by Mankiw and Summers (1986). The fiscal multiplier for government expenditure is positive if $L_G < (I_r L_I + C_r L_C + I_r)(I_r + C_r + k)$, as long as the government spending generates less money demand than a weighted average of consumption, investments and the capital mobility is rather low. As for the tax multiplier, higher taxes positively contribute to income only under the condition that the consumption-based demand for money is stronger in comparison to the investment-based demand for money, i.e. $L_C > L_I$, and if the money demand is sufficiently rigid with respect to the investment. However, a stimulating effect becomes not sensitive to structural features for the case of perfect capital mobility ($k = \infty$), as the tax multiplier becomes unambiguously positive: $dY/dt = L_C C_Y / (L_C C_Y + L_I I_Y)$.

As for stability of money demand being an important assumption behind viability of the Mamkiw–Summers model, the evidence for the stability of long-run demand functions for the M1 money aggregate is obtained for the U.S., Japan, Canada, U.K. and West Germany (Hoffman et al., 1995), as well as for seven East European countries (Bahmani and Kutan, 2010) and four South Asian countries (Narayan et al., 2009).

3 Data and statistical methodology

The data are quarterly observations from 2001Q1 to 2016Q2 made in the Ukraine's Ministry of Finance, which publishes quarterly time series on government finance statistics since 2000, and the IMF *International Financial Statistics* online database. Seasonally adjusted time series (in percent of GDP) for the current government expenditures on goods and services and the net revenues, G_t and REV_t respectively, are plotted in Fig. 1. Government expenditure has unevenly increased over the sample period, with local peaks in 2006, 2009, 2010 and 2013. The net revenue had been greater than the expenditure over the 2001–2007 period, but the budget balance deteriorated significantly in the wake of the world financial crisis of 2008–2009. Some fiscal consolidation efforts were made in 2011, but the budget deficit grew since then. Another financial crisis of 2014 brought about a steep decline in the level of both government expenditure and revenue, but the former recovered by the end of 2015 against the backdrop of a decline in the latter. GDP (Y_t) steadily increased in the 2001–2008 period, but financial crises of 2008–2009 and 2014–2015 brought it to the level of 2004, despite a steep depreciation of the real effective exchange rate (RER_t).



a) fiscal indicators b) GDP and real exchange rate

Fig. 1. Ukraine: selected macroeconomic indicators, 2001–2016.

Source: Ukraine's Ministry of Finance, IMF *International Financial Statistics*.

As revealed by the augmented Dickey-Fuller (ADF) test (results are available on request), for all series the null of the unit root cannot be rejected at 1 and 5 percent of statistical significance level for their levels, while it is the case for the previous differences. As endogenous variables are found to be integrated of order 1, i.e. I(1), it is necessary to investigate the co-integration relationship between them.. The results of the Johansen co-integration test are summarized in Table 1. Both the trace test and the maximum eigenvalue test suggest the co-integration rank r=1 with 5 percent confidence level.

Data trend	**None (I)**	**None (II)**	**Linear (III)**	**Linear (IV)**	**Quadratic (V)**
Trace	57.87 (0)	93.1^{**} (1)	85.6^{**} (1)	99.7^{**} (1)	89.6^{*} (1)
Max-Eng	27.89 (0)	48.1^{**} (1)	47.58^{**} (1)	49.0^{**} (1)	44.79^{**} (1)

Note: we use test types I (no intercept, no trend), II (intercept, no trend), III (intercept, no trend), IV (intercept, trend), V (intercept, trend); ** denotes rejection of the null hypothesis at the 5 percent level (* at the 10 percent level); number of co-integration vectors are in brackets.

Table 1. Johansen Co-integration Test.

As there is a co-integration of endogenous variables, the VAR system with error correction (VECM) should be used. If the endogenous variables are $I(1)$ and co-integrated with rank r ($0 < r < n$), then the VECM representation is as follows:

$$A(L)\Delta z_t = -\alpha\beta z_{t-1} + \delta D_t + u_t, \tag{6}$$

where $z_t = (REV_t, G_t, R_t, RER_t, Y_t)$ is the vector of endogenous variables, with R_t standing for lending rate, $A(L)$ is a matrix polynomial in the lag operator L, D_t is the vector of deterministic variables, u_t is a $k \times 1$ vector of reduced-form disturbances which are assumed to be normally distributed white noise $\mathrm{E}[u_t] = 0$ with a constant covariance matrix $\mathrm{E}[u_t u_t^{'}] = \Sigma_u$ and $\mathrm{E}[u_t u_s^{'}] = 0$ for $s \neq t$, Δ is the operator of the first differences. In addition to the lagged values of the endogenous variables, the VECM includes the level of external public debt (bn USD), the world metal and crude oil prices (index, 2010=100), and a crisis dummy (1 for 2008Q3-2009Q4, 2013Q4-2016Q2 and 0 otherwise).

The number of lags is set to two according to LR, FPE, AIC and HD tests. We use a constant in the VECM model, as it brings about better statistical properties of the residuals according to the tests of normality, serial correlation and homoscedasticity.

4 Estimation results

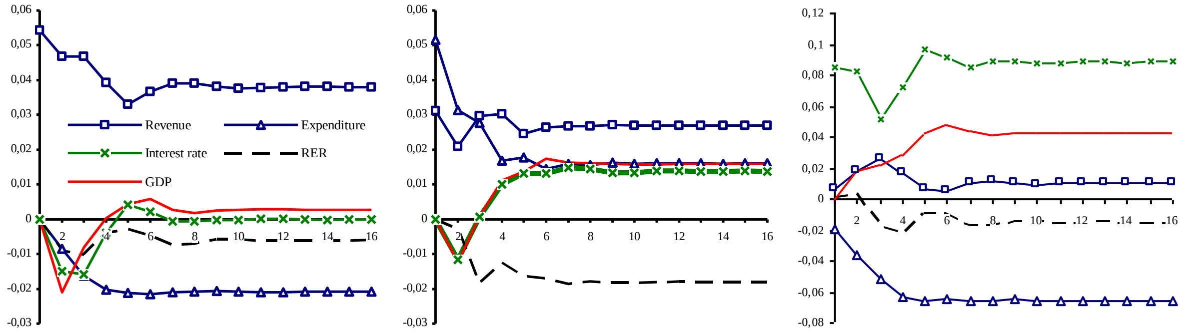
Estimates of the long-run co-integration relationships are as follows (the absolute values of standard deviations of parameter estimates are given in the brackets):

$$\begin{array}{rrrrr} REV_t = & -4.188G_t & +2.299R_t & +0.183RER_t & +7.732Y_t. \\ & (0.69) & (0.53) & (0.92) & (0.17) \end{array} \tag{7}$$

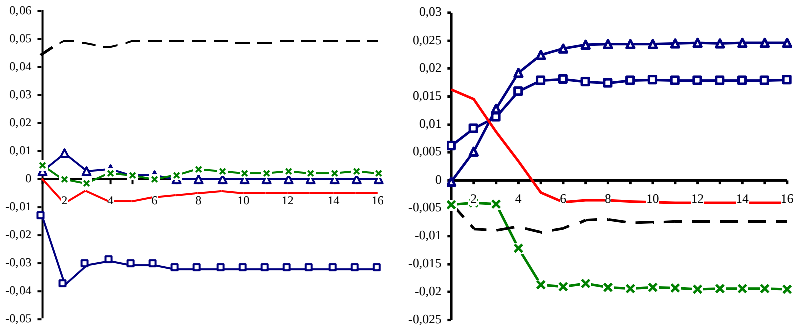
The co-integration relationship (7) implies that government revenue decreases in line with higher expenditures. A direct relationship between the interest rate and REV_t could reflect stronger tax-collection efforts in the high interest rate environment. Depreciation of the RER is not a strong factor behind higher government revenue, as the statistical significance of the coefficient on RER is rather low. The long-run estimates are in favor of a strong link between GDP and the government revenue.

Figure 2 presents the impulse-response functions for endogenous shocks. Table 2 reports the portion of the forecast error variance decomposition (FEVD) for endogenous variables.

Our main result is that both government expenditure and revenue shocks have symmetrical positive effects on output, being very persistent either. Impulse responses are consistent with the predictions of the Mankiw–Summers model. It seems that together fiscal shocks explain more than 50 percent of variation in the output. Among other fiscal policy effects, an increase in government revenue contributes to higher government expenditure and RER appreciation, with no significant impact upon the interest rate. A positive government expenditure shock brings about a reduction in government revenue and a decrease in the interest rate, both being not conventional outcomes. Shocks to REV_t explain up to 40 percent of changes in government expenditure, while the reverse causality is half that strong. The fraction of REV_t in decomposition of RER_t is as high as 29 percent, while G_t is more influential in respect to changes in the interest rate.



a) response of revenue; b) response of expenditure; c) response of interest rate;



d) response of RER; e) response of GDP;

Fig. 2. Impulse response functions of endogenous variables.

The government expenditure is likely to be pro-cyclical in the long-run, as higher output is associated with an increase in government spending on goods and services. The response of revenue to output shock is negative but low and short-lived. The effects of RER depreciation on the government revenue and expenditure are also negative.

An increase in the interest rate has no significant effect on the government revenue, while the effects on the expenditure turn out positive after three quarters. Shocks to the interest rate have negligible effects on the RER. However, there is a strong negative impact of the interest rate hikes upon output. It is worth noting that output shock is a factor behind an increase in the interest rate. Somewhat surprisingly, interest rate does not react to the RER shock.

Apart from the deterioration of the fiscal indicators, the depreciation of the RER has receding effect on the output. In the presence of the fiscal shocks, the RER does not react to changes in the output and the interest rate.

Responses of	**Innovations in**	**Forecast horizons**			
		4	**8**	**12**	**16**
REV	*REV*	82	78	77	77
	G	7	14	16	18
	R	4	3	2	2
	RER	2	2	2	2
	Y	5	3	2	2
G	*REV*	35	37	39	39
	G	53	35	30	26
	R	3	6	8	8
	RER	6	11	13	17
	Y	3	8	10	11
R	*REV*	4	2	1	1
	G	25	28	28	29
	R	64	59	57	56
	RER	3	2	2	2
	Y	5	10	11	12
RER	*REV*	27	28	29	29
	G	1	0	0	0
	R	0	0	0	0
	RER	70	70	69	69
	Y	1	1	1	1
Y	*REV*	24	24	24	24
	G	27	38	41	42
	R	10	22	25	26
	RER	12	7	6	5
	Y	27	8	5	4

Table 2. Forecast error variance decomposition.

Conclusions

The main results of the study can be summarized as follows. First, there is a robust positive impact of both the government expenditure and the revenue upon output in Ukraine. Such symmetry of fiscal policy effects is in accordance with the prediction of the

Mankiw–Summers model for a low capital mobility case that tax cuts could be restricted under (i) strong demand for money of consumption expenditure in comparison to the investment-based demand for money combined with (ii) significant inverse link between investments and interest rate. Second, there is an increase in government expenditure after a positive shock to the government revenue, with the budget deficit widening after an interest rate hike. Third, the real exchange rate depreciation brings about a decline in output and a symmetrical decrease in either government revenue or government expenditure. Fourth, there is a rather strong inverse relationship between the interest rate and the output. Contrary to recommendations by Alesina and Ardagna (2010) that spending cuts are more appropriate for stabilizing the sovereign debt than tax increases, our results suggest feasibility of revenue-based austerity policies in Ukraine, as higher tax rates and better tax collection may contribute to economic growth even in the short run.

References

Adam, C., & Bevan, D. (2005). Fiscal growth and deficits in developing countries. *Journal of Public Economics*, *89*(4), 571-597.

Afonso, A. (2010). Expansionary fiscal consolidations in Europe: new evidence. *Applied Economics Letters, 17*(2), 105-109.

Afonso, A., Nickel, C., & Rother, P. C. (2006). Fiscal consolidations in the Central and Eastern European countries. *Review of World Economics*, *142*(2), 402-421.

Ahtiala, P., & Kanto, A. (2002). Contractionary effects of expansionary fiscal policy: theory and tests. *Economic Modelling*, *19*(1), 137-152.

Alesina, A., & Ardagna, S. (2010). Large changes in fiscal policy: taxes versus spending. *Tax policy and the economy*, *24*(1), 35-68.

Alesina, A., Favero, C., & Giavazzi, F. (2015). The output effect of fiscal consolidation plans. *Journal of International Economics*, *96*, S19-S42.

Ardagna, S. (2004). Fiscal stabilizations: When do they work and why.*European Economic Review*, *48*(5), 1047-1074.

Bahmani, S., & Kutan, A. M. (2010). How stable is the demand for money in emerging economies?. *Applied Economics*, *42*(26), 3307-3318.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of economics*, *117*(4), 1329-1368.

De Castro, F., & de Cos, P. H. (2008). The economic effects of fiscal policy: the case of Spain. *Journal of Macroeconomics*, *30*(3), 1005-1028.

Duca, J. V., & VanHoose, D. D. (2004). Recent developments in understanding the demand for money. *Journal of Economics and Business*, *56*(4), 247-272.

Giavazzi, F., & Pagano, M. (1995). *Non-Keynesian effects of fiscal policy changes: international evidence and the Swedish experience* (No. w5332). Cambridge, MA: National Bureau of Economic Research.

Giavazzi, F., Jappelli, T., Pagano, M., & Benedetti, M. (2005). *Searching for non-monotonic effects of fiscal policy: new evidence* (No. w11593). Cambridge, MA: National Bureau of Economic Research.

Giudice, G., & Turrini, A. (2003). *Can fiscal consolidations be expansionary in the EU? Ex-post evidence and ex-ante analysis* (No. 195). Directorate General Economic and Financial Affairs (DG ECFIN), Brussel: European Commission.

IMF. (2015). Fiscal Policy and Long-Term Growth. *IMF Policy Paper*.

Hoffman, D. L., Rasche, R. H., & Tieslau, M. A. (1995). The stability of long-run money demand in five industrial countries. *Journal of Monetary Economics*, *35*(2), 317-339.

Mankiw, N. G., & Summers, L. H. (1986). Money demand and the effects of fiscal policies. *Journal of Money, Credit and Banking*, *18*(4), 415-429.

Narayan, P. K., Narayan, S., & Mishra, V. (2009). Estimating money demand functions for South Asian countries. *Empirical Economics*, *36*(3), 685-696.

Volatility spillovers between food, energy, US dollar, and equity markets. Evidence from Diebold-Yilmaz's approach

Sławomir Śmiech[1], Monika Papież[2], Kamil Fijorek[3]

Abstract

The causes of the surges in food prices in 2007 and 2012 are still a controversial issue. Literature offers their plausible explanations, which include e.g. financialization, depreciation of US dollars, or the tighter connection between the food market and the energy market (biofuels).

The aim of the study is to investigate volatility spillovers between food, energy, US dollar and equity markets. The analysis uses daily series of volatility of corn, soybean, wheat, rice, US dollar, crude oil, and SP500 futures covering the period from January 4, 2000 to December 30, 2016. We base our analysis on forecast-error variance decompositions in a generalized vector autoregressive framework, which are invariant to the ordering of variables, as proposed by Diebold and Yilmaz (2012). The data are studied in rolling subsamples, since the evolution of relationships is expected. Taking into account a large number of parameters in the models in single iterations, lasso estimation methods are used. The results of the study reveal that both total and directional volatility spillovers change in time. Most volatility transmissions are observed among the same category of instruments, i.e. the financial instrument group or the agricultural commodity group. Corn is the most important agricultural commodity, as it transmits vast volatility to other instruments in the food market.

Keywords: *volatility spillovers, agricultural commodity, lasso, financial markets*

JEL Classification:

1 Introduction

Volatility of food prices has a number of negative consequences. On a macro level, it influences public finances and the balance of trade for countries that are exporters or importers of food, while in developing countries it influences the level of inflation. At the micro level, it is an obstacle for food manufacturers, increasing the risk and the credit costs. Rising food prices have dramatic consequences for the poor in developing countries, as they spend a substantial part of their income (even 2/3 of disposable income) on food.

The recent food price surges (2008 and 2010-2011) covered many kinds of agricultural commodities grown in different places. That is why supply problems could not be the only reason for price co-movement. A set of factors that determine food price upsurges includes

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: smiechs@uek.krakow.pl.

[2] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: papiezm@uek.krakow.pl.

[3] Cracow University of Economics, Department of Statistics, Rakowicka 27 St., 31-510 Cracow, Poland, e-mail: fijorekk@uek.krakow.pl.

financial speculation in commodity futures markets, global economic growth (increased demand), trade restrictions, macroeconomic shocks to money supply, exchange rates (in relations to US dollars) (Abbott et al., 2009, 2011; Gilbert, 2010; Roache, 2010), competition for land (Harvey and Pilgrim, 2011), countries' aggressive stockpiling policies, and tightening relations between food prices and energy prices.

The food-energy-price nexus stems from several areas. First, modern food production requires more and more energy (to power agricultural machinery, to heat greenhouses, to power irrigation systems, for fertilizers, etc.). Second, food is also used as energy (biofuels). It the US, corn is used as the main feedstock to produce ethanol, which has another serious consequence connected with the competition for the harvested area: the area used for biofuels (corn) production increased, as fuel ethanol production grew eight-fold between 2000 and 2014 (from 233 trillion Btu in 2000 to 1 938 trillion Btu in 2014, EIA), and could not be used for other crops. Additionally, the links between oil and agricultural prices are connected with index investments. The instruments which are listed in agricultural commodity indexes are more strongly correlated with oil than the ones not listed in the indexes. The increase of the correlation between futures prices of agricultural commodities and oil after 2004, as observed by Tang and Xiong (2012), resulted from significant index investments which started to flow into commodity markets.

The aim of the paper is to analyse volatility spillovers between food prices, energy prices, the equity market and exchange rates, and to identify which of them are the main contributors to food price volatility in the last decades.

The study is based on daily series of volatility of futures price of corn, wheat, soyabean, rice, US dollar, crude oil and SP500 covering the period from January 4, 2000 to December 28, 2016. We base our analysis on forecast-error variance decompositions in a generalized vector autoregressive framework, as proposed by Diebold and Yilmaz (2012). This framework allows us to estimate total, net and directional volatility spillovers for each instrument. We also analyse forecast error variance decomposition of food prices in order to indicate the main contributors to food price volatility. The data are studied in rolling subsamples, since the evolution of relationships is expected. Taking into account a large number of parameters in the models in comparison to the number of observations, lasso estimation methods are used in single iterations.

Our study is not the first one which examines the role of financial instruments and energy prices as drivers of food prices (see: Diebold and Yilmaz, 2012; Chevallier and Ielpo, 2013; Jebabli et al., 2014; Awartani et al., 2016; Grosche and Heckelei, 2016). But we think that our approach is the most general. Two aspects differentiate our paper from other papers. First, we use a large number of instruments: the set we analyse in our study includes four main crops in the US, which helps understand the interrelations between them and different areas of financial markets. Second, we apply lasso estimation techniques, which allows us to use generous parametrization of VAR models with a relatively large number of lags. As a result, volatility spillovers could be observed in the extended horizon, which distinguishes our paper from previous studies.

2 Methodology

Diebold and Yilmaz (2009) introduce a volatility spillover measure based on forecast error variance decompositions (FEVD) from vector autoregressions (VARs). The Diebold and Yilmaz (2009) framework relies on the Cholesky-factor identification of VARs, and thus the resulting variance decompositions can be, and usually are, dependent on variable ordering. A spillover measure which is invariant to ordering should be preferred. Diebold and Yilmaz (2012) use a generalized vector autoregressive framework in which forecast-error variance decompositions are invariant to the variable ordering. This is especially important since it is rarely possible to justify one particular ordering of the variables under consideration. Additionally, Diebold and Yilmaz's (2012) approach allows for examining directional spillovers (from/to a particular market). Our study employs volatility spillover indexes as introduced by Diebold and Yilmaz (2012). The spillover indexes are constructed by performing a rolling-window forecast error variance decompositions. This methodology allows for identifying time-varying patterns. While the static FEVDs classify the variables of the study into transmitters and receivers, the dynamic FEVDs may identify episodes when the role of transmitters and receivers of spillovers is interrupted or even reversed .

During the first step, VARs are estimated using the lasso regression (see Tibshirani, 1996). The lasso is a shrinkage method for linear regression; it minimizes the sum of squared errors, with a bound on the sum of the absolute values of the individual regression coefficients. Particularly in the rolling-sample, estimated degrees of freedom are substantially limited, so the incorporation of pruning and shrinkage is appealing (Diebold and Yilmaz, 2015).

During the second step, the total and directional spillover indexes are obtained by the generalized forecast error variance decompositions of the moving average representation of the VAR model. Variance decompositions allow for parsing the forecast error variances of each variable into parts which are attributable to various system shocks. They allow for assessing the fraction of the H-step-ahead error variance in forecasting one variable caused by shocks from another variable.

It is assumed that volatility is fixed within periods (in our case, days) but variable across periods. Then, following Alizadeh et al. (2002), daily high and low prices are used to estimate daily volatility. Volatility proxy is defined as the logarithm of the difference between the future's highest and lowest log price. According to Alizadeh et al. (2002), the log range is better as a volatility proxy than log absolute or squared returns as it is more efficient and the distribution of the log range is closer to normality. This is particularly appealing as the generalized variance decomposition requires normality.

3 Data and empirical results

Data

The analysis of volatility spillovers is conducted using daily data from the period between January 4, 2000 and December 30, 2016, which yields 4225 observations. We examine the S&P 500 Index futures contract traded on the CME (SP500), the WTI crude oil futures contract traded on the NYMEX (WTI), the US Dollar Index futures contract traded on the ICE (USD), the corn, soybean, wheat and rough rice futures contract traded on the CBOT. The prices and values of index data are obtained from Bloomberg. Next, following Diebold and Yilmaz (2012), we calculate the range volatility proxy that is described in Alizadeh et al. (2002).

The table of the full-sample return spillovers

We have calculated the table of volatility spillovers with the full sample and report the results of average volatility spillovers in Table 1. Its *ij*th entry denotes the estimated contributions to the forecast error variance of market *i* coming from innovations to market *j*. Therefore, the off-diagonal column sums (labeled 'to others') or row sums (labeled 'from others'), are the "to" and "from" directional spillovers, and the "from minus to" differences are the net directional volatility spillovers. The last row in Table 1 represents the contribution to volatilities of all markets from this particular market (equity, energy, exchange rate and food). Similarly, the last column in the table represents the contribution to volatilities of the particular market from all markets. The table of return spillovers may be viewed as the "input–output" decomposition of the total return spillover index.

As Table 1 demonstrates, the contribution of the corn market to other markets equals 35.3%, the impact of the other markets on the corn market equals 31.4%, and the net spillover index of the corn market equals 3.9%. If we analyse only the food market, we find that the contribution of the corn, soybean, wheat and rice markets to other food markets equals respectively: 33.5%, 21.6%, 22.9%, and 1.7%, and the impact of other food markets on the corn soybean, wheat and rice markets equals respectively: 30.2%, 23.8%, 24.1%, and 1.6%. The net spillover index of the corn, soybean, wheat and rice markets equals respectively: 3.3%, -2.2%, -1,2% and 0.1%. It can be observed that the food market (corn, soybean, wheat) and financial instruments (SP500, USD, WTI) produce two separate clusters. The volatility transmission inside each of the clusters is significantly larger then between pairs of instruments from separate clusters. Rice is specific in this respect, as it seems to belong neither to “food” (which is surprising) nor “financial instruments” (which is natural). The different rice category can result from unique conditions required for rice production, which makes the problem of competitions for land invalid, as no other crop can be grown on the same land that is used for rice.

	SP500	**WTI**	**USD**	**Corn**	**Soybean**	**Wheat**	**Rice**	**From Others**
SP500	89.3	3.9	4.8	0.7	0.5	0.6	0.3	10.8
WTI	5.9	87.6	4.3	0.4	0.7	0.6	0.5	12.4
USD	6.0	3.8	87.8	0.7	0.8	0.4	0.5	12.2
Corn	0.4	0.3	0.5	68.6	14.5	15.3	0.4	31.4
Soybean	0.6	0.6	0.8	16.3	74.2	7.0	0.5	25.8
Wheat	0.7	0.6	0.7	16.7	6.6	73.9	0.8	26.1
Rice	0.2	0.4	0.9	0.5	0.5	0.6	96.9	3.1
To Others	13.8	9.6	12.0	35.3	23.6	24.5	3.0	x
Net spillovers	3.0	-2.8	-0.2	3.9	-2.2	-1.6	-6.1	x

Table 1. The direction of implied volatility spillovers.

Rolling test results

Next, we compute net volatility spillovers in a rolling window, as advised by Diebold and Yilmaz (2012). We use rolling subperiods of 250 days. The underlying VAR model has five lags, and the forecasting horizon is equal to 10 days. Net volatility spillovers estimated between the financial market and the food market are presented in Fig. 1. In every case there are short periods when food predominates (is the net transmitter, positive values in Fig. 1) over financial instruments, and short periods of opposite relations (food is the net receiver, negative values).

Also the volume of net volatility spillovers is usually low. All this suggests that relations between volatilities in the financial market represented by equity, exchange rate and oil prices and food market are not very strong, thus the results may not be reliable. Our results in this respect are similar to the ones obtained in studies by Diebold and Yilmaz, 2012; Chevallier and Ielpo, 2013; Jebabli et al., 2014; Awartani et al., 2016; Grosche and Heckelei, 2016.

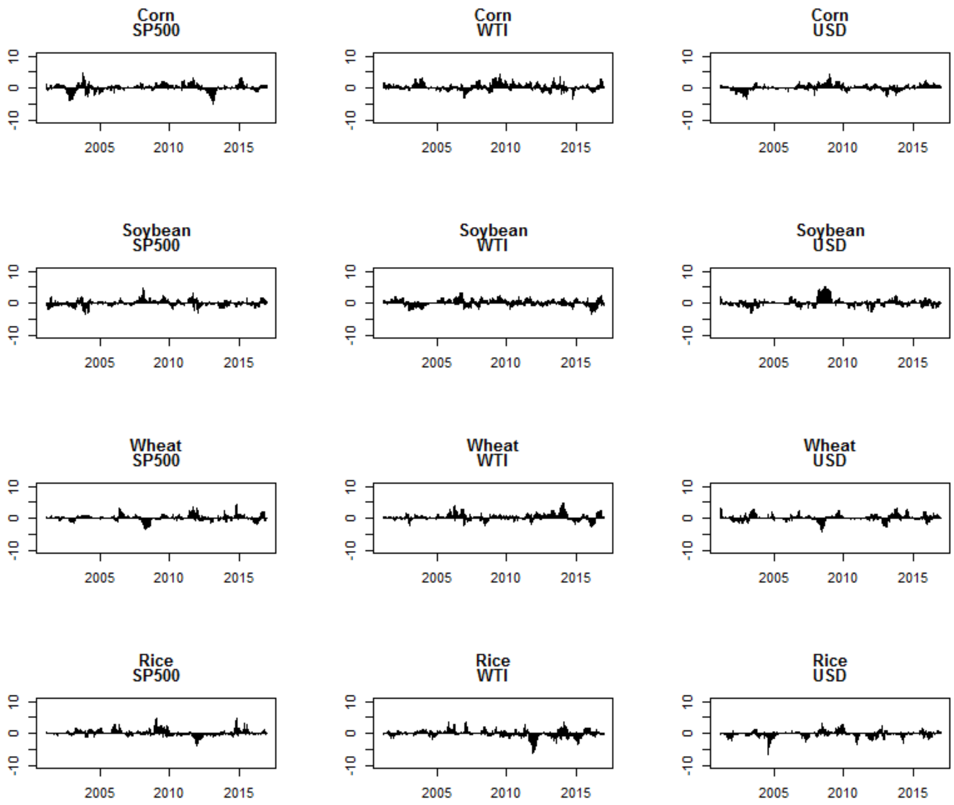


Fig. 1. Net volatility spillovers between the financial market and the food market.

Net volatility spillovers obtained for the food market are presented in Fig. 2. Corn is the net volatility transmitter to soybean and wheat. In this case net volatility spillovers are positive for most subperiods, their volume usually exceeds 1%, and in many subperiods even reaches 4 or 5%. Corn is also the volatility transmitter to rice. However, the net volatility index is smaller (less than 1%) in comparison to previously described corn-soybean relations. The patterns of the net pairwise connectedness of the net volatility index obtained for remaining pairs are not so clear. There are many subperiods in which a particular food is the net volatility transmitter, and a lot of subperiods in which the same instrument is the net volatility receiver. For example, soybean transmits more volatility to wheat in the period 2007-2011, while in the period 2011 – 2014 in most windows the relations are opposite.

A common factor in relations between soybean and rice and wheat and rice is that in windows covering 2009 (in 2009 rice prices surge to a record level, and rice transmits more volatility to both soybean and wheat then receives from them. But in the remaining superiods rice is the net volatility receiver).

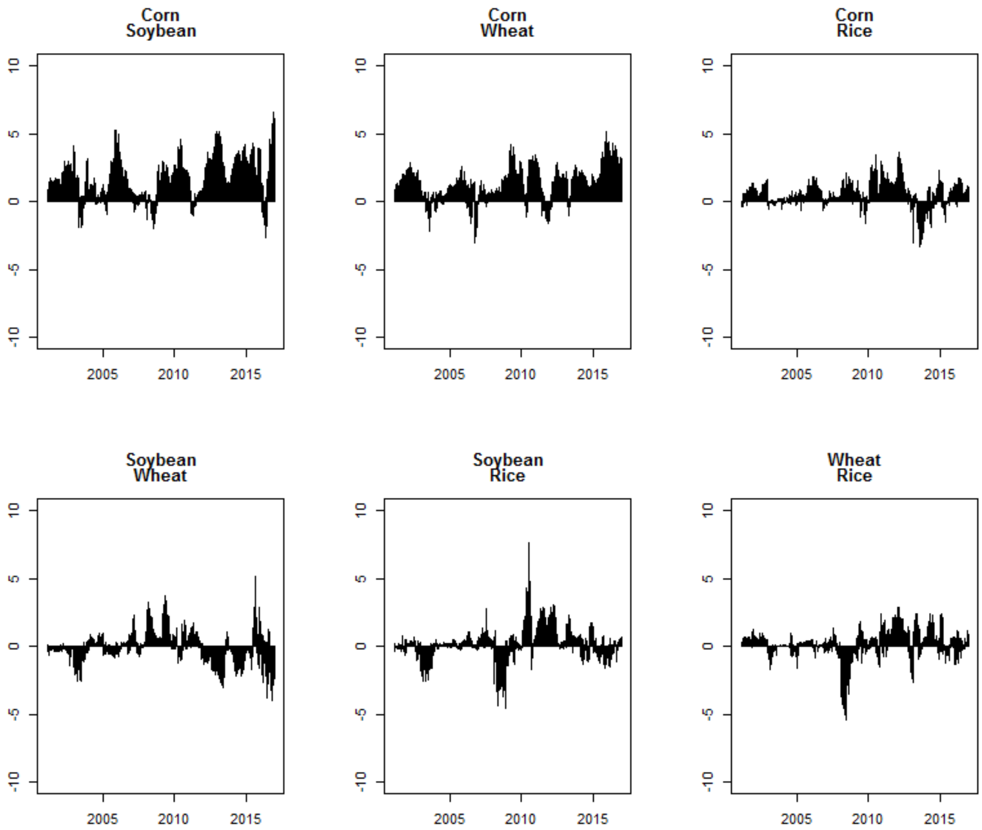


Fig. 2. Net volatility spillovers in the food market.

The results of forecast error variance decompositions obtained for each food are presented in Fig. 3. Different colours represent the share of volatility that comes from different markets. It can be noticed that for each food the greatest share of FEVD comes from its own, specific source. To be more precise, at least 50% FEVD for wheat around the year 2012 and 80% FEVD for rice result from own shocks.

In case of corn, a large proportion (about 20%, and in 2016 between 35% and 40%) of the forecast error variance (FEV) depends on volatility of soybean and wheat. What is worth noticing, soybean seems to be responsible for a similar proportion of FEV for corn in every subperiod, but the importance of wheat varies. The share of wheat in FEVD for corn varies between several percent in the subperiods covering 2005 and 2009 and about 20% between

2011-2014. The role of rice in the FEVD for corn, as well as financial instruments, is negligible.

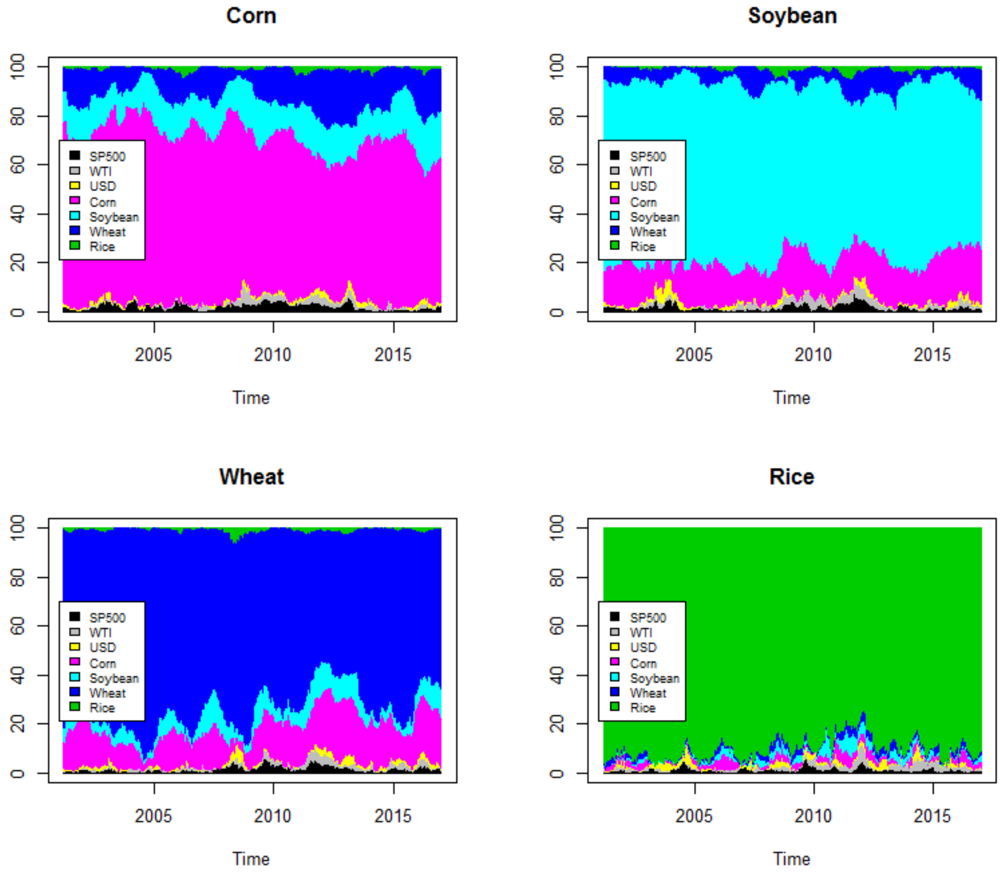


Fig. 3. The forecast error variance decompositions (FEVD) for food.

In case of FEVD for soybean and wheat, apart from the importance of their own FEV, the second most important factor is corn. The share of corn in FEV exceeds 20% in many subperiods. The third most important factor is wheat for soybean and soybean for wheat. The share of one of these agricultural commodities in FEV of the other ranges from 4 to 10%. Again, the role of financial instruments is not significant. FEVD obtained for rice demonstrates that no instrument transmits significant portions of volatility to rice. It is worth mentioning that around 2011 FEVD for rice depends on others factors in about 20%.

Conclusion

The objective of the study is to examine volatility spillovers for the food market and the financial market. Our main findings can be summarized in the following way.

Most volatility transmissions are observed among the same categories of instruments. We identify two groups which are interrelated in terms of volatility spillovers, i.e. the financial group (including equity, crude oil and US dollar exchange rates) and the food group (including corn, soybean and wheat). The food group transmits much more volatility within elements than the financial group (which can result from heterogeneity of the financial group). Rice volatility seems not to depend on other instruments and does not transmit volatility to other markets, with the exception of one episode, i.e. 2009, when rice reaches record values and significantly influences volatility of soybean and rice.

The sources of FEV for foods vary for different foods and for different supberiods. Corn, however, seems to be the most important agricultural commodity, as it transmits vast volatility to other instruments in the food group. Corn is the net volatility transmitter for soybean and wheat and is the second most important source of FEV for these two foods, representing up to 20% of FEV. On the one hand, the results demonstrate that the role of financial security in creating the food market volatility is limited. In particular, volatility of energy prices is insignificant for food prices. On the other hand, corn emerges as the most important commodity in the food market, as it is the net volatility transmitter to soybean and wheat. Since the share of corn production used for biofuels (ethanol) rises significantly during the analysed period, one can speculate that the relations between energy and agricultural commodities increase indirectly.

Acknowledgements

This paper is supported by funds from the National Science Centre (NCN, Poland) through grant No. 2012/07/B/HS4/00700.

References

Abbott, P. C., Hurt, C. A., & Tyner, W. E. (2009). *What's driving food prices?*. Farm Foundation.

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance, 57*(3), 1047-1091.

Chevallier, J., & Ielpo, F. (2013). Volatility spillovers in commodity markets. *Applied Economics Letters, 20*(13), 1211-1227.

Awartani, B., Aktham, M., & Cherif, G. (2016). The connectedness between crude oil and financial markets: Evidence from implied volatility indices. *Journal of Commodity Markets, 4*(1), 56-69.

Diebold, F. X., & Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, *119*(534), 158-171.

Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, *28*(1), 57–66.

Diebold, F. X., & Yilmaz, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press, USA.

Gilbert, C. L. (2010). How to understand high food prices. *Journal of Agricultural Economics*, *61*(2), 398-425.

Grosche, S. C., & Heckelei, T. (2016). Directional Volatility Spillovers Between Agricultural, Crude Oil, Real Estate, and Other Financial Markets. In *Food Price Volatility and Its Implications for Food Security and Policy* (pp. 183-205). Springer International Publishing.

Harvey, M., & Pilgrim, S. (2011). The new competition for land: Food, energy, and climate change. *Food Policy*, 36, S40-S51.

Jebabli, I., Arouri, M., & Teulon, F. (2014). On the effects of world stock market and oil price shocks on food prices: An empirical investigation based on TVP-VAR models with stochastic volatility. *Energy Economics*, *45*, 66-98.

Roache, S.K. (2010). What Explains the Rise in Food Price Volatility? International Monetary Fund (IMF) Working Papers 10/129. International Monetary Fund, Washington, DC

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, *68*(5), 54-74.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Separability index for cluster analysis

Andrzej Sokołowski[1], Sabina Denkowska[2], Kamil Fijorek[3]

Abstract

A new separability index for groups obtained in cluster analysis has been proposed in the paper. It is assumed that the number of groups and object assignments are given. The main idea is based on the squared Euclidean distance between each object and the closest one belonging to different group. Sum of these distances should be normalized, e.g. by within group sum of squares. The proposed measure can be use also for overlapping or fuzzy clusters.

The results of simulation studies under different models of separability are given. The proposed measure is compared to Calinski-Harabasz, Krzanowski-Lai and Rousseeuw indexes.

Keywords: *cluster analysis, cluster validity indices, simulation studies*

JEL Classification: C38, C15

1 Introduction

There are many measures of the quality of clustering (cluster validity). Migdał-Najman (2011) gave comprehensive bibliographic survey of this subject (see also: Everitt, 1995; Arabie, Hubert, De Soete, 1998; Everitt et al., 2011). Her paper has 178 references and she proposed some classifications of cluster validity measures. In this context, it can look strange that we propose another measure. The main idea is based on the squared Euclidean distance between each object and the closest one belonging to different group.

The proposed measure is compared to three popular measures: Caliński-Harabasz (CH), Krzanowski-Lai (KL) and Rousseeuw (S) indexes.

2 The new separability index AS

The new index AS is based on the squared Euclidean distance between each data point $\mathbf{x}$ and the closest one belonging to different group, $S(\mathbf{x})$. Sum of these distances SS is normalized by $WGSS$ which denotes the within-group sum of squares (it is a sum of squared Euclidean distances from each point to the center of its cluster).

[1] Corresponding author: Cracow University of Economics, Department of Statistics, 27 Rakowicka St., 31-150 Cracow, Poland, e-mail: sokolows@uek.krakow.pl.

[2] Cracow University of Economics, Department of Statistics, 27 Rakowicka St., 31-150 Cracow, Poland, e-mail: sabina.denkowska@uek.krakow.pl.

[3] Cracow University of Economics, Department of Statistics, 27 Rakowicka St., 31-150 Cracow, Poland, e-mail: kamil.fijorek@uek.krakow.pl.

Let X denotes a set of n observations (which are v-dimensional points), divided into k separated clusters C_i such that: $X = \bigcup_{i=1}^{k} C_i$. Let $\boldsymbol{d}(\cdot,\cdot)$ denotes squared Euclidean distance between two vectors.

The AS index is defined as:

$$\mathrm{AS} = \frac{SS}{WGSS}. \tag{1}$$

The SS in the numerator of index AS is the sum of $S(\mathbf{x})$:

$$SS = \sum_{\mathbf{x} \in X} S(\mathbf{x}) \tag{2}$$

where $S(\mathbf{x})$ is defined as:

$$S(\mathbf{x}) = \min_{\mathbf{y} \notin C(\mathbf{x})} \boldsymbol{d}(\mathbf{x},\mathbf{y}) \tag{3}$$

where $\mathbf{x},\mathbf{y}$ are the data points $(\mathbf{x},\mathbf{y} \in X)$ and $C(\mathbf{x})$ is a cluster to which $\mathbf{x}$ belongs to.

The overall within-cluster variance $WGSS$ is defined as:

$$WGSS = \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \boldsymbol{d}(\mathbf{x},\mathbf{m}_i) \tag{4}$$

where $\mathbf{x}$ is a data point $(\mathbf{x} \in X)$, $\mathbf{m}_i$ is the centroid of i-th cluster.

The optimal number of clusters is the solution with the highest AS index value.

3 Some other criteria for determining the number of groups in a data set

Let *k* denote the number of groups (clusters) in a data set. Assume that *n* denotes the number of cases (number of rows in a data set). Each case is described using *v* quantitative variables and each case is viewed as a *v*-dimensional point in the Euclidean space. Consequently for each pair of points a distance can be calculated.

3.1 Caliński-Harabasz criterion (1974)

Caliński and Harabasz (CH) in their 1974 paper introduced a criterion “to select the value of *k* at which the final partition appears to be best”. They call this criterion the *VRC* (*Variance Ratio Criterion*) and define it as:

$$VRC = \frac{BGSS/_{k-1}}{WGSS/_{n-k}}. \tag{5}$$

CH calculate distances between data-points using squared Euclidean metric. The *WGSS* denotes the within-group sum of squares and it is a sum of distances from each point to the center of a cluster it belongs to. The *BGSS* denotes the between-group sum of squares and it is a sum of distances, weighted by the cluster sizes, from each cluster centroid to the overall centroid.

When the clusters are well separated the *BGSS* will tend to be large and the *WGSS* small causing the *VRC* to be large. Consequently the k that gives the largest value of *VRC* is preferable. However in case of multiple local maxima the one with the smallest value of k should be chosen.

CH point out that the *VRC* is a heuristic that has no evident probabilistic foundation. However the *VRC* has some interesting properties. (1) When all points in a data set are equally distant from each other the *VRC* takes the value of one. (2) When all points are uniformly distributed in the space the relationship between k and *VRC* is monotonic and smooth. In this case CH suggest that each point should be considered its own cluster. (3) When the clear structure is present there should be a noticeable jump in the value of *VRC* when going from $k-1$ to k.

3.2 Krzanowski-Lai criterion (1985)

Krzanowski and Lai (KL) in 1985 introduced a criterion for determining the optimal number of groups in a data set. KL assume that the criterion will be used with clustering algorithms based on minimizing the within-group sum of squares.

KL showed that under the assumption that the data set consists of v independent variables distributed uniformly with equal variances the expression $k^{\frac{2}{v}} \cdot \frac{WGSS}{TSS}$ equals approximately one for any value of k. *TSS* denotes the total sum of squares. Unfortunately KL simulations showed that this result holds poorly in small samples typically encountered in practical data analysis. KL also point out that the criterion often gives multiple local optima. However they found out that despite this facts the criterion is still useful for determining the optimal number of groups.

3.3 Rousseeuw criterion (1987)

Rousseeuw in 1987 paper introduced a silhouette graph. This graph was meant to be a new useful tool aiding in the interpretation of clustering results. It was designed to differentiate the clear group structure from merely a data set partition to non-overlapping sets of points. As a side effect a criterion for determining the optimal number of groups in a data set was introduced.

For the *i*-th data point the average distance between the point and other points belonging to the same cluster is calculated and denoted $A(i)$. Next for the *i*-th data point the average distance between the point and other points belonging to some other cluster is calculated and repeated for all other clusters. The minimum of those averages is denoted $B(i)$. And finally for the *i*-th data point one gets the $S(i)$:

$$S(i) = \frac{B(i) - A(i)}{max\{A(i), B(i)\}}. \tag{6}$$

$S(i)$ takes values in range from -1 to +1. The higher the positive value the more certain it is that the point belongs to the correct cluster, since it is close to the rest of the cluster members and is far from the members of all other clusters. $S(i)$ close to zero shows uncertain cluster membership and the negative value suggests that the point is probably in the wrong cluster.

The global silhouette index (S) is calculated as an average value of all $S(i)$:

$$\mathrm{S} = \frac{1}{n}\sum_{i} S(i). \tag{7}$$

The *k* that reaches the maximum S is considered optimal. Rousseeuw points out that the advantage of the S index is that it was developed without any particular clustering algorithm in mind.

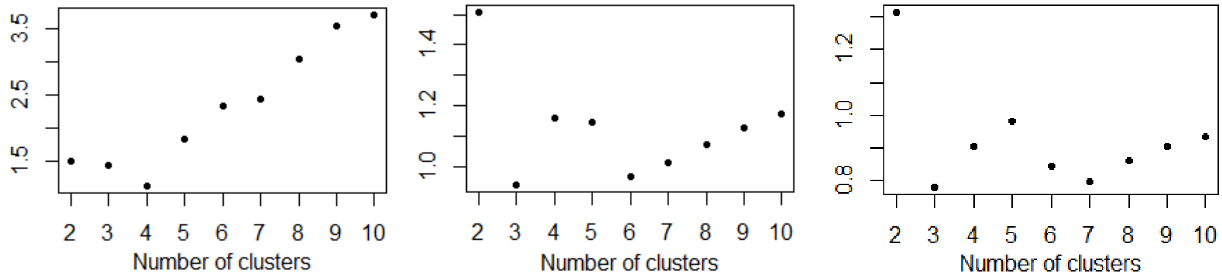
4 Simulation studies

The performance of the proposed separability index AS has been studied through simulations carried under some theoretical models. The accuracy of the number of groups specification, has been compared to three other indexes described in the previous section. Simulations were carried out in R using *clusterSim* package for Caliński-Hrabasz, Krzanowski-Lai and Rousseeuw criteria. On each run of the simulation data was clustered by *k*-medoids method (R package *cluster*).

Experiment 1

Equal samples for two groups have been generate from two two-dimensional normal distributions: $N(-d,0,1,1,0)$ and $N(d,0,1,1,0)$, taking $d = 1$ and $d = 3$. The sample sizes were $n = 20, 100, 200$.

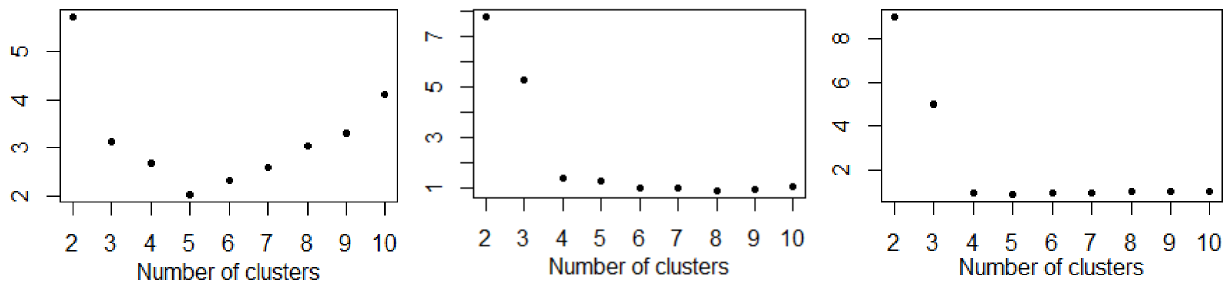
For $d = 3$ all criteria correctly identified the number of clusters, as two groups (See Fig. 1 for AS index). For small sample ($n = 20$) and $N(-1,0,1,1,0)$ and $N(1,0,1,1,0)$ model, all criteria got lost showing number of clusters bigger than 2. For large samples and d = 1 three criteria, all except Krzanowski-Lai identified two clusters. On Fig 1, values of AS index are shown against number of clusters.



a) *b*) *c*)

Fig. 1 a, b, c. AS indexes for different number of clusters with a) n=20, b) n=100, c) n=200 samples generated from N(-1,0,1,1,0) and N(1,0,1,1,0).

Source: own calculations.



a) *b*) *c*)

Fig. 2 a, b, c. AS indexes for different number of clusters with a) n=20, b) n=100, c) n=200 samples generated from N(-3,0,1,1,0) and N(3,0,1,1,0).

Source: own calculations.

With two groups of points with expected values lying in 6 standard deviations distance, the AS index definitely suggests the correct number of clusters.

Experiment 2

In this experiment samples were generated from four two-dimensional normal distribution, moving along coordinate axes. Thus expected values were equal to (d,0), (0,d), (-d,0), (0,-d). A unit variance-covariance matrix has been used. Distance d from the center of coordinate system has been taken as d = 1, 2, …, 100. We generated four samples of equal size n = 20 and n=100 objects. Partitions for 2, 3, …, 10 groups were found by k-medoids method.

All four indexes performed very well. With small samples of *n* = 20, Caliński-Harabasz and Silhuette identified 4 groups in 99% runs, Krzanowski-Lai in 97%, and AS in 96%. With large sample *n* = 100, all criteria (except Krzanowski-Lai – 98%) got the score of 99%.

Experiment 3

In this experiment we have been trying to estimate the probability of correct identification of the number of groups in case when they are rather distant to each other. Four groups have been generated with the following model: *N*(3,0,1,1,0), *N*(0,3,1,1,0), *N*(-3,0,1,1,0), *N*(0,-3,1,1,0). The probability has been estimated by 1000 simulation runs. The data has been also clustered for some "incorrect" number of clusters 2, 3, …, 10, with *k*-medoids method.

Four samples of equal size *n*	**AS**	**CH**	**KL**	**S**
30	0.907	0.995	0.664	1.000
50	0.984	1.000	0.870	1.000
100	1.000	1.000	0.992	1.000

Table 1. Probability of correct identification of the number of groups in Experiment 3. Source: own calculations.

The probability of correct identification increases with the sample size. Silhuette index looks ideal even for small samples.

Experiment 4

Four samples were generated from four two-dimensional normal distributions with unit variance-covariance matrix and expected values equal to (*d*,0), (0,*d*), (-*d*,0), (0,-*d*) with *d* = 1, 2, …, 100. Then random noise was added, as extra 5% or 10% of previously generated points. The noise came from uniform distribution based on rectangular area defined by $((x_{min},y_{min}),(x_{min},y_{max}),(x_{max},y_{min}),(x_{max},y_{max}))$. On Fig. 3 we can see one example of data generated in above described way.

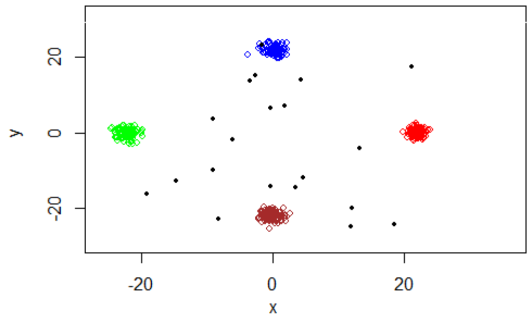


Fig. 3. Four clusters, each consisting of 100 points, d=22, with 5%.

Source: own calculations.

Random Noise	AS	CH	KL	S
5%	98%	16%	21%	77%
10%	99%	14%	11%	56%

Table 2. Percentage of correct identifications of number of clusters for d=1,…100.

Source: own calculations.

Simulation experiment with noise shows (See Table 2) the great advantage of our proposal over three other criteria.

Conclusions

Initial simulation studies reveals that the proposed separability index identifies the number of groups generally similarly to other three measures, expect the presence of random noise. In that case AS clearly outperforms the other three indexes.

Acknowledgement

The authors acknowledge support from research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Arabie, P., Hubert, L., & De Soete, G. (1978). Clustering and Classification, *Journal of Classification*, 15 (2), 286-286

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, *3*(1), 1-27.

Everitt, B.S. (1995). Classification and Cluster-Analysis – Commentary, *British Medical Journal*, 311(7004), 536-536.

Everitt, B. S., Landau, S., Leese M., & Stahl, D. (2011). Cluster analysis. 5th Edition, John Wiley & Sons Ltd, Chichester.

Krzanowski, W. J., & Lai, Y. T. (1985). A criterion for determining the number of groups in a data set using sum of squares clustering. *Biometrics*, *44*, 23-34.

Migdał-Najman, K. (2011). Ocena jakości wyników grupowania – przegląd bibliografii. *Przegląd Statystyczny*, R. LVIII, Zeszyt 3-4, 281-299.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

R packages used

cluster: M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, 2015, *''Finding Groups in Data'': Cluster Analysis Extended Rousseeuw et. al.*, R Package version 2.0.3, https://cran.r-project.org/package=cluster

clusterSim: M. Walesiak, A. Dudek, *Searching for Optimal Clustering Procedure for a Data Set*, R Package version 0.44-2, cran.r-project.org/web/packages/clusterSim/index.html

Statistical evaluation of sustainable development of Polish voivodeships in respect of social domain

Małgorzata Stec[1], Małgorzata Wosiek[2]

Abstract

Sustainable development is a strategic concept, in which the process of integrating political, economic and social actions occurs, taking into account natural balance and stability of basic natural processes, in order to guarantee possibilities of fulfilling basic needs of separate societies or citizens not only of the contemporary generation, but future generations as well.

This article presents a statistical evaluation of the level of sustainable development of Polish voivodeships in respect of social domain in the years 2005-2015. The measures of sustainable development proposed by CSOP (Central Statistical Office of Poland - GUS) were used for the purpose of the research. To construct synthetic measure, we employed the General Distance Measure (GDM) method by M. Walesiak in terms of dynamic approach. The obtained results allowed to form a rank of Polish voivodeships in respect of the level of sustainable development in terms of social domain. We also determined the position of the podkarpackie voivodeship in a regional structure of the country in terms of a problem in question.

Keywords: *sustainable development, synthetic measures, generalized distance measure (GDM), voivodeships of Poland*

***JEL Classification*:** Q01, O12, B23

1 Introduction

The notion of sustainable development is not univocal (Beckerman, 1994; Holden and Linnerud, 2007; Hopwood et al., 2005; Luke, 2005). In general, Sustainable development is a kind of development that meets the needs of the present generation without compromising the needs of future generations to meet their own needs (Borys, 2005; Grzebyk and Stec, 2015; Piontek, 2002; Waas et al., 2010). According to the World Bank definition, Sustainable development recognizes that growth must be both inclusive and environmentally sound to reduce poverty and must build shared prosperity for today's population and to continue to meet the needs of future generations. It is efficient with resources and carefully planned to deliver both immediate and long-term benefits for people, planet, and prosperity (http://www.worldbank.org/en/topic/sustainabledevelopment/overview#1).

[1] Corresponding author: University of Rzeszów, Faculty of Economics, M. Ćwiklińskiej 2, 35-601 Rzeszów, e-mail: malgorzata.a.stec@gmail.com.

[2] University of Rzeszów, Faculty of Economics, M. Ćwiklińskiej 2, 35-601 Rzeszów, e-mail: mwosiek@ur.edu.pl.

Sustainable development is a basic and superior aim of the European Union. Sustainable development is mentioned next to Smart growth and development supporting the Inclusive growth as one of the main priorities of the Europe 2020 Strategy (http://ec.europa.eu/eu2020; Stec and Grzebyk, 2016). In achieving the assumed objectives of 2020 Strategy, it is important to systematically monitor their implementation not only on the national level, but also on the regional or even local one.

The purpose of the article is a statistical evaluation of the level of sustainable development of Polish voivodeships in respect of a social domain, with particular emphasis on the podkarpackie region. Based on statistical indicators characterizing the sustainable development with reference to social domain, synthetic measures for Polish voivodeships were calculated. The rank is prepared using the dynamic approach of the General Distance Measure (GDM) method by M. Walesiak. The study was conducted in the years 2005-2015.

2 Sustainable development indicators of Polish voivodeships in a social aspect

The indicators of sustainable development are grouped in four main areas describing: Social domain, Economic domain, Environmental domain and Institutional and political domain (http://wskaznikizrp.stat.gov.pl). This article aims at evaluating a social aspect of sustainable development of Polish voivodeships. The set of 32 variables was used as a basis of the evaluation for the sustainable development of Polish voivodeships. The set consists of 10 areas determining: demographic changes, public health, poverty and living conditions, education, access to labour market, poverty and living conditions, sustainable consumption patterns, old-age income adequacy, determinants of health, road accidents, criminality.

Demographic changes:

X1- Natural increase per 1000 population (S), X2- Total fertility rate (S), X3- Ratio of balance of permanent migration person at working age (S), X4- Demographic dependency ratio (D),

Public health:

X5- Infant deaths per 1000 live births (D), X6- Life expectancy at birth by sex: males (S), X7- Life expectancy at birth by sex: females (S), X8- Deaths by selected causes of death in percentage of total: diseases of the circulatory system (D), X9- Deaths by selected causes of death in percentage of total: neoplasms (D), X10- Deaths by selected causes of death in percentage of total: diseases of the respiratory system (D), X11- Suicide rate for 10 thous. population (D),

Poverty and living conditions:

X12- The average number of people in a household receiving social benefits in relation to the average number of persons per household (D), X13- Average monthly available income per capita in private households (S),

Education:

X14- Adult persons participating in education and training (S), X15- Children covered by pre-school education in percentage of the total number of children at the age 3-5 (S), X16- Ratios the quality of education and the level of students' knowledge: At the secondary level - passing the exam maturity examination in relation to the national average (S),

Access to labour market:

X17- Employment rate of disabled persons S, X18- People at the age of 18-59 living in jobless households (D), X19- Long-term unemployed persons in registered unemployed persons total (D), X20- Unemployment rate (LFS) (D),

Sustainable consumption patterns:

X21- Number of passenger cars per 1000 population (D), X22- Consumption of electricity in households during the year per capita (D), X23- Consumption of gas in households during the year per capita (D), X24- Consumption of water in households during the year per capita (D), X25- Average monthly consumption of meat per capita (D), X26- Average monthly consumption of vegetables per capita (S),

Old-age income adequacy:

X27- Average monthly gross retirement pensions from non-agricultural social security system in relation to average monthly gross wages and salary (S),

Determinants of health:

X28- Entitled to practise doctors per 10 thous. population (S), X29- Persons injured in accidents at work per 1000 employed persons (D),

Road accidents:

X30- Fatal victimsof road accidents per 100 thous. registered motor (D),

Criminality:

X31- Ascertained crimes in completed preparatory proceedings per 1000 population (D), X32- Rate of detectability of the delinquents of ascertained crimes (S).

3 Methods applied

The initial stage of a taxonomic study is to choose diagnostic variables. Diagnostic variables should be characterised by a sufficiently high level of variability and low level of correlation.

Most frequently the classical coefficient of variation (v_j) is assumed as a measure for variability. To evaluate the correlation between variables, Pearson correlation coefficient is used. Diagnostic variables are then employed as a basis for a synthetic measure construct according to a chosen method. In this study General Distance Measure (GDM) method by M. Walesiak[3] was used. This method uses an idea of generalised correlation coefficient including Pearson correlation coefficient and Kendall rank correlation coefficient (Walesiak, 2011).

According to GDM, the procedure of linear ordering of objects comprises of the following stages (Walesiak, 2011):

- data matrix is a starting point:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{1}$$

where:

x_{ij} - value of j-th features *(j= 1,2,…,m)* in i-th object *(i=1,2,…, n)*.

- defining the type of variables (stimulants, destimulants)[4],
- a normalization of the variable values is conducted using the following formulas: (Wysocki, 2010):

$$z_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{R_j} \quad \text{for stimulating factors} \tag{2}$$

$$z_{ij} = \frac{\max_i\{x_{ij}\} - x_{ij}}{R_j} \quad \text{for non-stimulating factors} \tag{3}$$

Normalisation formulas are valuable because they provide diverse variability and simultaneously fixed range for normalised values of variables (Walesiak and Gatnar (eds.), 2009).

- determining coordinates of a model (high point of development) i.e. the most favourable values of the variables for stimulants and destimulants[5].

---

[3] This is one of the latest methodological solutions, computational procedure is available in the *R* software (Walesiak and Dudek, 2014).
[4] The notion of stimulant (S) and destimulant (D) was introduced by Hellwig (1968). Stimulants are features, whose high values are, from a given point of view, desirable phenomena (e.g. level of socio-economic development), while low values are undesirable. Destimulants on the other hand, are features whose low values are, from a given perspective, desirable occurrences, while its high values are undesirable.
[5] The coordinates of a model can be also determined with lower point of development, i.e. the least favourable values of the variables for stimulant and destimulant.

- determining distance of each objects from an object-model with the help of General Distance Measure (GDM) for metric data:

$$d_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ij} - z_{kj})(z_{kj} - z_{ij}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l \neq i,k}}^{n} (z_{ij} - z_{lj})(z_{kj} - z_{lj})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ij} - z_{lj})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(z_{kj} - z_{lj})^2\right]^{\frac{1}{2}}} \tag{4}$$

where:

i, *k*, *l = 1,…,n* – numbers of objects,

j=1,…,m – number of a variable,

m – number of variables,

z_{ij} (z_{kj}, z_{lj}) – normalised j value for a variable for *i* (*k*, *l*) object.

- Arranging the elements of a collection of objects according to GDM (high point of development). If a studied object gains lower values of synthetic measure, it has the higher level of development.

The value of the resulting synthetic measure may serve as the basis for allocating objects (e.g. voivodeships of Poland) into groups of similar levels of sustainable development in respect of social domain. An allocation scheme based on the arithmetic mean and standard deviation of the synthetic measure is also applicable in such circumstances (Nowak, 1990):

group I: $GDM_i < \acute{s}rGDM_i - S_{GDM_i}$ high level

group II: $\acute{s}rGDM_i > GDM_i \geq \acute{s}rGDM_i - S_{GDM_i}$ medium-high level (5)

group III: $\acute{s}rGDM_i + S_{GDM_i} > GDM_i \geq \acute{s}rGDM_i$ medium-low level

group IV: $GDM_i \geq \acute{s}rGDM_i + S_{GDM_i}$ low level

where: $\acute{s}rGDM_i$ - mean value of overall synthetic measure, S_{GDM_i} - standard deviation of overall synthetic measure.

4 Empirical results

Values for all variables proposed for evaluation of the level of sustainable development of Polish voivodeships in respect of social domain were collected between year 2005 and 2015, i.e. for 11 years. Then, the level of differentiation of each variable was checked and correlation coefficients were designated between them. The following variables had a low level of variability $(v_j \leq 0.10)$, during all these years taken into consideration: X2, X4, X6,

X7, X8, X9, X12, X16, X21, X25, X26, X27, X32. They were removed from the initial set of variables. The conducted analysis of correlation between the respective variables using Pearson's linear correlation did not indicated their strong correlation. The set of diagnostic variables used to evaluate the level of sustainable development of Polish voivodeships in respect of social domain included 19 variables. On its basis it was possible to determine synthetic measures with GDM for each voivodeship in 2005-2015. It should be emphasised that the calculations were done in a dynamic manner, using the so called 'object-periods'.

Table 1. shows value of synthetic measures calculated with GDM for Polish voivodeships in 2005-2015.

Voivode ships	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
DO	0.52	0.51	0.49	0.48	0.52	0.53	0.52	0.53	0.56	0.48	0.47
KP	0.40	0.44	0.44	0.41	0.44	0.40	0.46	0.44	0.43	0.45	0.46
LL	0.18	0.18	0.19	0.27	0.27	0,26	0,28	0,26	0,27	0,28	0,29
LS	0,49	0,51	0,51	0,49	0,52	0,53	0,57	0,53	0,56	0,50	0,50
LO	0,38	0,39	0,41	0,44	0,44	0,35	0,37	0,39	0,41	0,41	0,45
ML	0,27	0,25	0,29	0,27	0,29	0,28	0,26	0,27	0,25	0,25	0,25
MZ	0.41	0.39	0.33	0.36	0.36	0.36	0.35	0.35	0.37	0.37	0.37
OP	0.36	0.44	0.40	0.40	0.37	0.45	0.38	0.45	0.44	0.43	0.42
PK	0.26	0.27	0.28	0.30	0.29	0.33	0.34	0.36	0.34	0.39	0.39
PD	0.18	0.22	0.26	0.25	0.27	0.26	0.22	0.23	0.27	0.34	0.36
PM	0.43	0.42	0.42	0.38	0.39	0.35	0.34	0.28	0.33	0.32	0.29
SL	0.45	0.44	0.42	0.42	0.40	0.40	0.39	0.35	0.31	0.34	0.34
SW	0.29	0.29	0.31	0.30	0.25	0.29	0.34	0.28	0.35	0.36	0.39
WM	0.52	0.47	0.52	0.48	0.50	0.53	0.45	0.47	0.41	0.45	0.47
WK	0.38	0.42	0.43	0.43	0.45	0.41	0.43	0.48	0.43	0.45	0.43
ZP	0.61	0.60	0.59	0.54	0.59	0.54	0.56	0.52	0.51	0.51	0.49

Table 1. Values of Synthetic Measures (GDM) for Polish voivodeships[6] in 2005-2015.

[6] DO-dolnośląskie, KP-kujawsko-pomorskie, LL-lubelskie, LS-lubuskie, LO-łódzkie, ML-małopolskie, MZ-mazowieckie, OP-opolskie, PK-podkarpackie, PD-podlaskie, PM-pomorskie, SL-śląskie, SW-świętokrzyskie, WM-warmińsko-mazurskie, WK-wielkopolskie, ZP-zachodniopomorskie.

In years 2005-2015, the values of synthetic measure obtained by GDM for each voivodeship differ in regard to the value. During the researched period they do not show any significant uptrend or downward trend, they increase and decrease alternately. In 2005, the value of synthetic measure (GDM) for voivodeships oscillated from 0.18 to 0.61. In the rank of voivoideships, in relation to the calculated measure, among the best were: podlaskie, lubelskie, podkarpackie, małopolskie and świętokrzyskie; while at the end of the rank there were voievodeships such as: zachodniopomorskie, warmińsko-mazurskie, dolnośląskie, lubuskie and śląskie. In 2015, the value of synthetic measure (GDM) oscillated from 0.25 to 0.50. At the forefront of the rank, on the level of regional structure, in terms of analysed measure, there were voivodeships such as: małopolskie, pomorskie, lubelskie, śląskie and podlaskie. At the end of the rank there were: lubuskie, zachodniopomorskie, warmińsko-mazurskie, dolnośląskie and kujawsko-pomorskie voivodeships (table 1).

Voivode ships	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
DO	14	14	13	14	14	15	14	16	16	14	13
KP	9	12	12	9	11	9	13	10	11	11	12
LL	2	1	1	3	2	1	3	2	3	2	3
LS	13	15	14	15	15	14	16	15	15	15	16
LO	7	7	8	12	10	6	8	9	9	9	11
ML	4	3	4	2	4	3	2	3	1	1	1
MZ	10	6	6	6	6	8	7	6	8	7	6
OP	6	10	7	8	7	12	9	11	13	10	9
PK	3	4	3	5	5	5	6	8	6	8	8
PD	1	2	2	1	3	2	1	1	2	4	5
PM	11	9	10	7	8	7	4	4	5	3	2
SL	12	11	9	10	9	10	10	7	4	5	4
SW	5	5	5	4	1	4	5	5	7	6	7
WM	15	13	15	13	13	13	12	12	10	12	14
WK	8	8	11	11	12	11	11	13	12	13	10
ZP	16	16	16	16	16	16	15	14	14	16	15

Table 2. Ranks of Polish voivodeships in terms of values of synthetic measure calculated with General Distance Measure (GDM) in years 2005-2015.

Based on the data provided in table 2, it can be inferred that between years 2005-2015, the positions in the ranking of voivodeships in terms of synthetic measure obtained with GDM are quite stable. Improvement of ranks of voivodeships in 2015 in comparison to 2005 can be noticed for 7 voivodeships. The following voivodeships underwent the greatest positive changes: pomorskie (from position 11 in 2005 to position 2 in 2015), śląskie (from 12 to 4) and mazowieckie (from 10 to 6); whereas 9 voivodeships noted downgrade in the ranks, the highest drop for podkarpackie voivodeship (from position 3 in 2005 to position 8 in 2015), łódzkie (from 7 to 11), podlaskie (from 1 to 5).

In years 2005-2015, podkarpackie voivodeship recorded a systematic drop of its position in relation to other voivodeships, which is an alarming phenomenon. However, it should be acknowledged that the social domain of the sustainable development of podkarpackie voivodeship is quite a strong advantage. In other socio-economic domains it comes out worse than other voivodeships (Stec, 2012).

Using a classification outline (formula no 5), we provided a classification of Polish voivodships in relation to different levels of sustainable development in social domain in years 2005 and 2015 (table 3).

Level	**Year**	
	2005	**2015**
High	podlaskie, lubelskie, podkarpackie,	małopolskie, pomorskie, lubelskie,
Medium- high	małopolskie, świętokrzyskie, opolskie, łódzkie, wielkopolskie,	śląskie, podlaskie, mazowieckie, świętokrzyskie, podkarpackie,
Medium- low	kujawsko-pomorskie, mazowieckie, pomorskie, śląskie, lubuskie,	opolskie, wielkopolskie, łódzkie, kujawsko-pomorskie, dolnośląskie, warmińsko-mazurskie,
Low	dolnośląskie, warmińsko-mazurskie, zachodniopomorskie.	zachodniopomorskie, lubuskie.

Table 3. Groups of Poland voivodeships with similar levels of sustainable development in respect of social domain in 2005 and 2015.

Three Poland voivodeships, podlaskie, lubelskie, podkarpackie, attained a high level of sustainable development in respect of social domain in 2005. The group of medium-high level comprises 5 voivodships małopolskie, świętokrzyskie, opolskie, łódzkie, wielkopolskie. Medium- low levels showed by voivodeships kujawsko-pomorskie, mazowieckie, pomorskie, śląskie, lubuskie. Dolnośląskie, warmińsko-mazurskie, zachodniopomorskie belong to the group of low level.

High levels of sustainable development in respect of social domain was, in 2015, showed by voivodeships małopolskie, pomorskie, lubelskie. Most voivodeships of Poland belong to the group of medium-high (5 voivodeships) and medium-low levels (6 voivodeships). Zachodniopomorskie and lubuskie are the voivodeships which belong to the group of the lowest level of sustainable development in respect of social domain (table 3).

Conclusions

In this article, the comparison of Polish voivodeships in terms of the level of the sustainable development in years 2005-2015 was conducted. The measures of the sustainable development in social domain proposed by CSOP were used for analysis. The General Distance Measures (GDM) method by M. Walesiak was employed for the empirical studies. The results confirm the diversity of Polish regions in social terms of sustainable development. The position of podkarpackie voivodeship in the regional structure of the country in relation to the studied problem can be deemed as quite good. However, its systematic downgrade in the rank is alarming. The results of the research concerning the sustainable level in social domain can be used for the initial evaluation of sustainable level for all voivodeships. It would be interesting to conduct the similar research for other aspects of the sustainable growth, i.e. Economic domain, Environmental domain and Institutional and political domain.

References

Beckerman, W. (1994). 'Sustainable development': is it a useful concept?. *Environmental values*, *3*(3), 191-209.

Borys, T. (2005). Pomiar trwałego i zrównoważonego rozwoju [w:] Zrównoważony rozwój. *Wybrane problemy teoretyczne i implementacja w świetle dokumentów Unii Europejskiej*, 126-143.

Grzebyk, M., & Stec, M. (2015). Sustainable development in EU countries: concept and rating of levels of development. *Sustainable Development*, *23*(2), 110-123.

Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, *(4)*, 323-326.

Holden, E., & Linnerud, K. (2007). The sustainable development area: satisfying basic needs and safeguarding ecological sustainability. *Sustainable Development*, *15*(3), 174-187.

Hopwood, B., Mellor, M., & O'Brien, G. (2005). Sustainable development: mapping different approaches. *Sustainable Development*, *13*(1), 38-52.

Luke, T. W. (2005). Neither sustainable nor development: reconsidering sustainability in development. *Sustainable Development*, *13*(4), 228-238.

Nowak, E. (1990). *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, Państwowe Wydaw. Ekonomiczne, 93.

Piontek, B. (2002). *Koncepcja rozwoju zrównoważonego i trwałego Polski*. Wydaw. Naukowe PWN, 27.

Stec, M. (2012). Analiza rozwoju społeczno-gospodarczego województwa podkarpackiego w latach 1999-2009 [w:] M*odelowanie i prognozowanie zjawisk społeczno-gospodarczych. Aktualny stan i perspektywy rozwoju*. Wydaw. Uniwersytetu Ekonomicznego w Krakowie, 107-118.

Stec, M., & Grzebyk, M. (2016). The implementation of the Strategy Europe 2020 objectives in European Union countries: the concept analysis and statistical evaluation, *Quality & Quantity*, DOI 10.1007/s11135-016-0454-7

Waas, T., Verbruggen, A., & Wright, T. (2010). University research for sustainable development: definition and characteristics explored. *Journal of Cleaner Production*, *18*(7), 629–636.

Walesiak, M. (2011). *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wydawnictwo Uniwersytetu Ekonomicznego, 38; 74-76.

Walesiak, M., & Dudek, A. (2014). *ClusterSim Package*, *URL*, http://www.R-project.org.

Walesiak, M., & Gatnar, E. (Eds.). (2009). *Statystyczna analiza danych z wykorzystaniem programu R*. Wydawnictwo Naukowe PWN, 67.

Wysocki, F. (2010). *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*. Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań, 51.

The labour market situation in medium-sized urban centres of the Kujawsko-Pomorskie voivodeship and the problem of unemployment in the province

Elżbieta Szulc[1], Mateusz Jankiewicz[2]

Abstract

The aim of the paper is to analyse the spatial and spatio-temporal dependences in matters related to the situation in the labour market of the Kujawsko-Pomorskie voivodeship across municipalities in the period of 2004-2015. With reference to the guidelines of the 2020 development strategy for the province and the 2020+ modernisation plan, the analysis may constitute the basis for verifying whether, in the presence of a dependence, investing (e.g. through a growing number of enterprises) in the development of individual medium-sized urban centres with a high level of unemployment such as Włocławek, Grudziądz and Inowrocław can significantly improve the situation in the whole province. The assessment of the situation in the labour market for each of the municipalities is made with the use of two characteristics: the share of registered unemployed persons in the number of the working age population and the number of business entities per 10 000 working age population. The spatial and spatio-temporal tendencies and dependences are in turn investigated using the conception of spatial trends and spatial autocorrelation. The empirical analyses have been supplemented by simulations of the labour market situation in the province resulting from situation improvement in selected urban centres.

Keywords: *labour market, medium-sized urban centres, spatial and spatio-temporal dependence*

JEL Classification: C10, E24, R58

1 Introduction

With regard to labour market conditions, as stated by the data from the Central Statistical Office, the Kujawsko-Pomorskie voivodeship for many years has been located at the bottom of the Polish provinces rankings. The development of individual territorial units within provinces contributes to jobs creation, and thus, to a reduction of the unemployment rate. According to the 2020 development strategy for the Kujawsko-Pomorskie voivodeship and the 2020+ modernisation plan, an improvement in the labour market situation of the province should be the result of investments in the development of the medium-sized urban centres such as Włocławek, Inowrocław and Grudziądz, where the level of unemployment is high. The desire to verify the correctness of this supposition constitutes the main motivation for

[1] Corresponding author: Nicolaus Copernicus University, Faculty of Economic Sciences and Management, Gagarina 13A, 87-100 Toruń, Poland, e-mail: eszulc@.umk.pl.

[2] Nicolaus Copernicus University, Faculty of Economic Sciences and Management, Gagarina 13A, 87-100 Toruń, Poland, e-mail: mateuszj@doktorant.umk.pl.

conducting this research. The mentioned assumption is also the main research hypothesis of the study.

The labour market is a heterogeneous category in different aspects of the analysis (Szulc, 2011; Vega and Elhorst, 2014; Nosek and Netrdova, 2014; Pillet et al. 2014; Semerikova, 2015; Blinova et al. 2016). In particular, in addition to variability in time, the labour market situation is characterised by spatial differentiation occurring at different levels of data aggregation. Simultaneously, certain regularities, tendencies and dependences in this area can be observed.

The primary aim of the paper is to analyse the spatial and spatio-temporal dependences in matters concerning the situation in the labour market of the Kujawsko-Pomorskie voivodeship across municipalities in the period of 2004-2015. It is the basis to verify that, with the existing dependencies, investments (e.g. through a growing number of enterprises) in the development of individual medium-sized urban centres of the province can significantly improve the labour market situation in the whole province. The labour market situation in each of the municipalities has been assessed on the basis of variables which, on the one hand, represent an unexploited labour supply in the form of the number of the unemployed, and, on the other, the demand for labour which generates enterprises operating in the market.

2 Subject and the scope of the investigation

The study concerns the labour market situation in the Kujawsko-Pomorskie voivodeship across municipalities in the period of 2004-2015. The following indicators have been analysed: the share of registered unemployed persons in the number of the working age population (Y) and the number of business entities per 10 000 working age population (X). It was considered that these variables are the most important determinants of the market situation. The data availability at a fixed level of the spatial aggregation (NUTS-5 classification) is also of significance. The adopted level of spatial aggregation and time range of the research allowed to observe the spatial and temporal tendencies and regularities in shaping of the analysed variables, which requires a large number of observations.

The spatial and spatio-temporal tendencies and dependences in the behaviour of the analysed variables were in turn investigated using the conception of the spatial trend and spatial autocorrelation. Additionally, spatial autoregressive models were estimated and verified. The findings in this range were used for the specification of the econometric model for the pooled spatial and temporal data. The empirical analyses have been supplemented by simulations of the labour market situation in the province assuming an improvement of the situation in the selected urban centres. For this purpose the abovementioned model was used. The said model contained the spatio-temporal trend and spatially lagged variables.

The study verifies the hypothesis of an importance of job creation in certain medium-sized urban centres for the improvement of labour market situation in the whole region. In particular, an attempt was made to determine the extent of spatial influences of stimulation of labour demand in the selected spatial units. The question is whether a growing number of enterprises in selected centres will reduce unemployment in the whole region or in the neighbouring areas only.

3 Data

The data used in the analysis is retrieved from database of the Central Statistical Office (GUS) (https://bdl.stat.gov.pl/BDL/dane). The information regarding variable Y (the share of registered unemployed persons in the working age population) were drawn directly from the database, while the values of variable X (the number of entities per 10 000 working age population) were obtained through our own calculations.

Figure 1 shows the spatial differentiation of variable Y in the two extreme years of analysis, i.e. in the years 2004 and 2015. In Figure 2 the trend surfaces of the considered variable for the respective years are shown. It may be noted that in both years the central part of the province is characterised by a low level of unemployment compared to the remaining part of the region. A higher unemployment rate than that in the center of the region was noted, among others, in medium-sized urban centres, to which particular attention was paid in the investigation, i.e. in Grudziądz, Inowrocław and Włocławek.

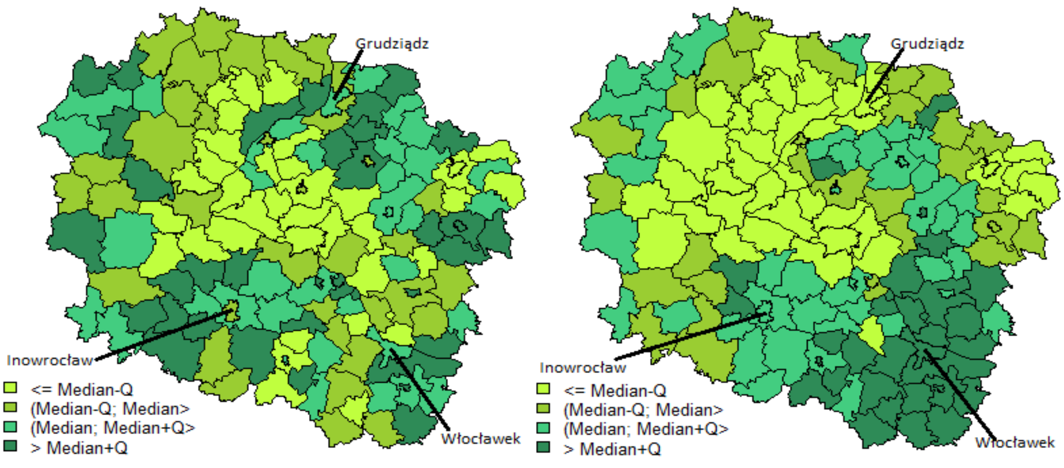


Fig. 1. Unemployment across the municipalities in the years 2004 and 2015.

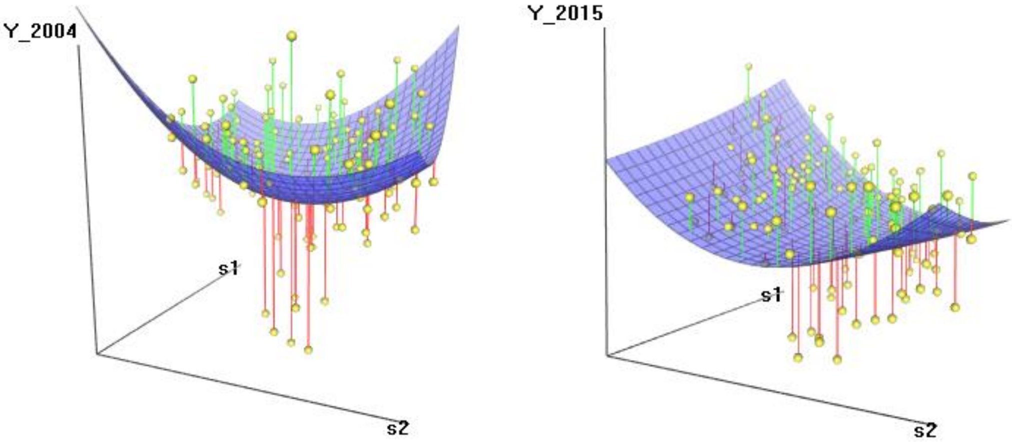
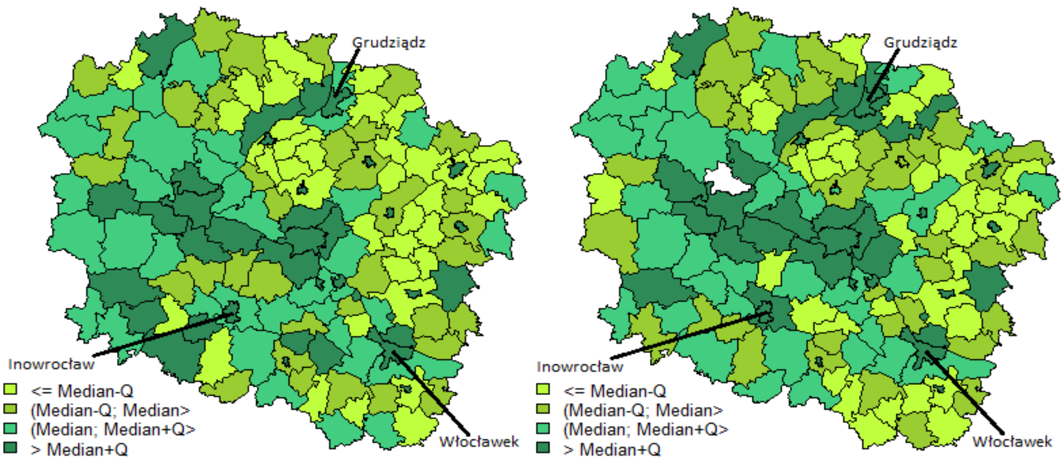


Fig. 2. Trend surfaces of unemployment across the municipalities in the years 2004 and 2015.



Note: The uncoloured municipality on the right map is the commune of Osielsko which constitutes the so-called outlier with a high value of the variable.

Fig. 3. The number of entities per 10 000 population at working age across the municipalities in the years 2004 and 2015.

In turn, Figure 3 shows spatial differentiation of variable X. In the case of the number of entities per 10 000 population at working age, the spatial distribution displays an opposite regularity as compared with the distribution of the variable characterising the level of unemployment. The municipalities located in the central part of the region are characterised by a relatively high level of variable X. This concerns both the first and the last year of the investigation. Furthermore, it should be emphasised that a relatively high number of enterprises in the central part of the province was observed also in urban municipalities, which are of particular interest to us, i.e. Grudziądz, Inwrocław and Włocławek. Figure 4 shows the trend surfaces observed in the formation of variable X in the extreme years of our analysis.

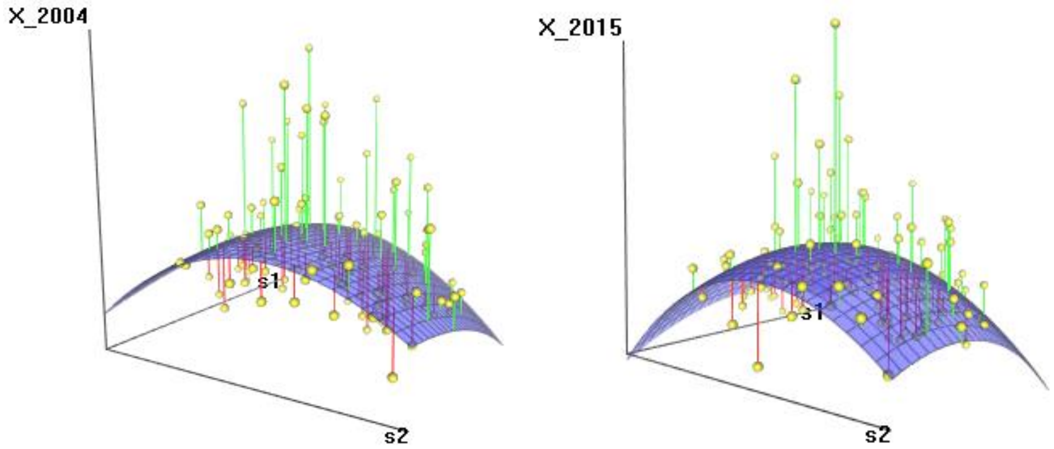


Fig. 4. Trend surfaces of the number of entities per 10 000 population at working age across the municipalities in the years 2004 and 2015.

Based on Figures 2 and 4 a presumption was formulated regarding the presence of the second-degree spatial trend for both the variables analysed, whereas the spatial directions of the increases and decreases in those variables are opposite.

4 Methodology

The primary tool of analysis conducted in the study is an econometric model specified for spatio-temporal stochastic process, i.e. a random function $Y(\mathbf{s},t)$, where non-random arguments are defined as $\mathbf{s}=[s_1,s_2]\in D\subset R^2$, $t\in T\subset R$ (Cressie, 1993; Schabenberger and Gotway, 2005). The two-dimensional argument **s** quantifies the locations of spatial units, and t indicates the time. Due to the discreet quantification of space and time in the analyses of economic phenomena the process $Y(\mathbf{s}_i,t)$ is considered, where $\mathbf{s}_i=[s_{i1},s_{i2}]$, i = 1, 2, …, N is the number of the spatial unit, t = 1, 2, ..., T $\subset$ C and C denotes the set of natural numbers.

The specification of the spatio-temporal model was made on the basis of the research of trend-autoregressive structures of the analysed spatial processes $Y(\mathbf{s}_i)$ and $X(\mathbf{s}_i)$, for each year of the period considered. For this purpose spatial trends were identified and spatial autocorrelation was tested. Moreover, the estimation and verification of the spatial autoregressive models was conducted. The models of spatial trend were used, whose general form can be written as follows (Cressie, 1993; Schabenberger and Gotway, 2005):

$$P(\mathbf{s}_i)=\sum_{k=0}^{p}\sum_{m=0}^{p}\theta_{km}s_{1i}^{\ k}s_{2i}^{\ m}, \tag{1}$$

where: $\mathbf{s}_i = [s_{1i}, s_{2i}]$ stands for location coordinates on the plane, i = 1,2, ..., N (indexes of spatial units), and p means the polynomial trend degree ($k + m \leq p$).

Respectively, the spatio-temporal trend model takes the form (Szulc, 2007):

$$P(\mathbf{s}_i, t) = \sum_{k=0}^{p} \sum_{m=0}^{p} \sum_{l=0}^{p} \theta_{mkl} s_{1i}^{k} s_{2i}^{m} t^{l}, \qquad (2)$$

where: $\mathbf{s}_i$, i, p are as above, wherein $k + m + l \leq p$.

The spatial autocorrelation was tested using the test based on the Moran statistic (Moran's I) which can be expressed as follows (Schabenberger and Gotway, 2005):

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} [z(\mathbf{s}_i) - \bar{z}][z(\mathbf{s}_j) - \bar{z}]}{\sum_{i=1}^{N} [z(\mathbf{s}_i) - \bar{z}]^2}, \qquad (3)$$

where: $z(\mathbf{s}_i)$ denotes the observation of the phenomenon in the region i, $\bar{z}$ is the average value of the phenomenon, w_{ij} represents components of the appropriate connectivity matrix. In the study the matrix **W** of connections based on the common border criterion was used.

With the designations adopted above, the spatial autoregressive models (Arbia, 2006; LeSage and Pace, 2009) including spatial trends can be written in the following form (Szulc, 2011):

$$Z(\mathbf{s}_i) = \sum_{k=0}^{p} \sum_{m=0}^{p} \theta_{km} s_{1i}^{\;k} s_{2i}^{\;m} + \rho \mathbf{W} Z(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \qquad (4)$$

where: $\varepsilon(\mathbf{s}_i)$ is the spatial white noise.

In turn, the hypothesis of space-time model for the studied process $Y(\mathbf{s}_i)$ was as follows:

$$Y(\mathbf{s}_i, t) = \sum_{k=0}^{p} \sum_{m=0}^{p} \sum_{l=0}^{p} \theta_{mkl} s_{1i}^{k} s_{2i}^{m} t^{l} + \rho \mathbf{W}^* Y(\mathbf{s}_i, t) + \gamma X(\mathbf{s}_i, t) + \delta \mathbf{W}^* X(\mathbf{s}_i, t) + \varepsilon(\mathbf{s}_i, t), \qquad (5)$$

where $\mathbf{W}^*$ denotes the block matrix of spatio-temporal connections which takes the form:

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_T \end{bmatrix},$$

wherein: $\mathbf{W}_1 = \mathbf{W}_2 = ... = \mathbf{W}_T$ represent standard spatial connectivity matrixes, such as in (4), the same for all years.

5 Specification of spatio-temporal econometric model – results of the empirical analysis

5.1 Analysis of the spatial trends and spatial autocorrelation

The analysis of the situation of the labour market in the municipalities of the Kujawsko-Pomorskie voivodeship started with identifying the spatial structure of both variables. For this purpose the spatial trend models were estimated and verified and next the spatial autocorrelation using Moran's *I* was tested. Table 1 reports the results of the investigation.

Year	**Degree of spatial trend**		**R^2**		**Spatial autocorrelation**		**Morans's *I***	
	$Y(\mathbf{s}_i)$	$X(\mathbf{s}_i)$	$Y(\mathbf{s}_i)$	$X(\mathbf{s}_i)$	$Y(\mathbf{s}_i)$	$X(\mathbf{s}_i)$	$Y(\mathbf{s}_i)$	$X(\mathbf{s}_i)$
2004	2	2	0.1494	0.0996	+	-	0.3746	0.0295
2005	2	2	0.1612	0,1016	+	-	0.3600	0.0337
2006	2	2	0.1580	0.1004	+	-	0.3807	0.0222
2007	2	2	0.2856	0.1021	+	-	0.4050	0.0339
2008	2	2	0.2736	0.1107	+	-	0.4363	0.0429
2009	2	2	0.2600	0.1343	+	-	0.3951	0.0585
2010	2	2	0.2891	0.1336	+	+	0.4488	0.0842
2011	2	2	0.2839	0.1397	+	+	0.4973	0.0979
2012	3	2	0.3400	0.1386	+	+	0.4722	0.1127
2013	3	2	0.3220	0.1469	+	+	0.4420	0.1320
2014	3	2	0.4309	0.1529	+	+	0.4339	0.1434
2015	2	2	0.4611	0.1504	+	+	0.4404	0.1480

Table 1. Spatial structure of the processes: $Y(\mathbf{s}_i)$ – share of the registered unemployed persons in the working age population and $X(\mathbf{s}_i)$ – entities per 10 000 working age population. *Note*: Symbol "+" means that spatial autocorrelation occurs (at the level of significance of at least 0.05), symbol "-" means that spatial autocorrelation does not occur.

The presence of spatial tendencies in each year of the investigation for both analysed variables has been identified. The considered variables dependence on their values reported in neighbouring municipalities was observed in each year for variable Y and from 2010 for variable X. Next, on the basis of the pooled time series and spatial data, the analysis of spatio-temporal trend was conducted for each variable. In the case of unemployment the presence of second-degree spatio-temporal trend was observed (with regard to both the spatial and temporal component). At the same time, the variable concerning the number of business entities did not show a significant tendency over time.

Based on the foregoing findings the estimation and verification of spatio-temporal SAR model was made where the level of unemployment in a given year in each of the municipalities of the province analysed is explained by the current level of unemployment in neighboring municipalities, the number of business entities in the municipality in a given year, and the number of entities in the neighboring municipalities in the same year.

Parameter	Estimate of parameter	Standard error	Statistic z	p-value
θ_{000}	1.00E+00	3.29E-01	3.0374	0.0024
θ_{100}	-8.39E-07	5.78E-07	-1.4526	0.1463
θ_{010}	-2.19E-06	7.90E-07	-2.7748	0.0055
θ_{200}	6.12E-13	4.42E-13	1.3829	0.1667
θ_{110}	3.47E-13	5.55E-13	0.6254	0.5317
θ_{020}	1.66E-12	5.52E-13	3.0060	0.0026
θ_{001}	-1.16E-02	1.35E-03	-8.5634	< 2.2E-16
θ_{002}	9.56E-04	1.23E-04	7.7413	9.77E-15
γ	-1.58E-05	1.96E-06	-8.0839	6.66E-16
δ	-3.27E-05	4.69E-06	-6.9785	2.98E-12
$\rho = 0.6574$				
Test LR: 505.3, p-value: < 2.22E-16				
Moran test: -1.2234, p-value: 0.1106				

Table 2. The results of the estimation and verification of SAR model with spatio-temporal trend.

In accordance with the above findings, the model contained an additional component, i.e. the spatio-temporal trend of degree 2. The model was estimated on the basis of data for the period from 2004 to 2012 in order to enable a simulation of changes in the level of unemployment on assumptive changes regarding the number of enterprises. Table 2 contains the results of the model estimation and verification.

5.2 Simulation of the level of unemployment in municipalities

Using the model described above, a simulation of the level of unemployment in municipalities of the Kujawsko-Pomorskie voivodeship was performed on the assumption that number of enterprises per 10 000 population at working age in individual medium-sized urban centres such as Grudziądz, Inowrocław and Włocławek increases (the assumed increase amounted to approximately 20%). Table 3 sets out the municipalities where a decrease of the level of unemployment was noted.

Among the municipalities, which showed a decrease in the unemployment as a result of an increase in the number of enterprises per 10 000 working age population are the three centres considered and their neighbouring municipalities.

Year	Municipalities: Urban – (1), Rural – (2), Urban-Rural (3)
2013	Inowrocław (1), Pakość (3), Inowrocław (2), Brześć Kujawski (3), Bobrowniki (2), Lubanie (2), Włocławek (2), Dobrzyń nad Wisłą (3), Fabianki (2), Włocławek (1), Grudziądz (2), Dragacz (2), Grudziądz (1), Rogóźno (2)
2014	Inowrocław (1), Pakość (3), Inowrocław (2), Brześć Kujawski (3), Bobrowniki (2), Lubanie (2), Włocławek (2), Dobrzyń nad Wisłą (3), Fabianki (2), Włocławek (1), Grudziądz (2), Dragacz (2), Grudziądz (1), Rogóźno (2)
2015	Inowrocław (1), Pakość (3), Inowrocław (2), Brześć Kujawski (3), Bobrowniki (2), Lubanie (2), Włocławek (2), Dobrzyń nad Wisłą (3), Fabianki (2), Włocławek (1), Grudziądz (2), Dragacz (2), Grudziądz (1), Rogóźno (2)

Table 3. List of the municipalities where a decrease in the unemployment rate was observed.

Conclusion

The analysis showed that with the existing spatial and spatio-temporal dependencies observable in the labour market in the Kujawsko-Pomorskie voivodeship across municipalities an improvement of the situation of this market may be expected as a result of an increase in the number of enterprises in medium-sized urban centres such as Grudziądz, Włocławek and Inowrocław. However, according to the results of the investigation, investments in new business entities in the indicated centres will improve the labour market situation only in the said centres and the neighbouring municipalities. The obtained results do not provide sufficient grounds to formulate a statement that the indicated actions will significantly improve the labour market situation in the entire province.

References

Arbia G. (2006). *Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin–Heidelberg.

Blinova, T., Markow, V., & Rusanovskiy, V. (2016). Empirical Study of Spatial Differentiation of Youth Unemployment in Russia. *Acta Oeconomica*, 66(3), 507-526.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, INC, New York.

LeSage, J., & Pace, K. R. (2009). *Introduction to Spatial Econometrics*. Champion & Hall/CRC, Boca Raton, London–New York.

Nosek, V., & Netrdova, P. (2014). Measuring Spatial Aspects of Variability. Comparing Spatial Autocorrelation with Regard Decomposition in Internal Unemployment Research. *Historical Social Research-Historische Socialforschung*, 39(2), 292-314.

Pillet, F., Cañizares, M. C., Ruiz, A. R., Martínez, H., Plaza, J., & Santos, J. F. (2014). Applying the European Spatial Development Perspective in low-density regions: a methodology based on mobility and labour market structure. *Urban Studies*, 51(3), 577-595.

Schabenberger, O., & Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Champion & Hall/CRC, Boca Raton, London–New York.

Semerikova, E. (2015). Spatial Patterns of German Labor Market: Panel Data Analysis of Regional Unemployment. In: Mussida, Ch., Pastore, F. (eds.), *Geographical Labor Market Imbalances: Recent Explanations And Cures*, Springer Berlin Heidelberg, 37-64.

Strategia rozwoju województwa kujawsko-pomorskiego do roku 2020 – Plan modernizacji 2020+ (The 2020 Development Strategy for the Kujawsko-Pomorskie Voivodeship and the 2020+ Modernisation Plan), http://www.kujawsko-pomorskie.pl/pliki/planowanie/planowanie_przestrzenne/strategia_rozwoju_wojewodztwa_plan_modernizacji_2020.pdf.

Szulc, E. (2007). *Ekonometryczna analiza wielowymiarowych procesów gospodarczych* (*Econometric Analysis of Multidimensional Economic Processes*), UMK, Toruń.

Szulc, E. (2011). Identification of the Structures of Spatial and Spatio-Temporal Processes and a Problem of Data Aggregation. *Dynamic Econometric Models*, 11, 4-20.

Vega, S. H., & Elhorst, J. P. (2014). Modelling Regional Labour Market Dynamics in Space and Time. *Papers in Regional Science*, 93(4), 819-841.

An application of Extreme Value Theory for modelling life expectancy

Grażyna Trzpiot[1], Justyna Majewska[2]

Abstract

In this paper we have applied new method (Medford, 2015) in order to project the life expectancy in Poland using life expectancy values from different European countries. This new approach combines concept of extreme value distribution and "best life expectancy" idea of Oeppen and Vaupel (2002), who demonstrated that the development of "best practice" life expectancy is best approximated by a linear trend, estimated over the past 160 years. Approach based on EVT has never been applied to life expectancy projections in Poland. The method produces results (life expectancy at birth and age 65) that are similar to those obtained by Central Statistical Office of Poland for period 2015-2050.

Keywords: *EVT, extreme value distribution, life expectancy*

JEL Classification: C130, C180

1 Introduction

Life expectancy at birth is defined as the mean number of years still to be lived by a person at birth, if subjected throughout the rest of his or her life to the current mortality conditions. Over the last 170-year period European countries have experienced a continuation of the pattern of falling mortality rates that began in the 19th century. Over long historical period we have observed linear trend in life expectancy. However~~,~~ today there is no consensus on trends over the very long term (European Commission 2014). Official projections generally assume that gains in future life expectancy at birth will slow down compared with historical trends.

Accuracy is crucial in a life expectancy projection. It allows governments and other institutions to plan wisely and helps individuals comprehend the likely futures for their countries and the world.

As Li and Lee (2005) indicated the convergence in mortality levels for closely related populations can lead to unsuitable mortality projections, if the projections for individual populations are obtained in isolation from one another. Similar historical trends in long-run life expectancy patterns can be useful for countries. Besides, analyses of the main determinants of life expectancy (the socio-economic, environmental or behavioural factors) of associated populations are crucial. Knowledge of existence of some common stochastic trends

[1] Corresponding author: University of Economics in Katowice, Department of Demography and Economic Statistics, 1 Maja 50, 40-287 Katowice, e-mail: justyna.majewska@ue.katowice.pl.

[2] University of Economics in Katowice, Department of Demography and Economic Statistics, 1 Maja 50, 40-287 Katowice, e-mail: grazyna.trzpiot@ue.katowice.pl.

in mortality rate in cluster of European countries (see Lazar et al. 2016) can be used for projections of mortality rates and life expectancy.

In this paper we have applied new method (Medford, 2015) in order to project the life expectancy in Poland using life expectancy values from different European countries. This new approach combines concept of extreme value distribution and “best practice“ life expectancy idea of Oeppen and Vaupel (2002). “Best practice” life expectancy reffers to the maximum life expectancy observed among national populations at a given age, and according to Oeppen and Vaupel (2002), the development of “best practice” life expectancy is best approximated by a linear trend, estimated over the past 160 years. The approach based on EVT has never been applied to life expectancy projections in Poland.

2 Basic of Extreme Value Theory

The EVT has been proved to be a powerful tool to study extreme event distributions and widely used in many applications in multidisciplinary areas. EVT is well documented in the literature (e.g. Coles, 20012; Gilli and Këllezi, 2006; Embrechts et al. at al, 1997).

The foundation of extreme value theory is a family of probability distributions known as extreme value distributions. For $\xi \neq 0$ the generalized extreme value GEV distribution function is given by

$$P(Y \leq y) = G(y;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \qquad (1)$$

where Y is some random variable and the three parameters $\mu(-\infty < \mu < \infty)$, $\sigma(\sigma > 0)$ and $\xi(-\infty < \xi < \infty)$ are referred to as the location, scale and shape parameters, respectively. The location parameter indicates the center of the distribution, the scale parameter indicates the size of deviations around the location parameter, and the shape parameter governs the tail behavior of the GEV distribution. All three parameters are important, but ξ is especially critical for extrapolation of extreme events. The shape ξ parameter determines the heaviness of the right tail and this leads to three types of distributions. When $\xi < 0$, the distribution has a bounded upper finite end point and is short-tailed and in this case the distribution is often called a Weibull distribution. When $\xi > 0$, there is polynomial tail decay leading to heavy tails and the GEV distribution is called then Fréchet. The case where ξ=0 is taken to be the limit of eq. (1) as $\xi \to \infty$, and there is exponential tail decay leading to light tails (usually referred to as the Gumbel type).

Generally there are two ways of identifying extremes in data: block maxima and peak over threshold method. Block maxima method considers the maximum the variable takes in successive periods. These selected observations constitute the extreme events, also called block maxima.

The limit law for the block maxima, which we denote by M_n, with n the size of the subsample (block), is given by the Extremal Types Theorem (Fisher and Tippett, 1928; Gnedenko, 1943): if there exists sequences of constants $\{a_n>0\}$ and $\{b_n\}$, such that as $n \to \infty$

$$P\left(\frac{M_n - b_n}{a_n} \leq y\right) \to G(y)$$

where $G(y)$ is a non-degenerate distribution function, then G must be a member of the GEV family of distributions.

In estimating the parameters μ, σ and ξ from observations, there are a number of popular methods including maximum likelihood estimation (MLE), Bayesian methods, and the L-moments technique.

Estimates of extreme quantiles of the maxima are obtained by solving for y in eq. (1):

$$y_p = \mu - \frac{\sigma}{\xi}\left[1 - \left[-\log(1-p)\right]^{-\xi}\right] \tag{2}$$

where the distribution function of the GEV, $G(y_p)=1-p$, and p is the tail probability or the probability of realising a value at least as large as y_p. When we focus on annual maxima then on average the quantile y_p is expected to be exceeded with probability p or on average once every $1/p$ years (Coles, 2001).

3 Application

Life tables were obtained from Human Mortality Dataset (HMD, 2017)[3]. Calculations were made in R with *spdep, demography, evd* and *segmented* packages[4].

Best life expectancies were calculated for the following European countries: Denmark, Finland, Norway, Sweden, Netherlands, Belgium, France, Italy, Ireland, Portugal, Spain, UK, Italy, Austria, and Czech Republic. Existence of similar historical trends in life expectancy

[3] http://mortality.org.

[4] Package *spdep* – a collection of functions to create spatial weights matrix objects from polygon contiguities, package *demography* – functions for demographic analysis, *demography* - functions for demographic analysis including lifetable calculations, *evd* – provides extends simulation, distribution, quantile and density functions to univariate and multivariate parametric extreme value distributions, *segmented* – for given a regression model, segmented „updates“ the model by adding one or more segmented (i.e., piecewise-linear) relationships.

was motivation for grouping countries into the clusters. SKATER algorithm[5] divided countries into homogeneous contiguous clusters according to the following variables: gross domestic product per capita, educational attainment, fertility and access to health care[6]. In this analysis we skipped cluster that contains most countries from Eastern and Southern Europe. The first cluster groups developed countries with the highest GDP per capita and very high fertility rate. The second cluster includes countries with the highest fertility rate and low education level. The third cluster groups countries with the lowest percentage of population aged 16 and over, that they had unmet needs for medical examinations or treatment. Lazar et al. (2016) proved the existence of some common stochastic trends determining the mortality in European countries.

Countries in cluster	Historical period
1. Denmark, Finland, Norway, Sweden, Netherlands	1960-2012
2. Belgium, France, Italy, Ireland, Portugal, Spain, UK	1950-2012
3. Italy, Austria, Czech Republic	1950-2012

Table 1. Clusters of European countries with common mortality trends.

Historical trends in life expectancy are observed in Fig. 2. and Fig. 3. For comparison purpose historical trend of life expectancy in Poland was added. Countries from the first cluster reached very high life expectancy sooner than the rest European countries, however the speed in increase between 1960 and 2010 was not as fast as it was in the countries from the second and the third cluster. Since 1960, in countries from the first cluster life expectancy at birth has typically increased by 6-9 years for female, meanwhile in the second cluster it was 8-12 years. In the third cluster trends are remarkably similar until ‘70s. This parallelism is rarely remarked later.

[5] Spatial Kluster Analysis by Tree Edge Removal is an alternative to other regionalization methods, based on minimum spanning trees (Assunção et al. 2006).
[6] Demographic and economic data (recorded in 2012) used for spatial clustering were obtained from Eurostat database.

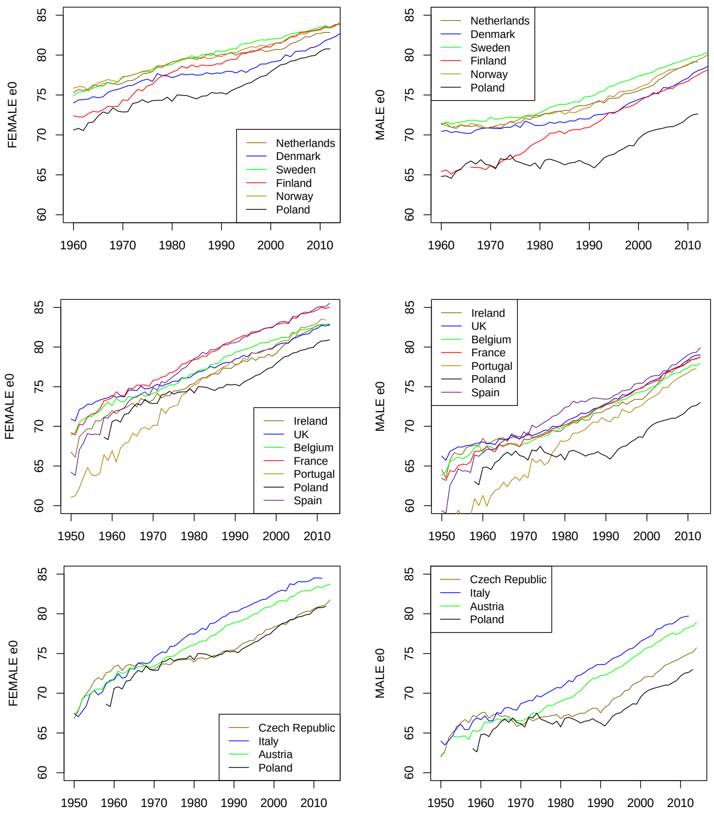


Fig. 1. Historical trends of life expectancy at birth in each cluster.

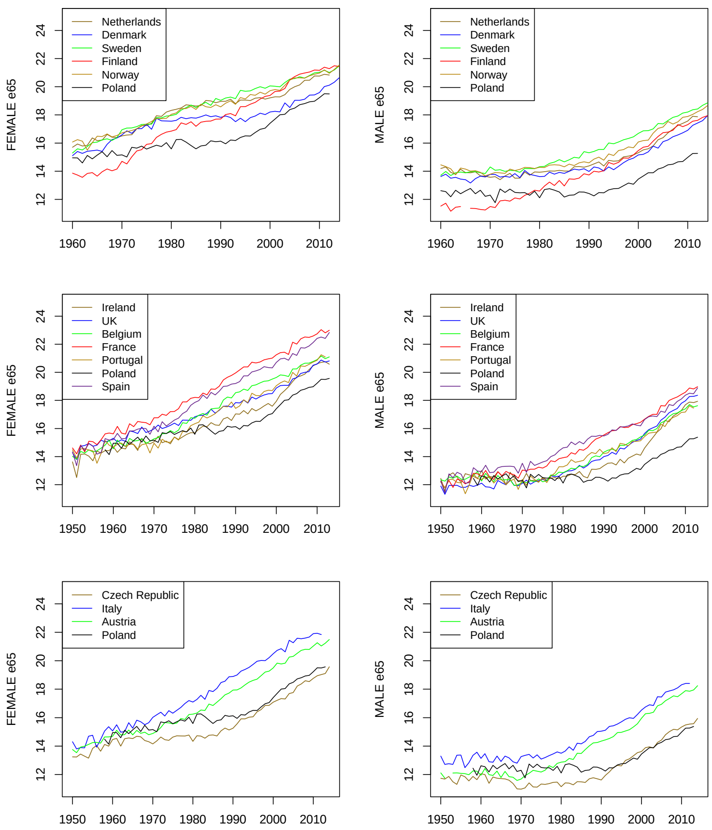


Fig. 2. Historical trends of life expectancy at 65 in each cluster.

Figures 3-5 show countries that has the highest period life expectancy by male and female, at birth[7]. Over much of the periods we observe very strong linear trends over time.

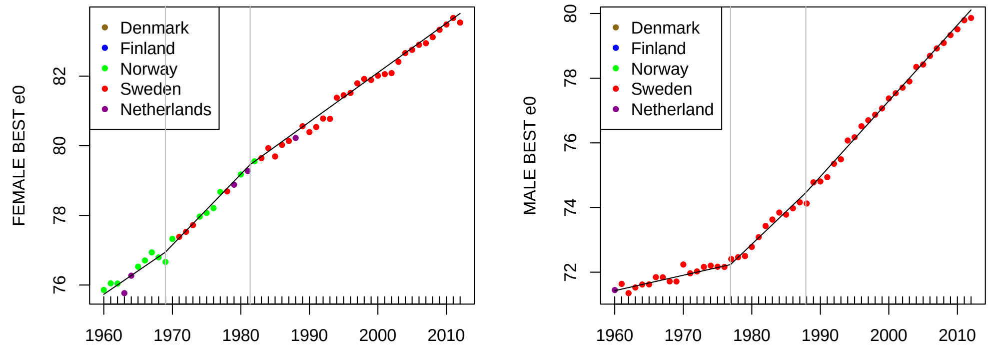


Fig. 3. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 1.

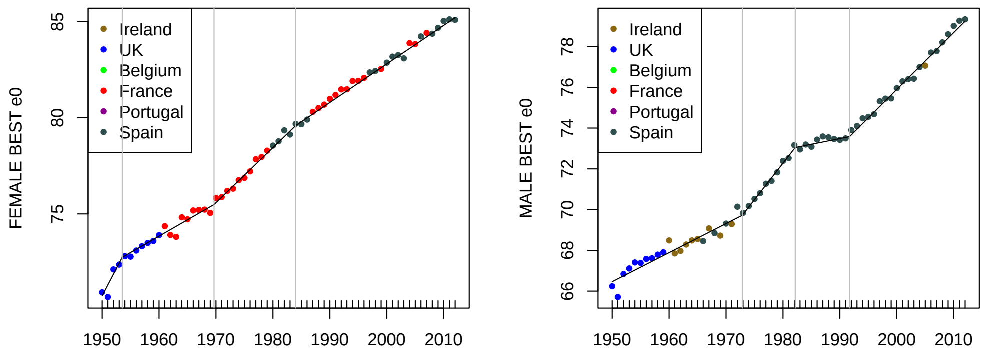


Fig. 4. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 2.

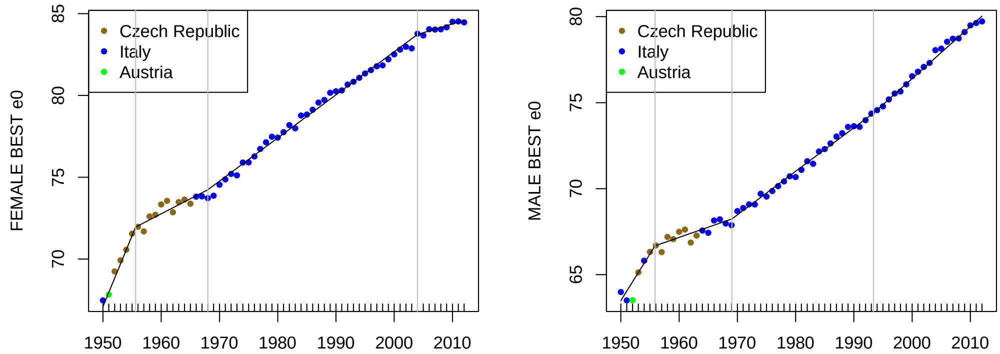


Fig. 5. Breakpoints in the trend of the highest countries life expectancies at birth and age 65, separately for males and females, in the cluster no. 3.

[7] Because the pages limitation, plots for the highest period life expectancy by male and female at age 65 are not presented in this paper.

Particularly interesting is the fact that life expectancy at birth for women has been slowing in the last decade, while life expectancy at birth for men has been accelerating. Acceleration in life expectancy is also observed by authors for people over age 65. In the first step the presence of differential rates of increase was investigated with Davies test (Davies, 2002). In the second step the break points were founded (marked as vertical lines in Fig. 4.-6.). As a result we got different fitted linear regressions between these break points. Followed by Medford (2015) the GEV model was fitted to the data beginning at the most recent break point, thus ensuring that the correct speed of life expectancy increase is captured as accurately as possible.

We fitted a time-dependent GEV model, GEV(μ_t,σ,ξ), where $\mu_t = \beta_0 + \beta_1 t$ is the location parameter, σ is the scale parameter, ξ is the shape parameter, and $t = 1...t_{max}$, where t_{max} is the last calendar year of the data set. Maximum likelihood method was used for parameters estimation. Results are presented in Table 2. – Table 4.

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	72.49 (0.11)	0.26 (0.01)	0.20 (0.03)	-0.07 (0.13)
Female e0	78.81 (0.03)	0.20 (0.00)	0.16 (0.02)	-0.57 (0.11)
Male e65	14.53 (0.05)	0.14 (0.00)	0.09 (0.02)	-0.16 (0.17)
Female e65	18.03 (0.03)	0.10 (0.00)	0.17 (0.05)	-0.80 (0.24)

Table 2. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 1).

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	73.39 (0.06)	0.28 (0.00)	0.18 (0.04)	-0.68 (0.29)
Female e0	80.71 (0.13)	0.11 (0.00)	0.91 (0.04)	-0.73 (0.21)
Male e65	15.30 (0.05)	0.16 (0.00)	0.21 (0.04)	-0.66 (0.30)
Female e65	20.88 (0.13)	0.15 (0.01)	0.20 (0.05)	-0.50 (0.22)

Table 3. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 2).

	$\hat{b}_0$	$\hat{b}_1$	$\hat{s}$	$\hat{x}$
Male e0	74.30 (0.07)	0.30 (0.01)	0.15 (0.03)	-0.11 (0.19)
Female e0	83.55 (0.12)	0.11 (0.00)	0.13 (0.01)	-0.82 (0.53)
Male e65	17.69 (0.04)	0.14 (0.02)	0.06 (0.01)	-1.03 (0.78)
Female e65	21.48 (0.00)	0.06 (0.00)	0.02 (0.01)	1.20 (0.65)

Table 4. Parameter estimates of the block maxima model (with standard errors in parentheses), at birth and age 65, separately for male and female (cluster no. 3).

Inspired by Medford (2015) projections of life expectancy for Poland were produced. These were done by updating the time-varying location parameter of the model μ_t for future values of time *t*. In order to make projections, the time varying GEV model was fitted and then the 50th percentile calculated (see eq. (2)). Projected values of the median of the fitted GEV distribution were used to approximate future values of life expectancy. Results are presented in Table 5. – Table 7.

Scenario 1	**2015***	**2025**	**2030**	**2035**	**2040**
Male e0	73.60	76.10	77.40	78.70	80.00
Female e0	81.60	81.55	82.55	83.55	84.55
Male e65	15.71	16.45	17.15	17.85	18.55
Female e65	20.06	19.46	19.96	20.46	20.96

Table 5. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Scenario 2	**2015***	**2025**	**2030**	**2035**	**2040**
Male e0	73.60	77.18	78.58	79.98	81.38
Female e0	81.60	82.87	83.42	83.97	84.52
Male e65	15.71	17.55	18.35	19.15	19.95
Female e65	20.06	23.01	23.76	24.51	25.26

Table 6. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Scenario 3	2015*	2025	2030	2035	2040
Male e0	73.60	78.37	79.87	81.37	82.87
Female e0	81.60	85.08	85.63	86.18	86.73
Male e65	15.71	19.55	20.25	20.95	21.65
Female e65	20.06	22.31	22.61	22.91	23.21

Table 7. Life expectancy projections (at birth and age 65) for Poland, separately for male and female (*actual value).

Projected life expectancy should not be discussed apart from official projections produced by Central Statistical Office of Poland (CSO, 2014) for 2015-2050. Mortality and life expectancy official projections are based on the target value (derived by comparison of life expectancy to selected countries from Western Europe), and were then calculated in three different scenarios of future development:

- medium scenario - "delay" of Polish mortality in relation to the developed countries will be maintained at the same level throughout the forecast period,
- low scenario - "delay" of Polish mortality will be remained at the same level until 2025; however, in subsequent years the reduction in mortality would be observed,
- high scenario - Polish mortality distance to the developed countries will gradually decline throughout the forecast period.

Under assumption that the future trend of life expectancy will be similar to trend observed for countries from cluster 1 (slow increase), we can expect that life expectancy at birth for female will increase from 81.60 in 2015 to 84.55 in 2040 (for male: from 73.60 to 80). According to the second scenario, where we assume that increase will be continued at a steady rate, life expectancy at birth for female will increase from 81.60 in 2015 to 84.52 in 2040 (for male: from 73.60 to 81.38). Acceleration in life expectancy (third scenario) will result in increase life expectancy at birth for female from 81.60 in 2015 to 86.73 in 2040 (for male: from 73.60 to 82,87). Meanwhile, according to the CSO of Poland, projected life expectancy at birth in 2040 increase to: (a) 85.9 (female) and 79.5 (male) under low scenario, (b) 86.5 (female) and 80.3 (male) under medium scenario, (c) 86.7 (female) and 80.9 (male) under high scenario.

Conclusions

We have applied a model that takes advantage of the past linear trends in life expectancy to make predictions about future life expectancy values. The model is based on block maxima method and uses generalized extreme distribution for modeling "best life expectancy" values.

Main advantage of this method is that it uses life expectancy levels for closely related populations. The method is worth further studying due to the reasonable results and – what is particularly interesting – comparable to the projections of NSO of Poland.

References

Assunção, R. M., Neves, M. C., Camara, G., & Freitas, C. D. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *Journal of Geographical Information Science, 20*, 797-811.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. *Springer Series in Statistics*.

Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika, 89*(2), 484-489.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling Extremal Events.

Fisher, R. A., & Tippett, L. H. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society, 24*(02), 180.

Gilli, M., & Këllezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics, 27*(2-3), 207-228.

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics, 44*(3), 423.

HMD. Retrieved February 04, 2017, from http://www.mortality.org/.

Lazar, D., Buiga, A., & Deaconu, A. (2016). Common stochastic trends in European mortality levels: testing and consequences for modeling longevity risk in insurance. *Romanian Journal for Economic Forecasting, 2*, 152-168.

Li, N., & Lee, R. D. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography, 42*(3), 575-594.

Medford, A. (2015). An Extremely different way to model and project Life Expectancy. Retrieved from http://epc2016.princeton.edu/uploads/160906

Oeppen, J., & Vaupel, J. (2002). DEMOGRAPHY: Enhanced: Broken Limits to Life Expectancy. *Science, 296*(5570), 1029-1031.

Population projection 2014-2050. (2014). Warsaw: CSO, Demographic Surveys and Labour Market Department. http://www.stat.gov.pl/

The 2015 Ageing Report. Underlying Assumptions and Projection Methodologies. (2014). *European Economy, 8*.

Foreign income of households and its determinants

Paweł Ulman[1]

Abstract

The purpose of this article is to determine and present the scale of foreign sources of income of Polish households and their significance for the budget and the financial situation of these households.

To achieve this purpose, individual data from the Household Budget Survey of 2015 will be used. In that edition of the survey, in addition to information about revenues from various sources (including 12 categories of foreign income), there were collected data concerning the direct assessment of the impact of foreign income on the financial situation of households.

To carry out a statistical analysis, theoretical models of income distribution have been used, including Burr distributions. They will allow to describe the distribution of income from abroad and the total income of households having income from abroad compared to households without such income. A censored data model has also been applied to identify determinants of income from foreign sources.

Keywords: *household, income, financial situation*

JEL Classification: C51, D31

1 Introduction

Over the last centuries, as well as the last decades, the history of Poland has been full of events that created favourable conditions for emigration or forced Polish people to emigrate to other countries. The loss of Poland's independence in the eighteenth century and unsuccessful attempts to regain the same in the nineteenth century led to mass emigration of Poles caused by political factors. The migration of Polish people for economic reasons began on a large scale in the late nineteenth century and the early twentieth century, when during the period from 1870 to 1914, approx. 4 million Poles emigrated to the USA (Biedka, 1995). The regaining of independence after the First World War did not put an end to emigration from Poland. As estimated by H. Janowska, approx. 2.1 million people left Poland during the interwar period (Janowska, 1981). The post-war period was characterized by a relatively low level of emigration. This phenomenon intensified in the 1980s. According to M. Okólski, the total number of those who left in that period amounted to approx. 2.2 million people (Okólski, 1994). The 1990s was a period of a decline in long-term migration in favour of short-term movements, mostly of economic nature. It is estimated that during that period (1990-1999), 216,000 people left Poland permanently (Bobrowska, 2013). After Poland acceded the

[1] Cracow University of Economics, Department of Statistics, 27 Rakowicka St., 31-510 Kraków, Poland, e-mail: ulmanp@uek.krakow.pl.

European Union, there was an increase in migration of Poles, mainly to EU countries. The migration of the population has become relatively easy, and the main objective of migrants was, and still is, to obtain better working conditions, improvement of material conditions and career development. Results of the last census carried out in Poland in 2011 showed that at the time of the census, more than 2,014 thousand permanent residents of Poland stayed abroad (for more than 3 months). The census also showed that 1,307 thousand households (9.6%) in the country included at least one person staying abroad for over 3 months. Among them, 52.2% were households in which not all of their members were abroad (GUS, 2015). Such a large scale of emigration, including that of economic nature, involves obtaining foreign income for the benefit of households in Poland. This income is transferred to Poland, and it often constitutes the basic source of income for families.

The purpose of this article is to determine and present the scale of foreign sources of income of Polish households and their significance for the budget and the financial situation of these households.

2 Remittances

Remittances are defined as household income received from abroad mainly as a result of gainful employment of those members of the household who stay temporarily or permanently in a country other than the country where the household resides (IMF, 2009). Remittances can be transferred in cash, via cash transfers, in kind or in other formal and informal ways.

For the purpose of an analysis of cash flows between a country from which funds are transferred and a country to which they flow, the term "personal remittances” is used that consists of: personal transfers and compensation of employees. Personal transfers include all current transfers in cash or in kind between resident and nonresident individuals, independent of the source of income of the sender (irrespective of whether the sender receives income from labor, entrepreneurial or property income, social benefits, and any other types of transfers; or disposes assets) and the relationship between the households (irrespective of whether they are related or unrelated individuals). Compensation of employees (wages and salaries in cash, wages and salaries in kind, and employers' social contributions) refers to the income of border, seasonal, and other short-term workers who are employed in an economy where they are not resident and of residents employed by nonresident entities. (IFM, 2009).

Papers addressing the remittance problem often consider it in the context of reasons that have led people to earning foreign income for the benefit of their households. As noted by Amuedo-Dorantes and Pozo (2011), in many developing countries, the changeability and volatility of income is a key reason for seeking such sources of revenues that will enable people to stabilize the economic situation of their households and raise the standard of living, in particular in the case of households experiencing poverty. Dean Yang and HwaJung Choi (2007) also showed that remittances are a response to one-time shocks (declines in income) that can help to stabilize the level of household income over time. However, there are more reasons for obtaining income from abroad. Remarks on migrants' motivation for transferring remittances can be found, for example, in (Amuedo-Dorantes and Pozo, 2011), (Bauer and Sinning, 2009) or (Ilahi and Jafarey, 1999). The problem of motivation for earning foreign income and the impact of such income on the situation of households and economic growth is widely discussed by Yang (2011).

Income generated abroad is of increasing importance. As Figure 1 shows, the level of personal remittances has grown dynamically, especially since 2000.

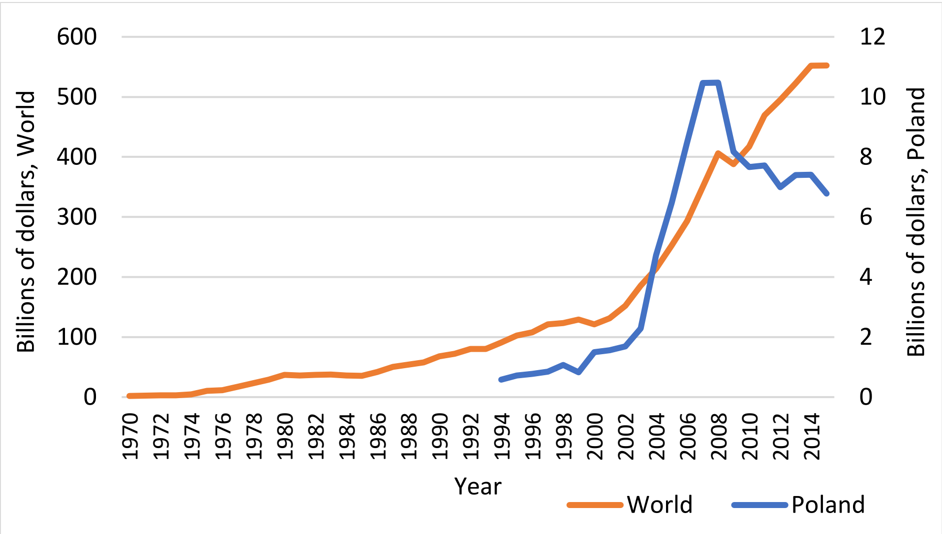


Fig. 1. Remittances received in billions of current US dollars for Poland and World.

Source: IMF.

In Europe, this has been a period of expansion of the European Union, and globally - of increased migration mobility resulting from enhanced communications possibilities, liberalization of administrative restrictions and, unfortunately, numerous conflicts in various parts of the world. In Poland, we could observe a noticeable increase in the level of remittances in the period after the accession to the European Union. The global economic crisis at the end of the first decade of the twenty-first century reduced the level of income sent to Poland from abroad. Over the last few years, a decreasing trend as regards the level of such income can be observed, which may result from the fact that a large part of migrants have put down roots in the country of their migration destination. A decision on the permanent stay abroad often involves the necessity for a family to join the person who has emigrated first, which naturally eliminates income remittances from abroad to Poland. Globally, the highest level of received remittances was recorded in 2015 in the case of India and China. These were amounts equal to 72.2 and 63.9 billion US dollars, respectively. Among countries sending remittances, the highest level of income flowed from the US and Saudi Arabia. In the case of Tajikistan, nearly 42% of GDP came from remittances, whereas Luxembourg was distinguished by the highest outflow of income in relation to GDP (19.6%). More than 25 billion US dollars were transferred only from the US to Mexico and another 16.3 billion from the US to China (World Bank Group, 2016).

3 Income from abroad on the basis of the Household Budget Survey in Poland

The empirical part of this paper has been developed using data from the HBS conducted in Poland in 2015. In total, 37,148 households were surveyed to acquire information on income, expenditure and characteristics of the households and their members. Under the study of sources of income earned by individual members of households, 12 categories of income received from abroad were distinguished, of which income from employment and income from self-employment are most important in terms of value and quantity. Comprehensively, income from working abroad in 1,023 surveyed households constitutes over 88% of all foreign income. The above 88% was achieved by 78% households. Among all the households, only 2.8% of them throughout Poland declared obtaining income from abroad. The largest share of such households was recorded in the Opolskie Voivodeship (8.0%), the Lubuskie Voivodeship (6.1%) and the Podkarpackie Voivodeship (5.9%), whereas the smallest - in the Mazowieckie Voivodeship (1.2%) and the Wielkopolskie Voivodeship (1.3%). Considering the category of the locality of residence, it can be noticed that along with an increase in the size of the locality, the share of households with income from foreign sources declines. The largest share of households with income from abroad, broken down by biological type, is on the part of households of other persons with dependent children (7.0%), households of married couples with at least one dependent child and other persons (5.8%) and married couples with three dependent children. The above-mentioned households of other persons with dependent children refer - as it can be assumed - to a situation where both parents work abroad and children in Poland are cared for by other persons (e.g. grandparents or other relatives). The share of households with income from abroad is also associated with evaluation of the financial situation and the manner of spending money. The higher the assessment of this situation, the more common are households with income from abroad. The largest share of households with income from abroad relates to a situation where a household declares that it can afford some luxuries.

As part of the HBS in 2015, for the first time there were collected data on the extent of income from abroad and its impact on the economic situation of households. In the first case, respondents were asked what part of household income is formed by income from work abroad, and in the second case - how income from work abroad influences the ability to purchase durable goods and real property. In general, slightly more than 29% of households with income from abroad stated that such income constituted their entire income, and nearly 47% - that it formed more than half of their income. The largest share of households declaring that their entire income came from work abroad (almost 43%) was observed in the Pomorskie Voivodeship, while the smallest (just over 20%) - in the Lubelskie Voivodeship. It is worth noting that along with an increase in the number of dependent children, income from work abroad becomes increasingly important for the household budget. It can also be noted that along with an improvement in the financial situation and positive evaluation of the manner of spending money, there is an increase in the share of households with budgets more dependent on income from work abroad. Very similar conclusions can be drawn from an analysis of distribution of answers to a question about the influence of income from work abroad on the ability to purchase durable goods and real property.

In the next part, results of an analysis of income distribution will be presented briefly. Such a distribution can be examined empirically or with the use of theoretical functions. The application of the second approach is beneficial in terms of relatively easy calculation of characteristics of the distribution; however, it requires the selection of an appropriate density function and fitting it to empirical data. The following part shows values of the said characteristics of distribution using the Burr Type III (Dagum) theoretical distribution, whose parameters were estimated using ML[2]. In each case, the estimated parameters were found to be statistically significant. Table 1 presents results of calculations of basic characteristics of income distribution of all households altogether, income distribution of households having income from abroad and those that do not have such income. Analogical calculations were carried out for per capita income and equivalent income, using the OECD scale. Furthermore, the distribution of income from abroad was analysed in these three categories.

[2] The subject of theoretical distributions and their application in an analysis of income is widely discussed in (Kleiber nad Kotz, 2003). On the other hand, McDonalda and Xu (1995) indicates that the Burr III distribution is one of the best distributions to approximate empirical distributions of income.

Descriptive statistics	Total income of household	Total income of household with remittances	Total income of household without remittances
Mean	3,973.46	5,279.88	3,933.12
Median	3,312.58	4,716.36	3,277.05
Mode	2,635.41	4,124.68	2,605.59
Coefficient of Variation	0.81	0.61	0.807
Relative mean deviation	0.244	0.206	0.244
Gini Coefficient	0.347	0.294	0.348
Skewness	0.418	0.361	0.418
Sen Index	2,592.82	3,726.53	2,566.27

Table 1. Descriptive statistics for the distribution of household total income.

An average level of household income is significantly higher in the case of households with income from abroad, and the level of inequality is lower compared to households without such income. This results in a significantly higher value of the Sen index, which, in a simplified manner, measures the level of social welfare resulting from income distribution[3]. When the distribution of per capita income is considered, the difference in the average level of income is not so significant any more, though the level of inequality is still lower in the case of income of households with income from abroad. The reduction in the difference between average values of per capita income results from the fact that households with income from abroad are characterized by a larger number of people, often dependent children. A diversified average level of equivalent income confirms this conclusion. The equivalence scale transforms the number of people in a household into a number of equivalent units the number of which is lower than the number of people in the household. It results in an increase in the equivalent income in terms of per capita income, especially in multi-person households with dependent children. In addition to the above categories of income, the distribution of income from abroad was also analysed. Its average level per household amounted to PLN 3,375, per persons - PLN 921 and per equivalent unit - PLN 1,235. The level of

[3] The Sen index (as the result of the following formula S = average income*[1-Gini coefficient]) allows to compare the level of social welfare, the higher the level of Sen index, the higher the level of social welfare.

inequality was relatively high in this case. The value of the Gini coefficient ranged from 0.367 to 0.387, which shows that income of Polish households generated abroad is relatively highly diversified.

In order to determine factors affecting the amount of income from abroad, the Heckman sample selection model was applied. In this model, it is assumed that the criterion on the basis of which observations are included in a sample depends on an additional regression equation (selection equation). If it does not depend on that, the proper method for estimating the regression equation (output equation) is OLS[4]. As an explanatory variable, a natural logarithm of household foreign income (log y_1) was adopted, whereas the following were considered as explanatory variables: 5 dummy variables identifying individual categories of localities where households reside (the reference category is the largest city), 3 dummy variables identifying the level of education of the household head (the reference category is lower secondary education), 3 variables defining the number of people in the household (a number of children up to the age of 14, a number of dependent children between 14 and 25 years of age, a number of adults), one dummy variable identifying whether the household head is disabled and one variable which is a logarithm of age of the household head. In the selection equation, the following variables were included: the category of localities where households reside (6 categories), the level of education of the household head (4 categories), an assessment of the financial situation of the household (5 categories) and the age of the household head. Results of the Heckman regression estimation with the use of ML are summarized in Table 2.

The parameter (rho) indicates that error terms of the selection equation and the output equation are strongly correlated, which justifies the use of a model with sample selection. All estimates of regression parameters were found to be statistically significant. Along with a decrease in the size of localities where households reside, an average level of income from foreign sources decreases (e.g. in the villages these revenues are lower by 40,9%[5] than in the largest cities), while it increases along with an increase in the level of education of the household head (in the case of higher education by 114% than for primary educated household head). An increase in the number of children under the age of 14 and between 14 and 25 years of age, as well as an increase in the number of adults in the household weighed in favour of an increase in the average foreign income in a household, which is consistent with the results presented above.

[4] More on the Heckman sample selection model can be found for example in (Green, 1993).
[5] exp(-0.5251) - 1 = -0.409.

Regressors	Coefficient	Standard error	Z	p-value	
Output equation: $\log y_1 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + \varepsilon$					
Constant	8.7022	0.4404	19.75	<0.0001	***
Town 200-500 thous. inhabitants	−0.3025	0.1487	−2.03	0.0420	**
Town 100-200 thous. inhabitants	−0.4222	0.1534	−2.75	0.0059	***
Town 20-100 thous. inhabitants	−0.3669	0.1434	−2.56	0.0105	**
Town less than 20 thous. inhabitants	−0.4953	0.1461	−3.39	0.0007	***
Village	−0.5251	0.1450	−3.62	0.0003	***
Vocational education	0.3799	0.0656	5.80	<0.0001	***
Secondary education	0.5161	0.0816	6.33	<0.0001	***
Higher education	0.7611	0.1164	6.54	<0.0001	***
Number of children age up to 14 years	0.0640	0.0221	2.89	0.0038	***
Number of children aged 14 to 25 years	0.0559	0.0314	1.77	0.0760	*
Number of adults	0.0630	0.0250	2.52	0.0116	**
Disabled person	−0.2229	0.1090	−2.05	0.0408	**
Log of age	0.6379	0.1240	5.14	<0.0001	***
lambda	−1.5911	0.0805	−19.76	<0.0001	***
Sample selection equation: $y_2 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_s z_s + \zeta$					
constant	−0.4774	0.1001	−4.77	<0.0001	***
Class town of residence	0.0521	0.0085	6.12	<0.0001	***
Education level	−0.1713	0.0165	−10.39	<0.0001	***
Financial situation	−0.2034	0.0142	−14.38	<0.0001	***
Age	−0.0139	0.0010	−14.44	<0.0001	***

rho1 = -0.96; Akaike'a criterion = 11055.23; Sigma (Se) = 1.65.

Table 2. The ML estimation results of the Heckit model.

In households in which their head is a disabled person, the level of income from abroad was lower on average by 20% than in other households, while an increase in the age of the household head by 1% involved an increase in the average level of the surveyed income by 0.64%, *ceteris paribus*. Moreover, results of the estimation of parameters of selection equation showed that the existence of income from abroad in household depends on class town of residence, educational level of household head and financial situation in the household.

Conclusions

The problem of income from abroad is an important research issue in both the macroeconomic scale of financial flows among countries and in the microeconomic scale of its impact on the economic and financial situation of the population (households). The presented research results indicate high significance of this income for the total level of income of households in Poland and, as a consequence, the financial situation of these households. This is supported by a lower level of poverty risk in the case of households with income from abroad than in the case of households without such income (33.7% of households with income from abroad and 37.8% of those without such income did not reach the subsistence level).

To sum up, the highest share of households receiving income from abroad can be observed in the case of the Opolskie Voivodeship (inhabited by a population having contacts with Germany in great measure), in the case of households living in rather small towns, households where parents do not attend directly to their children and households assessing positively their financial situation. The level of income of households with income from abroad is significantly higher than that of households without such income, though in the case of per capita income, the difference ceases to exist. However, when diversification of the household composition is considered (by using the equivalence scale), income from abroad leads to higher average income in total. The application of regression modelling basically confirmed the previously observed regularities.

Acknowledgements

Publication was financed from funds allocated to Faculty of Management (Cracow University of Economics) within grants to maintain research capacity.

References

Amuedo-Dorantes, C., & Pozo, S. (2011). Remittances and Income Smoothing. American Economic Review, 101(3), 582-587.

Bauer, T., & Sinning, M. (2009). The Purpose of Remittances: Evidence from Germany. Jahrbücher für Nationalökonomie und Statistik, 229(6).

Biedka, B. (1995). Osadnictwo polonijne w ośrodkach miejskich na północno-wschodnim wybrzeżu Stanów Zjednoczonych Ameryki Północnej. In Migracje i społeczeństwo. Zbiór Studiów (pp. 83-100). Warszawa: Instytut Historii PAN.

Bobrowska, A. (2013). Migracje Polaków po przystąpieniu do Unii Europejskiej. Colloquium Wydziału Nauk Humanistycznych i Społecznych, Kwartalnik,2, 49-64.

Green, W. H. (1993). Econometric Analysis (2nd ed.). Prentice-Hall International, Inc.

GUS. Gospodarstwa domowe i rodziny z osobami przebywającymi czasowo za granicą - notatka informacyjna (Issue brief). Retrieved February 10, 2017, from GUS website: http://stat.gov.pl/obszary-tematyczne/ludnosc/migracje-zagraniczne-ludnosci/gospodarstwa-domowe-i-rodziny-z-migrantami-nsp-2011,12,1.htm

IMF. (2009). International transactions in remittances: guide for compilers and users. Washington, D.C.: International Monetary Fund.

Ilahi, N., & Jafarey, S. (1999). Guestworker migration, remittances and the extended family: evidence from Pakistan. Journal of Development Economics,58(2), 485-512.

Janowska, H. (1981). Emigracja zarobkowa z Polski 1918–1939. Warszawa: PWN.

Kleiber, C., & Kotz, S. (2003). Statistical size distributions in economics and actuarial sciences (1st ed.). Hoboken (New Jersey): Wiley-Interscience.

Mcdonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. Journal of Econometrics,66(1-2), 133-152.

Okólski, M. (1994). Migracje zagraniczne w Polsce w latach 1980-1989. Zarys problematyki badawczej. Studia Demograficzne,3, 45-63.

World Bank Group. (2016). Migration and remittances. Factbook 2016. Retrieved February 10, 2017, from
https://openknowledge.worldbank.org/bitstream/handle/10986/23743/9781464803192.pdf

Yang, D. (2011). Migrant Remittances. The Journal of Economics Perspectives,25(3), 129-152.

Yang, D., & Choi, H. (2007). Are Remittances Insurance? Evidence from Rainfall Shocks in the Philippines. The World Bank Economic Review,21(2), 219-248.

Modelling the distribution of loan repayments of households in Poland

Agnieszka Wałęga[1], Grzegorz Wałęga[2]

Abstract

A rapid growth of household debt in Poland has been observed since the beginning of the twenty-first century. Households more often buy on credit to cover current expenditure and the purchase of durable goods. This situation makes the household budget appears the item "debt repayment".

The purpose of the article is the choice of theoretical models which reflect as closely as possible the distribution of debt repayments (credit and loans) by households. The study included theoretical distributions used in the analysis of the distribution of wages and income. The most widely used distributions selected for the purpose of the research were: Burr distribution (type III), log-logistic and log-normal distribution. The distributions were tested for consistence of estimation of measures of position, variability and inequality of the empirical distribution of debt repayments. The distributions were tested for consistence of estimation of measures of position, variability and inequality of the empirical distribution of the debt repayments.

The source data consisted of individual information relating to monthly debt repayments in obtained as part of the household budget survey in Poland in 2015.

The results indicate the usefulness of the application of theoretical distributions modelling income distribution in modelling distributions of debt repayments.

Keywords: *debt repayments distribution, household indebtedness, income distribution models*

JEL Classification: D12, D14, D31

1 Introduction

The twenty first century sees large increase in consumer credits in most European countries. Indebtedness and bankruptcy in households resulting from growing consumer credit use are rapidly increasing in Western societies (Kamleitner and Kirchler, 2006). In Denmark and the Netherlands, household indebtedness is at a record high (gross debt-to-income ratio of households exceeds 200 percent). The volume of household debt burden in new member states of the European Union has doubled, tripled or even quadrupled between 2004 and 2015 (OECD, 2017). Households are increasingly ready to use credits to meet the living standards in developed countries. The same situation is observed in Poland. Between 2004 and 2014, the amount of outstanding debt rose from PLN 98.7 billion to PLN 587.0 billion. In terms of average annual real growth, it is over 19 percent.

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27, 31–510 Kraków, e-mail: agnieszka.walega@uek.krakow.pl.

[2] Cracow University of Economics, Department of Microeconomics, Rakowicka 27, 31–510 Kraków, e-mail: grzegorz.walega@uek.krakow.pl.

Credit use by households has various consequences on economic and political levels. For many households, debt repayments become considerable items in the household budget. At the same time, a growing number of borrowers is reporting financial distress and problems with their debt repayments. Currently, personal indebtedness is enjoying public attention. Monetary policymakers have become concerned about the extent of household indebtedness and impact on aggregate economic indicators. While the outstanding amount of household debts can be easily monitored by using administrative data from the financial system, there is no detailed data on the distribution of household debt at the micro level. This paper is an attempt at a closer analysis of this issue.

2 Debt repayment and indebtedness

The phenomenon of widespread household indebtedness is caused by several factors. Consumers rely on credit arrangements because they can implement consumption plans, which at the current level of income are not available to them (Estelami, 2001). Generally, credit use has become socially acceptable. Moreover, in the era of globalization hedonic values as well as an increase in orientation towards the present are important driving forces of consumer behaviour in Western societies (Wood 1998; Watson 2003). Consequently, credit use no longer means investing in one's personal future, but rather has made it acceptable and accessible to acquire luxury goods and other services (Kamleitner and Kirchler, 2006). Also, easy availability of credits facilitates compulsive consumption, which is a characteristic of modern societies (for instance, see Neuner et al., 2005).

The understanding of household repayment distribution is a central question for researchers and policy makers. The economic slowdown in Europe and subprime mortgage crisis in the USA demonstrated that large segments of households are unable to repay their debts (Demyanyk and Hemert, 2009). Naturally, a revived interest in household default behaviour has appeared. The key to the better understanding of the process of excessive debt is to know the distribution of debt repayments. This will recognize appropriate instruments to ensure macroeconomic stability and allow for supporting effectively the households having difficulty repaying their debts. To understand arrears we must analyse both market sides: the households which take loans and the lender's decision to grant the loan (Grant and Padula, 2016). Political debates have also emphasized an important role which is likely to be played by financial market institutions (Duygan and Grant, 2009).

There are two ways of examining household debt. The use of administrative data that are provided by lenders allows for more reliable estimation of debt repayments since the lenders keep accurate records of credit arrangements, arrears and default. However, the high level of data aggregation has some limitations. These collections are difficult to obtain and they do not allow for a detailed analysis of the behaviour of households on the credit market. Antczak (2013) writes that in order to understand household consumption, the possibility of consumer reaction to the financial shock and the ability to service the debt by households is necessary to analyse the data at the micro level. The best sources of this type of data are cross-sectional surveys, in which one of the modules (sections) is indebtedness. Similar observations for the USA are formulated by Dynan et al. (2003).

In our paper, we focus on debt repayments which reflect the burden of income and determine the current consumption level. The analysis of debt repayment distribution can also provide guidelines on how to identify a household that has too much debt. There have been a few studies that have examined this issue by constructing various indicators of over-indebtedness. In the literature, there are four common indicators that are used to test over-indebtedness: making high repayments relative to income, being in arrears, making heavy use of credit and subjective perception of debt burden (see: Kempson et al., 2004; D'Alessio and Iezzi, 2013). The first two capture the burden imposed by debt repayments and put arbitrary limits on repayments relative to gross income, beyond which they are thought to represent a significant burden for households. As a complement to the above set of indicators, debt repayment distribution is proposed.

Finding the best possible model of debt repayment distribution allows one to carry out a detailed and reliable evaluation of the debt repayment distribution, examine its characteristics thoroughly and compare it with other repayment distributions.

3 Selected models of income distribution

Research conducted by the authors on the shape of household debt repayments in Poland has confirmed that the repayment distribution is characterized by unimodality and positive skewness. Therefore, it seems reasonable to apply theoretical models used for describing distribution of wages (income) to model the household debt repayment distribution.

Research on income distribution has over a hundred years of history in economic theory, which is why this problem has received many theoretical solutions. Using a mathematical function to describe income distribution was first proposed by Pareto (1895). In literature on the subject, the following distributions can often be met: the log-normal distribution popularized in the area of research on income by Aitchison and Brown (1957), distributions of the Burr type - the Burr Type XII (Singh-Maddala) distribution described by Singh and

Maddala (1976) and the Burr III (Dagum) distribution described by Dagum (1977), as well as the log-logistic distribution (Fisk, 1961). Other models that have been used to describe the income distribution included the Gamma distribution, 2-parameter Weibull distribution, Beta distribution and generalized Beta distribution of the second kind described by McDonald (1984), special cases of which are the Burr Type XII and Burr Type III distributions.

In this paper, to describe the household debt repayment distribution, the log-normal distribution, the log-logistic distribution and the Burr type III distribution were used, a brief description of which is presented below[3].

The density function of the log-normal distribution is recorded as follows:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(\ln(y)-\mu)^2}{2\sigma^2} \right\}, \tag{1}$$

where: μ, σ - distribution parameters (μ - location parameter, σ - shape parameter).

In the case of both empirical and theoretical distributions, their analysis is based primarily on the determination of values of distribution characteristics, which are measures of the distribution position, measures of variability, measures of asymmetry and measures of concentration. The expected value can be determined using the formula:

$$E(Y) = \exp\left(\mu + \frac{\sigma^2}{2} \right), \tag{2}$$

and the standard deviation:

$$D(Y) = \sqrt{\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1]}\ . \tag{3}$$

Another theoretical distribution used in the analysis of the shape of the debt repayment distribution was the Fisk log-logistic distribution. The density function in the log-logistic distribution can be recorded as follows:

$$f(y) = \frac{\beta\exp(-(\alpha+\beta\ln y))}{y[1+\exp(-(\alpha+\beta\ln y))]^2}, \tag{4}$$

where α and β are distribution parameters.

The expected value in this distribution is obtained using the formula:

$$E(Y) = \frac{\pi}{\beta\sin\left(\frac{\pi}{\beta}\right)} \exp\left(-\frac{\alpha}{\beta} \right), \tag{5}$$

[3] The model description was prepared based on: (Kot, 1999), (Ulman, 2011) and (Salamaga, 2016).

and the standard deviation:

$$D(Y) = \sqrt{(E(Y))^2 \left[\frac{\beta}{\pi} tg\left(\frac{\pi}{\beta} \right) - 1 \right]}. \tag{6}$$

Another theoretical distribution frequently used to describe the shape of the income distribution is the Dagum (Burr III) distribution. The density function of this distribution can be recorded as follows:

$$f(y) = \frac{cb\exp(-a)y^{-(b+1)}}{[1+\exp(-a)y^{-b}]^{c+1}} , \tag{7}$$

where *a, b, c* are distribution parameters.

The ordinary moment of order *r* in this distribution is expressed as follows:

$$m_r = B\left(1-\frac{r}{b}, c+\frac{r}{b}\right)\exp\left(-\frac{a}{b}r\right)c, \tag{8}$$

where: $B(p,q)$ is the Euler's beta function. Using the formula (10), the average value of income can be determined as the ordinary moment of order $r = 1$, and the standard deviation - using the formula:

$$D(Y) = \sqrt{m_2 - (m_1)^2}. \tag{9}$$

The fundamental issue for the practical application of theoretical functions as models of wage distributions is the knowledge of distribution parameters. Among many estimation methods, the most frequently used is the maximum likelihood estimation (MLE), which provides consistent, asymptotically unbiased, asymptotically efficient and asymptotically normal estimators of parameters.

To assess the matching rate of the theoretical distribution and the empirical debt repayment distribution, there were applied values of the Akaike information criterion.

It should be emphasized that the fact that theoretical distribution fits specific empirical data better does not mean that this distribution will also be the best model in case of other data (for a different community, in a different period, for a different phenomenon). For this reason, it is worth considering various theoretical distributions to describe empirical debt repayment distributions.

4 Data and research results

The research used data from a household budget survey conducted by the Central Statistical Office in 2015. Members of households participating in the study declared revenues that had involved taking out a credit or a loan and expenses (expenditure) on the repayment of debts.

Since 2005, debt service expenses have been provided broken down by mortgage loans, loans relating to the use of credit cards, as well as other bank borrowings, repayments of debts contracted at institutions other than banks and from private individuals.

Amounts of household income and expenditure that relate to loans and credits are given in the total amount, i.e. without breakdown by the principal part and interest part.

For the purpose of the study, it was assumed that an indebted household was a household that met at least one of the following conditions:

- during the examined period, it took out a mortgage loan, a loan in a bank using a credit card, another loan in a bank, a loan from another institution or a cash loan from private individuals;
- during the examined period, it repaid a principal instalment of and/or interest on a mortgage loan, a loan in a bank taken out using a credit card, another loan taken out in a bank, a loan taken out from another institution, a cash loan taken out from private individuals;

In the study, there was used a repayment amount of the principal instalment of and/or interest on the loan. The analysis included only those households whose amount of debt repayments did not exceed income obtained in the month of the study. In 2015, such households constituted 25.3% of the surveyed households. On the other hand, households whose debt repayments exceeded the amount of the obtained income amounted to 1.2% of the indebted households.

In 2015, the surveyed households spent PLN 748.24 per month, on average, on debt repayment, while half of the households spent not more than PLN 500 per month on debt repayment. However, it should be emphasized that the amount of debt repayments is characterized by a great diversity and significant right-tailed asymmetry (Table 2). There can also be observed considerable disparities of the distribution, measured with the use the Gini coefficient.

In order to estimate parameters, the GRETL econometric package was used. All the estimated parameters of distributions were found to be statistically significantly different from zero – in each case, the test probability is significantly lower than any reasonably assumed level of significance (Table 1).

Based on the value of the likelihood function and the Akaike criterion, it can be stated that for the examined set of data, the distribution of loan repayments was best reflected by the Dagum (Burr III) theoretical distribution, followed by the Fisk (log-logistics) distribution.

Considering the above matching rates of the model and the actual data, the log-normal distribution proved to be worst fitting.

parameter	estimator	st. error	z	p-value
Log-normal distribution				
(Log-likelihood=-79960.96; Akaike=159926; Schwarz=159940)				
μ	6.1650	0.0104	594.22	<0.0001
σ	0.9915	0.0073	136.63	<0.0001
Log-logistic distribution				
(Log-likelihood=-71448.38; Akaike=142901; Schwarz=142915)				
α	-10.8106	0.1090	-99.19	<0.0001
β	1.7467	0.0177	98.60	<0.0001
Burr III distriution				
(Log-likelihood=-71432.22; Akaike=142870; Schwarz=142892)				
a	-17.1700	0.0185	-927.46	<0.0001
b	2.5180	0.0024	1031.79	<0.0001
c	0.4301	0.0007	638.04	<0.0001

Table 1. Estimation of parameters of the log-normal, log-logistic and Burr Type III distributions.

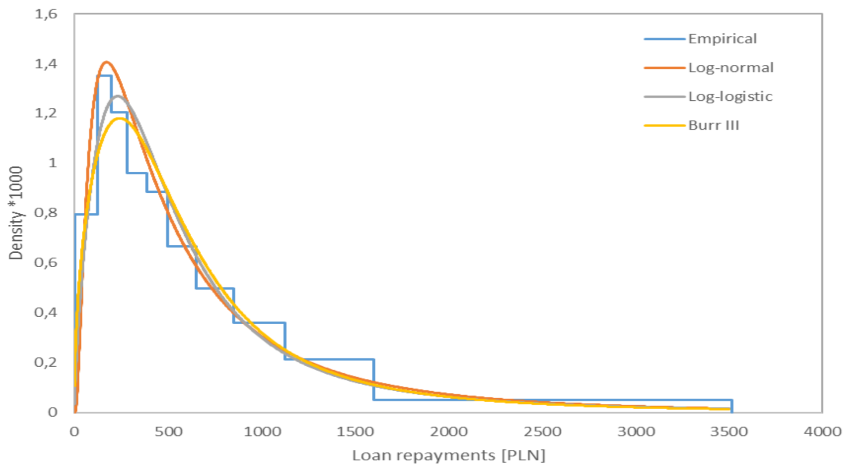


Fig. 1. Empirical distribution and theoretical the log-normal, log-logistic and Burr type III distributions of loan repayments.

Figure 1 shows the course of the density function of the individual theoretical distributions in relation to the empirical decile distribution. In the empirical distribution, the first class of loan repayments was closed on the left side with the zero value, while the final class with the value of 3515.20 determined based on the knowledge of the average value.

specification	**distribution**			
	empircal	**log-normal**	**log-logistic**	**Burr III**
Mean	748.24	777.85	900.06	747.44
Median	500.00	475.80	487.49	527.05
Mode	184.61	178.02	231.26	206.58
Coef. of variation	1.10	1.29	0.76	1.40
Gini Index	0.488	0.517	0.573	0.489
Skewnees	0.683	0.596	0.983	0.517

Table 2. Numerical characteristics of the empirical distribution and theoretical distributions of debt repayments.

In Table 2, there are results of calculations of relevant characteristics of the empirical debt repayment distribution. In this case, the mode was determined based on the grouped data (decile distribution), while other characteristics were determined on the basis of individual data. The comparison of characteristics of the empirical distribution and individual theoretical distributions enabled the authors to evaluate how these distributions fit the actual data.

When analysing the numerical characteristics calculated on the basis of the estimated theoretical models, it is hard to state unequivocally which model gives the best results. Undoubtedly, as regards the mean and the Gini coefficient, the Burr III distribution has values most similar to the empirical distribution. In the case of the evaluation of the coefficient of variation level, the log-normal distribution is closer to the empirical value. General principle states that the relative differences between descriptive parameters should not exceed 5%. In our analysis mean, median and mode of log-normal distribution and mean, median and Gini index of decomposition Burr type III comply only with the above principle. The largest differences in characteristics can be found in the log-logistic distribution.

Conclusions

The rising level of household debt and the burden on their budgets repayment of loans affects the long-term processes of consumption. In the context of emerging, not only in the period of crisis, insolvency households particularly interesting is the analysis of the wider distribution of repayment of loans. Calculated numerical characteristics indicate on positive skewness of repayments of loans in 2015, with an average value repayment in the amount of 750 PLN.

Comparison of characteristics of particular theoretical distributions with the empirical distribution shows that for, each of them has certain drawbacks. In the case of log-normal and log-logistic distributions was overstated the value of the mean, while in the distribution of log-logistic and Burr type III inflated value is mode. In turn, the distribution log-normal and log-logistic undercut the value of the median. The evaluation of model fitting based on descriptive parameters indicates that the best model from the proposed ones is Burr III distribution. However, it should be noted, that methods of assessing goodness of fit yielded inconclusive results. It all makes the research on modelling the distribution of debt repayments need to continue. It would be appropriate to compare methods for parameter estimation and the inclusion of analysis of other models used for example in the analysis of distribution of income.

Acknowledgements

This paper was supported by funds from the National Science Centre, Poland through grant No. 2015/19/D/HS4/02569.

References

Aitchison, J., & Brown, J. A. (1957). *The lognormal distribution with special reference to its uses in economics*. Cambridge: University Press.

Antczak, R. (2013). Koncepcja i pomiar zasobności gospodarstw domowych. *Wiadomości Statystyczne*, GUS, *11*(630).

Dagum, C. (2008). A New Model of Personal Income Distribution: Specification and Estimation. *Modeling Income Distributions and Lorenz Curves*, 3-25.

D'Alessio, G., & Iezzi, S. (2013). Household Over-Indebtedness: Definition and Measurement with Italian Data, Bank of Italy Occasional Paper No. 149.

Demyanyk, Y., & Hemert, O. V. (2009). Understanding the Subprime Mortgage Crisis. *Review of Financial Studies*, *24*(6), 1848-1880.

Duygan-Bump, B., & Grant, C. (2009). Household debt repayment behaviour: what role do institutions play? *Economic Policy*, *24*(57), 107-140.

Dynan, K., Johnson, K., & Pence K. (2003). Recent changes to a measure of US household debt service, *Federal Reserve Bulletin*, *89*(10).

Estelami, H. (2001). Determinants of discount rates in consumer credit decisions. *Journal of Marketing Theory and Practice*, *9*(1), 63-73.

Fisk, P. R. (1961). The Graduation of Income Distributions. *Econometrica*, *29*(2), 171.

Grant, C., & Padula, M. (2016). The repayment of unsecured debt by European households. *Journal of the Royal Statistical Society: Series A (Statistics in Society).*

Kamleitner, B., & Kirchler, E. (2007). Consumer credit use: a process model and literature review. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *57*(4), 267-283.

Kempson, E., McKay, S., & Willitts, M. (2004). Characteristics of Families in Debt and the Nature of Indebtedness, Corporate Document Services, DWP Research Report.

Kot, S. (1999). Teorie genezy rozkładu płac. In *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji* (Kot S.M. ed.). Warszawa-Kraków: Wydawnictwo Naukowe PWN.

Mcdonald, J. B. (1984). Some Generalized Functions for the Size Distribution of Income. *Econometrica*,*52*(3), 647.

Neuner, M., Raab, G., & Reisch, L.A. (2005). Compulsive buying in maturing consumer societies: an empirical re-inquiry. *Journal of Economic Psychology*, *26*(4), 509-522.

OECD (2017), Household debt (indicator). (Accessed on 08 February 2017)

Salamaga M. (2016). *Badanie wpływu metody estymacji teoretycznych modeli rozkładu dochodów na jakość aproksymacji rozkładu dochodów mieszkańców Krakowa*, Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie *3*(951), 63-79.

Sierminska, E. (2014). Indebtedness of households and the cost of debt by household type and income group, Research note 10/2014, European Commission.

Singh, S. K., & Maddala, G. S. (2008). A Function for Size Distribution of Incomes. *Modeling Income Distributions and Lorenz Curves*, 27-35.

Ulman, P. (2011). *Sytuacja ekonomiczna osób niepełnosprawnych i ich gospodarstw domowych w Polsce*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

Watson, J. J. (2003). The relationship of materialism to spending tendencies, saving, and debt. *Journal of Economic Psychology*, *24*(6), 723-739.

Wood, M. (1998). Socio-economic status, delay of gratification, and impulse buying. *Journal of Economic Psychology*, *19*(3), 295–320.

Uncertainty of the dependence structure and risk diversification effect in Solvency II

Stanisław Wanat[1], Krzysztof Guzik[2]

Abstract

This paper is an attempt at estimating the "actual level" of the diversification effect in the process of determining Solvency Capital Requirements (SCR) in Solvency II. At first, the method of determining SCRs in Solvency II is briefly characterised and the role of dependences for the correct specification of diversification effect is presented. This is followed by an analysis of the sensitivity of diversification effect to the dependence structure based on the example of life and health underwriting risk. Cases of the lack of knowledge on the structure, of partial knowledge (of a correlation coefficient only) and of total knowledge are considered (it was assumed that variables are independent and comonotonic).

Keywords: *Solvency Capital Requirements, risk aggregation, VaR bounds, diversification effect*

JEL Classification: C150, C580, G220

1 Introduction

The estimation of the diversification effect on the "actual level" is depended by the proper modelling of dependences among risk factors. Misidentified dependence structure leads to the estimation of the incorrect level of diversification effect, which may cause overestimation or underestimation of capital requirements and may have a considerable impact on the functioning of an insurer and its solvency. In the standard model proposed in the Solvency II system, the variance-covariance method is used to aggregate capital requirements. In this method, dependence is modelled only with the use of linear correlation coefficients. The influence of such a solution on the level of solvency capital requirement of insurers in European Union states was assessed in the fifth Quantitative Impact Study (QIS5). It indicates (see: (*EIOPA Report*..., 2017)) that the diversification effect obtained as a result of applying the aggregation method considerably influences the reduction of solvency capital requirements. In total, such requirements for solo insurers and groups of insurance companies participating in the study were lower by 35.1% (EUR 466 billion).

From the methodological point of view, the variance-covariance method is correct when capital requirements are determined for risk factors subject to multivariate normal (elliptical)

[1] Corresponding author: Cracow University of Economics, Department of Mathematics, 27 Rakowicka St. 31-510 Cracow, Poland, e-mail: wanats@uek.krakow.pl.

[2] Cracow University of Economics, Department of Mathematics, 27 Rakowicka St. 31-510 Cracow, Poland, e-mail: guzikk@uek.krakow.pl.

distribution. When analysing the risk of insurer, this assumption is rarely met (and the creators of the proposed solutions are aware of this). It means that in the analysed standard Solvency II model, the diversification effect is estimated using dependence structures that can describe relations between risk factors in an incorrect way. An obvious question arises: To what extent is the diversification effect estimated in this way reliable?

At first, the method of determining solvency capital requirements used in Solvency II is briefly characterised in the paper and the role of dependences for the correct specification of diversification effect is presented. This is followed by an analysis of the sensitivity of diversification effect to the dependence structure based on the example of life and health underwriting risk. The diversification ratio was estimated applying the standard Solvency II approach, i.e. using the variance-covariance method with the life and health underwriting risk correlation coefficient of 0.5. With a copula, a number of examples were provided to demonstrate that the same correlation coefficient value may correspond to various dependence structures, thus various diversification ratios. The value of the diversification ratio was calculated for the case when the considered risks are independent and comonotonic, and the ratio’s minimum and maximum value was calculated assuming no information about the dependence structure between these risks.

2 Risk aggregation and diversification effect in Solvency II

In Solvency II, the principal role in the process of evaluating the solvency of an insurer is played by a solvency capital requirement (SCR). This capital is considered as a cushion against significant deviations from expected loss, whereas coverage for expected loss is provided through provisions. Therefore, it is defined as economic capital, which should guarantee security for the insured if unpredictable losses occur. It is calculated at least once a year and when a considerable change occurred in the risk profile of an insurer. It is assumed that SCR should guarantee with a 0.995 probability that the insurer will be able to meet its obligations within 12 months. It must provide for all measurable risk kinds to which the insurer is exposed.

In the standard Solvency II approach, overall Solvency Capital Requirement for the insurer is calculated with the use of the following formula:

$$SCR = BSCR + Adj + SCR_{Op}, \tag{1}$$

where: *BSCR*- Basic Solvency Capital Requirement, *Adj* - adjustment for the risk absorbing effect of technical provisions and deferred taxes, SCR_{Op} - the capital requirement for operational risk.

BSCR value is determined when aggregating SCRs designated for main risk modules (i.e. market risk, counterparty default risk, life underwriting risk, non-life underwriting risk, health underwriting risk, intangible assets risk); SCRs for modules are determined by aggregating SCRs for sub-modules whereas the later result from the aggregation of SCRs for risk drivers[3]. Thus, the process involves 3 aggregation levels which are presented in detail, for example, in (*QIS5 Technical ...*, 2017; Wanat, 2014). It is assumed in the process that not all risks occur simultaneously, so SCR determined for a specific level (e.g. module) is generally not greater than the sum of SCRs set at the -1 level (e.g. for sub-modules). The resulting difference is referred to as the diversification effect (benefit) and it is a key element in the risk management process of an insurer.

If it is formally assumed that on *l*-th ($l = 1,...,3$) aggregation level the capital requirement for *j*-th risk $Y_j^{(l)}$ (insurer[4], module, sub-module) dependent on *k* risks $X_{j1}^{(l-1)},..., X_{jk}^{(l-1)}$ from $l-1$ level (modules, sub-modules, drivers) is determined, the diversification effect can be measured with the use of the diversification ratio (see: e.g. Embrechts et al., 2015):

$$d_j^{(l)} = \frac{\kappa\left(Y_j^{(l)}\right)}{\sum_{i=1}^{k} \kappa\left(X_{ji}^{(l-1)}\right)} \tag{2}$$

where: $\kappa\left(X_{ji}^{(l-1)}\right)$ - capital requirement for risk $X_{ji}^{(l-1)}$, $\kappa\left(Y_j^{(l)}\right)$ - capital requirement for the aggregate risk $Y_j^{(l)}$.

The above formula (2) suggests that the diversification effect depends on the manner of determining capital requirements (henceforth, for the purpose of preserving the transparency of notation, superscripts *l* and subscripts *j* will be omitted) $\kappa(X_i)$ for individual risks X_i and capital requirement $\kappa(Y)$ for aggregated risk *Y*. As already mentioned, these requirements should correspond to economic capital determined for one year, at the confidence level of 0.995. Therefore, in accordance with its definition (cf. e.g. Lelyveld, 2006), $\kappa(X_i)$ and $\kappa(Y)$ should be equal:

$$\kappa(X_i) = VaR_{0.995}(L_i) - \mu_i \tag{3}$$

[3] The manner of determining *Adj* and SCR_{Op} values is presented in (*QIS5 Technical ...*, 2017).
[4] It is obviously BSCR.

$$\kappa(Y) = VaR_{0.995}(L) - \mu \tag{4}$$

where: L_i , μ_i – loss distributions for X_i risks and their expected values, respectively,
L, μ- loss distribution for aggregated risk Y and its expected value, respectively,
$VaR_{0.995}(\cdot)$ - Value-at-Risk at the confidence level of 0.995.

It is, thus, clear that the procedure of estimating capital requirement for aggregated risk Y, which mainly depends on the modelling of dependence structure among variables $L_1,...,L_k$ (so $X_1,...,X_k$ risks) is of key importance to the correct evaluation of the diversification ratio. In the standard Solvency II solution, in case of the aggregation of solvency capital requirements on the second and third level, the variance-covariance method is proposed. The method involves:

- Determining capital requirements for individual risks: $\kappa(X_1),...,\kappa(X_k)$.
- Determining capital requirement $\kappa(Y)$ for aggregated risk Y based on the linear correlation coefficients ρ_{ij} among all pairs of risks $X_1,...,X_k$, in accordance with the following formula:

$$\kappa^{(solv)}(Y) = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}\rho_{ij}\kappa(X_i)\kappa(X_j)}. \tag{5}$$

As stated above, $\kappa^{(solv)}(Y)$ should correspond to economic capital necessary for securing potential losses (higher than expected) related to risk Y, in an annual time horizon and at the security level of 0.995. Thus, value $\kappa^{(solv)}(Y)$ obtained as a result of applying the standard procedure (5) should be equal to $\kappa(Y)$ value obtained with the use of formula (4). With the application of formulas (4) and (5), we obtain the same value of solvency capital requirement (i.e. $\kappa^{(solv)}(Y) = \kappa(Y)$) only when (cf. e.g. Dhaene et al., 2005):

i. Capital requirements for individual risks $\kappa(X_1),...,\kappa(X_k)$ are determined in accordance with formula (3).
ii. $L = L_1 + ... + L_k$ and $L_1,...,L_k$ have multivariate normal (elliptical) distribution, which means, in particular, that each variable L_i has normal distribution.

In the process of estimating solvency capital requirements of the insurer, those risks are aggregated which due to their essence are modelled with the use of different distributions and different methods. Therefore, the assumption that they are subject to multivariate normal distribution is ill-founded. A question arises concerning the estimations of the diversification ratio in case we do not know the multivariate distribution of random vector $(L_1,...,L_k)$. An insurer may estimate quite precisely the distribution of losses related to individual risks $X_1,...,X_k$, that is marginals of this vector. Let us attempt to answer the question by assuming that the distributions of variables $L_1,...,L_k$ are known whereas the dependence structure among them is unknown. With this assumption, the diversification effect will only depend on the dependence structure among variables $L_1,...,L_k$ which determines $VaR_{0.995}(L)$. The problem is discussed in more detail in the following section.

3 Value-at-Risk bounds with fixed marginal distributions

Let us assume that on the determined aggregation level, the distributions of random variables $L_1,...,L_k$ have known cumulative distribution functions $F_1,...,F_k$. Then, the distribution of the sum $L = L_1 + ... + L_k$, and the value of $VaR_\alpha(L)$ depend only on the dependence structure of the vector $(L_1,...,L_k)$. Based on Sklar's Theorem, all information on it is in copula *C*. Thus, *k*-th dimensional random vector with fixed marginals $F_1,...,F_k$ and dependence structure in the form of copula *C* will be designated by $(L_1^C,...,L_k^C)$. If the dependence structure between $L_1,...,L_k$ is unknown, we are unable to determine the exact $VaR_\alpha(L)$, and it can only be assumed that it fulfils the following inequalities (see: e.g. Embrechts et al., 2013):

$$\underline{VaR_\alpha}(L) \leq VaR_\alpha\left(L_1^C + ... + L_k^C\right) \leq \overline{VaR_\alpha}(L) \tag{6}$$

where:

$$\underline{VaR_\alpha}(L) = \inf_{C \in \Xi_k} \left\{VaR_a\left(L_1^C + ... + L_k^C\right)\right\} \tag{7}$$

$$\overline{VaR_\alpha}(L) = \sup_{C \in \Xi_k} \left\{VaR_a\left(L_1^C + ... + L_k^C\right)\right\} \tag{8}$$

whereas Ξ_K denotes the class of all *k*-dimensional copulas.

The issue of seeking bounds (7) and (8) is extremely important from the perspective of risk management and has a long history. The first results in this respect for the sum of two random variables are presented in papers by (Makarov, 1981) and independently in (Rüschendorf, 1982). In recent years, it has been discussed, for example, in (Puccetti and Rüschendorf, 2012a, 2012b, 2014; Embrechts et al. 2013; Puccetti et al., 2013; Bernard et al., 2015, 2016).

The natural outcome of research on respecting *VaR* boundaries was the creation of RA algorithm (*Rearrangement Algorithm*), proposed by Puccetti and Rüschendorf (2012a) and Embrechts et al. (2013). It allows for designating *VaR* boundaries for known but not necessarily identical marginal distributions. In brief, RA algorithm is about constructing dependence functions between random variables L_i by regrouping properly the columns made of random variables so that the distribution of the sum of random variables in convex order would be as low as possible. Research conducted so far indicates that estimation ranges of VaR obtained by AR in various cases are quite broad.

The modification of RA algorithm, also known as ARA (*Adaptive Rearrangement Algorithm*), was proposed by Hofert et al. (2017). The main impulse for the creation of the algorithm was a huge (even if the use of computers is taken into account) number of operations which needed to be performed with the use of RA algorithm in case of a large number of variables L_i. Bernard et al. (2015) constructed ERA algorithm (*Extended Rearrangement Algorithm*) also on the basis of RA algorithm. In relation to RA algorithm, the extension involves determining minimum elements of the distribution of a sum of random variables in the sense of convex order, in the upper and lower part of the distribution separated by the given α with the limitation of variance taken into account. ERA algorithm aims at making the distribution of L as flat as possible on the upper and lower part by applying the RA algorithm on both parts and by moving through the domains in a systematic way in order to satisfy the variance constraint. The examples presented by the authors prove that ERA algorithm works well and an additional condition of limited variance leads to better (as compared to RA algorithm) estimations of *VaR* boundaries. On the basis of ERA algorithm, the authors prove that models used by participants and regulators of the capital market can underestimate *VaR* whereas values-at-risk designated in this manner may be incomparable. They additionally claim that the determination of capital requirements at a high confidence level, e.g. 99.5%, is justified.

4 Diversification effect for life and health underwriting risk – empirical example

This section presents the results of the analysis of the impact of selected dependence structures on the diversification effect in the case of the aggregation of capital requirements for life and health underwriting risk. In the study, it was assumed that losses (in million euros) related to life underwriting risk and health underwriting risk are modelled with the use of random variables of normal distribution[5]: $L_1 \sim N(0, 392)$ and $L_2 \sim N(0, 248)$, respectively. Capital requirements $\kappa(Y)$ and diversification ratio d have been determined:

[5] The parameters adopted were the same as in (Bernard et al. 2016).

- in accordance with the standard procedure proposed in Solvency II, i.e. with the use of formula (5) with correlation coefficient[6] $\rho_{12} = 0.25$;
- with the assumption that the dependence structure between L_1 and L_2 is modelled with the use of the Student copula (df=2), the Student copula (df=5), the Gumbel copula, the Frank copula, the Clayton copula and the Galambos copula. Copula parameters are determined in such a way that the linear correlation coefficient ρ between marginal distributions should be equal to 0.25 in each of the analysed structures. The values of these parameters and additional information on dependences in the lower (λ_L) and upper (λ_U) tail are given in the first column in table 1;
- with the assumption that L_1 and L_2 are independent;
- with the assumption that L_1 and L_2 are comonotonic.

Then, it was assumed that there was no information on the dependence structure between L_1 and L_2, and lower and upper estimations $\kappa(Y)$ and d were determined. $\underline{VaR_{0.995}}(L)$ and $\overline{VaR_{0.995}}(L)$ values necessary for this purpose were obtained with the use of ARA algorithm. The results are given in the second and third column of table 1 and in Fig. 1.

The conducted study indicates that in the process of determining solvency capital requirements in Solvency II, familiarity with only distributions of aggregated risks X_i without familiarity of the dependence structure between them is insufficient. The range of possible values $\kappa(Y)$ from EUR 274.7 to 1791.5 million obtained in this way is useless from a practical perspective. Considering the above, in the process of aggregating capital requirements one should take into consideration dependence between risks. The standard Solvency II solution proposes the use of linear correlation coefficients only. However, as the results of the analysis indicate, the method does not guarantee that the capital will be determined unequivocally. The same correlation coefficients between marginal distributions may correspond to different dependence structures. This results in the estimation of capital $\kappa(Y)$ and the corresponding diversification ratio on different levels. However, it seems natural to expect that additional information on the dependence structure in the form of correlation coefficients between risks should largely narrow down the range of potential values for $\kappa(Y)$ and d. The presumption was confirmed in studies only in the case of several

[6] See: DIRECTIVE 2009/138/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II), ANNEX IV, p. 124.

selected dependence structures (i.e. the Student copula (df=2), the Student copula (df=5), the Gumbel copula, the Frank copula, the Clayton copula and the Galambos copula). Capital values $\kappa(Y)$ from 1234.4 to 1468.1 million were obtained for them, which corresponded to the diversification ratio from the range (74.9, 89.1). Generally, it can be stated that the greater the dependence in the upper tail, the greater the capital requirement $\kappa(Y)$, and thus, the lower the diversification effect. It is just the opposite in case of dependences in the lower tail – the stronger the dependence, the lower the value $\kappa(Y)$ and the greater the diversification effect. The purpose of further research to be undertaken by the authors will be to determine the lower and upper limit for $\kappa(Y)$ and d for any dependence structures.

Dependence structure	**Capital requirement**	**Diversification ratio in %**
Solvency II standard formula	1322.9	80.2
Student copula (df=2, $\rho = 0.265$, $\lambda_L = \lambda_U = 0.278$)	1468.1	89.1
Student copula (df=5, $\rho = 0.253$, $\lambda_L = \lambda_U = 0.107$)	1395.9	84.7
Gumbel copula ($\theta = 1.186$, $\lambda_L = 0$, $\lambda_U = 0.206$)	1445.1	87.7
Frank copula ($\theta = 1.631$, $\lambda_L = \lambda_U = 0$)	1279.1	77.6
Clayton copula ($\theta = 0.370$, $\lambda_L = 0.154$, $\lambda_U = 0$)	1234.4	74.9
Galambos copula ($\theta = 0.426$, $\lambda_L = 0$, $\lambda_U = 0.196$)	1450.6	88.0
Independence structure	1194.8	72.5
Comonotonic structure	1648.5	100.0
Unknown dependence structure	(274.7, 1791.5)	(16.6, 108.7)

Table 1. Research results.

Conclusions

The results of QIS5 presented in (*EIOPA Report...*,2017) show that the diversification effect may have a considerable impact on the decrease of solvency capital requirement of an insurer. The solution proposed as part of Solvency II, where the requirement is taken into account when determining SCR, on the one hand, belongs to the elements of awarding good risk management systems but, on the other hand, it requires that managers develop the right risk aggregation methods.

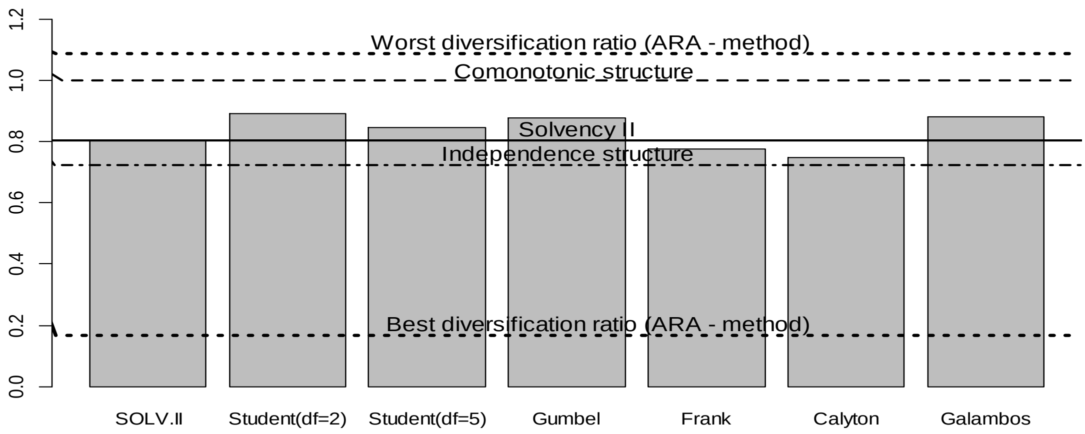


Fig. 1. Diversification ratio for the analysed dependence structures.

It should be emphasised here that the diversification effect is closely related to (or results from) the dependence structure between risks for which capital requirements are aggregated. Therefore, in order to estimate correctly the diversification effect, the structure must be identified properly. If only linear correlation coefficients are used for this purpose, this may cause errors in results as they unequivocally describe linear dependences only. In general, dependences between risks may be so complex that several numbers in the correlation matrix may not suffice for describing them (they may be non-linear, characterised by stronger dependences in tails, etc.). Furthermore, due to insufficient reliable data, correlation coefficients used in standard formulas depend to a considerable extent on the individual opinions of experts. Therefore, there is a need to carry out research that will focus on seeking new, more precise methods of recognising and modelling dependence structures as well as ways of including them in solvency models. In internal models (full or partial) or own parameters, Solvency II Directive allows or even encourages such research and implementation of nonstandard solutions into solvency models. New methods must be accepted by the market regulator.

Acknowledgements

The study is supported with subsidies for maintaining research capacity granted to the Faculty of Finance and Law at the Cracow University of Economics by the Polish Ministry of Science and Higher Education.

References

Bernard, C., Denuit, M., & Vanduffel, S. (2016). Measuring Portfolio Risk Under Partial Dependence Information. *Journal of Risk and Insurance.*

Bernard, C., Rüschendorf, L., & Vanduffel, S. (2015). Value-at-Risk Bounds With Variance Constraints. *Journal of Risk and Insurance*.

Dhaene, J., Goovaerts, M., Lundin, M., & Vanduffel S. (2005), *Aggregating economic capital*, Belgian Actuarial Bulletin, 5(1), 14-25.

EIOPA Report on the fifth Quantitative Impact Study (QIS5 ... (n.d.). Retrieved February 24, 2017 , from https://eiopa.europa.eu/Publications/Reports/QIS5_Report_Final.pdf.

Embrechts, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. Journal of Banking & Finance, 37(8), 2750-2764.

Embrechts, P., Wang, B., & Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4), 763-790.

Hofert, M., Memartoluie, A., Saunders, D., & Wirjanto, T. (2017). Improved algorithms for computing worst Value-at-Risk. *Statistics & Risk Modeling*, 0(0).

Lelyveld, I. V. (2006). *Economic capital modelling: concepts, measurement and implementation*. London: Risk Books.

Makarov, G.D. (1981). Estimates for the distribution function of the sum of two random variables when the marginal distributions are fixed. *Theory of Probability and its Applications*, (26), 803–806.

Puccetti, G., & Rüschendorf, L. (2012a). Bounds for joint portfolios of dependent risks. *Statistics & Risk Modeling*, 29(2), 107-132.

Puccetti, G., & Rüschendorf, L. (2012b). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7), 1833-1840.

Puccetti, G., & Rüschendorf, L. (2014). Asymptotic equivalence of conservative value-at-risk- and expected shortfall-based capital charges. *The Journal of Risk*, 16(3), 3-22.

Puccetti, G., Wang, B., & Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. Insurance: Mathematics and Economics, 53(3), 821-828.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, (14), 623–632.

QIS5 Technical Specifications - European Commission. (n.d.). Retrieved February 24, 2017, from https://eiopa.europa.eu/publications/qis/insurance

Wanat S. (2014). Ocena efektu dywersyfikacji ryzyka w Solvency II - aspekt praktyczny i metodologiczny. *Wiadomości Ubezpieczeniowe*, (3), 31-48.

A market-implied approach to measuring corporate diversification

Florian Zechser[1], Joachim Rojahn[2]

Abstract

The standard literature provides mixed results on the effects of corporate diversification on shareholder value with a tendency towards a diversification discount. Among other factors, differences in the measurement techniques may explain these mixed results on the valuation effects. In industrial organization, corporate diversification is commonly operationalized using business count measures, such as the number of business units or Berry-Herfindahl indices. Irrespective of their advantages for quantitative research, these measures are widely criticized for their dependency on segment data and Standard Industrial Classification codes.

In this study, we introduce a market-implied diversification measure which uses standardized regression coefficients instead of industry classification schemes to identify the business activities that a firm is engaged in. The coefficients are obtained through forward stepwise regressions within which a firm's stock market return is regressed against a set of ten STOXX® EUROPE 600 sector indices. Thereby, we assume that the degree of diversification is likely to be a negative function of the amount of unsystematic variation. Using a representative sample of firms listed in the STOXX® EUROPE 600 index over the years 2010 to 2015, we compare our measure to commonly used business count measures.

Keywords: *Corporate diversification, business count measure, market-implied diversification measure*

JEL Classification: G32, G12

1 Introduction

Ever since the seminal work of Rumelt (1974), the nature of the relationship between corporate diversification and shareholder value has been at the center of research studies of strategic management and industrial organization (Instead of many, see Dess et al., 1995; Martin and Sayrak, 2003). On the one hand, corporate diversification can provide additional shareholder value through debt coinsurance effects or an increased efficiency of internal capital markets (Hann et al., 2013; Lewellen, 1971; Stein, 1997). On the other hand, diversification is said to amplify existing agency problems and impose further coordination costs on the firm (Amihud and Lev, 1981; Jensen, 1986; Rajan et al., 2000; Rawley, 2010).

Empirical studies on the relationship between corporate diversification and shareholder value provide mixed results ranging from "Diversification destroys shareholder value" to "Diversification creates shareholder value" (Instead of many, see Erdorf et al., 2013; Martin and Sayrak, 2003). The reasons for the mixed evidence on the valuation effects of corporate

[1] isf Institute for Strategic Finance, FOM Hochschule, Leimkugelstraße 6, 45141 Essen, Germany, e-mail: florian.zechser1@fom-net.de.

[2] isf Institute for Strategic Finance, FOM Hochschule, Leimkugelstraße 6, 45141 Essen, Germany, e-mail: joachim.rojahn@fom.de.

diversification include time-variant effects and the use of various performance indicators (Hoskisson and Hitt, 1990; Keats, 1990; Ramanujam and Varadarajan, 1989). Additionally, the proliferation of diversification measures and the limited information on a single measure's construct validity constrain our understanding of the valuation effects of diversification (Chatterjee and Blocher, 1992; Hoskisson et al., 1993, Robins and Wiersema, 2003).

In industrial organization, the total diversity of a firm's business units is commonly estimated using the business count approach which determines diversity either by simply counting the business units which a firm operates in or by employing comprehensive indices. Using objective, secondary data to allocate a firm's reporting units to well-established industry classification schemes, they benefit from high reliability and objectivity (Montgomery, 1982). Moreover, the ease of computation and the possibility to handle large data samples are further arguments for business count measures. However, using business count measures based on industrial classification schemes for the determination of the relatedness may lead to low levels of construct validity as they are unable to simultaneously reflect the level and the type of diversification (Hoskisson et al., 1993). By way of construction, the SIC classes do not adequately differentiate between industries (Fan and Lang, 2000; Montgomery, 1982) and, therefore, offer "only limited information on the types of strategic interrelationships" (Robins and Wiersema, 1995). Additionally, due to the "implicit assumption of equal 'dissimilarity' between distinct SIC classes" (Rumelt, 1974), numerical differences between SIC codes cannot be interpreted on an interval or ratio scale (Montgomery, 1982; Nayyar, 1992). Moreover, business count measures require a somewhat arbitrary decision about the level of refinement, which is likely to offset the advantage of higher objectivity received from publicly available data. Last, they are likely to be exposed to the risk of strategic accounting as they build on segment data (Villalonga, 2004). As Robins and Wiersema (1995) note: "These assumptions can be relaxed only by going beyond the SIC system and employing additional sources of information".

We contribute to the growing body of literature devoted to the measurement of diversification by promoting an alternative measure of diversification. Our market-implied diversification measure utilizes stock market data to group a firm's business activities into homogenous groups instead of relying on an industry classification system. This way, we can avoid the limitations inherent in the SIC system and, at the same time, take advantage of the benefits of quantitative measures. The remainder of the paper is organized as follows: Section 2 presents the methodology underlying the market-implied diversification measure. Section 3 describes the data. Section 4 presents the results of a comparison between MDIV and traditional business count measures. Section 5 concludes.

2 Market-based approach to valuing corporate diversification

We build on the stock-market based measure of corporate diversification by Barnea and Logue (1973). Following Sharpe's (1963) single-index model, Barnea and Logue (1973) show that "the degree of diversification is a direct function of the amount of residual unsystematic variation that remains in a combination of risky assets" and promote the proportion of explained variance (R^2) and the standard error of the estimate as measures of diversification. We further refine the measurement approach of Barnea and Logue (1973) in two aspects: First, instead of using a broad market portfolio, we employ a set of ten STOXX® EUROPE 600 sector indices to gather information about the extent of diversification. Second, we introduce a weighting vector based on a Herfindahl index.

The starting point for constructing the market-implied diversification measure is the following regression model in which the stock market return of firm i in year t (r_{it}) is regressed against the multivariate return series of ten STOXX® EUROPE 600 sector indices during the period commencing 250 days before and ending on the last trading day prior to the individual firm's fiscal year end:

$$r_{it} = \sum_{j=1}^{10} \beta_{jt} * X_{jt} + \varepsilon \tag{1}$$

MDIV_{it} is then the maximum of the proportion of unexplained variance that remains after controlling for systematic valuation factors ($1-R^2$) and the value of a Herfindahl index based on standardized regression coefficients:

$$MDIV_{it} = \max(1 - R^2, \sum_{j=1}^{J} \beta_{jit}^2 / B_{it}^2) \tag{2}$$

Analogous to the stock-market based measure by Barnea and Logue (1973), the first element of the maximum function of equation (2) represents the level of unsystematic variance in the regression model. It contains information about the relative use of diversification effects in the corporate portfolio. The sector indices stand proxy for the market portfolio and, thus, provide information about the extent to which equity risks are diversifiable. Consequently, higher values of R^2 indicate a greater level of diversification. Statistical inferences about R^2 are based on Huber-White standard errors. The second element of the maximum function of equation (2) is a Herfindahl measure made of regression coefficients that are obtained through forward stepwise regressions of equation (1). The boundaries for the removal and the addition of a sector index are $p \geq .1$ and $p \leq .05$, respectively. More formally, the Herfindahl index is equal to the sum of the squares of the standardized regression coefficients divided by the squared sum of the absolute regression coefficients B_{it} Standardized regression coefficients contain information about the relative strength of the predictors of the model and, henceforth, are predestined to take over the role of the primary weighting vector from the SIC industry segments or groups. MDIV converges towards one as the firm becomes more focused. The maximum function, thereby, ensures that the sector indices jointly explain a high proportion of the variance of the stock market return of the firm in question. Assuming all else is constant, the level of corporate diversification is a positive function of r-squared. For instance, consider the case of three significant indices with homogenous beta-coefficients which would per the second element of equation (2) mean a high degree of diversification (33%). Nevertheless, r-squared could be relatively small, resulting in higher values for MDIV and, thus, a lower level of corporate diversification.

The market-implied diversification measure goes beyond the limitations of the SIC system. It reflects the full range of the diversification concept by considering both the extent of diversification and the type of relatedness. Instead of assuming equal dissimilarities between distinct SIC classes, MDIV exclusively relies on the market mechanisms to determine the coherence between a firm's business units. Consequently, MDIV returns the *market-implied* level of corporate diversification. Moreover, MDIV is not inversely influenced by strategic accounting and accounts for interaction effects between different lines of operations, irrespective of whether they are reported as segregated segments (Davis and Duhaime, 1992; Martin and Sayrak, 2003). Given its low data requirements compared to traditional diversification measures, MDIV offers a simplified and less cumbersome measure of corporate diversification.

3 Sample and Data

We construct our sample from all firms included in the STOXX® EUROPE 600 index as of 2010. During the validation process, we will compare MDIV with traditional business count measures, the estimation of which requires firm-specific segment data. For this reason, we restrict our sample to the most recent periods of 2010 to 2015. Although the membership in the STOXX® EUROPE 600 index is reviewed on a quarterly basis, there will be now rebalancing to avoid a survivorship bias. Firms falling under the sector "financials" per the ICB sector classifications will be removed from the sample data. We justify the removal by the differences in capital structure and restrictive regulatory requirements that apply to financial firms. Moreover, we stipulate that the sum of the segment sales is within 10% of the total sales of the firm and that the sum of common equity and preferred stock is positive. Last, the sample is truncated at the 95% confidence level to remove outliers from the exogenous variables. The sample selection process is illustrated in Table 1.

STOXX® EUROPE 600 index	2010	2011	2012	2013	2014
Basic population:	600	600	600	600	600
- financials	136	136	136	136	136
- segment sales / neg. equity firms	3	1	12	10	9
- lack of data / outliers	43	53	66	73	81
Sample population:	418	410	386	381	374

Table 1. Sample selection process.

4 The relationship between MDIV and SIC-based diversification measures

We concentrate on two dimensions of the validity of the market-implied diversification measure: Convergent validity and predictive validity.

Measure	P50	max	min	sd	Spearman
MDIV	0.670	1.000	0.217	0.211	1.000
COUNT2	2	7	1	1.11	-0.074***
COUNT4	3	9	1	1.41	-0.071***
H2DIV	0.898	1.000	0.182	0.228	0.087***
H4DIV	0.648	1.000	0.149	0.259	0.077***
LARGEST2	0.946	1.000	0.243	0.185	0.085***
LARGEST4	0.778	1.000	0.173	0.223	0.076***

Note: Spearman correlation regarding MDIV, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2. Summary statistics and convergent validity.

Convergent validity. Convergent validity refers to the extent to which indicators of the same construct coincide with one another (Hoskisson et al., 1993). Besides MDIV, the following tests comprise a selection of widely used diversification measures. The most straightforward measures involve numerically counting the number of business units defined through two- and four-digit SIC codes (COUNT2, COUNT4). Moreover, we consider two revenue-based Herfindahl indices which can formally be expressed as the sum of the firm's squared output in the gth industry group (H2DIV) and jth industry segment (H4DIV), respectively, to the firm's squared total output across all business units. Finally, we include the share of the largest business defined as the ratio of a firm's primary industry output to total firm output, where the primary industry is again identified by two-digit (LARGEST2) and four-digit SIC codes (LARGEST4). Table 2 reports summary statistics and Spearman correlation coefficients among the business count measures that were employed in this study. MDIV varies from a maximum of 1 (single-segment) to a minimum of .22 (high degree of diversification), indicating a broad range of different diversification strategies among the sample firms. The median diversification level is 0.67 and falls in between the median levels of H2DIV and H4DIV. Regarding standard deviation, the results of the various business count measures are also very similar. A different picture is obtained by looking at the Spearman correlation coefficients between MDIV and the continuous measures. Although there is a significant relationship for most of the SIC-based measures, the coefficients are close to zero, which suggests a rather low level of convergent validity. To underpin the results from the Spearman correlation coefficients, factor analysis is performed on all diversification measures. From omitted Shapiro-Wilk tests, we infer that the assumption of normal distribution for all measures is breached. Accordingly, we will use the natural logarithm of MDIV (positively skewed) as well as the power transformation 1/x for variables H2DIV, H4DIV, LARGEST2, and LARGEST4 (negatively skewed). The number of relevant factors are extracted using the Very Simple Structure Criterion (VSS). The results indicate a two-factor solution from a varimax rotation which accounts for 73% of the variance. Diversification measures focusing on industry groups (two-digit SIC codes) have high factor loadings on the first factor, while measures on industry segments (four-digit SIC codes) load heavily on the second factor. MDIV, however, does not load on any of the two factors and its commonality is close to zero, again indicating low convergent validity.

Predictive validity. Predictive validity concerns the measure's ability to predict future performance of the construct and is commonly examined using correlation analysis. Instead of pure correlation coefficients, we investigate regression coefficients obtained from two-way fixed effect regressions (within-estimator) to gather information on MDIV's predictive validity. The domain of interest is the financial performance, which we operationalize through return on assets (ROA), return on equity (ROE), and return on sales (ROS).

Diversification measure	Factor 1	Factor 2	Uniqueness
MDIV	-0.035	-0.096	0.990
COUNT2	0.662	0.345	0.443
COUNT4	0.386	0.697	0.365
H2DIV	0.927	0.368	0.005
H4DIV	0.420	0.905	0.005
LARGEST2	0.921	0.334	0.040
LARGEST4	0.387	0.899	0.042
Proportional variance explained	0.375	0.355	
Cumulative variance explained	0.375	0.730	

Note: Likelihood Chi Square = 5256.715 ($p < 0$), RMSR = 0.079, RMSEA = 0.2819.

Table 3. Uniqueness: Confirmatory factor analysis.

Panel regression analysis		Regression coefficient	Hausman Test	F-Stat.	Within R²
Return on asset (ROA)	MDIV	-0.02	655.08***	3.15***	0.34
	H2DIV	-0.01	653.15***	2.75***	0.34
	H4DIV	0.01	652.95***	2.80***	0.34
Return on equity (ROE)	MDIV	-0.04**	187.04***	10.45***	0.06
	H2DIV	0.01	189.52***	8.98***	0.06
	H4DIV	0.01	189.08***	8.96***	0.06
Return on sales (ROS)	MDIV	0.00	715.75***	1.62	0.37
	H2DIV	-0.01	716.08***	1.55	0.37
	H4DIV	0.00	715.48***	1.58	0.37

Note: P-values based on robust standard errors, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. Predictive validity: Panel regression analysis.

Therefore, our study provides only some support for the prevalent finding of no or possibly a negative relationship between diversification and accounting performance as concluded by Hoskisson and Hitt (1990). This may be due to an increased efficiency of internal capital allocation during the financial crisis (Kuppuswamy and Villalonga, 2015).

Table 4 presents the panel regression results for MDIV and both Herfindahl measures. The other diversification measures are not included in this analysis as they are highly correlated with all of the Herfindahl measures (> 0.6). The natural logarithm of both total assets as a proxy for firm size and the debt to equity capital ratio are included as control variables. A significant relationship is found only for the combination of MDIV and ROE, indicating higher returns for diversified firms. Although not significant, the signs for the other performance indicators are in the same direction. Notwithstanding the lack of significance, while the coefficient estimates for H2DIV indicate that focused firms have higher accounting returns compared to diversified firms, the sign for H4DIV shows an opposite result.

Summary and conclusions

This study introduces a continuous diversification measure that solely builds on stock market data to operationalize the corporate portfolio strategy. Formally, the market-implied diversification measure is the maximum of the proportion of unexplained variance that remains after controlling for systematic valuation factors and the value of a Herfindahl index based on standardized regression coefficients. By focusing on stock market returns, both the limitations of standardized industrial classification schemes, such as the SIC system and distortions induced by segment data can be reduced.

Using a sample of non-financial firms included in the STOXX® EUROPE 600 index during the period 2010 through 2015, we investigate the characteristics of MDIV regarding convergent validity and predictive validity. The correlation coefficients between MDIV and traditional diversification measures indicate a low level of convergent validity. The results from a two-factor analysis, where MDIV shows a commonality close to zero, support the findings of the correlation analysis. Moreover, the results of the panel regression analyses suggest a higher predictive validity of MDIV compared to traditional business count measures, although its predictive validity is rather low. The coefficient estimates, thereby, indicate significantly higher returns for diversified firms when performance is measured through return on equity.

The theoretical arguments for MDIV (content validity) as well as the tests for convergent and predictive validity suggest that MDIV offers an alternative measurement approach. However, the results also demonstrate that there is a need for further investigations. For instance, the causes for the low correlation coefficients between traditional business count measures and MDIV remain unanswered. Additionally, the results of the panel regression models call for further assessments of the coherence between MDIV and accounting as well as market based performance measures. Moreover, we focus on a rather small sample of European non-financial firms over a recent period. To reinforce the findings, further investigations with an increased firm sample and for different time periods may provide additional insights.

References

Amihud, Y., & Lev, B. (1981). Risk Reduction as a Managerial Motive for Conglomerate Mergers. *Bell Journal of Economics, 12*(2), 605-617.

Barnea, A., & Logue, D. E. (1973). Stock-Market Based Measures of Corporate Diversification. *The Journal of Industrial Economics, 22*(1), 51-60.

Chatterjee, S., & Blocher, J. D. (1992). Measurement of Firm Diversification: Is It Robust?. *The Academy of Management Journal, 35*(4), 874-888.

Davis, R., & Duhaime, I. M. (1992). Diversification, Vertical Integration, and Industry Analysis: New Perspectives and Measurement. *Strategic Management Journal, 13*(7), 511-524.

Dess, G. G., Gupta, A., Hennart, J.-F., & Hill, C. W. L. (1995). Conducting and Integrating Strategy Research at the International, Corporate, and Business Levels: Issues and Directions. *Journal of Management, 21*(3), 357-393.

Erdorf, S., Hartmann-Wendels, T., Heinrichs, N., & Matz, M. (2013). Corporate diversification and firm value: a survey of recent literature. *Financial Markets and Portfolio Management, 27*(2), 187-215.

Fan, J. P. H., & Lang, L. H. P. (2000). The Measurement of Relatedness: An Application to Corporate Diversification. *The Journal of Business, 73*(4), 629-660.

Hann, R. N., Ogneva, M., & Ozbas, O. (2013). Corporate Diversification and the Cost of Capital. *The Journal of Finance, 68*(5), 1961-1999.

Hoskisson, R. E., & Hitt, M. A. (1990). Antecedents and Performance Outcomes of Diversification: A Review and Critique of Theoretical Perspectives. *Journal of Management, 16*(2), 461-509.

Hoskisson, R. E., Hitt, M. A., Johnson, R. A., & Moesel, D. D. (1993). Construct Validity of an Objective (Entropy) Categorical Measure of Diversification Strategy. *Strategic Management Journal, 14*(3), 215-235.

Jensen, M. C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review, 76*(2), 323-329.

Keats, B. W. (1990). Diversification and Business Economic Performance Revisited: Issues of Measurement and Causality. *Journal of Management, 16*(1), 61-72.

Kuppuswamy, V., & Villalonga, B. (2016). Does Diversification Create Value in the Presence of External Financing Constraints? Evidence from the 2007–2009 Financial Crisis. *Management Science*, 62(4), 905-923.

Lewellen, W. G. (1971). A Pure Financial Rationale for the Conglomerate Merger. *The Journal of Finance, 26*(2), 521-537.

Martin, J. D., & Sayrak, A. (2003). Corporate diversification and shareholder value: a survey of recent literature. *Journal of Corporate Finance, 9*(1), 37-57.

Montgomery, C. A. (1982). The Measurement of Firm Diversification: Some New Empirical Evidence. *Academy of Management Journal, 25*(2), 299-307.

Nayyar, P. R. (1992). On the Measurement of Corporate Diversification Strategy: Evidence from Large U.S. Service Firms. *Strategic Management Journal, 13*(3), 219-235.

Rajan, R., Servaes, H., & Zingales, L. (2000). The Cost of Diversity: The Diversification Discount and Inefficient Investment. *The Journal of Finance, 55*(1), 35-80.

Ramanujam, V., & Varadarajan, P. (1989). Research on Corporate Diversification: A Synthesis. *Strategic Management Journal, 10*(6), 523-551.

Rawley, E. (2010). Diversification, coordination costs, and organizational rigidity: evidence from microdata. *Strategic Management Journal, 31*(8), 873-891.

Robins, J., & Wiersema, M. F. (1995). A Resource-Based Approach to the Multibusiness Firm: Empirical Analysis of Portfolio Interrelationships and Corporate Financial Performance. *Strategic Management Journal, 16*(4), 277-299.

Robins, J. A., & Wiersema, M. F. (2003). The Measurement of Corporate Portfolio Strategy: Analysis of the Content Validity of Related Diversification Indexes. *Strategic Management Journal, 24*(1), 39-59.

Rumelt, R. P. (1974). *Strategy, structure, and economic performance*. Cambridge, Mass.: Harvard University Press.

Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. *Management Science, 9*(2), 277-293.

Stein, J. C. (1997). Internal Capital Markets and the Competition for Corporate Resources. *The Journal of Finance, 52*(1), 111-133.

Villalonga, B. (2004). Diversification Discount or Premium? New Evidence from the Business Information Tracking Series. *The Journal of Finance, 59*(2), 479-506.