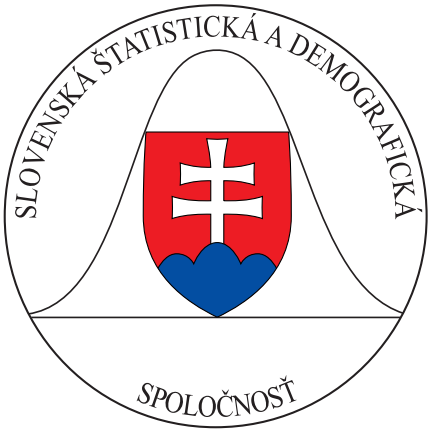
7/2013

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420

Slovenská štatistická a demografická spoločnosť Miletičova 3, 824 67 Bratislava
www.ssds.sk

Naše najbližšie akcie:

(pozri tiež www.ssds.sk, blok Organizované akcie)

Pohľady na ekonomiku Slovenska 2014
8. apríl 2014, Aula Ekonomickej univerzity v Bratislave

MedStat 2014
24. – 25.apríl 2014, Aula VN SNP Ružomberok

Aplikácie metód na podporu rozhodovania
STU Bratislava

Nitrianske štatistické dni 2014
máj 2014, Nitra

Ekomstat 2014
25.-30.5.2014, Trenčianske Teplice

Slovenská štatistická konferencia
September 2014, 2 dni, Prešovský kraj

Výpočtová štatistika 2014
December 2014, Bratislava

Regionálne akcie
priebežne

Slávnostná konferencia 50 rokov Slovenskej štatistickej a demografickej spoločnosti
marec 2018, Slovenská republika

FOREWORD

Dear colleagues,

we present the seventh issue of the nineth volume of the scientific peer-reviewed journal Forum Statisticum Slovacum published by the Slovak Statistical and Demographical Society (SSDS). This issue comprises contributions that are in its content oriented on the thematic scope „Computational statistics“.

Editors: Iveta Stankoviļ ov§, Jozef Chajdiak, Ján Luha, Tomáš Želinský

Reviewers: Jozef Chajdiak, Bohdan Linda, Ján Luha, Iveta Stankovičová, Tomáš Želinský

Assoc. Prof. Ing. Jozef Chajdiak, CSc.

Editor in chief

International Year of Statistics ("Statistics2013") is a global reminder of the importance of statistics. The Slovak Statistical and Demographical Society joined the International Year of Statistics and will be mentioned at its professional events in the year 2013.

PREDHOVOR

Vážené kolegyne, vážení kolegovia,

predkladáme siedme číslo deviateho ročníka vedeckého recenzovaného časopisu Slovenskej štatistickej a demografickej spoločnosti (SŠDS). Toto číslo je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou „Výpočtová štatistika“.

Editori: doc. Ing. Iveta Stankovičová, PhD., doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., doc. Ing. Tomáš Želinský, PhD.

Recenzenti: doc. Ing. Jozef Chajdiak, CSc., doc. RNDr. Bohdan Linda, CSc., RNDr. Ján Luha, CSc., doc. Ing. Iveta Stankovičová, PhD., doc. Ing. Tomáš Želinský, PhD.

doc. Ing. Jozef Chajdiak, CSc.

Šéfredaktor

STATISTICS2013.ORG

MEDZINÁRODNÝ ROK ŠTATISTIKY

ÚČASTNÍCKA ORGANIZÁCIA

Medzinárodný rok štatistiky ("Statistics2013") je celosvetové pripomenutie významu štatistiky. Slovenská štatistická a demografická spoločnosť sa pripojila k Medzinárodnému roku štatistiky a bude ho pripomínať pri svojich odborných akciách v roku 2013.

Lorenzova křivka a odvozené míry příjmové nerovnosti
Lorenz curve and derived inequality indicators

Jitka Bartošová, Vladislav Bína

Abstract: The contribution brings the most commonly used measures of income inequality, analyzes their relationship and presents a comparison of these measures. Particularly it focuses on the well known Gini index, Robin Hood index, income quintile share ration, variation coefficient, mean to median ratio, quartile skewness coefficient and Moors' quantile kurtosis coefficient. Their properties and relations are illustrated in the case of EU SILC 2011 data describing the income inequality situation in most European countries.

Abstrakt: Příspěvek se zabývá nejčastěji užívanými mírami příjmové nerovnosti, analyzuje jejich vztah a přináší porovnání těchto měr. Konkrétně se zaměřuje na dobře známý Giniho index, index Robina Hooda, podíl krajních kvintilů, variační koeficient, podíl průměru a mediánu, koeficient kvartilové šikmosti a Moorsův koeficient kvantilové špičatosti. Jejich vlastnosti a vztahy jsou ilustrovány na případě dat z průzkumu EU SILC 2011 popisujicích situaci nerovnosti příjmů ve většině evropských států.

Key words: Income inequality, Gini index, Robin Hood index, nonparametric measures, EU SILC 2011.

Kľúčové slová: Príjmová nerovnost, Giniho index, index Robina Hooda, neparametrické míry, EU SILC 2011.

JEL classification: C14, D31, D63

1. Introduction

Analysis of the income distribution is a useful tool for the decision making in different fields of social policy and is important for the estimation of the consumption of households. There are numerous contributions devoted to the analysis of income inequality, risk of monetary poverty, material deprivation and to the modelling of income distribution and forecasting. Let us mention few papers concerning the situation particularly in the Czech republic and Slovakia. These are the works of Bílková (2012), Fiala and Langhamrová (2009), Labudová etal. (2010), Löster and Langhamrová (2011), Malá (2013), Marek (2010), Pacáková etal. (2012),Řezanková and Loster (2011), Stankovičová (2010), Večerník (2013), Želinský (2012) and contributions of many other authors.

2. Methods of income inequality measurement

The income (and expense) inequality is usually presented using the Lorenz curve (L) and related basic measures of income (or expense) differentiation derived by Pareto, Bresciani, Gini, Pietro, etc. Yet another approach to the measurement of income inequality was developed in 1975 by Gastwirth and further enriched by Dagum (1978). These measures express the differences between population varying from the social–economic or geographic viewpoint. The last – decompositional – approach is focused on the contributions of particular subpopulations to the inequality in the entire heterogeneous population. Each subpopulation is specified by the value of some social–economic characteristic and the amount of its contribution can be expressed according to the total inequality or according to the total income of the population. This approach first utilized by Theil (1967) was further elaborated in the models of Dagum.

For the quantification of the income (or expense) inequality many indicators were developed. Commonly used is, e.g., Gini index (G), Robin Hood index (RHI), Atkinson index (A), Theil indices (L and T), income quintile share ration (x_{80}/x_{20}) or the mean to median

ratio (x_{80}/x_{20}). Inequality in the income (expense) distribution directly stems also from the density estimate of the empirical distribution and from its characteristics of shape. The values of variation coefficient ($s_x/\bar{x}$), quantile coefficient of skewness and kurtosis, etc., provide evidence of the inequality of distribution. The income inequality corresponds also to the risk of monetary poverty and projects into the position of the poverty line (*PL*), values of risk (*RP*), depth (*DP*) and severity of poverty (*SP*). More detail, e.g., in Foster etal. (1984), Bartošová and Bína (2009) or Bartošová and Želinský (2013).

3. Lorenz curve and derived inequality indicators

Lorenz curve is commonly used for the graphical presentation of the inequality of income (or expense) distribution. It depicts cumulative shares of incomes or expenses (axis *y*) in dependence on the cumulative count of individuals (or households) on the *x* axis, sorted according to the increasing incomes (expenses). Lorenz curve *L(p)* can be expressed as an integral of mean value of incomes (or expense). If *f(x)* denotes the density of income or expense distribution, $F(x) = \int_0^x f(t)dt$ is the corresponding cumulative density function and $E(X) = \int_0^\infty xf(x)dx$ is the mean (expected) value, then the Lorenz curve in the given quantile $p \in (0,1)$ is given by

$$L(p) = \frac{1}{E(X)} \int_0^{F^{-1}(p)} xf(x)dx,$$

or equivalently by

$$L(p) = \frac{1}{E(X)} \int_0^{p} F^{-1}(x)dx.$$

Argument *p* stands for cumulative percentages of individuals (or households) and *F*(*x*) is cumulative distribution function of incomes (expenses). The graph of Lorenz curve lies in the interior of a square between its bottom (*L*(*p*) = 0) and diagonal (*L*(*p*) = *p*). These boundaries correspond to the Lorenz curves for the absolute inequality (bottom of the square) and absolute equality (diagonal). The higher values of *L*(*p*) in *p* indicate more egalitarian distribution of incomes (expenses).

The Gini coefficient is elicited from the Lorenz curve and measures the deviation of the income distribution of individuals or households from the perfectly uniform distribution. Its value is given by the ratio of the area between the line of absolute equality (diagonal *y* = *p*) and the Lorenz curve *L*(*p*) and the entire area under the diagonal. According to the fact that the area under the diagonal is equal to the half of area of the unit square we can obtain the Gini coefficient using the numerical integration of estimated Lorenz curve

$$G = 1 - 2\int_0^1 L(p)dp.$$

The Gini index takes values from the interval $\langle 0; 1\rangle$ – value approaching to 0 indicates more egalitarian distribution of incomes in the considered society and vice versa.

Yet another inequality measure derived from the Lorenz curve is the Robin Hood index (*RHI*) representing the amount of incomes necessary to distribute in order to achieve absolute uniformity in the distribution. Its value is equivalent to the maximal vertical distance between the line of absolute equality and the Lorenz curve ($argmax[y - L(y)]$).

4. Inequality indicators and properties of income distribution

The inequality in the distribution of incomes (or expenses) corresponds to the properties of the income distribution. Schematic depiction of this interconnection is presented in the Figure 1.

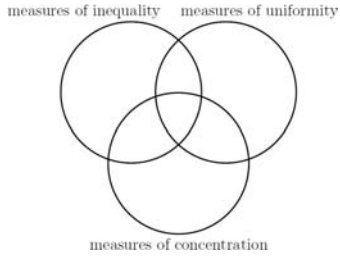


Fig. 1: Measures of inequality, uniformity and concentration.

The influence of the shape of income distribution, non-uniformity, skewness and kurtosis on the Lorenz curve and thus also Gini index and Robin Hood index is illustrated on Figures 2 – 5.

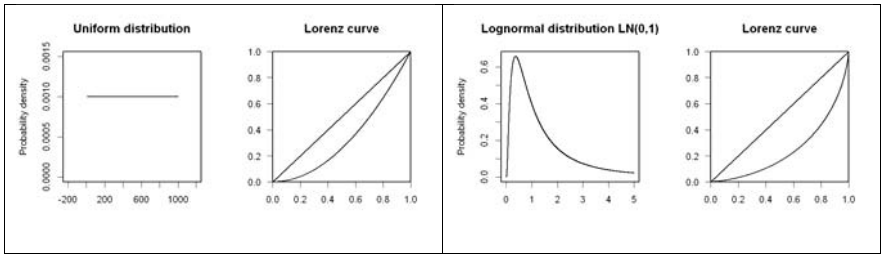


Fig. 2: Influence of (non)uniformity of income distribution.

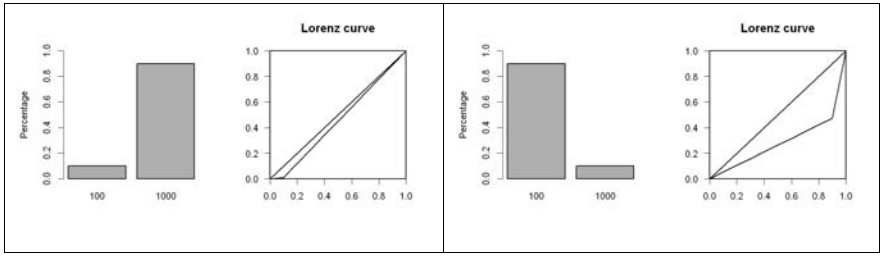


Fig. 3: Influence of skewness of the income distribution.

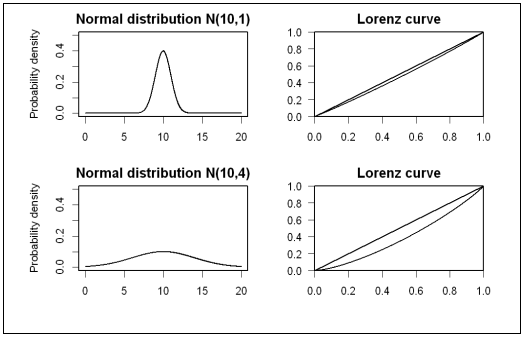


Fig. 4: Influence of the kurtosis of income distribution.

5. Correlation of inequality measures with properties of income distribution in Europe

The well known and most frequently published measure of the income inequality is the Gini index (*G*).

Tab. 1: Measures of income inequality in 2011 in European countries[1].

State	$\bar{x}$	$\hat{x}$	x_{20}	x_{80}	$s_x/\bar{x}$	quartile skewness	Moors' kurtosis	$\hat{x}/\bar{x}$	x_{80}/x_{20}	*G*	*RHI*
AT	23407.5	20928.0	14151.7	30000.0	0.603	0.118	1.292	0.894	2.120	0.277	0.192
BE	21305.2	18961.0	12542.0	27840.0	1.096	0.126	1.141	0.890	2.220	0.279	0.196
CZ	10122.3	8392.3	4536.2	14259.2	0.761	0.154	1.334	0.829	3.143	0.363	0.260
DE	21281.1	16836.0	8461.5	31038.7	0.908	0.197	1.301	0.791	3.668	0.400	0.287
DK	25634.5	19889.9	10082.2	37872.1	0.994	0.257	1.329	0.776	3.756	0.421	0.301
EE	6597.9	4852.1	2336.3	9939.8	0.865	0.294	1.281	0.735	4.254	0.427	0.312
ES	16468.5	12800.0	6308.0	24526.7	0.818	0.238	1.345	0.777	3.888	0.414	0.298
FI	24089.3	18982.0	9782.0	34816.0	0.900	0.213	1.266	0.788	3.559	0.389	0.281
FR	25092.6	19173.3	10580.0	35080.0	1.096	0.233	1.367	0.764	3.316	0.400	0.287
HU	6096.4	4949.9	2693.3	8764.2	0.736	0.208	1.305	0.812	3.254	0.365	0.263
IS	19991.2	18077.2	13241.6	25388.8	0.609	0.171	1.280	0.904	1.917	0.254	0.175
IT	18233.7	15956.0	9645.2	24374.0	0.773	0.105	1.308	0.875	2.527	0.325	0.225
LT	4406.5	3699.7	2335.3	6042.1	0.687	0.220	1.413	0.840	2.587	0.339	0.240
LU	37398.6	33138.7	21984.0	49581.3	0.593	0.178	1.233	0.886	2.255	0.280	0.198
LV	4907.4	3930.3	2532.9	6843.5	0.726	0.323	1.416	0.801	2.702	0.356	0.255
NL	22200.7	19741.3	14224.0	28683.3	0.592	0.192	1.316	0.889	2.017	0.265	0.184
NO	37303.0	34819.4	24330.0	47626.3	0.591	0.071	1.355	0.933	1.958	0.252	0.172
PL	5878.8	4976.8	3132.9	7773.3	0.728	0.164	1.339	0.847	2.481	0.322	0.226
PT	10421.9	8255.2	5118.0	13582.1	0.780	0.234	1.441	0.792	2.654	0.352	0.251
SE	22859.3	21183.1	13890.3	30498.4	0.592	0.065	1.163	0.927	2.196	0.263	0.185
SI	12265.4	11301.2	7422.5	16143.1	0.510	0.096	1.326	0.921	2.175	0.259	0.181
SK	6898.4	6133.8	4474.6	8766.3	0.929	0.199	1.413	0.889	1.959	0.262	0.182
UK	20358.5	16728.1	10757.3	27045.2	0.854	0.224	1.338	0.822	2.514	0.334	0.236

[1] Encoding of European countries: AT – Austria, BE – Belgium, CZ – Czech Republic, DE – Germany, DK – Denmark, EE – Estonia, ES – Spain, FI – Finland, FR – France, HU – Hungary, IS – Iceland, IT – Italy, LT – Lithuania, LU – Luxembourg, LV – Latvia, NL –Netherlands, NO – Norway, PL – Poland, PT – Portugal, SE – Sweden, SI – Slovenia, SK – Slovakia, UK – United Kingdom

For the appraisal of the temporal evolution of inequality within the particular states and for the mutual comparison among different states, regions, education and age groups of inhabitants very often also Robin Hood index (*RHI*), income quintile share ration (x_{80}/x_{20}) or the mean to median ratio (x_{80}/x_{20}) are used. The correspondence between the income inequality and selected properties of income distribution can be demonstrated on equalized income in the sample of European states (see Table 1).

The influence of variation and shape of income distribution on the mentioned inequality measures is measured by the means of their correlation. Among the basic characteristics we will choose a correlation coefficient ($s_x/\bar{x}$), quartile skewness coefficient and Moors' quantile kurtosis coefficient (see Table 2). The statistically significant correlations are typed in boldface.

Tab. 2: Pearson and Spearman correlation coefficients for selected measures.

Pearson correlation coefficient	$\hat{x}/\bar{x}$	x_{80}/x_{20}	*G*	*RHI*
$s_x/\bar{x}$	**-0.612**	0.532	**0.591**	**0.587**
quartile skewness	**-0.830**	**0.606**	**0.704**	**0.711**
Moors' kurtosis	-0.321	0.068	0.239	0.224
Spearman correlation coefficient	$\hat{x}/\bar{x}$	x_{80}/x_{20}	*G*	*RHI*
$s_x/\bar{x}$	**-0.681**	**0.587**	**0.636**	**0.633**
quartile skewness	**-0.852**	**0.684**	**0.744**	**0.738**
Moors' kurtosis	-0.248	0.093	0.118	0.115

6. Conclusion

We can infer that the indicators of incomes inequality are strongly interconnected with the statistical measures of (non)uniformity and concentration of the distribution. Particularly there exists a strong (statistically significant) dependence among the variation coefficient as relative measure of variability and mean to median ratio, Gini and Robin Hood indices, the relation to income quintile share ration is significant only in case of Spearman correlation coefficient. The quartile skewness coefficient is signifficantly correlated with all the aforementioned inequality indicators, whereas the Moors' quantile kurtosis coefficient appears to be rather independent on the mean to median ratio, income quintile share ration, Gini and Robin Hood indices.

References

BARTOŠOVÁ, J. – BÍNA, V. 2012. Sensitivity of monetary poverty measures on the setting of parameters concerning equalization of household size. In: *Proceedings of 30th International Conference Mathematical Methods in Economics 2012*, ed. J. Ramík a D. Stavárek, pp. 25 – 30.

BARTOŠOVÁ, J. – ŽELINSKÝ, T. 2013. Extent of poverty in the Czech and Slovak Republics fifteen years after split. In: *Post-Communist Economies* 25(1), pp. 119 – 131. BÍLKOVÁ, D. 2012. Recent Development of the Wage and Income Distribution in the Czech Republic. In: *Prague Economic Papers* 21(2), pp. 233 – 250.

DAGUM, C. 1978. A measure of inequality between income distributions. In: *Economie Appliquée* 30(3) – (4), pp. 401 – 413.

FIALA, T. – LANGHAMROVÁ, J. 2009. Human resources in the Czech republic 50 years ago and 50 years after. In: *IDIMT-2009 System and Humans – A Complex Relationship*. Linz: Trauner Verlag, J. Hradec, 9. 8. 2009 – 11. 08., pp. 105 – 114.

FOSTER, J. – GREER, J. – THORBECKE, E. 1984. A Class of Decomposable Poverty Measures. In: *Econometrica* 52(3), pp. 761 – 766.

LABUDOVÁ, V. – VOJTKOVÁ, M. – LINDA, B. 2010. Application of multidimensional methods to measure poverty. In: *E+M Ekonomie a management* 13(1), pp. 6 – 21.

LÖSTER, T. – LANGHAMROVÁ, J. 2011. Analysis of Long-Term Unemployment in the Czech Republic. In: *International Days of Statistics and Economics*, ed. T. Löster a T. Pavelka, pp. 228 – 234.

MALÁ, I. 2013. Použití konečných směsí logaritmicko-normálních rozdělení pro modelování příjmů českých domácností. In: *Politická ekonomie* 61(3), pp. 356 – 372.

MAREK, L. 2010. Analýza vývoje mezd v ČR v letech 1995–2008. In: *Politická ekonomie* 58(2), pp. 186 – 206.

PACÁKOVÁ, V. – LINDA, B. – SIPKOVÁ, Ľ. 2012. Rozdelenie a faktory najvyšších miezd zamestnancov v Slovenskej republike, In: *Ekonomický časopis* 60(9), pp. 935 – 948.

ŘEZANKOVÁ, H. – LÖSTER, T. 2011. Analysis of the Dependence of the Housing Characteristics on the Household Type in the Czech Republic. In: *APLIMAT – Journal of Applied Mathematics* 4(3), pp. 351 – 358.

STANKOVIČOVÁ, I. 2010. Regional Aspects of Monetary Poverty in Slovakia. In: *Social Capital, Human Capital and Poverty in the Regions of Slovakia*, ed. I. Pauhofová, O. Hudec a T. Želinský, pp. 67 – 75.

THEIL, H. 1967. *Economics and Information Theory*. Chicago: Rand McNally and Company.

VEČERNÍK, J. 2013. The changing role of education in the distribution of earnings and household income: the Czech Republic in 1988–2009. In: *Economics of Transition* 21(1), pp. 111 – 133.

ŽELINSKÝ, T. 2012. Changes in Relative Material Deprivation in Regions of Slovakia and the Czech Republic. In: *Panoeconomicus* 59(3), pp. 335 – 353.

Authors' adresses:

Jitka Bartošová, doc. RNDr., Ph.D.
Fakulta managementu, VŠE v Praze
Jarošovská 1117/II, 37701 Jindřichův Hradec
bartosov@fm.vse.cz

Vladislav Bína, Ing., Ph.D.
Fakulta managementu, VŠE v Praze
Jarošovská 1117/II, 37701 Jindřichův Hradec
bina@fm.vse.cz

Analýza závislosti medzi medzinárodnou migráciu a Legatum prosperity indexom[1]
The analysis of the international migration and the Legatum prosperity index

Jana Bednáriková

Abstract: The Legatum prosperity index has been created as an alternative to Human development index. Based on these indexes we can divide countries into more and less prosperous. The main aim of this paper is to analyse the correlation between crude rate of net migration plus statistical adjustment and Legatum prosperity index. We assume that its eight dimensions, as Economy, Entrepreneurship & Opportunity, governance, education, or health; represent socioeconomic variables which have statistically significant impact on the migration within the European area.

Abstrakt: Legatum prosperity index vznikol ako alternatívny index k Indexu ľudského rozvoja. Na základe týchto indexov môžeme rozdeliť krajiny na menej alebo viac prosperujúce. Cieľom tohto príspevku je analýza závislosti medzi štatisticky upravenou hrubou mierou čistej migrácie a práve Legatum prosperity indexom, pričom predpokladáme, že jeho osem dimenzii, medzi ktoré patria napríklad ekonomika, podnikanie a príležitosti, vláda, vzdelanie, či zdravie, predstavujú také socioekonomické premenné, ktoré majú štatisticky významný vplyv na migráciu obyvateľstva v európskom priestore.

Key words: Legatum prosperity index, crude rate of net migration plus statistical adjustment, path analysis.

Kľúčové slová: Legatum prosperity index, štatisticky upravená hrubá miera čistej migrácie, úseková analýza.

JEL classification: F22, O15

1. Úvod

Medzinárodná migrácia v rámci európskeho priestoru nie je ničím novým či neočakávaným, aj napriek voľnému pohybu osôb sa krajiny snažia monitorovať príliv cudzincov na svoje územie v jednotlivých rokoch a takto získané výsledky následne pretransformujú do svojich migračných politík.

Na rozhodovanie jednotlivcov alebo skupín o prípadnej emigrácii z ich domovskej krajiny vplýva veľké množstvo ako ekonomických, tak socioekonomických faktorov, ktoré je potrebné analyzovať, aby krajina získala potrebné informácie na predikovanie migračných tokov. Medzi socioekonomické faktory, ktoré môžu ovplyvniť migračné rozhodovanie patria napríklad, úroveň HDP v krajine, zdravie, vzdelanie, podnikateľské prostredie, osobná sloboda, istota a bezpečnosť a mnohé ďalšie.

V súčasnosti existujú dva indexy, ktoré s takýmito faktormi pracujú a na základe ich dôkladnej analýzy môžeme krajiny rozdeliť na menej a viac prosperujúce. Prvým indexom je Index ľudského rozvoja (HDI), ktorý sa zaoberá troma dimenziami – životnou úrovňou, vzdelaním a zdravím. Tento index vznikol ako alternatíva k hodnoteniu krajín iba na základe ich HDP. Z našich predchádzajúcich analýz je zrejmé, že medzi HDI a štatisticky upravenou hrubou mierou čistej migrácie (CRN) existuje nelineárna závislosť. HDI však čelí kritike pre malý počet dimenzií, ktorými sa zaoberá, a preto bol vytvorení nový Legatum prosperity

[1] Príspevok vznikol ako súčasť grantového projektu VEGA 1/0127/11

index (LPI), ktorý delí krajiny na prosperujúce a neprosperujúce na základe analýzy ôsmich dimenzii.

Cieľom tohto príspevku je preto analyzovať závislosť medzi CRN a LPI a využitím úsekovej analýzy analyzujeme zároveň závislosť medzi jednotlivými dimenziami a CRN, pričom skúmame ako priamu, tak nepriamu závislosť. Na skúmanie závislosti medzi CRN a LPI využívame taktiež neparametrické metódy.

2. Materiál a metódy

Štatistické údaje, ktoré sú v príspevku použité pochádzajú z Eurostatu a z reportu Legatum Institutu a boli spracované v štatistickom programe SAS a v programe Microsoft Excel. Pod pojmom *CRN* rozumieme podiel medzi štatisticky upravenou čistou migráciou v danom roku k priemernému počtu populácie v tom istom roku. Hodnota je vyjadrená na 1 000 obyvateľov. CRN sa teda rovná rozdielu medzi hrubou mierou zmeny populácie a hrubou mierou prirodzenej zmeny (Eurostat, 2013).

Pod pojmom *Legatum prosperity index* rozumieme multidimenzionálny kompozitový index, ktorý analyzuje prosperitu 142 krajín sveta. Na tento účel využíva 89 premenných, pričom existuje štatisticky významná závislosť medzi nimi a príjmami a blahobytom, ktorá sa premieta do každého z ôsmich sub-indexov, ktorými sú: ekonomika (EK), podnikateľské prostredie a príležitosti (PP), vláda/štýl riadenia (V), vzdelanie (VZ), zdravie (Z), istota a bezpečnosť (IaB), osobná sloboda (OS) a sociálny kapitál (SC). Výsledný celkový Prosperity index vznikne ako geometricky priemer získaného skóre v každom sub-indexe, pričom v tejto fáze má každé získané skóre rovnakú váhu. (Prosperity, 2013)

Matematicky môžeme celkový index prosperity zapísať ako:

$$PI_T(S) = \left[\left(\frac{1}{8}\right) S1,T + \cdots + \left(\frac{1}{8}\right) S8,T\right] \quad (1)$$

Úseková analýza nám umožňuje skúmať závislosť medzi niekoľkými exogénnymi (nezávislými) a endogénnymi (závislými) premennými, pričom skúmané premenné môžu byť závislé vo vzťahu k niektorým premenným a nezávislé vo vzťahu k iným analyzovaným premenným. Grafickým znázornením úsekovej analýzy je úsekový diagram. Pri úsekovej analýze využívame na analýzu lineárnej závislosti medzi jednotlivými premennými Pearsonov korelačný koeficient. P-hodnota pre Pearsonov korelačný koeficient je vypočítaná pomocou online štatistickej kalkulačky (Danielsoper), pričom sa zameriavame na p-hodnotu pre dvojstranný test významnosti. Matematický zápis pre úsekovú analýzu môžeme nájsť napríklad v práci Mosesa (2006).

Nech k je počet premenných X a Y je analyzovaná závislá premenná, r_{ij} (i, j = 1, 2, ..., k) sú Pearsonové korelačné koeficienty medzi premennými X_i a X_j, r_iY (i = 1, 2, ... , k) je Pearsonov korelačný koeficient medzi premennými X_i a Y. Úsekový koeficient p_iY (i = 1, 2, ..., k) je potom kalkulovaný prostredníctvom nasledujúcej lineárnej rovnice:

$$\begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ \vdots & & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} p_{1Y} \\ \cdots \\ p_{kY} \end{pmatrix} = \begin{pmatrix} r_{1Y} \\ \cdots \\ r_{kY} \end{pmatrix} \quad (2)$$

Úsekový koeficient je niekedy uvádzaný aj ako priamy efekt. Úsekový koeficient faktora U predstavujúceho neznámy faktor, ktorý nie je zahrnutý v modeli sa vypočíta nasledovne:

$$p_{UY} = \sqrt{1 - \sum_{i=1}^{k} r_{iY} p_{iY}} \quad (3)$$

Nepriamy vplyv premennej X_i cez premennú Xj na závislú premennú Y je vypočítaná na základe vzťahu:

$$r_{ij}P_{jY} \text{ (i, j = 1, 2, …, k)} \quad (4)$$

Celkový efekt T všetkých faktorov X, ktoré priamo ovplyvňujú závisle premennú Y je vyjadrený nasledovne:

$$T = \sum_{i=1}^{k} r_{iY} p_{iY} \quad (5)$$

Štatistická významnosť hodnoty T je overená pomocou f štatistiky:

$$f=T(N-k-1)/(1-T) \quad (6)$$

kde N reprezentuje rozsah súboru. Testovacia štatistika, ktorú takto získame, je porovnaná s kritickou hodnotou pre Fisherovú distribúciu:

$$F_{\alpha}\,(k, N-k-1) \quad (7)$$

pričom ak je testovacia štatistika väčšia ako kritická hodnota F_{α} (k, N-k-1), potom má hodnota T na hladine významnosti α štatisticky významný vplyv na skúmanú závislú premennú Y.

Závislosť medzi CRN a LPI sme analyzovali zároveň prostredníctvom Spearmanovho koeficientu a Heoffdingovho testu závislosti.

Vzorec na výpočet Searmanovho koeficientu je nasledovný:

$$\theta = \frac{\sum_i \left((R_i - \bar{R})(S_i - \bar{S}) \right)}{\sqrt{\sum_i (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}} \quad (8)$$

kde R_i je poradie x_i, S_i poradie y_i, $\overline{R}$ je priemer R_i hodnôt a $\overline{S}$ priemer S_i hodnôt.

Hoeffdingov test závislosti môžeme matematicky vyjadriť nasledovne:

$$D = 30\,\frac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)} \quad (9)$$

kde $D_1 = \sum_i (Q_i - 1)(Q_i - 2)$, $D_2 = \sum_i (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2)$ a $D3 = \sum_i (R_i - 2)(S_i - 2)(Q_i - 2)$. R_i je poradie x_i, S_i je poradie y_i, a Q_i je 1 plus a počet bodov oboch hodnôt x a y menších od *i*-teho bodu. (SAS)

Koeficienty regresnej priamky boli odhadnuté prostredníctvom neparametrickej lineárnej regresnej analýzy. Model regresnej priamky môžeme zapísať ako[2]:

$$y_i = a + bx_i - e_i,\ i = 1, 2, \ldots n \quad (10)$$

kde: xi (i = 1, 2, …, n) dané konštanty
a, b neznáme konštanty
ei (i= 1, 2, … n) navzájom nezávislé spojité náhodné premenné

Neparametrický odhad regresného koeficientu b je:

$$\hat{b} = \text{median}\,\{Sij, i < j\} \quad (11)$$

Neparametrický odhad regresného koeficientu a je:

$$\hat{a} = \text{median}\,\{qij, i < j\} \quad (12)$$

3. Výsledky a diskusia

Výsledky prezentované v tejto časti príspevku boli vypočítané využitím vyššie uvedených vzorcov (1) – (12).

Korelácia medzi CRN a LPI a jeho ôsmimi dimenziami je testovaná pomocou Pearsonovho korelačného koeficientu úsekovou analýzou. Korelačnú maticu zachytáva tabuľka 1.

[2] Pre podrobnejšie vysvetlenie matematického vyjadrenia koeficientov a, b regresnej priamky prostredníctvom neparametrickej lineárnej regresie sa nachádza v práci: Stehlíková, B. and Žofajová, A. (1987): Neparametrická lineární regresia. Genetika a šlechtění, 23(2), 1-5.

Tab. 3 Matica: Pearsonov korelačný koeficient

	LPI	CRN	EK	PP	V	VZ	Z	IaB	OS	SC
LPI	**1**	0,3526	0,9344	0,9703	0,9563	0,5670	0,8437	0,8538	0,9273	0,9407
CRN	0,3526	**1**	0,5010	0,3545	0,4089	-0,0932	0,4659	0,1751	0,3595	0,1958
EK	0,9344	0,5010	**1**	0,8878	0,8943	0,4403	0,8965	0,7349	0,8638	0,8199
PP	0,9703	0,3545	0,8878	**1**	0,9572	0,5084	0,7483	0,8094	0,8899	0,9292
V	0,9563	0,4089	0,8943	0,9572	**1**	0,5133	0,7891	0,7416	0,8376	0,9145
VZ	0,5670	-0,0932	0,4403	0,5084	0,5133	**1**	0,3524	0,4762	0,4374	0,5842
Z	0,8437	0,4659	0,8965	0,7483	0,7891	0,3524	**1**	0,7170	0,7733	0,6982
IaB	0,8538	0,1751	0,7349	0,8094	0,7416	0,4762	0,7170	**1**	0,7878	0,7845
OS	0,9273	0,3595	0,8638	0,8899	0,8376	0,4374	0,7733	0,7878	**1**	0,8133
SC	0,9407	0,1958	0,8199	0,9292	0,9145	0,5842	0,6982	0,7845	0,8133	**1**

Zdroj: Vlastné výpočty z údajov Eurostatu a Legatum institute.

V súvislosti so skúmaním závislosti medzi LPI a CRN sme si položili hypotézu H_0: $\varrho = 0$, že medzi danými premennými nie je štatisticky významná lineárna závislosť. Na základe porovnania Pearsonovho korelačného koeficientu pre skúmané premenné LPI a CRN z vyššie uvedenej korelačnej matice, t.j. $\mathbf{r}_{LPI,\ GMI} = 0{,}3526$ a p-hodnoty 0,0649 pre dvojstranný test významnosti musíme na hladine významnosti $\alpha = 0{,}05$ prijať H_0 a teda medzi premennými LPI a CRN neexistuje štatisticky významná lineárna závislosť. Na základe daného výsledku však nemôžeme tvrdiť, že medzi skúmanými premennými neexistuje žiadna korelačná závislosť, aby sme túto hypotézu mohli potvrdiť alebo vyvrátiť, uskutočnili sme aj ďalšie analýzy založené na výpočte Spearmanovoho korelačného koeficientu a Hoeffdingovom teste závislosti. Výsledky týchto analýz sú uvedené v ďalšej časti textu. Výsledky pre úsekovú analýzu, ktorou skúmame závislosť medzi CRN a jednotlivými dimenziami LPI zachytáva tabuľka 2 nižšie. V ľavej časti tabuľky je uvedený priamy vplyv jednotlivých sub-indexov na CRN a v jej pravej časti sú zachytené nepriame vplyvy jednotlivých sub-indexov na CRN.

Tab. 4 Výsledky úsekovej analýzy

Priamy vplyv jednotlivých sub-indexov na CRN		CRN	Nepriamy vplyv sub-indexov na CRN		EK	PP	V	VZ	Z	IaB	OS	SC
	EK	0,7268		**EK**		0,6453	0,6501	0,3200	0,6516	0,5342	0,6278	0,5960
	PP	0,0732		**PP**	0,0650		0,0701	0,0372	0,0548	0,0592	0,0651	0,0680
	V	0,7266		**V**	0,6499	0,6956		0,3730	0,5734	0,5389	0,6087	0,6645
	VZ	-0,2432		**VZ**	-0,1071	-0,1236	-0,1248		-0,0857	-0,1158	-0,1064	-0,1421
	Z	-0,0156		**Z**	-0,0140	-0,0117	-0,0123	-0,0055		-0,0112	-0,0121	-0,0109
	IaB	-0,1562		**IaB**	-0,1148	-0,1264	-0,1158	-0,0744	-0,1120		-0,1230	-0,1225
	OS	-0,0102		**OS**	-0,0088	-0,0091	-0,0086	-0,0045	-0,0079	-0,0081		-0,0083
	SC	-0,8489		**SC**	-0,6960	-0,7887	-0,7763	-0,4959	-0,5927	-0,6659	-0,6904	

Zdroj: Vlastné výpočty z údajov Eurostatu a Legatum institute.[3]

Z tabuľky 3 je zrejmé, že našu hypotézu o existencii štatisticky významnej lineárnej závislosti medzi sub-indexmi a CRN môžeme prijať iba v troch prípadoch a to v prípade sub-indexov ekonómia, vláda/riadenie a zdravie. V ostatných prípadoch nemôžeme zamietnuť hypotézu H_0 a teda medzi danými premennými a CRN neexistuje štatisticky významná lineárna závislosť.

[3] Poznámka 1: Výpočet p-hodnoty pre Pearsonov koeficient bol uskutočnený cez online kalkulačku <http://www.danielsoper.com/statcalc3/calc.aspx?id=44>

Poznámka 2: pre lepšiu prehľadnosť v tabuľke uvádzame hodnoty zaokrúhlené čísla, pri výpočtoch sme však počítali s plným počtom desatinných miest.

Tab. 5 Pearsonove korelačné koeficienty a p-hodnoty

	$r_{EK,\ CRN}$	$r_{PP,\ CRN}$	$r_{V,\ CRN}$	$r_{VZ,\ CRN}$	$r_{Z,\ CRN}$	$r_{IaB,\ CRN}$	$r_{OS,\ CRN}$	$r_{SC,\ CRN}$
Pearsonov korelačný koeficient	0,50099	0,3545	0,4089	-0,0932	0,4659	0,1751	0,3595	0,1958
Príslušná P-hodnota	0,0066	0,0642	0,0308	0,6371	0,0125	0,3729	0,0603	0,3180

Zdroj: Vlastné výpočty z údajov Eurostatu a Legatum institute.

Tabuľka 4 uvádza výsledky pre celkový efekt (T) všetkých sub-indexov a neznámeho faktora U na CRN. Štatistická významnosť celkového efektu T bola vypočítaná na základe vzorca (6).

Tab. 6 Celkový efekt jednotlivých sub-indexov na CRN a neznámeho faktora U

	Hodnota:
Celkový efekt T všetkých sub-indexov	0,5054
Efekt neznámeho faktora U	0,7033

Zdroj: Vlastné výpočty z údajov Eurostatu a Legatum institute.

Hodnota testovacej štatistiky je 13,6522. Kritická hodnota pre Fisherovú distribúciu vypočítaná na základe vzorca (7) je 2,06. Z uvedeného vyplýva, že keď 13,6522 > 2,06, potom je celkový vplyv všetkých sub-indexov na CRN štatisticky významný.

Tabuľka 5 zachytáva výsledky analýzy závislosti medzi LPI a CRN Spearmanovým koeficientom a Hoeffdingovým testom závislosti.

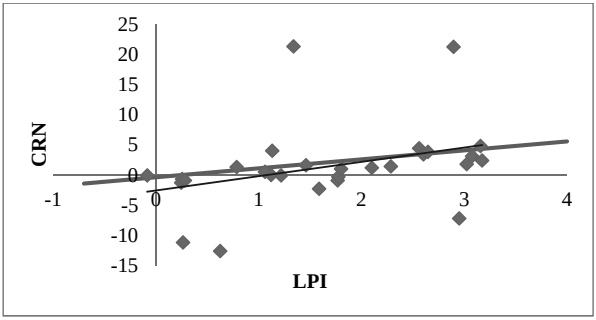
Tab. 7 Korelačné koeficienty a p-hodnoty pre $r_{LPI,\ CRN}$

Spearman Correlation Coefficients,		
Prob > |r| under H0: Rho=0		
	CRN	
LPI	0,56099	
	0,0019	
Hoeffding Dependence Coefficients,		
Prob > D under H0: D=0		
	CRN	
LPI	0,14245	
	0,0003	<.0001

Zdroj: Vlastné výpočty z údajov Eurostatu a Legatum institute.

Hodnota Spearmanovho koeficientu je 0,56099, p-hodnota je 0,0019. Nakoľko 0,0019 < 0,05, môžeme zamietnuť hypotézu H_0 a teda medzi premennými LPI a CRN existuje štatisticky významná korelačná závislosť. Hodnota Hoeffdingovho testu závislosti je 0,14245, p-hodnota je 0,0003. Aj na základe tohto testu sa nám potvrdila závislosť medzi skúmanými premennými.

Obrázok 1 nižšie ilustruje regresnú priamku, ktorej regresné koeficienty a, b boli odhadnuté neparametrickou metódou (hrubšia priamka). Regresná priamka má tvar y = -0,3852 + 1,4859x. Z obrázku je zrejmé, že imigrácia rastie v krajinách s vyšším skóre LPI. Regresná priamka, ktorej regresné koeficienty a,b sú vypočítané lineárnou regresiou je y = -2,5719 + 2,3698x, priamka (tenšia priamka) je strmšia. Na jej trend však vplývali extrémne krajiny ako Cyprus, Luxembursko, Lotyšsko, Litva a Írsko, tento vplyv sa využitím neparametrickej lineárnej regresie odstránil a čím sme získali reálnejší pohľad na závislosť medzi danými premennými.



Obr. 1 Grafické znázornenie závislosti medzi LPI a CRN

4. Záver

Z vyššie uvedených výsledkov je zrejmé, že medzi LPI a CRN neexistuje štatisticky významná lineárna závislosť. Z výsledkov úsekovej analýzy je zrejmé, že štatisticky významná lineárna závislosť je medzi premennými Ekonomika, vláda/riadenie a zdravie. Celkový efekt nezávislých premenných (sub-indexov) na závislú premennú (CRN) je štatisticky významný. Zároveň sa nám na základe Spearmanovho korelačného koeficientu a hoeffdingovho testu závislosti potvrdila štatisticky významná korelačná závislosť medzi LPI a CRN. Z grafického znázornenia regresnej priamky, ktorej regresné koeficienty a, b boli odhadnuté neparametrickou metódou je zrejmé, že imigrácia rastie v krajinách s vyšším indexom LPI.

Literatúra

DANIELSOPER.COM. Statistic calculators. Dostupné na internete. 14.11.2013. <http://www.danielsoper.com/statcalc3/default.aspx>

EUROSTAT.EU. 2013. Demographic balance and crude rates - NUTS 3 regions. Dostupné na internete. 14.11.2013. <http://appsso.eurostat.ec.europa.eu/nui/show.do>

MOSES, E. O. 2006. A User´s Guide to Path Analysis. United States: University Press of America, 2006, 171 s.

PROSPERITY.COM. 2013. Legatum prosperity index. Dostupné na internete. 14.11.2013. <http://prosperity.com/#!/ranking>

SAS. 2013. Base sas(r) 9.2 procedures guide: Statistical procedures, third edition. Dostupné na internete. 14.11.2013. <http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm>

STEHLÍKOVÁ, B. – ŽOFAJOVÁ, A. 1987. Neparametrická lineární regresia. Genetika a šlechtění, 23(2), 1-5.

Adresa autora:

Jana Bednáriková, Ing.
FEP Paneurópska vysoká škola
Tematínska 10, 851 05 Bratislava
janka.bednarikova@gmail.com

Poznámka ku gaussovskej frekvenčnej krivke
A note on the Gaussian frequency curve

Martin Boďa

Abstract: In line with the founders of statistical methodology and under the treatment of descriptive statistics, the paper exposits the Gaussian frequency curve and their use in statistical analysis of data.

Abstrakt: Nadväzujúc na zakladateľov štatistickej metodológie, článok približuje gaussovskú frekvenčnú krivku a jej použitie pri štatistickej analýze dát v rámci výkladu deskriptívnej štatistiky,

Key words: the Gaussian frequency curve, the theory of errors.

Kľúčové slová: gaussovská frekvenčná krivka, teória chýb.

JEL classification: C02, C08.

1. Úvod

Súčasný výklad štatistiky a jej metód sa dosť podstatne odlišuje od spôsobu, akým približovali štatistické skúmanie sveta a jeho zákonitostí autori, ktorých možno označiť bez akéhokoľvek hyperbolizácie za zakladateľov či otcov štatistickej metódy. Publikácie Yuleho (1924), Janka (1942), Yuleho a Kendalla (1950), Kendalla (1954) a prípadne aj Ezekiela a Foxa (1959) sa vo vysvetľovaní značne odchyľujú od súdobých učebníc, akou je napr. kniha Kanderovej a Úradníčka (2005). Novšie publikácie dôsledne (a správne) odlišujú deskriptívnu štatistiku od pojmoslovia teórie pravdepodobnosti a induktívnej štatistiky. Kým v deskriptívnej štatistike sa dáta chápu ako reprezentácia (merania) štatistického znaku (nejakej premennej), cez pravdepodobnostnú prizmu induktívnej štatistiky sú interpretované ako realizácie náhodnej premennej. Pôvodné a normatívne učebnice štatistiky (napr. Yule, 1924; Janko, 1942) prezentovali súvislosti induktívnej štatistiky a jej ciele v úzkej nadväznosti na výklad deskriptívnej štatistiky (a zase celkom správne). Súvisí to pravdepodobne s tým, že až v roku 1933 publikoval Andrej Kolmogorov svoju axiomatickú teóriu pravdepodobnosti, ktorá umožnila rozvinúť teóriu pravdepodobnosti do solídneho matematického rámca a vybudovať pevné premostenie medzi deskriptívnou štatistikou a induktívnou štatistikou. Týmto premostením je, prirodzene, teória pravdepodobnosti. Toto rezultovalo do súčasnej podoby učebníc, v ktorých sú striktne oddelené deskriptívna štatistika a induktívna štatistika a v ktorých sa uplatňuje spravidla silnejší matematický aparát.

Iné umiestnenie malo v počiatočnom výklade štatistickej metódy aj gaussovské rozdelenie, ktoré sa, pochopiteľne, neuvádzalo ako rozdelenie náhodnej premennej s určitou pravdepodobnostnou hustotou, ale ako frekvenčné rozdelenie dát. Cieľom tohto článku je oživiť a pripomenúť pôvodný výklad gaussovského rozdelenia a charakterizovať ho prostredníctvom jeho frekvenčnej krivky. Článok pozostáva z dvoch nosných častí. Nasledujúca časť uvádza pojem gaussovská frekvenčná krivka a osvetľuje súvislosti jej vzniku, kým ďalšia časť na priemerných denných teplotách v Banskej Bystrici počas vykurovacej sezóny 2012/2013 ukazuje, ako skorší autori pristupovali k zisťovaniu, či rozdelenie dáta korešponduje s gaussovskou frekvenčnou krivkou.

2. Gaussovská frekvenčná krivka

Pri skúmaní zákonitostí niektorých prírodných a spoločenských javov sa ukázalo, že výskyt (resp. rozdelenie početnosti) niektorých štatistických znakov možno opísať spoločnou frekvenčnou krivkou. Najvýznamnejšou frekvenčnou krivkou je *gaussovská frekvenčná krivka* označovaná azda častejšie ako *normálne rozdelenie početnosti*. Jej funkčný tvar závisí

iba od dvoch parametrov: lokačného parametra (μ) a od disperzného parametra (σ^2), ktoré majú bezprostredný vzťah k prvým momentovým charakteristikám príslušného štatistického znaku. Jej graf je dobre známy, zvonovitej podoby symetrický okolo lokačného parametra μ.

Ak sa rozdelenie početnosti spojitého štatistického znaku riadi gaussovskou frekvenčnou krivkou, potom pre absolútnu početnosť n_i (resp. analogicky aj relatívnu početnosť f_i) hodnoty x_i platí nasledujúci vzťah

$$n_i \propto \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}, \tag{1}$$

kde μ je lokačný parameter určujúci umiestnenie frekvenčnej krivky a σ^2 je disperzný parameter určujúci tvar a rozšírenie zvona. Symbol e označuje základ prirodzeného logaritmu a symbol $\propto$ znamená „je úmerné“. Možno tiež písať

$$n_i \approx c_0\, \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}, \tag{2}$$

kde c_0 je nejaká konštanta, ktorá vyplýva zo skutočnosti, že početnosti určené krivkou musia zodpovedať celkovému počtu pozorovaní n.

Normálne rozdelenie početnosti (gaussovská frekvenčná krivka) „sa získava ako postupnosť veľkého počtu náhodných alebo nepredvídateľných príčin, pričom každá má malý podiel na celku. Teda rozdelenie hodnôt získaných hodom viacerých kociek a spočítaním bodiek po každom hode frekvenčne inklinuje k normálnemu rozdeleniu [gaussovskej krivke]. Premenné zložené z veľkého počtu malých nezávislých prvkov zvyknú mať tiež mať normálne rozdelenie.“ (Ezekiel a Fox, 1959, s. 9). Pokiaľ ide o druhú situáciu, kedy vzniká normálne rozdelenie, typickým prípadom sú napr. chyby meraní (či pozorovaní). Totiž aj „o chybách pozorovaní sa možno vo všeobecnosti domnievať, že vznikajú súčtom väčšieho počtu prvkov následkom pôsobenia viacerých príčin“ (Yule, 1924, s. 307).

Normálne rozdelenie početnosti vzniklo pôvodne ako rozdelenie chýb nejakého procesu alebo javu. V teórii chýb sa predpokladá, že odchýlky (tzn. chyby meraní) sú výsledkom nekonečne veľkého množstva minoritných príčin, z ktorých každá vedie k malej perturbácii daného procesu alebo javu. Navyše sa vyžaduje, že tieto malé perturbácie sú všetky frekvenčne rovnaké a že kladné aj negatívne perturbácie sú rovnako pravdepodobné. Dá sa ukázať, že frekvenčné rozdelenie je potom gaussovské a relatívna početnosť f_i výskytu hodnoty x_i je zhodná s

$$f_i \approx \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\frac{x_i^2}{2\sigma^2}}, \tag{3}$$

kde parameter σ určuje presnosť meraní. (Yule a Kendall, 1950, s. 187.) Ide teda o gaussovskú frekvenčnú krivku s nulovým lokačným parametrom $\mu = 0$. O gaussovskej frekvenčnej krivke sa preto hovorí nielen ako o „normálnom (gaussovskom) zákone“, ale tiež aj ako o „zákone výskytu chýb“. V literatúre sa ale gaussovská frekvenčná krivka najbežnejšie odvodzuje ako limitný prípad binomickej frekvenčnej krivky (binomického rozdelenia početnosti), ako je to napr. u Janka (1942, s. 77-79).

Normálne rozdelenie početnosti bolo objavené už v roku 1753 Abrahamom de Moivrem, ale sa naň zabudlo a znova bolo objavené o niekoľko rokov neskôr výskumníkmi Carl Friedrich Gauss (1809), Robert Adrian (1809) a Pierre-Simon Laplace (1810) v oblasti teórie pravdepodobnosti a teórie chýb. Obvykle sa celá „ťarcha“ objavu prisudzuje práve Carlovi Friedrichovi Gaussovi, po ktorom sa označuje ako „Gaussovo“ alebo „gaussovské“. Význam tohto prominentného nemeckého matematika, astronóma, geodéta a fyzika dokladá aj skutočnosť, že od roku 1989 do roku 2001 bolo jeho portrét a gaussovskú frekvenčnú krivku možné nájsť na averznej strane nemeckej desaťmarkovej bankovky série BBK3. Toto

zobrazenie Carla Friedricha Gaussa a gaussovskej frekvenčnej krivky možno nájsť na obrázku 1. Označenie „normálne“ je zásluhou Karla Pearsona používa sa variantne s atribútom „gaussovské“. Janko (1942) gaussovskú frekvenčnú krivku však označuje ako „Laplaceovu-Gaussovu“ a v učebnici Kanderovej a Úradníčka (2005, s. 85) sa spomína alternatívny názov normálneho rozdelenia početnosti „Gaussovo-Laplaceovo“.

Obr. 1: Carl Friedrich Gauss a gaussovská frekvenčná krivka na nemeckej desať markovke[1]

Hoci normálne rozdelenie početnosti je jedným z najdôležitejších prostriedkov štatistickej analýzy reálnych javov, nie je úplne ideálnou reprezentáciou mnohých štatistických znakov. Gaussovská frekvenčná krivka bola spopularizovaná v 19. storočí, keď sa zistilo, že množstvo populácií (najmä biometrické populácie výšky a váhy) má hodnoty rozdelené symetricky okolo priemeru s frekvenciou zvonovitého tvaru zhodujúcou sa s gaussovskou krivkou. Takisto sa zistilo, že normálna krivka viac alebo menej kopírovala skutočne napozorované rozdelenia chýb, hoci nie vždy úplne zhodne. (Yule a Kendall, 1950, s. 187.)

V druhej polovici 19. storočia sa preukázalo na mnohých dátach, že normálne rozdelenie je obvyklé rovnako ako ľubovoľné iné rozdelenie. Dokonca „výskyt normálneho rozdelenia [početnosti] sa začal javiť ako čosi abnormálne. V teórii chýb paradigma normálneho rozdelenia pretrvala Normálne rozdelenie početnosti začalo byť vnímané skôr ako vhodná aproximácia, a nie ako objektívny empirický fakt. V podstate je blízko ľubovoľnému rozdeleniu zvonovitého tvaru a možno ho použiť ako prvú aproximáciu. Častokrát sa ukazuje, že postačuje na deskripciu (bežného) symetrického frekvenčného rozdelenia. (Yule a Kendall, 1950, s. 188.) Treba ale poznamenať, že v oblasti aplikovanej a finančnej ekonómie sa používa niekedy nie náležite a iba s cieľom podstatného zjednodušenia, aj na úkor kvality výsledkov.

Napriek tomu normálne rozdelenie početnosti zostalo rozhodujúcim frekvenčným rozdelením. Yule a Kendall (1950, s. 188-189) systemizujú štyri základné dôvody:

a. Normálne rozdelenie početnosti má množstvo matematických vlastností a jeho teória je veľmi podrobne rozpracovaná, čo uľahčuje jeho použitie.

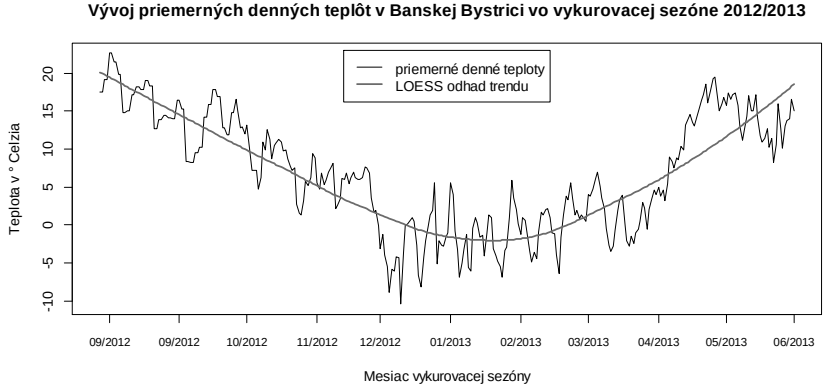
[1] Prevzatý zo stránky Deutsche Bundesbank http://www.bundesbank.de/Redaktion/EN/Bilder/Bilderstrecken/Banknoten_Serie3_BBK/banknoten_bdl_10_deutsche_mark_vs.jpg?__blob=poster4&v=3 (08-11-2013).

b. Väčšina symetrických dát je exaktne alebo približne popísateľných gaussovskou frekvenčnou krivkou.
c. Teória normálneho rozdelenia početnosti bola aplikovaná aj na aproximáciu frekvenčných kriviek, ktoré nie sú gaussovské.
d. Je nezriedka možné použiť (normalizujúcu) transformáciu a dosiahnuť, že rozdelenie štatistického znaku sa priblíži normálnemu rozdeleniu početnosti.

Yule (1924, s. 305-308) odporúčal pri overovaní, či sú dáta kompatibilné s gaussovskou frekvenčnou krivkou, nakresliť do jedného grafu skutočné rozdelenie početnosti a preložiť ho gaussovskou frekvenčnou krivkou. Postup demonštroval na výške dospelých mužov na Britských ostrovov z roku 1883. Dáta evidentne nepredstavujú náhodný výber (pozri Yule, 1924, s. 87-88), a proponovaná metóda si túto vlastnosť nevyžaduje, na rozdiel od v súčasnosti používaného QQ-plotu. Yuleho dáta boli roztriedené do intervalov, pre stredy intervalov Yule spočítal frekvenčnú hodnotu na ľavej strane vzorca (1) a konštantu c_0 vo vzorci (2) stanovil tak, aby početnosti určené krivkou musia zodpovedať celkovému počtu pozorovaní n. Lokačný parameter μ stanovil pritom ako priemernú hodnotu a disperzný parameter ako σ^2 momentový rozptyl. Aj Yule si bol však vedomý, že tento postup je iba „hrubým testom“ (Yule, 1924, s. 308).

3. Aplikácia gaussovskej frekvenčnej krivky na reálne dáta

Aplikácia Yuleovej superpozičnej metódy prekrytia skutočného rozdelenia početnosti gaussovskou frekvenčnou krivkou je ukázaná na dátovej vzorke predstavujúcej denné teploty v Banskej Bystrici vo vykurovacej sezóne 2012/2013 za obdobie od 01-09-2012 do 05-06-2013[2]. Dátovú vzorku o 308 pozorovaniach nemožno, prirodzene, považovať za náhodný výber. Vývoj denných priemerných hodnôt v stupňoch Celzia je znázornený na obrázku 2.

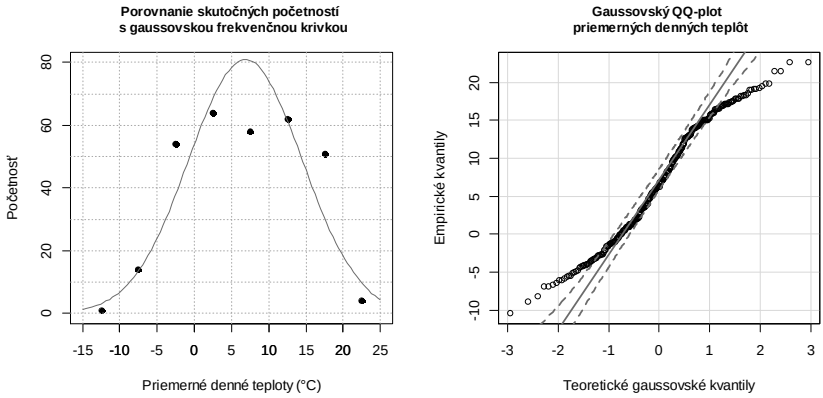


Obr. 2: Vývoj priemerných denných teplôt a odhad ich trendu metódou lokálnych polynómov

Nie je prekvapivé, že na obrázku 2 je zreteľná prítomnosť trendového klimatického vývoja. Na obrázku je znázornený aj odhad trendu denných priemerných denných teplôt získaný Clevelandovou metódou lokálnych polynómov LOESS s voľbou troch štvrtín dát pre lokálne odhadovanie a voľbou kvadratických lokálnych polynómov.

[2] Získané zo stránky spoločnosti STEFE Banská Bystrica, a. s., http://www.stefe.sk/menu/priemerne_denne_teploty/ (08-11-2013).

Je možné, že výskumníci na začiatku 20. storočia by nezobrali do úvahy nestacionaritu priemerných denných teplôt a skúmali by, či priemerné denné teploty v Banskej Bystrici vo vykurovacej sezóne *talis qualis* sa riadia gaussovskou frekvenčnou krivkou. Priemer dát je 6.86 °C a momentová smerodajná odchýlka je 7.59 °C. Výsledky dotazovania sú uvedené na obrázku 3. V ľavej časti sú čiernou bodkou znázornené skutočné početnosti a zvonovitá funkcia zase gaussovskú frekvenčnú krivku, v pravej časti sa nachádza gaussovský QQ-plot, hoci dáta nepredstavujú náhodný výber. Už obrázok 2 poukazuje na to, že rozdelenie početnosti teplôt nebude unimodálne, čo dokumentuje dobre aj ľavá časť obrázka 3. O dátach nemožno usudzovať, že sa riadia gaussovskou frekvenčnou krivkou.



Obr. 3: Gaussovskosť priemerných denných teplôt

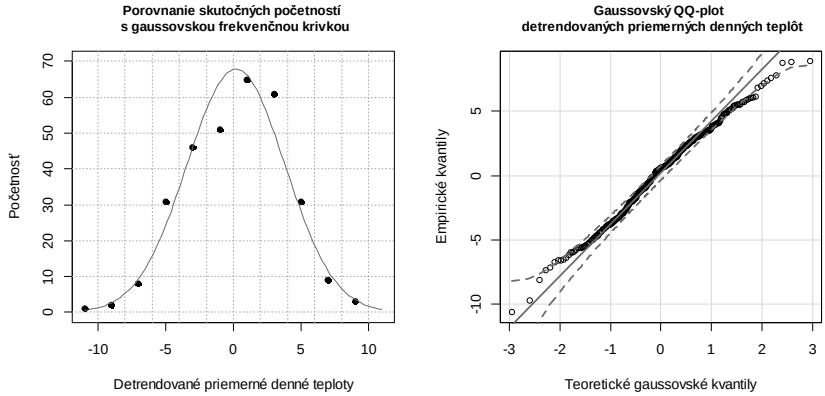
Rozumnejší model pre priemerné denné teploty je zrejme nasledovný. Nech $\{Y_t\}_t$ reprezentuje časový rad priemerných denných teplôt, nech $\{T_t\}_t$ je jeho deterministický trendový komponent a $\{e_t\}_t$ je gaussovský biely šum $WN(\mu,\sigma^2)$. Bude sa predpokladať, že pre každé pozorovanie platí $Y_t = T_t + e_t$. Trendový komponent bol odhadnutý, ako bolo uvedené, Clevelandovým LOESS estimátorom. Priemerné denné teploty boli následne detrendované a z rezultujúcich rezíduí bol spočítaný priemer 0.20 a smerodajná odchýlka 3.62. Výstupy overovania normálneho zákona sú prezentované na obrázku 4. Je zrejmé, že gaussovská frekvenčná krivka je vhodnou aproximáciou frekvenčného rozdelenia reziduálnej zložky. Pri splnení predpokladov je použitie QQ-plotu náležité.

Spracovanie dát a ich grafická prezentácia prebiehali v programe R 3.0.1 (R Core Team, 2013).

4. Záver

Na pozadí článku je záujem o čo najlepší výklad štatistickej metódy a o gaussovskú frekvenčnú krivku. Článok dokazuje, že gaussovské rozdelenie je možné vysvetľovať nielen z pozície náhodnej premennej, ale aj z pozície štatistického znaku („iba“ premennej) v rámci deskriptívnej štatistiky. Gaussovská frekvenčná krivka vznikla v teórii chýb a skutočnosť, že má viacero štatistických konotácií, pripodobnení v reálnom svete a vyznačuje sa veľkým množstvom výhodných štatistických vlastností, z nej robí jedným zo základných stavebných prvkov štatistiky minulosti a súčasnosti, ale nepochybne sa jeho význam nezmení ani pre budúce a modernejšie poňatie štatistickej metódy. Spomedzi veľkého množstva dostupných fyzikálnych, chemických, biometrických demografických, sociologických a ekonomických

dát bola v článku Yuleho superpozičná metóda prekladania skutočných početností (vlastne histogramu) hodnôt spojitého štatistického znaku odhadnutou gaussovskou frekvenčnou krivkou ilustrovaná na dátach o priemerných denných teplotách v Banskej Bystrici počas vykurovacej sezóny 2012/2013. Článok pritom vznikol z osobného záujmu autora o túto problematiku.



Obr. 4: Gaussovskosť detrendovaných priemerných denných teplôt

Literatúra

EZEKIEL, M., FOX, K. A. 1959. *Methods of correlation and regression analysis. Linear and curvilinear*. 3. vyd. New York: Wiley, 1959. Bez ISBN.

JANKO, J. 1942. *Jak vytváří statistika obrazy světa a života. I. díl.* Praha: Jednota českých matematiků a fysiků v Praze, 1942. Bez ISBN.

KANDEROVÁ, M., ÚRADNÍČEK, V. 2005. *Štatistika a pravdepodobnosť pre ekonómov. 1. časť.* Banská Bystrica: OZ Financ, 2005. ISBN 80-968702-9-7.

KENDALL, M. G. 1945. *The advanced theory of statistics.* Vol. I. 2. vyd. Londýn: Griffin, 1945. Bez ISBN.

R CORE TEAM 2013. *R: A language and environment for statistical computing.* Viedeň: R Foundation for Statistical Computing, http://www.R-project.org/.

YULE, G. Y. 1924. *An introduction to the theory of statistics.* 7. vyd. Londýn: Griffin, 1924. Bez ISBN.

YULE, G. Y., KENDALL, M. G. 1950. *An introduction to the theory of statistics*. 14. prepr. a rozš. vyd. Londýn: Griffin, 1950. Bez ISBN.

Adresa autora:

Martin Boďa, Mgr. Ing., PhD.
Univerzita Mateja Bela v Banskej Bystrici
Ekonomická fakulta
Tajovského 10, 975 90 Banská Bystrica
martin.boda@umb.sk

Využitie radiálnych bázických funkcií pre modelovanie interpolačných plôch zrážkových intenzít
Using radial basis functions for modelling of the interpolation surfaces of the spatial rainfall

Róbert Bohdal, Mária Bohdalová

Abstract: This paper evaluates various interpolation methods based on radial basis functions. The aim of this paper is the modelling of spatial and temporal scaling exponent of rainfall over a range of scales. We will compare two most used methods: Thin Plate Spline method and Hardy`s multi quadric function. Moreover, the second method will be evaluated for various parameters. These interpolation methods are employed, and examples of the results are given. Both modelling approaches are used to predict the rainfall intensity over all places in Slovakia. These model approaches give acceptable forecasts. Their accuracy will be evaluated by bootstrapping statistical approach. The models can be used to predict in real time the spatial rainfall.

Abstrakt: V príspevku vyhodnocujeme dve rôzne interpolačné metódy založené na radiálnych bázických funkciách s cieľom modelovať priestorový a časový škálovací exponent za účelom predpovedania zrážkovej intenzity pre ľubovoľné miesto na Slovensku. Porovnávame dve najpoužívanejšie metódy: tenkosplajnovú a Hardyho. Ich presnosť vyhodnocujeme boot-strappingovým štatistickým prístupom. V závere stanovíme vhodný model, ktorý môže byť použitý na stanovenie intenzity zrážok v priestore a čase.

Key words: Thin Plate Spline, Hardy`s Multiquadric Function, rainfall, scaling exponent

Kľúčové slová: tenkostenný splajn, Hardyho multikvadratická funkcia, dažďové zrážky, škálovací exponent.

JEL classification: C02, C13, C63, C65

1. Úvod

V príspevku riešime problém interpolovania nerovnomerne rozložených bodov, ktoré predstavujú namerané hodnoty zrážok z jednotlivých merných staníc na Slovensku. Pre tento účel budeme skúmať vhodnosť použitia interpolačných metód založených na radiálnych bázických funkciách. Na záver vyhodnotíme presnosť jednotlivých metód bootstrapingovým prístupom a metódou Jack knife.

2. Metódy radiálnych bázických funkcií

Radiálne bázické funkcie získali obrovskú popularitu len nedávno. Ukázalo sa, že sú vhodné pre viacrozmernú interpoláciu nerovnomerne rozmiestnených údajov (Dyn, 1987, 1989), (Buhmann, 2000), (Iske 2003), (Powell, 1992). Sú jednoduché na implementáciu a vytvárajú interpolačnú plochu s dostatočnou hladkosťou. Ako prvý zaviedol tieto interpolačné funkcie na začiatku 70-tych rokov minulého storočia Hardy (Hardy, 1971). V uvedenej práci prvý raz použil metódu multikvadrík, ktorá opisuje konštrukciu multikvadratickej plochy pomocou súčtu častí rotačných plôch 2-dielnych hyperboloidov (resp. rotačných paraboloidov, v závislosti od hodnoty μ).

Radiálne bázické funkcie (RBF) sú definované pre danú množinu bodov $\boldsymbol{P} = \boldsymbol{p}_i[x_i, y_i] \in \mathbb{E}^2$, pre ktoré poznáme hodnoty $z_i \in \mathbb{R}$, $i = 1, \ldots, n$. Našou úlohou je nájsť takú interpolačnú funkciu f: $\mathbb{E}^2 \rightarrow \mathbb{R}$, pre ktorú platí $f(\boldsymbol{p}_i) = z_i$, pre $i = 1, 2, \ldots, n$.

Pomocou metódy radiálnych bázických funkcií môžeme interpolačnú funkciu $f(x, y) = f(\boldsymbol{x})$ zapísať v nasledujúcom tvare (Hoschek, Lasser, 1993):

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i R\left(\left\|\boldsymbol{x} - \boldsymbol{p}_i\right\|\right) + \Phi_m(\boldsymbol{x}), \tag{1}$$

pričom $$\Phi_m(\boldsymbol{x}) = \sum_{k=1}^{l} c_k \Phi_k(\boldsymbol{x}), \tag{2}$$

kde $\Phi_k(\boldsymbol{x}) \in \pi_m^2$, $l = \dim(\pi_m^2) = \binom{m-1+2}{2}$. Symbolom π_m^d označujeme lineárny priestor obsahujúci všetky polynómy nad poľom $\mathbb{R}$ s d premennými a stupňa nanajvýš $m-1$.

Funkcie $R(r_i) = R\left(\left\|\boldsymbol{x} - \boldsymbol{p}_i\right\|\right)$, $r_i \geq 0$ vyjadrujú euklidovskú vzdialenosť bodov $\boldsymbol{x}, \boldsymbol{p}_i$ a v literatúre (Hardy, 1971), (Iske, 2003), (Hoschek, Lasser, 1993), (Fogel, Tinney, 1996) sú známe pod názvom radiálne bázické funkcie. Spomedzi všetkých RBF sú najpoužívanejšie tzv. polyharmonické splajny. Do tejto triedy funkcií patria aj tenkostenné splajny. Ak potrebujeme obmedziť globálny vplyv transformačnej funkcie zvolíme také RBF, ktoré nepoužívajú polynomický člen $\Phi_m(x)$ a ktorých vplyv v danom bode $\boldsymbol{p}_i$ klesá s rastúcou vzdialenosťou (napr. gaussovské funkcie resp. recipročné multikvadriky (Iske, 2003)). Veľkosť vplyvu týchto funkcií môžeme riadiť pomocou parametrov σ a c, prípadne μ (pozri tabuľku 1), avšak žiadna voľba týchto parametrov nezabezpečí čisto lokálny vplyv. Nevhodný výber parametra c môže viesť k riešeniu sústavy so zle podmienenou maticou vo vzťahu (3).

Tab. 8 Prehľad radiálnych bázických funkcií R(r) (Iske, 2003)

Radiálna bázická funkcia	***R(r)***	***Hodnoty parametrov***[1]	***m***
Polyharmonické splajny v priestore $\mathbb{R}^d$	r^{2k-d} $r^{2k-d}\log(r)$	d je nepárne, $2k>d$ ($k = 3$, $d = 3$) d je párne, $2k>d$ ($k = 2$, $d = 2$)	$m = k - \left\lceil \frac{d}{2} \right\rceil + 1$
Gaussovské funkcie	$\exp\left(\frac{-r^2}{2\sigma^2}\right)$	$\sigma > 0, \left(\sigma = \sqrt{\frac{1}{2}}\right)$	$m = 0$
Multikvadriky	$(c^2+r^2)^{\mu}$	$c, \mu > 0$ ($c = 1$, $\mu = 1/2$)	$m = \lceil \mu \rceil$

Interpolačnú funkciu $f(\boldsymbol{x})$ nájdeme tak, že vypočítame neznáme koeficienty λ_i pre jednotlivé radiálne bázické funkcie $R(r_i)$ a podobne aj neznáme koeficienty c_k pre polynomický člen. Neznáme hodnoty $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)^{\mathrm{T}}$ vo vzťahu (1) a $\mathbf{c} = (c_1, ..., c_l)^{\mathrm{T}}$ vo vzťahu (2) určíme riešením nasledujúcej sústavy rovníc:

$$\begin{pmatrix} \mathbf{A} & \mathbf{P} \\ \mathbf{P}^{\mathbf{T}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{z} \\ \mathbf{0} \end{pmatrix} \tag{3}$$

kde $\mathbf{A}_{i,j} = R\left(\left\|\boldsymbol{p}_j - \boldsymbol{p}_i\right\|\right)$, pre $i, j = 1, ..., n$ a $\mathbf{P}_{i,k} = \Phi_k(\boldsymbol{p}_i)$, pre $i = 1, ..., n$ a $k = 1, ..., l$. Poznamenajme, že sústava rovníc (3) má riešenie práve vtedy, keď sú body $\boldsymbol{p}_i$ nekolineárne.

1.1 Metóda tenkostenných splajnov

Tenkostenné splajny (thin plate splines) patria do triedy polyharmonických splajnov. Samotný názov „tenkostenné splajny“ zaviedol v roku 1977 Duchon (Duchon, 1977). Tento názov je odvodený zo vzťahu, v ktorom sa hľadá minimum integrálu opisujúceho rozloženie tzv. energie ohybu (bending energy) na nekonečne tenkej elastickej doske.

[1] Vhodné hodnoty parametrov uvádzame v zátvorke

Duchon ukázal, že pre konečnú množinu bodov $\boldsymbol{P} \subset \mathbb{R}^d$ je interpolant $f(\boldsymbol{x})$ určený vzťahom:

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i R_{d,k}\left(\left\|\boldsymbol{x} - \boldsymbol{p}_i\right\|\right) + \sum_{|\boldsymbol{\alpha}|<k} c^{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}, \tag{4}$$

kde $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}$, $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_d$.

Po dosadení $d = 2$ (dimenzia priestoru $\mathbb{E}^2$), $k = 2$ (pozri tabuľku 1) môžeme predchádzajúci vzťah prepísať do tvaru:

$$f(x,y) = \frac{1}{2} \sum_{i=1}^{n} \lambda_i r_i^2 \log(r_i^2) + c_1 + c_2 x + c_3 y, \tag{5}$$

kde $r_i^2 = (x - x_i)^2 + (y - y_i)^2$

Využitím vzťahu (3) môžeme neznáme $\lambda_1, ..., \lambda_n$ a c_1, c_2, c_3 vypočítať zo sústavy rovníc vyjadrených v maticovom tvare:

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & x_1 & x_2 & \cdots & x_n \\ 0 & 0 & 0 & y_1 & y_2 & \cdots & y_n \\ 1 & x_1 & y_1 & 0 & r_{21}^2 \log\left(r_{21}^2\right) & \cdots & r_{n1}^2 \log\left(r_{n1}^2\right) \\ 1 & x_2 & y_2 & r_{12}^2 \log\left(r_{12}^2\right) & 0 & \cdots & r_{n2}^2 \log\left(r_{n2}^2\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & y_n & r_{1n}^2 \log\left(r_{1n}^2\right) & r_{2n}^2 \log\left(r_{2n}^2\right) & \cdots & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \lambda_1/2 \\ \lambda_2/2 \\ \vdots \\ \lambda_n/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}. \tag{6}$$

1.2 Metóda Hardyho multikvadrík

Táto metóda je veľmi podobná predchádzajúcej metóde. Rozdiel je len v tom, že používa iné RBF a pre $d = 2$ (dimenzia priestoru $\mathbb{E}^2$) nepoužíva polynomický člen. Pre náš interpolačný problém dostávame nasledujúcu interpolačnú funkciu:

$$f(x,y) = \sum_{i=1}^{n} \lambda_i \sqrt{r_i^2 + c^2}, \tag{7}$$

kde $r_i^2 = (x - x_i)^2 + (y - y_i)^2$.

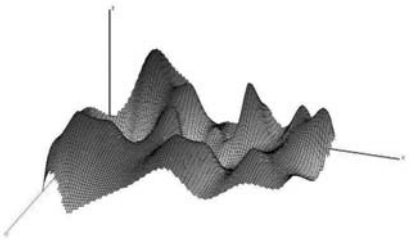
Hodnota c určuje tvar výslednej funkcie. Vo všeobecnosti platí, že menšia hodnota parametra c vytvára v grafe funkcie tzv. „ostré extrémy“ (pozri obrázok 2), zatiaľ čo jeho väčšia hodnota „vyhladzuje“ funkciu (pozri obrázok 4). V literatúre sa uvádzajú viaceré možnosti ako ho vhodne zvolit’ (Hoschek, Lasser, 1993), (Fogel, Tinney, 1996). Tu uvedieme niektoré z nich:

- $c = 0.815d$, kde d je priemerná vzdialenosť bodov $\boldsymbol{p}_i$ množiny $\boldsymbol{P}$ k ich najbližším susedom,
- $c = 1.25D/n$, kde D je priemer najmenšej kružnice, ktorá obsahuje všetky body množiny $\boldsymbol{P}$,
- $c = \sqrt{\frac{1}{10} \max_{i,j} \left\|x_i - x_j\right\|}$
- $c = \sqrt{\frac{3}{5} \min_{i,j} \left\|x_i - x_j\right\|}$. (8)

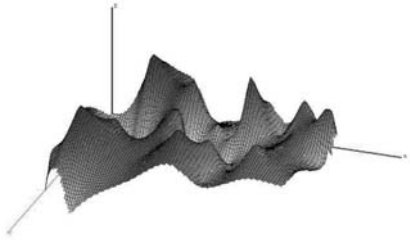
2 Aplikácia

Cieľom nášho príspevku bolo porovnať dve interpolačné metódy na reálnych údajoch a určiť vhodnejšiu z nich pomocou známych štatistických mier. Ako vstupné údaje sme použili maximálne intenzity zrážok zo 63 zrážkomerných staníc z celého územia Slovenska, pre trvania dažďov 5 až 180 min. V týchto zrážkomerných staniciach bola použitá metóda jednoduchého škálovania na určenie návrhových dažďových intenzít pre celé Slovensko (pozri (Látečková, 2013)). Pre dané údaje sme určili jednotlivé modely, ktoré sme overovali nami navrhnutou metodikou. Zo 63 zadaných bodov (zrážkomerných staníc) sme vytvorili 100 testovacích vzoriek (ďalej vzorka 1). V každej testovacej vzorke sme náhodne vylučovali merania, pričom ich počet bol určený náhodným celým číslom z intervalu od 1 po 5. Na overenie modelov sa zvyčajne používa metodika Jack knife, známa tiež ako bumerangový test, v ktorej sa zo vzorky systematicky vylučuje vždy len jeden bod a preto je možné získať v našom prípade len 63 testovacích vzoriek (ďalej vzorka 2).

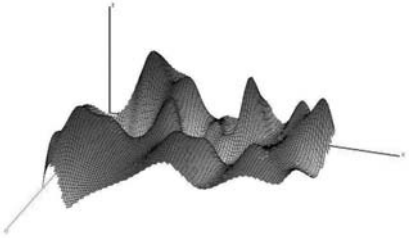
Pre obe vzorky sme použili dve vyššie uvedené interpolačné plochy založené na RBF (pozri kapitolu 1). Na rozdiel od tenkostenných splajnov metóda Hardyho multikvadrík používa i vstupný parameter *c*. Hodnoty tohto parametra sme postupne menili od 0.2 do 63 (Tab. 2). Hodnotu $c = 63$ sme určili pomocou vzťahu (8). Ostatné vzťahy viedli k vysokej hodnote parametra *c*, preto sme ich nezahrnuli do testovania.



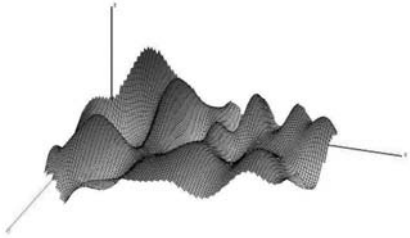
Obr. 2: Interpolácia Tenkosplajnovou plochou



***Obr. 3: Interpolácia Hardyho multikvadrikou pre c* = 0.2**



Obr. 4: ***Interpolácia Hardyho multikvadrikou pre c* = 3**



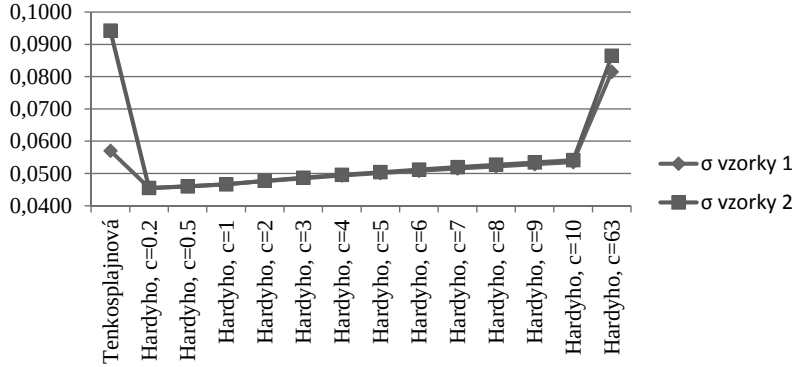
***Obr. 5: Interpolácia Hardyho multikvadrikou pre c* = 63**

3 Záver

Z grafu na obrázku 5 vyplýva, že tenkosplajnové interpolačné plochy sú citlivé na výber vzorky, pretože obsahujú polynomický člen, ktorý zabezpečuje globálny vplyv[2]. Odhady parametra *c* pre interpolačné plochy vytvorené Hardyho multikvadrikami uvedené v kapitole

[2] Globálny vplyv znamená, že zmena jedného interpolovaného bodu spôsobí zmenu na veľkej časti plochy.

1.2 sú pre naše údaje nevyhovujúce, keďže aj najmenšia hodnota parametra $c = 63$ už dáva horšie výsledky v porovnaní s tenkosplajovou plochou. Heuristickou metódou sme odhadli, že hodnota parametra $c = 3$ pre našu dátovú množinu dáva uspokojivé výsledky s dostatočne vizuálne hladkou plochou. Príspevkom sme ukázali, že ak určíme vhodnú hodnotu parametra c dostaneme oveľa presnejšie výsledky ako u tenskosplajnových plôch, ktoré sú vo všeobecnosti používanejšie, pretože nevyžadujú odhad žiadneho parametra. V problematike interpolovania škálovacieho exponentu za účelom predpovedania zrážkovej intenzity nebola metóda Hardyho multikvadrík zatiaľ použitá a rovnako nebola zatiaľ použitá nami navrhnutá metóda vytvárania vzoriek pre testovanie vhodnosti správnej metódy pre reálne údaje.



Obr. 6: Porovnanie smerodajných odchýlok interpolačných metód

Tab. 9: Porovnanie σ, AIC a BIC interpolačných metód

Metóda	σ vzorky 1	σ vzorky 2	AIC	BIC
Tenkosplajnová	0.0570	0.0942	-3.7296	-0.0325
Hardyho, *c*=0.2	0.0455	0.0455	-4.1787	-0.4816
Hardyho, *c*=0.5	0.0460	0.0460	-4.1574	-0.4603
Hardyho, *c*=1	0.0466	0.0466	-4.1303	-0.4332
Hardyho, *c*=2	0.0477	0.0477	-4.0871	-0.3900
Hardyho, *c*=3	0.0485	0.0487	-4.0503	-0.3532
Hardyho, *c*=4	0.0494	0.0496	-4.0170	-0.3199
Hardyho, *c*=5	0.0501	0.0504	-3.9867	-0.2896
Hardyho, *c*=6	0.0508	0.0512	-3.9578	-0.2607
Hardyho, *c*=7	0.0515	0.0519	-3.9310	-0.2339
Hardyho, *c*=8	0.0522	0.0527	-3.9050	-0.2079
Hardyho, *c*=9	0.0529	0.0534	-3.8802	-0.1831
Hardyho, *c*=10	0.0535	0.0541	-3.8563	-0.1592
Hardyho, *c*=63	0.0815	0.0864	-3.7413	-0.0442

4 Poďakovanie

Príspevok bol podporený grantom SPINKLAR-3D (Project VEGA No. 1/1106/11)

Literatúra

BARA, M., GAÁL, L., KOHNOVÁ, S., SZOLGAY, J., HLAVČOVÁ, K., 2008. Simple scalingof extreme rainfall in Slovakia: a case study. In: *Meteorological Journal*. 4(11), str. 153–157.

BOHDAL, R., BOHDALOVÁ, M., 2009. Scaling exponent of rainfall modeling by interpolation methods. In: *Forum Statisticum Slovacum* 3, str. 1-6.

BUHMANN, M.D., 2000. Radial basis functions. *Acta Numerica*, str. 1–38.

DUCHON, J., 1977. *Lecture Notes in Mathematics* 571. Springer–Verlag, Berlin, str. 85–100.

DYN N., 1987. Interpolation of scattered data by radial functions, In: *Topics in Multivariate Approximation*, (Eds. Chui C.K., Schumaker L.L. and Utreras F.I.), Academic Press, New York, str. 47–61.

DYN N., 1989. *Interpolation and approximation by radial and related functions*, (Eds. Chui C.K., Schumaker L.L. and Ward J.D.), Academic Press, New York, str.211–234.

FOGEL, D., TINNEY, L., 1996. *Image Registration using Multiquadric Functions, the Finite Element Method, Bivariate Mapping Polynomials and the Thin Plate Spline*. National Center for Geographic Information and Analysis, str. 1–63.

FRANKE, R., Nielson, G., 1980. Smooth interpolation of large sets of scattered data. *Intern. Journal for Numerical Methods in Engineering* (15), str. 1691–1704.

FRANKE, R. 1982. Scattered data interpolation: Test of some methods. *Mathematics of Computation* 38(157), str. 181–200.

HARDY, R., 1971. Multiquadric equations of topography and other irregular surfaces. In: *Journal Geophysical Research* U(76), str. 1905–1915.

HOSCHEK, J., LASSER, D., 1993. *Fundamentals of Computer Aided Geometric Design*. A K Peters, Wellesley, MA, str. 388–421.

ISKE, A., 2003. *Radial basis functions: basics, advanced topics and meshfree methods for Transport Problem*. Seminar of Mathematics, str. 247–274.

LÁTEČKOVÁ, J. 2013. *Škálovanie intenzít krátkodobých dažďov v jednotlivých mesiacoch a sezónach na Slovensku*. Dizertačná práca, SvF STU v Bratislave, 126s.

LÁTEČKOVÁ, J., KOHNOVÁ, S., GAÁL, L., SZOLGAY, J. 2011. Odvodenie škálovacích exponentov intenzít dažďov pre jednotlivé mesiace teplého polroku vo vybraných staniciach oblasti severovýchodného Slovenska. In: *Acta Hydrologica Slovaca*, špeciálne číslo, 12, 47–54.

MENABDE, M., SEED, A., PEGRAM, G. 1999. A simple scaling model for extreme rainfall. *Water Resour. Res*., 35 (1), 1999, s. 335–339

POWELL, M.J.D.,1992. The theory of radial basis function approximation in 1990. In: *Advances in numerical analysis II*: wavelets, subdivision and radial basis functions, (Ed. Light W.A.), Clarendon Press, Oxford, str. 105–210.

Adresa autorov:

Róbert Bohdal, RNDr. PhD.
Fakulta matematiky, fyziky a informatiky
Univerzity Komenského
Mlynská Dolina, 842 48 Bratislava
Robert.bohdal@fmph.uniba.sk

Mária Bohdalová, doc., RNDr., PhD.
Fakulta managementu
Univerzity Komenského
Odbojárov 10, 820 05 Bratislava
maria.bohdalova@fm.uniba.sk

Pravdepodobnostná analýza na časových škálach a jej aplikácie na modelovanie riadenia kvality výroby firiem
Probability analysis on time scales and some applications to the modelling of firms

Eva Brestovanská

Abstract: The article deals with some new concepts of the mathematical calculus (mathematical analysis on time scales), which was first published in dissertation by Stefan Hilgera in 1988. The probability is the perfect discipline, in which this theory certainly has its application. It unifies the concept of the standard notion of the discrete and continuous random variable and opens up new possibilities for interesting examples in the field of economics and finance.

Abstrakt: Článok sa zaoberá niektorými pojmami nového matematického kalkulu (matematická analýza na časových škálach), ktorý bol prvýkrát publikovaný v dizertačnej práci Stefana Hilgera v roku 1988. Pravdepodobnosť je ideálna disciplína, v ktorej má rozhodne táto teória svoje uplatnenie. Zjednocuje štandardný pojem diskrétnej a spojitej náhodnej premennej a otvára nové možnosti riešenia zaujímavých príkladov v oblasti ekonómie a financií.

Key words: time scales, graininess function, Δ - probability, Δ - measure, binomial and Poisson random variables on $h\mathbb{N}$.

Kľúčové slová: časová škála, funkcia zrnitosti, Δ - miera, Δ - pravdepodobnosť, binomické a Poissonovo rozdelenie.

1. Úvod

Základy teórie miery na časových škálach môžeme nájsť v práci M. Bohnera a A. Petersona (rok 2004). Medzi ďalších autorov, ktorí sa venujú tejto problematike, zameranej na teóriu pravdepodobnosti a aplikácie týchto výsledkov v oblasti financií a ekonómie patrí T. Matthews (rok 2011), svoje výsledky publikoval v dizertačnej práci, vypracovanej pod vedením M. Bohnera. Riemannov Δ - integral na time scales bol definovaný G. Guseinovom v roku 2002. Pojem Δ - miera a Lebesguevov a Δ - integral bol zavedený G. Guseinovom v roku 2003 a tiež v prácach A. Cabada v roku 2006, U. Ufuktepe a A. Deniza v roku 2009 a T. Rzezuchowského v roku 2005.

V kapitole 2. sú uvedené definície pojmov Δ - miera a Δ - pravdepodobnosť na časových škálach a následne vety z nich vyplývajúce. V kapitole 3. sú popísané pravdepodobnostné rozdelenia binomické a Poissonové na časovej škále $h\mathbb{N}$.

2. Miera na časových škálach

Definícia 1.: Časová škála (time scales) je ľubovoľná neprázdna uzavretá podmnožina reálnych čísel $\mathbb{R}$. V ďalšom texte ju budeme označovať písmenom $\mathbb{T}$.

Definícia 2.: Pre $t\in\mathbb{T}$ definujeme predný a spätný operátor skoku $\sigma(t)$; $\rho(t)$: $\mathbb{T}\to\mathbb{T}$, $\sigma(t):= \inf\{s\in\mathbb{T}: s > t\}$, $\rho(t):= \sup\{s\in\mathbb{T}: s < t\}$.

Definícia 3.: Zobrazenie $\mu : \mathbb{T} \to [0,\infty)$ nazveme funkciou zrnitosti, resp. mierou zrnitosti časovej škále $\mathbb{T}$, ktorá je definovaná vzťahom $\mu(t):= \sigma(t) - t$, alebo $\mu(t):= t - \rho(t)$ pre $t\in\mathbb{T}$.

Definícia 4.: Ak $\rho(t)=t=\sigma(t)$, t je hustý bod, ak $\sigma(t) > t$, t je sprava riedky bod, $\rho(t) < t$, t je zľava riedky bod, ak $\sigma(t)=t$, t je sprava hustý bod, ak $\rho(t)=t$, t je zľava hustý bod, ak $\rho(t) < t < \sigma(t)$, t je izolovaný bod.

Nech $\mathbb{T}$ je časová škála a $\sigma(t)$ je predný operátor skoku. Nech $\mathcal{F}$ je systém zľava uzavretých a sprava otvorených intervalov na $\mathbb{T}$: $\mathcal{F} = \{[a, b) \cap \mathbb{T} : a, b \in \mathbb{T}, a \leq b\}$. Potom m: $\mathcal{F} \rightarrow [0, \infty]$ je množinová funkcia, ktorá priraďuje každému intervalu jeho dĺžku: $m([a, b)) = b - a$. Množinová funkcia m je spočitateľná aditívna miera na $\mathcal{F}$.

- $[a, a) = \emptyset$, $m([a, a)) = 0$ pre $a \in \mathbb{T}$.
- Pre všetky dvojice dizjunktných intervalov [a, b), kde $a, b \in \mathbb{T}$ platí:

$$m(\bigcup_{i=1}^{\infty} [a,b)) = \sum_{i=1}^{\infty} m([a,b)) .$$

- $m([a, b)) = b - a \geq 0$ pre $b \geq a$.

Z predchádzajúcich vlastností vyplýva, že ak $\{I_n\}$ je postupnosť dizjunktných intervalov z $\mathcal{F}$, potom platí:

$$m(\bigcup_{n=1}^{\infty} I_n) = \sum_{n=1}^{\infty} m(I_n) .$$

Nech $E \subset T$. Ak existuje najmenší konečný alebo spočitateľný systém intervalov $I_n \in \mathcal{F}$, taký že $E \subset \bigcup_n I_n$, potom $m^*(E)$ nazývame vonkajšou mierou množiny E.

$$m^*(E) = \inf_{E \subset \bigcup_n I_n} \sum_n m(I_n) .$$

Ak neexistuje žiadne také pokrytie E potom $m^*(E) = \infty$.

Definícia 5.: Ak množina $E \subset \mathbb{T}$, potom hovoríme, že E je m^*- merateľná alebo Δ - merateľná ak nasledujúca rovnica $m^*(A) = m^*(A \cap E) + m^*(A \cap E^c)$, kde $E^c = \mathbb{T} - E$, platí pre všetky podmnožiny A z $\mathbb{T}$. Ak E je Δ - merateľná, potom E^c je tiež Δ - merateľná. Je zrejmé, že prázdna množina $\{\emptyset\}$ a $\mathbb{T}$ sú Δ - merateľné. Nech $\mathcal{M}(m^*) = \{ E \subset \mathbb{T}: E$ je Δ - merateľná$\}$ je systém merateľných množín, $\mathcal{M}(m^*)$ je σ - algebra.

Definícia 6.: Zúženie m^* do $\mathcal{M}(m^*)$ sa nazýva Lebesgueová Δ - miera a označuje sa μ_Δ.

$$m^*(E) = \mu_\Delta(E) \text{ ak } E \subset \mathcal{M}(m^*).$$

Tvrdenie1.: Nech $\{E_n\}$ je nekonečná, klesajúca postupnosť Δ - merateľných množín, postupnosť $E_1 \supset E_2 \supset \ldots \supset E_n \supset \ldots \in \mathbb{F}$, $m^*(E) < \infty$. Potom

$$m^*(\bigcap_{n=1}^{\infty} E_n) = \lim_{n \to \infty} m^*(E_n) .$$

Tvrdenie2.: (vlastnosti m^*)

- $m^*(\emptyset) = 0$;
- Ak $E \subset F$ potom $m^*(E) \leq m^*(F)$;
- $\{E_n\}$ je postupnosť prvkov z $\mathcal{F}$, potom $m^*(\bigcup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} m^*(E_n)$.

Veta1.: Pre každý izolovaný bod $t_0 \in \mathbb{T} - \{\max \mathbb{T}\}$ je množina $\{t_0\}$ Δ - merateľná a pre Δ - mieru platí: $\mu_\Delta(\{t_0\}) = \sigma(t_0) - t_0$.

Veta2.: Ak a, b ∈ $\mathbb{T}$, a ≤ b, potom $\mu_\Delta([a,b)) = b - a$; $\mu_\Delta((a,b)) = b - \sigma(a)$ a ak a, b ∈ $\mathbb{T}$ – {max $\mathbb{T}$}, potom $\mu_\Delta((a,b]) = \sigma(b) - \sigma(a)$; $\mu_\Delta([a,b]) = \sigma(b) - a$.

3. Pravdepodobnosť na časových škálach

Definícia 7.: Nech $\Omega_{\mathbf{T}}$ je množina všetkých možností, $E \subset \Omega_{\mathbf{T}}$, potom $P_\Delta(A)$ sa nazýva Δ - pravdepodobnosť na A.

$$P_\Delta(A) = \frac{\mu_\Delta(A)}{\mu_\Delta(\Omega_T)} .$$

Tvrdenie3.: Ak $A \subset \Omega_{\mathbf{T}}$, ak $0 \le \mu_\Delta(A) \le \mu_\Delta(\Omega_{\mathbf{T}})$, potom dostávame $0 \le P_\Delta(A) \le 1$, $P_\Delta(\Omega_{\mathbf{T}}) = 1$. Nech $A_1, A_2, \ldots$ sú spočitateľné, dizjunktné podmnožiny $\Omega_{\mathbf{T}}$, potom

$$P_\Delta\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} P_\Delta(A_n).$$

Tvrdenie4.: Ak $A, B \subset \Omega_{\mathbf{T}}$, $P_\Delta(A) \le P_\Delta(B)$, ak $A \subset B$.

Definícia 8.: Náhodná premenná $X_{\mathbf{T}}$ je reálna funkcia definovaná na $\mathcal{F}$.

Príklad 1.:

- Časová škála $\mathbb{T}$ = {0, 1/4 , 2/4 , 3/4, 1, 5/4 , ... } a A = {0, 1, 3, 4} potom $\mu_\Delta(A) = 1$.
- Nech $\mathbb{T}$ = {0, 1, 2, 3, 4, ... } a A = {1, 3, 5, 6,7}, potom $\mu_\Delta(A) = 5$.
- Nech ($\Omega_{\mathbf{T}}$ = {1, 2, 3, 4, 5 ... , 120} a A = {1, 2, 3, 4, 8, 9}, potom $P_\Delta(A) = \mu_\Delta(A) / \mu_\Delta(\Omega_{\mathbf{T}}) = 6 / 120$.

4. Binomické a Poissonovo rozdelenie pravdepodobnosti

V tejto časti uvažujme časovú škálu $\mathbb{T} = h\mathbb{N}$ h>0. V pokuse označme ho pokus_A, sledovaný jav A nastane s pravdepodobnosťou p_h. Sledujeme m - krát takýto pokus, pričom pokusy musia byť štatisticky nezávislé. Potom pravdepodobnosť, že sledovaný jav A nastane práve x- krát vyjadruje binomické rozdelenie pravdepodobnosti.

Napríklad, výrobok nespĺňa určité požadované parametre (je zlý, nekvalitný), s pravdepodobnosťou p_h a spĺňa dané parametre s pravdepodobnosťou $1-p_h$, vyberieme m výrobkov zaujíma nás pravdepodobnosť, že z nich je x zlých. Avšak na rozdiel od klasického binomického rozdelenia pravdepodobnosti, každý pokus pozostáva opäťz h-krát opakovaného pokusu. Pokusy sú štatisticky nezávislé, pričom sledovaný jav B nastane s pravdepodobnosťou p, najmenej k krát. Celkovo nastátie javu B sledujeme n = m*h krát. Napríklad každý výrobok sa skladá z h kusov. Výrobok považujeme za nekvalitný ak aspoň k kusov je chybných. Jav B je sledovaný kus je zlý, nastane s pravdepodobnosťou p a sledovaný kus je dobrý s pravdepodobnosťou q=1-p.

Pravdepodobnosť p_h vyjadruje vzťah (najmenej k kusov z h kusov je zlých)

$$p_h = \sum_{i=k}^{h} \binom{h}{i} p^i q^{h-i} \qquad (1)$$

potom pravdepodobnosť nastátia x-krát javu A v n pokusoch možno vyjadriť:

$$P(x) = P(X_{\mathbf{T}} = x_t) = \binom{m}{x} p_h^{\,x} q_h^{\,m-x} \qquad (2)$$

kde $m = \frac{n}{\mu(x)} = \frac{n}{h}$, $q_h = 1-p_h$, $\mu(x) = x + h - x = h$.

Príklad 2.: Vianočné ozdoby sú balené po troch kusoch do škatule (hviezdička, zvonček, srdiečko). Výrobok považujeme za nekvalitný ak aspoň jedna ozdoba nespĺňa požadované parametre, je poškodená. Pravdepodobnosť poškodenia ľubovoľnej ozdoby je 0,2. Hľadáme pravdepodobnosť, že v zásielke 9 výrobkov bude najviac jeden výrobok poškodený.

Riešenie.: V našom prípade máme dokopy n=9*6 ozdôb, balených po h=6 kusoch, čiže m=9. Dosadením do vzťahov (1) a (2) dostávame:

$$P_A = \sum_{i=0}^{1} \binom{9}{i} p_h^i (1 - p_h)^{9-i}, \text{ kde } p_h = \sum_{i=2}^{6} \binom{6}{i} p^i q^{6-i}.$$

Na výpočet pravdepodobností p_h a P_A použijeme software EXCEL, konkrétne funkciu BINOMDIST. Najskôr vypočítame pravdepodobnosť p_h=0,34464.

Tabuľka 1: Výpočet pravdepodobnosti p_h funkciou BINOMDIST v systéme EXCEL.

2	0,24576	BINOMDIST(A1;6;0,2;0)
3	0,08192	BINOMDIST(A2;6;0,2;0)
4	0,01536	BINOMDIST(A3;6;0,2;0)
5	0,001536	BINOMDIST(A4;6;0,2;0)
6	6,4E-05	BINOMDIST(A5;6;0,2;0)
	0,34464	SUM(A1:A5).

Teraz vypočítame hľadanú pravdepodobnosť P_A=0,18987 (z deviatich výrobkov bude najviac jeden poškodený, nekvalitný).

Tabuľka 2: Výpočet pravdepodobnosti P_A funkciou BINOMDIST v systéme EXCEL.

0	0,022300745	BINOMDIST(A7;9;0,34464;0)
1	0,167573044	BINOMDIST(A8;9;0,34464;0)
	0,189873789	SUM(A7:A8).

Binomické rozdelenie pre n $\to \infty$ sa približuje k Poissonovmu rozdeleniu s parametrom $\lambda = \frac{n}{h} p_h = m.p_h$. Aproximovaná formula pre binomické rozdelenie má tvar:

$$P(x) = P(X_{\mathbf{T}} = x_t) = e^{-\lambda} \frac{\lambda^{\frac{t}{\mu(t)}}}{\mu(t)(\frac{t}{\mu(t)})!} = e^{-\lambda} \frac{\lambda^{\frac{t}{h}}}{h(\frac{t}{h})!}. \quad (3)$$

Príklad 2. možno riešiť aj pomocou Poissonovho rozdelenia nasledovne:

$\lambda = m * p_h = 9*0,18987 = 3,1$. Po dosadení do vzťahu (3) dostávame : $P_A = \sum_{i=0}^{1} e^{-9*ph} \frac{9 * p_h{}^i}{i!}$.

Na výpočet pravdepodobností p_h a P_A použijeme software EXCEL, konkrétne funkciu POISSON.

Tabuľka 3: Výpočet pravdepodobnosti P_A funkciou POISSON v systéme EXCEL.

0	0,04497	POISSON(A10;0,189873789;0)
1	0,139486	POISSON(A11;0,189873789;0)
	0,184456	SUM(A10:A11)

5. Záver

Hlavným cieľom práce bolo poukázať na možnosti aplikácie matematickej analýzy na časových škálach v štatistickom skúmaní úrovne kvality výrobného procesu, ktorá predurčuje ekonomickú silu firmy. Príklad uvedený v práci je ukážkou možnosti využitia nového pohľadu na teóriu pravdepodobnosti. Konkrétne rozširuje definíciu binomického a Poissonvho rozdelenia na všeobecnejšiu na časovej škále $\mathbb{T} = h\mathbb{N}$. V prípade že položíme h=1 dostávame klasické definície týchto pravdepodobnostných rozdelení. Pravdepodobnostné Poissonové rozdelenie zadefinované na $\mathbb{T} = h\mathbb{N}$ má uplatnenie hlavne v poisťovníctve, kde centrálna poisťovňa má pobočky. Predpokladáme h>0, ďalej je to celé číslo, ktoré znamená počet uzavretých zmlúv. V každej z pobočiek je uzavretých h poistných zmlúv. S pravdepodobnosťou p_h je poisťovňa zaviazaná vyplatiť poistku najviac k poistencom. Zaujíma nás aká je pravdepodobnosť, že napríklad x pobočiek z m pobočiek muselo zaplatiť poistku. Hlavným cieľom článku bolo poukázať ako je možné teóriu, časových škál použiť na pravdepodobnostné modelovanie rôznych ekonomických javov a procesov.

Literatúra

BOHNER, M. – PETERSON, A. 2004. Advances in Dynamic Equations on Time Scales. In: *Birkhauser Boston.*

CABADA, A.– VIVERO, D. R. 2006. Expression of Lebesgue Delta Integral on Timescales as a usual Lebesgue integral application to the calculus of delta antiderivative. In: *Mathematical and Computer Modelling*, roč. 43, č.1-2, s. 194 – 207.

HILGER, S. 1990. Analysis on measure chains a unified approach to continuous and discrete calculus. In: *Results Math*. roč. 18, s. 18-56.

MATTHEWS, T. 2011. Probability theory on time scales and application to finance and inequalities. In: *Missouri university of science and technology*.

GUSEINOV, G. 2003. Integration on time scales. In: *J. Math. Anal. Appl*, roč. 285, s. 107–127.

GUSEINOV, G. – KAYMAKCALAN, B. 2002. Basics of Riemann delta and nabla integration on time scales. In: *J. Diff. Equ. Appl.*, roč. 8, 11, s. 1001–1017.

RZEZUCHOWSKI, T. 2005. *A Note on Measure on Time Scales, Demonstratio Mathematica*, roč.V:38, N:1, s. 79-84.

UFUKTEPE, U – DENIZ, A. 2009. Lebesgue-Stieltjes Measure on Time Scales. In: *Turk J. Math*, roč. 32, s. 1-8.

Adresa autora :

Eva Brestovanská, Mgr., PhD
FM UK , KEF, Odbojárov 10
820 05 Bratislava 25
Eva.Brestovanska@fm.uniba.sk

Vypracovanie štatistickej charakteristiky súboru vinárskych podnikov v SR v roku 2010 v Exceli
Statistical characteristics of wine companies in Slovakia in 2010 in Excel

Lucia Coskun

Abstract: Statistical characteristics of wine companies in Slovak republik in 2010 were studied. Correlation analysis and regression modeling in excel were used. The results have proved that the turnover is profit, property, added value and assets depending. Using regression modeling the turnover dependance on personnel costs and assets has been defined.
Abstrakt: V súbore vinárskych podnikov v SR v roku 2010 sme študovali závislosti hodnôt ukazovateľov pomocou korelačnej analýzy a regresného modelovania. Výsledky dokazujú, že najväčšie závislosti obratu sa dosahujú od zisku, vlastného imania, pridanej hodnoty a majetku. Regresným modelovaním sme určili vhodnú regresnú závislosť obratu (tržieb) od osobných nákladov a majetku.

Key words: wine companies, correlation analysis, regression modeling, turnover.

Kľúčové slová: vinárske podniky, korelačná analýza, regresné modelovanie, tržby.

JEL Classification: C00

1. Úvod

V minulosti bola konzumácia vína spojená s požívaním „vzácneho produktu“, v súčasnosti patrí medzi tovary každodennej spotreby s rastúcou tendenciou. Od trhu orientovaného predovšetkým na mužov, vzrástol záujem kúpy tohto produktu aj u žien. Víno našlo uplatnenie nie len ako nápoj, ale aj ako súčasť prípravy pokrmov pri pečení a varení (Quinton, 2003).

Tržby sú pre podniky rozhodujúcim faktorom prežitia na trhu. Tvoria ich tržby z predaja tovaru a tržby z predaja vlastných výrobkov a služieb. Tržby sú ekonomickým cieľom výrobného procesu, základom, na ktorom sa pohybuje činnosť firiem, slúžia na ekonomické pokrytie nákladov – spotreby vstupných zdrojov (materiál, pracovná sila, majetok, ostatné zdroje), sú príjmovou čiastkou bilancie výsledku hospodárenia (Chajdiak, 2013).

2. Opis súboru

Použité údaje sú zo súboru firiem, ktoré vyrábajú víno v SR v roku 2010. Údaje sú v eurách a získali sme ich z výkazu súvaha a výkazu ziskov a strát. Počet podnikov v súbore je 56. V01 predstavuje tržby z predaja tovaru, V05 tržby z predaja vlastných výrobkov a súčet V01 a V05 celkový obrat podniku. V tabuľke 1 je zobrazené ukazovatele minumim, maximux, priemer, median, smerodajná odchylka (Std) a suma.

Tab. 1: Štatistický rozbor ukazovateľov tržieb

	V01	V05	V01+V05
n	56	56	56
Min	0	0	0
Max	22119983	4783167	26903150
Priemer	870774	136625	1007399
Median	14813	173	36228
Std	3319833	653191	3911867
Sum	48763340	7650992	56414332

3. Miery vzájomnej závislosti hodnôt ukazovateľov vyjadrené v korelačnej matici

Pri výpočte korelačnej matice pomocou excelu sme využili údaje z tabuľky 2: tržby z predaja tovaru (V01), tržby z predaja vlastných výrobkov (V05), súčet V01 a V05 tvorí obrat podniku, Z pred – zisk pred zdanením (V59), zisk po – zisk po zdanení (V61) Maj-majetok (S001) a Vl.im.-vlastné imanie a záväzky S067 z výkazu súvaha a výkazu ziskov a strát. PH - pridanú hodnotu sme vypočítali pomocou excelu ako V01 +V04-V02-V08. V04 predstavuje výrobu, V02 náklady vynaložené na obstaranie predaného tovaru a V08 výrobnú spotrebu. Obrat/ON je v exceli vypočítaný podiel obratu a osobných nákladov (V12), kedže osobné náklady podnikov č. 39-56 sú nulové, výsledky týchto podnikov nedávajú význam. PH/ON je podiel pridanej hodnoty a osobných nákladov, taktiež výpočet pre podniky 39-52 nedáva význam. Q/Maj je podiel obratu a majetku, Z pr/Q je podiel zisku pred zdanením a obratu a Zisk/VI je podiel zisku po zdanení a vlastného imania.

Tab. 2: Súbor vinárskych podnikov

ID	V05	V01	Obrat	PH	Z pred	zisk po	Maj	Vl. im.	Obrat/ ON	PH/ ON	Q/Maj	Z pr/ Q	Zisk/ VI
1	2,2E+07	5E+06	3E+07	-9566333	2817023	2256219	29869932	20992366	14,703505	4,73861	0,9007	0,1047	0,10748
2	1,2E+07	426252	1E+07	-852502	2228698	1799648	13680128	9274039	8,9332657	2,95753	0,8871	0,1837	0,19405
3	1956953	0	2E+06	3	33608	12537	6274899	3701565	2,3939141	0,36224	0,3119	0,0172	0,00339
4	1934214	783689	3E+06	-1567374	-161496	-161496	2938251	-2622628	6,6287892	1,62276	0,925	-0,0594	0,06158
5	476851	528	477379	-1051	-1249615	-1237887	9466700	8865184	1,6481805	-1,9506	0,0504	-2,6177	-0,13963
6	1649700	0	2E+06	6	25779	20950	2174996	262724	6,8275785	2,02417	0,7585	0,0156	0,07974
7	2576512	0	3E+06	7	257265	208968	6983945	744020	12,795932	3,67502	0,3689	0,0999	0,28086
8	1309948	7392	1E+06	-14776	133957	124380	2706932	1067467	7,8403761	3,11628	0,4867	0,1017	0,11652
9	306944	1824	308768	-3639	-253694	-253694	3355185	90141	2,0329065	0,05053	0,092	-0,8216	-2,81441
10	256846	0	256846	10	-118532	-118532	2326464	1038756	1,8243588	0,69339	0,1104	-0,4615	-0,11411
11	126227	0	126227	11	-193636	-193636	573077	536647	0,9309188	-0,2177	0,2203	-1,534	-0,36083
12	215763	0	215763	12	-354197	-357514	2329274	-1773259	2,1066285	-2,1104	0,0926	-1,6416	0,20161
13	367407	0	367407	13	730	528	460692	283899	3,7086719	1,60668	0,7975	0,002	0,00186
14	1343094	7712	1E+06	-15410	174460	139795	4838742	429402	14,061814	7,00649	0,2792	0,1292	0,32556
15	921079	356717	1E+06	-713419	67571	54203	1945757	863844	13,872651	1,80973	0,6567	0,0529	0,06275
16	462073	595	462668	-1174	39784	32224	355329	78390	6,970096	1,79658	1,3021	0,086	0,41107
17	100196	28715	128911	-57413	-91045	-92116	378414	259260	1,9874963	0,0471	0,3407	-0,7063	-0,3553
18	85649	2152	87801	-4286	18504	18504	369716	209658	1,9095892	-0,3509	0,2375	0,2107	0,08826
19	172129	0	172129	19	62776	47960	330860	232584	8,2658951	4,51786	0,5202	0,3647	0,20621
20	4396	19654	24050	-39288	-167846	-167846	983056	-194353	1,2490911	-4,8511	0,0245	-6,979	0,86361
21	25362	5386	30748	-10751	8699	8699	43314	41356	2,0054787	1,62438	0,7099	0,2829	0,21034
22	232787	0	232787	22	-9153	-9153	453870	60399	16,023334	0,35841	0,5129	-0,0393	-0,15154
23	11	822233	822244	-1644443	-58499	-58114	705203	28348	59,039563	-1,1918	1,166	-0,0711	-2,05002
24	8834	121370	130204	-242716	-432	-432	8612	5376	10,934162	1,01369	15,119	-0,0033	-0,08036
25	0	110289	110289	-220553	7039	5482	24980	16875	9,9413196	1,55832	4,4151	0,0638	0,32486
26	41461	246	41707	-466	11809	9565	101419	76954	4,1495374	2,68083	0,4112	0,2831	0,1243
27	16587	0	16587	27	245	245	8093	4549	1,7498681	1,03935	2,0495	0,0148	0,05386
28	26391	58123	84514	-116218	16881	16880	64106	-11917	11,563005	3,66931	1,3183	0,1997	-1,41646
29	9043	3455	12498	-6881	-18436	-18436	66233	-163106	1,9617015	-0,8432	0,1887	-1,4751	0,11303
30	46190	15098	61288	-30166	-47226	-47226	561703	357952	12,228252	-0,6562	0,1091	-0,7706	-0,13193

31	0	640	640	-1249	858	564	1213333	818445	0,1447309	8,34012	0,0005	1,3406	0,00069
32	0	18437	18437	-36842	-14362	-14362	28065	-29059	6,4419986	-2,0692	0,6569	-0,779	0,49424
33	13039	9756	22795	-19479	-10995	-10995	85067	46624	8,10633	-0,3563	0,268	-0,4823	-0,23582
34	0	8000	8000	-15966	-19491	-19491	256742	229441	3,087611	0	0,0312	-2,4364	-0,08495
35	0	3289	3289	-6543	-260112	-260112	21955	-277520	1,5156682	-118,57	0,1498	-79,085	0,93727
36	3016	0	3016	36	4242	4242	14992	12645	1,7433526	-3,4572	0,2012	1,4065	0,33547
37	1829	0	1829	37	-1101	-1101	1591	1045	1,9252632	0,14	1,1496	-0,602	-1,05359
38	365	0	365	38	-1084	-1084	4179	3916	0,6612319	-0,9638	0,0873	-2,9699	-0,27681
39	0	0	0	39	-61	-61	13955	13495	#DIV/0!	#DIV/0!	0		-0,00452
40	0	0	0	40	-1310	-1310	4218	3690	#DIV/0!	#DIV/0!	0		-0,35501
41	0	0	0	41	-379	-379	4621	4621	#DIV/0!	#DIV/0!	0		-0,08202
42	0	100	100	-158	-199	-200	8102	8102	#DIV/0!	#DIV/0!	0,0123	-1,99	-0,02469
43	0	22497	22497	-44951	2952	2372	26801	7372	#DIV/0!	#DIV/0!	0,8394	0,1312	0,32176
44	0	0	0	44	-346	-346	48463	-92	#DIV/0!	#DIV/0!	0		3,76087
45	0	0	0	45	0	0	6938	6938	#DIV/0!	#DIV/0!	0		0
46	2980	0	2980	46	22	15	182332	7233	#DIV/0!	#DIV/0!	0,0163	0,0074	0,00207
47	157944	6297	164241	-12547	68134	55189	127208	64963	#DIV/0!	#DIV/0!	1,2911	0,4148	0,84955
48	43	0	43	48	-480	-480	7353	7353	#DIV/0!	#DIV/0!	0,0058	-11,163	-0,06528
49	0	0	0	49	-6903	-6908	80107	80107	#DIV/0!	#DIV/0!	0		-0,08623
50	5000	0	5000	50	300	243	293307	126142	#DIV/0!	#DIV/0!	0,017	0,06	0,00193
51	400	0	400	51	-32420	-32420	159330	-20514	#DIV/0!	#DIV/0!	0,0025	-81,05	1,58038
52	0	0	0	52	-243	-243	5357	4757	#DIV/0!	#DIV/0!	0		-0,05108
53	81	27379	27460	-54705	2657	2499	22225	5050	#DIV/0!	#DIV/0!	1,2355	0,0968	0,49485
54	47106	0	47106	54	-34638	-34638	171577	-32898	#DIV/0!	#DIV/0!	0,2745	-0,7353	1,05289
55	29657	0	29657	55	9487	9487	18889	18489	#DIV/0!	#DIV/0!	1,5701	0,3199	0,51312
56	0	0	0	56	-20513	-20513	790	-39931	#DIV/0!	#DIV/0!	0		0,51371

Medzi premennými udanými v tabuľke sme hľadali závislosť. Podľa Chajdiaka, na analýzu vzájomnej závislosti viacerých premenných možno použiť korelačnú analýzu. Mieru závislosti udáva koeficient korelácie. Nadobúda odnoty -1 až +1.Čím sa koeficient korelácie viac približuje k hodnote +1 alebo -1 tým je miera závislosti dvoch premenných väčšia. Čím viac sa blíži k nule, tým je závislosť premenných menšia (Chajdiak, 2013).

Premenné podnikov 1-38 sme analyzovali korelačnou analýzou pomocou excelu. Výsledky sú zobrazené v tabuľke č.3. Z výsledkov je zrejmé, že závislosti od zisku sú podstatne variabilné. Závislosť obratu (tržieb) od zisku podnikov vykazuje vysokú mieru závislosti. Zisk pred zdanením dosahuje hodnoty 0,899 a po zdanení 0,872. Závislosť obratu od vlastného imania dosahuje hodnotu až 0,902. Najväčšie štatisticky významné závislosti preukazuje závislosť obratu od pridanej hodnoty, dosahuje hodnotu -0,920 a od majetku, dosahuje hodnotu až 0,945.

Tab. 3: Korelačná analýza súboru vinárskych podnikov pre podniky č.1-38.

	ID	*V05*	*V01*	*Obrat*	*PH*	*Z pred*	*zisk po*	*Maj*	*Vl. im.*	*Obrat/ ON*	*PH/ ON*	*Q/Maj*	*Z pr/ Q*	*Zisk/ Vl*
ID	1,000													
V05	-0,473	1,000												
V01	-0,323	0,887	1,000											
Obrat	-0,456	0,997	0,920	1,000										
PH	0,323	-0,887	-1,000	-0,920	1,000									
Z pred	-0,287	0,912	0,744	0,899	-0,744	1,000								
zisk po	-0,256	0,885	0,720	0,872	-0,720	0,998	1,000							
Maj	-0,589	0,947	0,842	0,945	-0,842	0,759	0,718	1,000						
Vl. im.	-0,445	0,903	0,810	0,902	-0,810	0,722	0,679	0,945	1,000					
Obrat/ ON	-0,056	0,136	0,285	0,163	-0,285	0,173	0,179	0,114	0,082	1,000				
PH/ ON	-0,261	0,096	0,072	0,093	-0,072	0,140	0,150	0,115	0,091	0,106	1,000			
Q/Maj	0,076	-0,028	0,020	-0,021	-0,020	0,027	0,036	-0,080	-0,046	0,132	0,063	1,000		
Z pr/ Q	-0,238	0,068	0,053	0,067	-0,053	0,116	0,126	0,083	0,066	0,118	0,994	0,072	1,000	
Zisk/ Vl	0,003	0,098	-0,016	0,081	0,016	0,113	0,114	0,057	0,060	-0,383	-0,238	0,000	-0,255	1,000

4. Regresné modelovanie závislosti obratu (tržieb) od osobných nákladov a majetku

Tržby sú funkciou osobných nákladov a majetku. Závislosť tržieb od osobných nákladov a majetku sme vyšetrovali regresným modelovaním závislosti. Výpočet sme robili pomocou excelu. Ako vstup Y sme zadali oblasť s hodnotami tržieb – závisle premennej, do vstupu X sme zadali hodnoty osobných nákladov a majetku – nezávisle premenných. Prvý výstup (tab. 4) zobrazuje hodnoty všetkých 56 podnikov, druhý výstup (tab. 5) zobrazuje hodnoty len tých 38 podnikov, ktorých osobné náklady boli nenulové.

Tab. 4: Regresné modelovanie tržieb od osobných nákladov a majetku v súbore vinárskych podnikov pre podniky č.1-56.

Regression Statistics	
Multiple R	0,952419713
R Square	0,90710331
Adjusted R Square	0,903597775
Standard Error	1214584,06
Observations	56

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	7,63462E+14	3,817E+14	258,76312	4,48752E-28
Residual	53	7,81864E+13	1,475E+12		
Total	55	8,41649E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	-399603,1	173731,0602	-2,300125	0,0254075	-748063,47	-51142,72971	-748063,47	-51142,72971
Maj	0,522928669	0,099803542	5,2395803	2,841E-06	0,322748119	0,723109219	0,322748119	0,723109219
ON	4,319953314	1,437896279	3,0043567	0,0040578	1,435898677	7,20400795	1,435898677	7,20400795

Tab. 5: Regresné modelovanie tržieb od osobných nákladov a majetku v súbore vinárskych podnikov pre podniky č.1-38.

Regression Statistics	
Multiple R	0,953556179
R Square	0,909269387
Adjusted R Square	0,90408478
Standard Error	1454042,785
Observations	38

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	7,41585E+14	3,708E+14	175,37867	5,76358E-19
Residual	35	7,39984E+13	2,114E+12		
Total	37	8,15584E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	-620650,41	261126,1529	-2,376822	0,0230622	-1150764,68	-90536,14024	-1150764,68	-90536,14024
Maj	0,531611176	0,119678201	4,4420051	8,541E-05	0,288651511	0,77457084	0,288651511	0,77457084
ON	4,426795997	1,723850727	2,5679694	0,0146571	0,927192992	7,926399003	0,927192992	7,926399003

V oboch prípadoch sa P-value významne líši od nuly. Oba modely môžme považovať za štatisticky vhodné. Avšak hodnota štatistickej vhodnosti significance F je oveľa vyššia pri modelovaní s 38 podnikmi bez nulových hodnôt osobných nákladov (5,76358E-19). Pri modelovaní so všetkými 56 podnikmi dosahovala hodnota significance F len (4,48752E-28).

Rovnica závislosti obratu (tržieb) od osobných nákladov a majetku bude mať tvar:

Obrat = -620650,41 + 4,427 ON + 0,532 Maj (1)

5. Záver

Výsledky korelačnej analýzy dokazujú, že najväčšie závislosti obratu sa dosahujú od zisku, vlastného imania, pridanej hodnoty a majetku. Závislosť obratu od pridanej hodnoty, dosahuje hodnotu -0,920 a od majetku, dosahuje hodnotu až 0,945.

Regresným modelovaním sme určili regresnú závislosť obratu (tržieb) od osobných nákladov a majetku a dokázali, že štatisticky vhodnejšie je modelovanie s podnikmi, ktorých osobné náklady sú nenulové.

Na porovnanie odporúčame prácu Chajdiaka (2013), kde bola závislosť popísana Cobb-Douglasovou funciou.

Literatúra

CHAJDIAK, J. 2013. *Štatistika jednoducho v Exceli*. Bratislava Statis.

CHAJDIAK, J. 2013. Koeficient konkurencieschopnosti firmy v súbore firiem. *Forum Statistikum Slovacum.*č.2, s. 56-59.

QUINTON S., HARRIDGE-MARCH S. 2003, Strategic interactive marketing of wine – a case of evolution. *Marketing Intelligence & Planning*, Vol. 21 Iss: 6, pp.357 – 362, ISSN: 0263-4503.

Adresa autora:
Lucia Coskun, Ing.
Ústav manažmentu, externá doktorandka
Vazovova 5, 812 43 Bratislava
luciacoskun@stuba.sk

Aplikácia neurónových sietí vo finančnej analýze podniku s využitím SPSS
The application of neural networks in financial analysis using SPSS

Stanislav Cút

Abstract: The paper focuses on the evaluation of the classification ability of Multilayer Perceptron neural network, using the statistical software SPSS Statistics 20, for the selected dataset containing information on tax subjects, requiring, for the tax period January 2013, refund of excess of VAT.

Abstrakt: Príspevok sa zameriava na zhodnotenie klasifikačnej schopnosti viacvrstvovej neurónovej siete typu Multilayer perceptron s využitím štatistického softvéru SPSS Statistics 20 na vybranej dátovej množine obsahujúcej informácie o daňových subjektoch, požadujúcich za zdaňovacie obdobie január 2013 vrátenie nadmerného odpočtu DPH.

Key words: neural network, Multilayer Perceptron, classification analysis, activation function, synaptics weights, input layer, hidden layer, output layer.

Kľúčové slová: neurónová sieť, viacvrstvová neurónová sieť, klasifikačná analýza, aktivačná funkcia, synaptické váhy, vstupná vrstva, skrytá vrstva, výstupná vrstva.

JEL classification: C38, C45, H26

1. Úvod

Nestabilita súčasného ekonomického prostredia a z toho vyplývajúca strata kontinuálnej prolongácie informácií obsiahnutých v historických časových radoch do súčasnosti má za následok irelevantnosť hodnotenia finančnej situácie podnikov pomocou štandardných lineárnych nástrojov používaných v ekonomickej prognostike a vychádzajúcich z dostatočného množstva historických údajov.

Potreba hľadať nové možnosti, ktoré by boli dostatočne robustné, rýchle, výpočtovo jednoduché a schopné predikovať vývoj a prípadné problémy podnikov v súčasnom dynamicky sa meniacom ekonomickom prostredí, je preto stále aktuálnejšou.

Jedným z efektívnych riešení sa hlavne vďaka schopnosti aproximovať zložité a nelineárne väzby v údajoch a neexistencii obmedzujúcich predpokladov pre ich relevantnú aplikáciu, javí použitie neurónových sietí patriacich do kategórie datamingových metód.

Cieľom predkladaného príspevku je vysvetlenie podstaty vybraného typu neurónovej siete a verifikácia jej klasifikačných schopností na vybranej dátovej množine s využitím štatistického softvéru SPSS Statistics 20.

2. Podstata neurónových sietí

Neurónovú sieť je možné definovať ako biologicky inšpirovaný analytický nástroj z oblasti umelej inteligencie, ktorý je prostredníctvom napodobňovania kognitívnej schopnosti neurónov v ľudskom mozgu schopný modelovať priebeh výrazne nelineárnych vzťahov. (Haykin, 1998)

Štruktúra najjednoduchšej neurónovej siete typu perceptron (pozri obr. 1) pozostáva z konečného počtu skalárnych vstupov p_i, multiplikovaných hodnotou príslušnej synaptickej skalárnej váhy w_i *(weight)*, ktorá určuje mieru citlivosti, akou príslušný vstup *wp* pôsobí prostredníctvom aktivačnej resp. transferovej funkcie neurónu *f* (*activation function*, *transfer function*) na skalárny výstup z neurónu *n*, pre ktorý platí:

$$n = f(a) \tag{1}$$

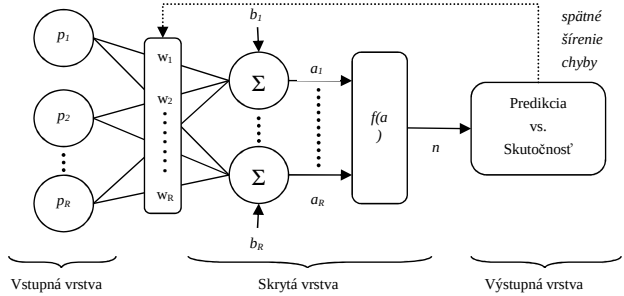
Vplyv na správanie neurónu môže mať tiež skalárna veličina označovaná ako prah neurónu *b* (*bias*), ktorý predstavuje argument transferovej funkcie určujúci mieru aktivity príslušného

neurónu. Hodnota prahu sa pridáva k váženému vstupu *wp* prostredníctvom sumujúceho uzlu, čím zabezpečuje zvýšenie vstupu do aktivačnej funkcie *a*, pre ktorý platí:

$$a = w_1 p_1 + w_2 p_2 + ... + w_R p_R + b = \sum_{i=1}^{R} w_i p_i + b \qquad (2)$$

V prípade, že hodnota vstupného signálu neurónu je nižšia ako stanovená prahová hodnota, výsledkom je pasívny stav neurónu. K zmene výstupného signálu z neurónu, resp. jeho aktivizácii tak dochádza až po prekročení stanovenej prahovej hodnoty.

Informačný tok generujúci predpovede plynie od vstupnej vrstvy (*input layer*), cez jednu alebo viaceré skryté vrstvy neurónov (*hidden layer*) až do výstupnej vrstvy, ktorá produkuje výstup zo siete (*output layer*). Platí, že výstupy z každej medzivrstvy sú vstupmi do nasledujúcej vrstvy.



Obr. 1: Architektúra jednovrstvovej neurónovej siete. [1]

Vďaka svojmu výkonu, jednoduchosti a flexibilite použitia aj v prípade, že vstupné dáta obsahujú zašumené informácie, vysokokorelované vysvetľujúce premenné, prípadne sa jedná o neúplné časové rady, nachádzajú neurónové siete široké uplatnenie v oblasti prognózovania spotrebiteľského dopytu, reakcií domácností na zaslané ponuky formou direct mail marketingu, pri hodnotení rizikovosti žiadateľa o poskytnutie bankové úveru, pri detekcii podvodných transakcií v databázach poistných udalostí a pod. (Wallace, 2008)

Z veľkého množstva existujúcich typov neurónových sietí sa historicky pri aplikácii v ekonomických vedných disciplínach ako najvhodnejšie osvedčili neurónové siete typu Linear Network (*LN*), Generalized Regression Neural Network (*GRNN*), Probabilistic Neural Network (*PNN*), Multilayer percepton (*MLP*), Radial Basis Function (*RBF*) a Kohen Network (*SOFM*). (Klieštik, 2011)

3. Aplikácia neurónovej siete typu MLP v programe SPSS Statistics20

V nasledujúcej kapitole sa zameriame na praktickú aplikáciu neurónovej siete typu Multilayer Perceptron (*MLP*) dostupnej v analytickom balíku štatistického softvéru SPSS Statistics 20, s cieľom priblížiť prebiehajúce algoritmy budovania uvedeného typu neurónovej siete a zhodnotiť jej klasifikačnú schopnosť na vybranej dátovej množine.

Typ architektúry neurónovej siete bol zvolený na základe dostupných možností analytického balíka štatistického softvéru SPSS Statistics 20 a na základe dosahovania vyššej

[1] Vlastné spracovanie na základe OLSON, D. L. – DELEN, D. – MENG, Y. 2012. Comparative analysis of data mining methods for bankruptcy prediction. In *Elsevier*, s. 466.

klasifikačnej schopnosti v porovnaní s druhou dostupnou architektúrou neurónovej siete typu Radial Basis Function (*RBF*).

Východiskovú dátovú maticu, očistenú o chýbajúce údaje, predstavuje spolu 17 ukazovateľov vyhodnocovaných pre každý z 754 daňových subjektov, prevažne malých a stredných podnikov, ktoré za zdaňovacie obdobie január 2013 podali daňové priznanie pre daň z pridanej hodnoty (DPH), na ktorom v dôsledku prebytku nároku na odpočítanie dane nad hodnotu vlastnej daňovej povinnosti, požadovali v sledovanom období vrátenie nadmerného odpočtu DPH.

S využitím neurónovej siete *MLP* sa pokúsime vytvoriť model, resp. klasifikačné pravidlo, ktoré bude schopné čo najpresnejšie klasifikovať vybraný daňový subjekt do skupiny rizikových platcov DPH.

Rizikové daňové subjekty použité pre učenie predstavuje presne 377 subjektov, ktoré za zdaňovacie obdobie január 2013 podali daňové priznania DPH, na ktorom požadovali vrátenie nadmerného odpočtu DPH a u ktorých bola v horizonte troch mesiacov od uplynutia sledovaného zdaňovacieho obdobia začatá daňová kontrola.

Označenie a stručnú charakteristiku uvažovanej vysvetľovanej premennej a jednotlivých vysvetľujúcich premenných zahrnutých do modelu prezentuje tabuľka 1.

Tab. 1.: Charakteristika premenných vstupujúcich do modelu

Označenie premennej	Popis
SEKCIA	Sekcia hospodárstva (podľa SK NACE)
POC_DOV_SLED	Počet dôvodov sledovania subjetku
OBCH_KOM	Sledovaná obchodná komodita
NK_NO_DPH	Neumožnená kontrola vrátenia NO DPH
NK_NO_DPH_PS	Neumožnená kontrola vrátenia NO DPH – prepojené subjekty
PS_NAL_NAD_10	Prepojený subjekt s nálezom nad 10 000 EUR
PS_ZRUS_REG_UM	Prepojený subjekt so zrušenou registráciou DPH z úradnej moci
RIZ_DP	Miera rizikovosti daňového priznania
PH_DPH	Pridaná hodnota DPH (obrat – náklady)
OBRAT_DPH	Obrat DPH
VELKOST_DS	Veľkosť daňového subjektu
DM1	Rizikovosť - Dataminingový model 1
DM3	Rizikovosť - Dataminingový model 2
POC_ZAM	Počet zamestnancov
KOEF_RIZ	Pomer celkovej dane k odpočtu dane v rozsahu koeficientu 0, 95 – 1,05
POZ_NO_DPH	Požadovaný NO DPH
ZAC_KONT	Začaté kontroly za zdaňovacie obdobie január 2013

Zdroj: vlastné spracovanie

Vysvetľovaná premenná *ZAC_KONT* nadobúda charakter kategoriálnej premennej s hodnotou 1 pre rizikové daňové subjekty a hodnotou 0 pre nerizikové daňové subjekty. Vysvetľujúce premenné (prediktory) sú spojité alebo kategoriálne premenné zakódované pomocou umelých premenných.

Keďže vybrané spojité vysvetľujúce premenné sa vyznačujú značne rozdielnymi mierkami, pre zaručenie porovnateľnosti sme ich hodnoty normovali.

Pôvodnú dátovú množinu obsahujúcu presne 754 pozorovaní sme v záujme zreálnenia klasifikačnej schopnosti modelu následne pomocou generátorom náhodných čísiel vytvorenej deliacej premennej rozdelili na trénovaciu, testovaciu a hodnotiacu podmnožinu. Pozorovania s kladnou hodnotou deliacej premennej boli zaradené do trénovacej podmnožiny používanej na budovanie neurónovej siete, pozorovania s nulovou hodnotou do testovacej podmnožiny používanej na účely hodnotenia chyby v procese učenia sa, a napokon pozorovania so zápornou hodnotou deliacej premennej do hodnotiacej podmnožiny používanej na posúdenie

výslednej klasifikačnej schopnosti neurónovej siete. Absolútne a relatívne početnosti uvedených podmnožín prezentuje tabuľka 2.

Tab. 2.: Absolútne a relatívne početnosti podmnožín vstupujúcich do modelu

		N	Percent
Podmnožina	Trénovacia	405	53,70%
	Testovacia	99	13,10%
	Hodnotiaca	250	33,20%
Zahrnuté pozorovania		754	100%
Vylúčené pozorovania		0	
Pozorovania celkom		754	

Zdroj: vlastné spracovanie v programe SPSS Statistics 20

Za účelom nastavenia vhodnej architektúry siete, bolo konštruovaných spolu 10 modelov líšiacich sa počtom skrytých vrstiev, typom aktivačnej funkcie používanej pre skryté vrstvy a následne aj výstupnú vrstvu neurónovej siete, typom učenia sa, a napokon aj optimalizačným algoritmom pre odhad synaptických váh.

Proces učenia sa neurónovej siete typu *MLP* využíva algoritmus spätného šírenia chýb (*back propagation of error*), ktorý pracuje v dvoch krokoch. V prvom kroku (doprednom) sú vstupy šírené cez aktivačné funkcie neurónov s cieľom stanoviť chybu *E*, v druhom kroku (spätnom) sú následne prepočítané hodnoty synaptických váh w_{ij}.[2]

Cieľom učiacej fázy neurónovej siete bolo dosiahnuť také nastavenie synaptických váh w_{ij}, pri ktorom je odchýlka (chyba) *E* medzi skutočnými a cieľovými výstupmi na hodnotiacej podmnožine minimálna a zároveň dochádza k optimalizácii klasifikačnej schopnosti modelu.

Ako môžeme vidieť z tabuľky 3, v prípade výsledného modelu bola na základe uvedených kritérií zvolená dvojvrstvová architektúra siete s počtom neurónov 8 v prvej skrytej vrstve a s počtom neurónov 6 v druhej skrytej vrstve. Ako aktivačná funkcia spájajúca vážené sumy neurónov v jednotlivých skrytých vrstvách $f(a)_{HIDDEN}$, bola defaultne zvolená hyperbolická tangentná funkcia nadobúdajúca hodnoty z intervalu (-1, 1). Rovnaký typ bol zvolený aj v prípade aktivačnej funkcie výstupnej vrstvy $f(a)_{OUT}$. Počet neurónov v jednotlivých vrstvách neurónovej siete bol stanovený automatickým algoritmom a na výpočet chyby *E* bola použitá suma štvorcov chýb.

Tab. 3.: Architektúra skrytých vrstiev a výstupnej vrstvy neurónovej siete

Skrytá vrstva	Počet skrytých vrstiev	2
	Počet neurónov v 1. skrytej vrstve	8
	Počet neurónov v 2. skrytej vrstve	6
	Typ aktivačnej funkcie $f(a)_{HIDDEN}$	Hyperbolická tangentná
Výstupná vrstva	Závislá premenná	*ZAC_KONT*
	Počet neurónov	2
	Typ aktivačnej funkcie $f(a)_{OUT}$	Hyperbolická tangentná
	Chybový algoritmus	Suma štvorcov chýb

Zdroj: vlastné spracovanie v programe SPSS Statistics 20

Spôsob učenia neurónovej siete bol, vzhľadom na výskyt vysokej korelácie medzi vybranými dvojicami prediktorov a uvažovaný stredne veľký rozsah dátovej množiny, definovaný pomocou skupinovej metódy (*mini-batch*), ktorej podstata spočíva v rozdelení

[2] Pre lepšie pochopenie pozri PASW Modeler 13. 2009. *Algorithms Guide*. Chicago : SPSS Inc., s. 267 - 269.

pozorovaní tréningovej podmnožiny do približne rovnako veľkých podskupín a následnej aktualizácii váh w_{ij} zvlášť po načítaní každej z vytvorených podskupín. Na odhad synaptických váh bola napokon použitá gradientná metóda (*gradient descent*). [3]

Informácie o priebehu učenia sa, resp. trénovania finálnej podoby neurónovej siete prezentuje tabuľka 4. Ako pozitívny signál vypovedajúci o kvalite neurónovej siete môžeme hodnotiť fakt, že percento nesprávne klasifikovaných pozorovaní nadobúda pre jednotlivé sledované podmnožiny zhruba podobné hodnoty a tiež fakt, že došlo k pomerne výraznému poklesu sumy štvorcov chýb. Z tabuľky 4 tiež vidíme, že pravidlom na základe ktorého bola ukončená fáza učenia sa neurónovej siete bolo dosiahnutie maximálneho počtu iterácií bez poklesu chyby, v našom prípade obmedzený defaultnou hodnotou 1.

Tab. 4.: Sumarizácia učiacej fázy modelu

Trénovacia podmnožina	Suma štvorcov chýb	40,322
	Percento nesprávnych predikcií	12,30%
	Kritérium zastavenia algoritmu	počet iterácií bez poklesu chyby [a]
Testovacia podmnožina	Suma štvorcov chýb	11,458
	Percento nesprávnych predikcií	15,20%
Hodnotiaca podmnožina	Percento nesprávnych predikcií	13,20%

Závislá premenná: ZAC_KONT

a. Výpočet chyby založený na testovacej podmnožine

Zdroj: vlastné spracovanie v programe SPSS Statistics 20

Na základe prehľadu o štruktúre správnosti resp. chybovosti klasifikácie modelu uvedeného v tabuľke 5 môžeme konštatovať, že z celkového počtu 125 rizikových daňových subjektov dokázala neurónová sieť *MLP* správne zaradiť 110 subjektov, čo v relatívnom vyjadrení predstavuje 88%. V rámci kategórie nerizikových klientov dokázala sieť správne klasifikovať 85,6% subjektov. Celková klasifikačná schopnosť modelu na hodnotiacej podmnožine sa tak pohybuje na úrovni 86,8%.

Tab. 5.: Klasifikačná schopnosť neurónovej siete typu MLP

Podmnožina	Skutočnosť	Predikcia		
		0	1	Správne zaradené
Trénovacia	0	176	30	85,4%
	1	20	179	89,9%
	Celkom %	48,4%	51,6%	87,7%
Testovacia	0	39	7	84,8%
	1	8	45	84,9%
	Celkom %	47,5%	52,5%	84,8%
Hodnotiaca	0	107	18	85,6%
	1	15	110	88,0%
	Celkom %	48,8%	51,2%	86,8%

Závislá premenná: *ZAC_KONT*

Zdroj: vlastné spracovanie v programe SPSS Statistics 20

Fakt, že v rámci testovacej a hodnotiacej podmnožiny nedošlo v porovnaní s hodnotami správne zaradených subjektov v trénovacej podmnožine k výraznejšiemu poklesu

[3] Pre lepšie pochopenie pozri SPSS Statistics 19. 2010. *Neural Networks*. Chicago : SPSS Inc., s. 10 – 15.

klasifikačnej schopnosti naznačuje, že sieť počas fázy učenia nebola pretrénovaná (*overfitting*). K pretrénovaniu dochádza vtedy, ak si sieť zapamätá šumy, špecifické pre jednotlivé typy konkrétnych tréningových dát, bez schopnosti generalizovať nové (lepšie) dáta.

4. Záver

Cieľom predkladaného príspevku bolo zhodnotiť klasifikačnú schopnosť viacvrstvovej neurónovej siete typu Multilayer Perceptron na dátovej množine obsahujúcej údaje o rizikovosti daňových subjektov požadujúcich za zdaňovacie obdobie január 2013 vrátenie nadmerného odpočtu DPH, s využitím štatistického softvéru IBM SPSS Statistics 20.

Z celkového počtu 250 daňových subjektov v hodnotiacej podmnožine, dokázala výsledná dvojvrstvová neurónová sieť pri zvolených nastaveniach správne klasifikovať 86,8% daňových subjektov. Senzitivita neurónovej siete resp. schopnosť správne klasifikovať rizikové daňové subjekty, u ktorých bola v horizonte 3 mesiacov od podania daňového priznania začatá daňová kontrola sa pohybuje na úrovni 88%. Špecificita neurónovej siete sa pohybuje na úrovni 85,6%.

Ako nevýhodu aplikovania uvedenej dataminingovej metódy môžeme hodnotiť fakt, že kvalita klasifikácie, resp. predikcie produkovanej modelom je medzi iným výrazne determinovaná zvolenou architektúrou siete, spôsobom učenia sa a stanovenia typu aktivačnej funkcie a tiež zvoleným optimalizačným algoritmom pre odhad synaptických váh.

Na základe uskutočnenej analýzy sa však s ohľadom na existujúce nedostatky javí použitie neurónových sietí ako aproximátora zložitých nelineárnych vzťahov vhodnou alternatívou niektorých štandardne využívaných štatistických a dataminingových metód.

Literatúra

HAYKIN, S. 1998. *Neural Networks: A Comprehensive Foundation*. Singapore: Pearson Education Pte. Ltd., 2001. 823 s. ISBN 81-7808-300-0.

IBM SPSS Statistics 19. 2010. *Neural Networks*. Chicago : SPSS Inc..

KARDOŠ, J. 2009. Aplikácia finančnej analýzy využitím neurónových sietí [online]. Bratislava : Ekonomická Univerzita, Národohospodárska fakulta, 2009 [cit. 2013-10-25]., 12 s. Dostupné na internete: http://www.derivat.sk/files/konferencia_forfin2009/Kardos.pdf.

KLIEŠTIK, T. 2011. Neurónové siete a umelá inteligencia v riadení podnikov. In *Modelování, simulace a optimalizace podnikových procesů – Sborník z konference konané dne 29. března 2011*. Praha : ČSOP, 2011. 519 s. ISBN 978-80-260-0023-5. s. 174-182.

OLSON, D. L. – DELEN, D. – MENG, Y. 2012. Comparative analysis of data mining methods for bankruptcy prediction. In *Elsevier*, 2012, s. 454 - 473.

PASW Modeler 13. 2009. *Algorithms Guide*. Chicago : SPSS Inc., s. 288 - 305.

WALLACE, M. P. 2008. Neural Networks and their application to finance. In *Business Intelligence Journal*, 2008, č. 7, s. 67 – 76.

Adresa autora:
Stanislav Cút, Ing.
EF UMB v Banskej Bystrici, KKMaIS
Tajovského 10, 975 10 Banská Bystrica
stanislav.cut@umb.sk

Odhady intervalově cenzorovaných dat v R
Estimates of interval censored data using R

Adam Čabla

Abstract: The article deals with the topic of interval censored data and their estimates both nonparametric and parametric using free software for statistical computing and graphics – R. It introduces reader into the issue of interval censored data and then shows how to work with this kind of data using only the available packages, i.e. no programming is needed. In the article the Turnbull estimate (NPMLE), the log-rank test, the Cox proportional hazards model and the accelerated failure time model are presented. The example of modelling time of unemployment from the Labour Force Surveys is used for the demonstration of these procedures.

Abstrakt: Článek se zabývá problematikou intervalově cenzorovaných dat a jejich parametrických i neparametrických odhadů dostupných při užití výpočetního prostředí R. Seznamuje čtenáře s tématem intervalově cenzorovaných dat a následně ukazuje jak s nimi pracovat jen za použití dostupných balíčků, pro tyto odhady tedy není potřeba umět programovat. V článku jsou představeny Turnbullův odhad, log-rank test, Coxův model proporcionálních rizik a model s urychleným selháním. Pro demonstraci těchto technik je užit příklad modelování doby nezaměstnanosti z Výběrového šetření pracovních sil.

Key words: R, Turnbull, log-rank test, Cox PH model, AFT model, LFS

Kľúčové slová: R, Turnbull, log-rank test, Coxův model, model s urychleným selháním, VŠPS

JEL classification: C13, C14, C24, J64

1. Úvod

Cenzorovaná data jsou typem pozorování, pro které není známa jeho přesná hodnota, ale pouze interval, ve kterém se tato hodnota nachází. Podle toho, jestli je tento interval otevřený zprava, zleva nebo uzavřený pak rozlišujeme cenzorování zprava, zleva nebo intervalové.

Nejpropracovanější metodologie je pro směs dat s přesnými hodnotami a hodnotami cenzorovanými zprava, které jsou typické v mnohých zdravotních follow-up studiích, kdy některé jednotky přeruší svou účast a tak je u nich známo pouze to, že sledovaná událost u nich nenastala do okamžiku přerušení.

Intervalové cenzorování většinou nastává v případech, kdy je jednotka sledována jednou za daný interval a je zaznamenán stav určité proměnné. To je typické především pro longitudinální studie, přičemž v článku jsou užita data z výběrového šetření pracovních sil, kde je každá jednotka dotazována jednou za čtvrtletí a to pět po sobě jdoucích čtvrtletí (ČSÚ, 2013).

Hlavním tématem článku jsou intervalově cenzorovaná data, pro která nejsou ve většině programů naprogramovány ani základní odhady, motivací pro článek je tedy poskytnout zájemcům možnost vytvořit alespoň základní odhady a modely.

Při řešení problematiky cenzorovaných dat se většinou klade důraz na odhad funkce přežití S(x), která je definována jako pravděpodobnost, že náhodná veličina nabyde hodnoty větší než konkrétní hodnota, je to tedy doplněk distribuční funkce.

$$S(x) = P(X > x) = 1 - F(x) \tag{1}$$

2. Data a řešený příklad

Pro řešený příklad byla vybrána část problému řešeného v příspěvku na konferenci AMSE (Čabla, 2013). Cílem je zjistit, zda se liší doba nezaměstnanosti v době před současnou ekonomickou krizí (období Q4/2007 – Q4/2008) a během této krize (Q1/2010 – Q1/2011) a tento očekávaný rozdíl kvantifikovat.

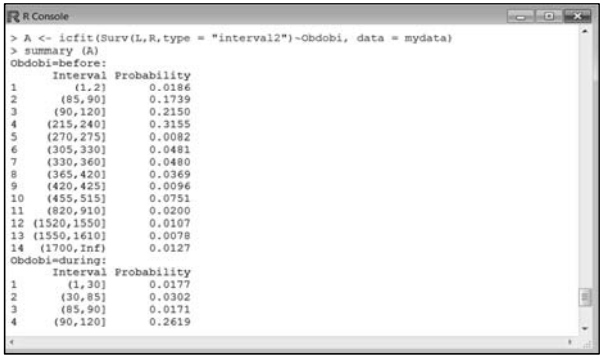
V každém čtvrtletí je u dotazovaných zjišťováno, zda jsou zaměstnaní, nezaměstnaní či ekonomicky neaktivní a jak dlouho jsou nezaměstnaní. Jelikož je každý dotazován v průběhu času několikrát, lze najít v datech lidi, kteří byli nezaměstnaní, následně zaměstnaní a na základě dalších proměnných určit dobu nezaměstnanosti jako intervalově cenzorované pozorování. Takovýchto pozorování bylo celkem nalezeno 352 před krizí a 803 během krize.

3. Turnbullův odhad

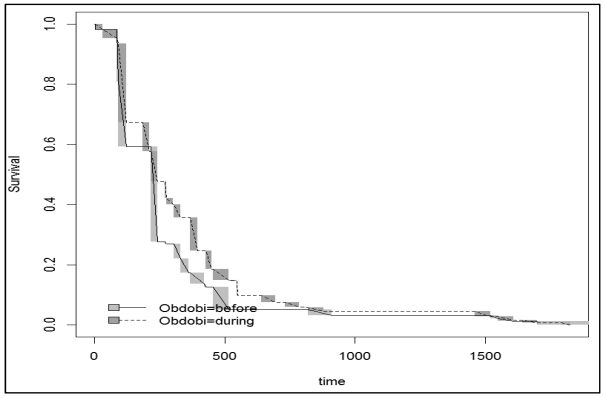
Turnbullův odhad je neparametrickým odhadem funkce přežití vytvořený primárně pro intervalově cenzorovaná data (Turnbull, 1976). Jedná se o vcelku jednoduchou iterativní proceduru popsanou např. v (Klein, a další, 1997). Výsledkem této procedury je odhad pravděpodobností, že náhodná veličina nabyde hodnoty z určitých intervalů. Mimo tyto intervaly je pravděpodobnost nulová a uvnitř těchto intervalů je její rozložení neznámé, přičemž grafické vyjádření v programu R pracuje s rovnoměrným rozdělením této pravděpodobnosti uvnitř intervalu.

Pro provedení tohoto odhadu v R je potřeba mít stažený a aktivní balíček „interval“, jedná se konkrétně o příkaz icfit, který je podrobně popsán v (Fay, 2013). V článku pro AMSE (Čabla, 2013) byl užit příkaz A <- icfit (Surv(L, R, type = “interval2”)~Obdobi, data = mydata), kde „A“ je název nově uloženého objektu typu „icfit“, se kterým budeme dále pracovat, „L“ je vektor spodních hranic intervalů pozorování, „R“ je vektor horních hranic intervalů pozorování, „type“ udává typ cenzorování při tvorbě objektu typu Surv, což je v našem případě „interval2“, „Obdobi“ je proměnná, podle které se člení odhady na dvě období – před a během krize, „data“ pak odkazuje na datovou matici načtenou v prostředí R, zde jsou tedy použita data načtená pod názvem „mydata“, což jsou spojená pozorování lidí, kteří našli zaměstnání před i během krize.

Samotný odhad je potom možné vyvolat příkazem summary (A), program R zobrazí intervaly a pravděpodobnosti výskytu v těchto intervalech, jak je zobrazeno v obrázku 1. Grafické znázornění odhadu funkce přežití poskytne funkce plot (A), což je uvedeno v obrázku 2. Za povšimnutí stojí šedě vyobrazená pole, která vyjadřují nejistotu ohledně rozdělení pravděpodobnosti uvnitř intervalu.

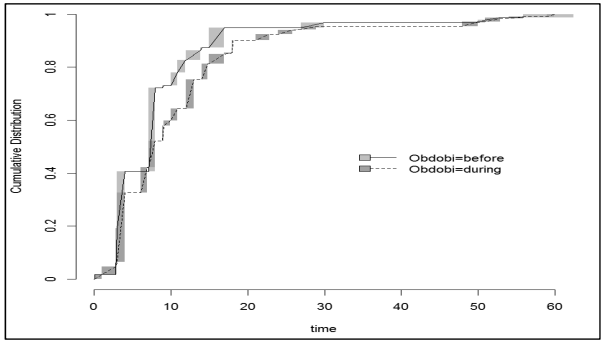


Obr. 5: Výstup příkazu „summary“ pro „icfit“ objekt (A)



Obr. 6: Výstup příkazu „plot“ pro „icfit“ objekt (A)

Funkce „plot“ pro „icfit“ objekt je opět podrobně popsána v (Fay, 2013), za zmínku stojí hlavně možnost zobrazení odhadu nejen jako funkce přežití, ale i jako distribuční funkce pomocí argumentu “dtype = „cdf““. Druhým užitečným argumentem je „xscale“, pomocí kterého jsou přepsány hodnoty na ose x, tedy je možné např. místo zobrazení funkce přežití ve dnech (viz obr. 2) ji zobrazit v měsících pomocí argumentu „xscale = 1/30,4375“. Tedy při použití příkazu plot (A, dtype = „cdf“, xscale = 1/30.4375, XLEG = 1000, YLEG = 0.5) vznikne graf z obrázku 3, který zobrazuje distribuční funkci doby nezaměstnanosti v měsících. Argumenty „XLEG“ a „YLEG“ změnily souřadnice legendy.



Obr. 7: Alternativní výstup příkazu „plot“ pro „icfit“ objekt (A)

Průměr, medián a další hodnoty je možno následně ručně dopočítat s užitím předpokladu o rovnoměrném rozdělení pravděpodobnosti uvnitř odhadnutých intervalů.

Druhou možností jak získat neparametrický odhad funkce přežit je užít funkci „survfit“ z balíčku „survival“ pro objekt typu „Surv“ stejně jako v případě funkce „icfit“: F <- survfit (Surv (L,R,type = "interval2")~Obdobi, data = mydata). Výhodou této funkce je automatický výpočet intervalu spolehlivosti, druhou výhodou je pak možnost automaticky vypočítat medián včetně intervalu spolehlivosti a ohraničený průměr. Obojí lze zobrazit funkcí „print“ pro typ objektu Surv, argument „rmean“ udává horní omezení intervalů pro výpočet průměru: např. print (F, rmean = 365) vypíše průměr pro hodnoty omezené shora na maximálně jeden rok. V obrázku 4 jsou demonstrovány odhady průměru pro různá omezení, hodnota argumentu „individual“ počítá průměr jako plochu pod křivkou od nuly do nejvyšší

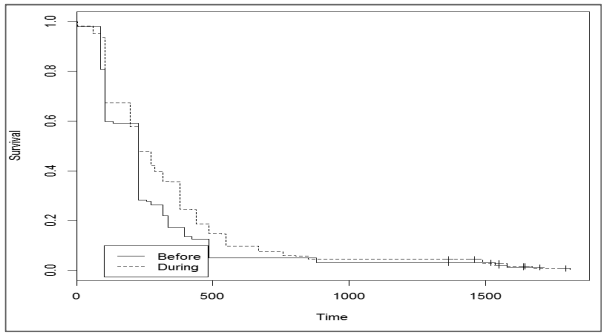
pozorované hodnoty, hodnota „common“ bere jako horní omezení nejvyšší nalezenou hodnotu.

```
> F <- survfit (Surv (L,R,type = "interval2")~Obdobi, data = mydata)
> print (F, rmean = 180)
Call: survfit(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)

                records n.max n.start events *rmean *se(rmean) median 0.95LCL 0.95UCL
Obdobi=before     352   352     352    345    145       2.40    220     220     220
Obdobi=during     803   803     803    786    152       1.51    228     228     272
    * restricted mean with upper limit =  180
> print (F, rmean = 365)
Call: survfit(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)

                records n.max n.start events *rmean *se(rmean) median 0.95LCL 0.95UCL
Obdobi=before     352   352     352    345    206       5.67    228     228     228
Obdobi=during     803   803     803    786    238       4.07    228     228     272
    * restricted mean with upper limit =  365
> print (F, rmean = "common")
Call: survfit(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)

                records n.max n.start events *rmean *se(rmean) median 0.95LCL 0.95UCL
Obdobi=before     352   352     352    345    267       15.5    228     228     228
Obdobi=during     803   803     803    786    333       11.6    228     228     272
    * restricted mean with upper limit =  1755
> print (F, rmean = "individual")
Call: survfit(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)

                records n.max n.start events *rmean *se(rmean) median 0.95LCL 0.95UCL
Obdobi=before     352   352     352    345    266       15.3    228     228     228
Obdobi=during     803   803     803    786    333       11.7    228     228     272
   * restricted mean with variable upper limit
```

Obr. 8: Výstup příkazu „print“ pro „survfit“ objekt (F)

Podle popisu má funkce vypsat Turnbullův odhad, ale výpis odpovídá odhadu Kaplan-Meiera. Grafické znázornění odhadu je v obrázku 5, funkce „plot“ nemá pro tento typ objektu žádné přednastavení, popisy a rozlišení funkcí musí být vloženo ručně, legenda chybí a je dodána dalším příkazem: plot (F, xlab = „Time“, ylab = „Survival“, lty = c (1, 2)) a legend (x = 100, y = 0.1, legend = c („Before“, „During“), lty = c (1, 2)).

Balíček „survival“ je podrobně popsán v (Therneau, 2013).



Obr. 9: Výstup příkazu „plot“ pro „survfit“ objekt (F)

4. Log-rank test

Kromě neparametrického odhadu funkce přežití umožňuje balíček „interval“ ještě neparametrický test závislosti doby přežití na hodnotě kategoriální proměnné s H_0: funkce přežití mají stejné rozdělení. Tento test se obecně nazývá k-výběrový log-rank test a k jeho provedení slouží funkce „ictest“, jež se provádí na stejném typu objektu (Surv) jako funkce „icfit“.

Kromě samotného výsledku testu vyjádřeného hodnotou testového kritéria a p-hodnotou udává log-rank test taktéž statistiku, jež ukazuje směr závislosti. U k-výběrového testu je obdobná rozdílu pozorované a očekávané hodnoty, tedy její pozitivní hodnota naznačuje

menší hodnoty vysvětlované proměnné pro danou skupinu, tj. kratší dobu nezaměstnanosti. U trendového testu pak pozitivní hodnota říká, že vyšší hodnoty vysvětlující proměnné vedou k menším hodnotám vysvětlované proměnné.

Pomocí argumentu „score“ lze zvolit jedno z pěti skóre, přednastaveno je Sunovo. Pomocí argumentu „exact“ lze zvolit asymptotickou nebo přesnou formu, přičemž u přesné formy je p-hodnota počítána buď přesně permutací, nebo za pomoci zvoleného množství Monte-Carlo simulací.

V případě, že je vysvětlující proměnná kvantitativní provede funkce „ictest“ test o nulovém trendu. Podrobný popis jednotlivých možností tohoto testu je k nalezení v článku autorů balíčku „interval“ (Fay, a další, 2010).

Pro provedení testu v základním nastavení stačí zapsat funkci ictest (Surv(L, R, type = “interval2”)~Obdobi, data = mydata), výsledky jsou zobrazeny v obrázku 6. Obrázek 7 pak zachycuje výsledky funkce ictest (Surv(L, R, type = “interval2“)~ISCED, exact = TRUE, scores = “wmw“, data = mydata), ve které je nastaveno jiné použité skóre a přesná forma testu o trendu pro vysvětlující proměnnou vzdělání dle ISCED, která je ordinální na škále 1 – 5.

Ve výsledku pak zjišťujeme, že před krizí byla doba nezaměstnanosti obecně kratší a že čím vyšší vzdělání, tím kratší doba nezaměstnanosti.

```
> ictest (Surv(L, R, type = "interval2")~Obdobi, data = mydata)

        Asymptotic Logrank two-sample test (permutation form), Sun's scores

data:  Surv(L, R, type = "interval2") by Obdobi
Z = 4.3801, p-value = 1.186e-05
alternative hypothesis: survival distributions not equal

                 n Score Statistic*
Obdobi=before 352          57.66694
Obdobi=during 803         -57.66694
* like Obs-Exp, positive implies earlier failures than expected
```

Obr. 10: Výstup příkazu „ictest“ pro dvouvýběrový test

```
> ictest (Surv(L, R, type = "interval2")~ISCED, exact = TRUE, scores = "wmw", data = mydata)

        Exact Wilcoxon trend test(permutation form)

data:  Surv(L, R, type = "interval2") by ISCED
p-value = 0.002
alternative hypothesis: survival distributions not equal

        n Score Statistic*
[1,] 1155          75.0624
* postive so larger covariate values give earlier failures than expected
p-value estimated from 999 Monte Carlo replications
99 percent confidence interval on p-value:
 0.00000000 0.01057916
```

Obr. 11: Alternativní výstup příkazu „ictest“ pro test trendu

Funkce „surfdiff“ v balíčku „survival“ taktéž počítá log-rank testy, ale neumožňuje pracovat s intervalově cenzorovanými daty.

5. Model s urychleným selháním

Modely s urychleným selháním jsou typem regresního modelu, ve kterém se předpokládá, že vysvětlující proměnná mění rychlost plynutí času. Obecně lze zapsat tento model následovně (Klein, a další, 1997):

$$S(t|\mathbf{x}) = S_0 * \left[\left(t / \alpha(\mathbf{x}) \right)^{\delta} \right]. \tag{2}$$

Ze (2) je vidět, že hodnota parametru $\alpha(\mathbf{x})$ určuje zrychlení či zpomalení času – pokud platí $\alpha(\mathbf{x}) > 1$, potom vektor vysvětlujících proměnný čas zpomaluje a obráceně. Základní funkce přežití S_0 může mít libovolné zvolené pravděpodobnostní rozdělení. Odhadnutý parametr regresní funkce se v případě užití Weibullova rozdělení musí upravit, protože se odhaduje

logaritmická linearizující transformace (rozdělení extrémní hodnoty), takže skutečný odhad pak získáme přirozeným exponentem odhadnutého parametru.

Pro odhad parametrického regresního modelu s urychleným selháním lze užít funkce „survreg“ z balíčku „survival“, která vychází z objektu typu „Surv“ jako ostatní užívané funkce: D <- survreg (Surv (L,R,type = "interval2")~Obdobi, data = mydata). Funkce „survreg“ umožňuje použít libovolné rozdělení pomocí argumentu „dist“, přičemž předdefinovany jsou rozdělení Weibullovo, exponenciální, normální, lognormální, logistické a loglogistické. Weibullovo rozdělení je přednastaveno. Další rozdělení si může uživatel definovat pomocí funkce „survreg.distributions“, základní popis je v (Therneau, 2013). V obrázku 8 je ukázka dvou odhadů AFT modelu s rozděleními Weibullovým (D) a lognormálním (D2).

```
> D <- survreg (Surv (L,R,type = "interval2")~Obdobi, data = mydata)
> D
Call:
survreg(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)

Coefficients:
 (Intercept) Obdobiduring
   5.6146923    0.2493376

Scale= 0.8680484

Loglik(model)= -1681   Loglik(intercept only)= -1689.9
        Chisq= 17.73 on 1 degrees of freedom, p= 2.5e-05
n= 1155
> D2 <- survreg (Surv (L,R,type = "interval2")~Obdobi, data = mydata, dist = "lognormal")
> D2
Call:
survreg(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata,
    dist = "lognormal")

Coefficients:
 (Intercept) Obdobiduring
   5.1926363    0.2987729

Scale= 0.8434165

Loglik(model)= -1628.6   Loglik(intercept only)= -1641.4
        Chisq= 25.65 on 1 degrees of freedom, p= 4.1e-07
n= 1155
```

Obr. 12: Výstup z AFT modelu

Z výstupu je patrné, že v obou případech je p-hodnota testu o modelu velmi malá a tedy můžeme říci, že doba nezaměstnanosti se významně liší v období před krizí a během krize. Logaritmus věrohodnosti (Loglik) je vyšší v případě lognormálního rozdělení, použijeme tedy tento odhad pro vykreslení grafu v obrázku 9.

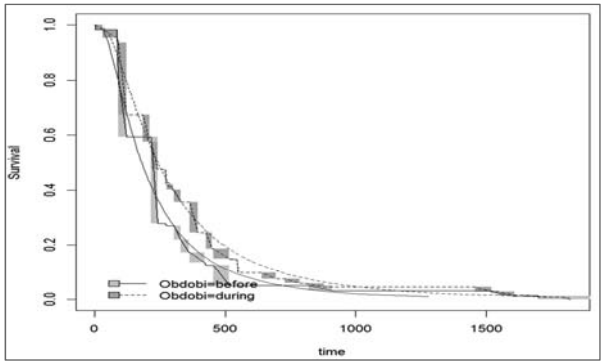
Pro zakreslení regresního odhadu do stejného grafu spolu s neparametrickým odhadem nejdříve vytvoříme nový datový list s odhady pomocí funkce „predict“: D2before <- predict (D2, newdata = list (Obdobi = „before“), type = „quantile“, p = seq (0.01, 0.99, by = 0.01)) a analogicky pro období během krize. Argument „p“ udává pro jaké kvantily je odhad vytvořen, zde jsme vytvořili odhad kvantilů pro p = 0.01 až 0.99 vždy po jedné setině.

Samotná funkce „predict“ s typem „quantile“ nám pak vypíše odhady kvantilů dané funkce, což může být užitečné pro odhad těchto kvantilů.

Odhady zaneseme do vytvořeného grafu neparametrického odhadu funkce přežití (plot (A)) pomocí funkce lines (D2before, seq (0.99, 0.01, by = 0.01), col = red, lty = 1) a analogicky pro období během krize s argumentem lty = 2, který zajistí vytvoření přerušované čáry analogicky k základnímu nastavení funkce „plot“ pro neparametrický odhad.

```
> D2before <- predict (D2, newdata = list (Obdobi = "before"), type = "quantile", p = seq (0.01, 0.99, by = 0.01))
> D2during <- predict (D2, newdata = list (Obdobi = "during"), type = "quantile", p = seq (0.01, 0.99, by = 0.01))
> plot (A)
> lines (D2before, seq (0.99, 0.01, by = -0.01), col = "red", lty = 1)
> lines (D2during, seq (0.99, 0.01, by = -0.01), col = "red", lty = 2)
```

Obr. 13: Příkazy k tvorbě obrázku 10



Obr. 14: Neparametrický odhad spolu s AFT modelem

Z výstupů lze potvrdit, že podle parametrické regrese s lognormálním rozdělením se doba nezaměstnanosti statisticky významně zvýšila a to exp(0,2987729) = 1,35 násobně, tedy o 35 %. Z grafu je vidět, že regresní funkce pro obě období vcelku dobře odpovídá neparametrickému odhadu.

6. Coxův model proporcionálních rizik

Coxův model vychází z odhadu rizikové funkce (Klein, a další, 1997)

$$h(t|\mathbf{x}) = h_0(t)r(\mathbf{x}), \tag{3}$$

kde platí základní vztah

$$H(t) = \int_0^t h(u)du = -\ln\left[S(t)\right] \tag{4}$$

a obvykle

$$r(\mathbf{x}) = \exp(\beta'\mathbf{x}) \tag{5}$$

H(t)se nazývá kumulativní riziková funkce a je výstupem odhadu pomocí funkce „intcox“. Riziková funkce h_0 se v kontextu Coxova modelu nazývá základní riziková funkce a může být definována parametrickým rozdělením, v tom případě hovoříme o parametrickém modelu, nebo není takto definována a model je pak semiparametrickým s vektorem parametrů **β**. Dále budeme pracovat se semiparametrickými modely, protože jejich implementace je výrazně jednodušší a propracovanější.

Pro odhady Coxova modelu pro intervalová data lze zatím použít pouze balíček „intcox“ se stejnojmennou funkcí, jež užívá iterativní minorant convex algoritmus (Huang, a další, 1993), (Henschel, a další, 2013). Obdobná funkce „coxph“ z balíčku „survival“ neumí pracovat s intervalově cenzorovanými daty. Ačkoliv funkce „intcox“ vytvoří objekt stejného typu jako funkce „coxph“, je její použitelnost v navazujících funkcích (např. survfit) omezená či žádná.

Samotný odhad Coxova modelu je přímočarý s užitím objektu typu Surv : G <- intcox (Surv (L, R, type = „interval2“)~Obdobi, data = mydata). Ve výstupu v obrázku 11 je vidět odhad exp(coef), což je násobek rizikové funkce pro danou proměnnou ve srovnání se základní funkcí hazardu, tedy r(**x**) (viz (5)). Tento koeficient můžeme vyvolat a uložit funkcí Coef <- G$coefficients [“Obdobiduring”]. Dále je vidět, že ve funkci zatím není přítomno testování významnosti těchto koeficientů, což omezuje využitelnost této informace na pouhý doplněk analýzy.

```
> G <- intcox (Surv(L,R,type = "interval2")~Obdobi, data = mydata)
no improvement of likelihood possible, iteration =  1
> G
Call:
intcox(formula = Surv(L, R, type = "interval2") ~ Obdobi, data = mydata)


                coef exp(coef) se(coef)  z  p
Obdobiduring -0.327     0.721       NA NA NA

Likelihood ratio test=NA  on 1 df, p=NA  n= 1155
```

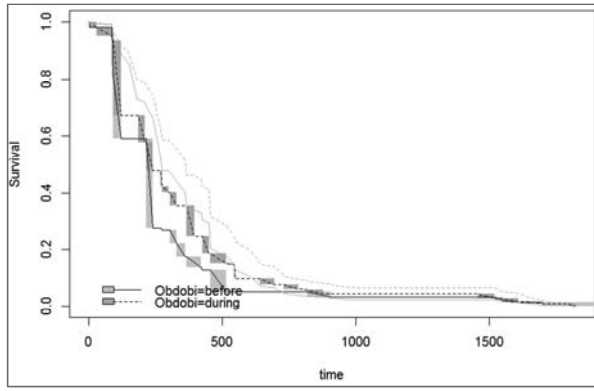
Obr. 15: Odhad Coxova modelu

Odhad základní kumulativní rizikové funkce lze vyvolat funkcí G$lambda0. Pomocí vztahu (4) pak můžeme tento odhad převézt na odhad základní funkce přežití Cox <- exp (-G$lambda0), tedy pro období před krizí. Odhady dalších funkcí přežití pro jednotlivé skupiny pak vyplývají ze vztahů (3) a (4), tedy pro období během krize použijeme Cox2 <- Cox^exp(Coef).

Pro vynesení těchto odhadů do grafu funkce přežití musíme ještě definovat hodnoty na ose x jako hodnoty časových bodů, ve kterých je proveden daný odhad rizikové funkce, tedy x <- G$time.point. Graf potom je funkce plot (x, Cox), přidání odhadu do již vytvořeného grafu pak pomocí funkce lines (x, Cox). Příkazy k tvorbě grafu jsou v obrázku 12, samotný graf pak v obrázku 13.

```
> plot (A)
> Coef <- G$coefficients ["Obdobiduring"]
> Cox <- exp (-G$lambda0)
> Cox2 <- Cox^exp(Coef)
> x <- G$time.point
> lines (x, Cox, lty = 1, col = "green")
> lines (x, Cox2, lty = 2, col = "green")
```

Obr. 16: Tvorba grafu v obrázku 13



Obr. 17: Turbnullův odhad a Coxův model

Coxův model proporcionálního rizika říká, že riziko nalezení práce je v období krize 0,721 násobné oproti období před krizí. Lze tedy říci, že pravděpodobnost nalezení práce podle tohoto modelu je v každém okamžiku o 27,9 % nižší. To vcelku odpovídá zjištění parametrické regrese, ale není potvrzeno statistickým testem významnosti. Z grafu je navíc

vidět, že odhadnuté funkce přežití jsou oproti neparametrickým odhadům výrazně vyšší a zřejmě nadhodnocené.

7. Závěr

V článku byly popsány a na příkladu demonstrovány základní techniky pro intervalově cenzorovaná data, jejich neparametrický i parametrický odhad a testování a zobrazování odhadů funkce přežití v grafech. To vše za pomoci volně dostupných a popsaných balíčků – „survival“, „interval“ a „intcox“. V rámci demonstrativního příkladu jsme si ukázali, že doba nezaměstnanosti se v České republice během krize prodloužila dle parametrického lognormálního modelu o 35 % a dle semiparametrického Coxova modelu a 27 %. Taktéž jsme si otestovali, že vzdělání jako ordinální proměnná má nepřímý vztah k době nezaměstnanosti.

Poděkování

Tento článek vznikl za finanční podpory grantu IG410062 z IGA VŠE.

Literatura

ČABLA, A. 2013. *Unemployment Duration in the Czech Republic Before and During the Crisis (Abstrakt).* [CD-ROM] Gerlachov : Banská Bystrica : Faculty of Economics, 2013.

ČSÚ. 2013. Výběrové šetření pracovních sil (VŠPSú. *CZSO.* [Online] 17. 1 2013. [Citace: 17. 11 2013.] http://www.czso.cz/vykazy/vykazy.nsf/i/vyberove_setreni_pracovnich_sil.

FAY, M. P., SHAW, P. A. 2010. Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R package. *R Project.* [Online] 2010. [Citace: 17. 11 2013.] http://cran.r-project.org/web/packages/interval/vignettes/intervalCensoring.pdf.

FAY, M. P. 2013. Package "interval". *R Project.* [Online] 06. 05 2013. [Citace: 17. 11 2013.] http://cran.r-project.org/web/packages/interval/interval.pdf.

HENSCHEL, V., MANSMANN, U., HEISS, C. 2013. Package "intcox". *R Project.* [Online] 18. 02 2013. [Citace: 17. 11 2013.] http://cran.r-project.org/web/packages/ intcox/intcox.pdf.

HUANG, J. A., WELLNER, J. A. 1993. Regression Models with Interval Censoring. *University of Washington.* [Online] 06. 10 1993. [Citace: 17. 11 2013.] https://www.stat.washington.edu/jaw/JAW-papers/NR/jaw-huang-96ProbThMathStat.pdf.

KLEIN, J. P., MOESCHBERGER, M. L. 1997. *Survival Analysis: Techniques for Censored and Truncated Data.* New Yourk : Springer - Verlag, 1997.

THERNEAU, T. 2013. Package "survival". *R Project.* [Online] 26. 02 2013. [Citace: 17. 11 2013.] http://cran.r-project.org/web/packages/survival/survival.pdf.

TURNBULL, B. W. 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society B38.* 1976**.**

Adresa autora:

Adam Čabla, Ing.
KSTP, Fakulta statistiky a informatiky, VŠE
Nám. W. Churchilla 4, 130 67, Praha 3
adam.cabla@vse.cz

Vývoj střední délky života a pravděpodobné délky života v České republice v letech 1960 - 2011
The development of life expectancy and probable length of life in the Czech Republic from 1960 to 2011

Petra Dotlačilová, Jitka Langhamrová

Abstract: This article will deal with the evolution of mortality in the Czech Republic from 1960 to 2011. For this analysis is the most frequently used life expectancy at exact age *x*. As the other indicator will be used probable length of life. The aim of this article is to analyze the evolution of life expectancy at birth. The other aim is to compare the value of life expectancy at exact age *x* with probable length of life at this age. At the end we would like to analyze the evolution of these characteristics during the whole period.

Abstrakt: Článek se zabývá vývojem úmrtnosti v České republice od roku 1960 do roku 2011. Obvykle se k vyhodnocení vývoje úmrtnosti používá jako souhrnná charakteristika úmrtnosti střední délka života v přesném věku, nejčastěji při narození. Zde využijeme také další souhrnnou charakteristiku úmrtnosti – pravděpodobnou délku života novorozenců. Spolu s vývojem střední délky života také vyhodnotíme vývoj pravděpodobné délky života u novorozenců.

Key words: mortality, life expectancy, probable length of life.

Klíčová slova: úmrtnost, střední délka života, pravděpodobná délka života.

JEL classification: C02, J11

1. Úvod

Při popisu vývoje úmrtnosti se nejčastěji používá ukazatel střední délka života při narození. Není to však jediný ukazatel, který charakterizuje úmrtnostní poměry populace. Kromě tohoto ukazatele lze zkoumat normální nebo pravděpodobnou délku života. V tomto článku bude zkoumán vývoj střední délky života v přesném věku *x* a pravděpodobné délky v témže věku. Dále bude zjišťováno, v jakém věku převýší střední délka života osoby v přesném věku *x* pravděpodobnou délku života osoby přesné *x*-leté. Analýza bude prováděna zvlášť pro muže a pro ženy.

2. Metodika

Jak již bylo uvedeno, k hodnocení vývoje úmrtnosti se nejčastěji používá jako syntetický ukazatel střední délka života osoby v přesném věku *x*. Její hodnota pro novorozence udává, jak dlouho bude v průměru na živu právě narozená osoba, pokud po celou dobu jejího života bude úmrtnost stejná a bude odpovídat úmrtnosti popsané úmrtnostními tabulkami (Fiala, 2005). Střední délka života je tedy syntetickou charakteristikou úmrtnosti v celém věkovém rozmezí. Pro analýzu úmrtnosti osob pouze od určitého věku se používá střední délka života ještě v dalších věcích. Proto také bude v tomto příspěvku definována střední délka života osoby přesně *x*-leté. Její hodnota udává, jak dlouho bude ještě v průměru na živu osoba v přesném věku *x*, pokud se po celou dobu jejího zbývajícího života nezmění úmrtnostní poměry a zůstanou na úrovni úmrtnosti popsané úmrtnostními tabulkami.

Hodnoty střední délky života v přesném věku *x* nalezneme v úmrtnostních tabulkách. Nejčastěji se počítají tzv. průřezové úmrtnostní tabulky, které charakterizují úmrtnost v poměrně krátkém časovém období, zpravidla v jednom roce. Základem pro jejich výpočet jsou specifické míry úmrtnosti v daném roce či období (Fiala, 2005):

$$m_x = \frac{M_x}{\frac{S_{1.1.t,x} + S_{31.12.t,x}}{2}}, \quad (1)$$

kde M_x je počet zemřelých v dokončeném věku x, $S_{1.1.t,x}$ je počet žijících x-letých k počátku roku t a $S_{31.12.t,x}$ je počet žijících x-letých na konci roku t.

Jejich hodnoty jsou však zatížené jak náhodnými, tak (zejména pro vyšší věk) též systematickými chybami, pro nejvyšší hodnoty věku dokonce nemusejí být specifické míry úmrtnosti k dispozici. Proto se provádí vyrovnání těchto hodnot, pro nižší věk zpravidla klouzavými průměry, pro vyšší věk (obvykle nad 60 let) nějakou funkcí. V tomto článku bude pro vyrovnání použita *Gompertzova–Makehamova* funkce (Koschin, 2002, Burcin et al., 2010, Makeham, 1860):

$$\mu_x = a + bc^x, \quad (2)$$

kde μ_x je intenzita úmrtnosti, a, b, c jsou parametry *Gompertzovy–Makehamovy* funkce, x je věk.

Po vyrovnání specifických měr byl proveden výpočet úmrtnostních tabulek podle všeobecně známých vzorců (Fiala, 2005).

Dalším ukazatelem je pravděpodobná délka života x-leté osoby. Hodnota pravděpodobné délky života pro novorozence udává, za jak dlouho se (za předpokladu zachování dané úmrtnosti popsané úmrtnostními tabulkami) výchozí počet živě narozených zmenší na polovinu, tj. za jak dlouho bude z výchozího počtu narozených naživu pouze polovina osob (Koschin, 2002). V souladu s předchozím tvrzením pravděpodobná délka života osoby přesně x-leté pak udává, za jak dlouho se počet původně x-letých osob sníží na polovinu. Pravděpodobná délka života je mediánovou charakteristikou. Hodnotu pravděpodobné délky života novorozence můžeme také interpretovat jako věk, kterého se (při zachování úmrtnosti) dožije novorozenec s pravděpodobností přesně ½ (a s pravděpodobností ½ se jej nedožije), pravděpodobná délka života osoby v přesném věku x je pak doba, za kterou bude osoba v přesném věku x s pravděpodobností ½ ještě naživu (Cipra, 1990).

Výpočet pravděpodobné délky života pro novorozence provádí podle vzorce:

$$\tilde{e}_0 = x_D + \frac{l(x_D) - 0{,}5 \cdot l(0)}{l(x_D) - l(x_D + 1)}, \quad (3)$$

kde x_D je poslední věk, ve kterém je ještě počet dožívajících se vyšší než polovina výchozího souboru živě narozených v tabulkové populaci, $l(x_D)$ je počet dožívajících se přesného věku x_D, $l(x_D+1)$ je počet dožívajících se následujícího věku (tj. věku x_D zvýšeného o jednotku).

Výpočet pravděpodobné délky života u osob přesně x-letých se provádí podle analogického vzorce (Koschin, 2002):

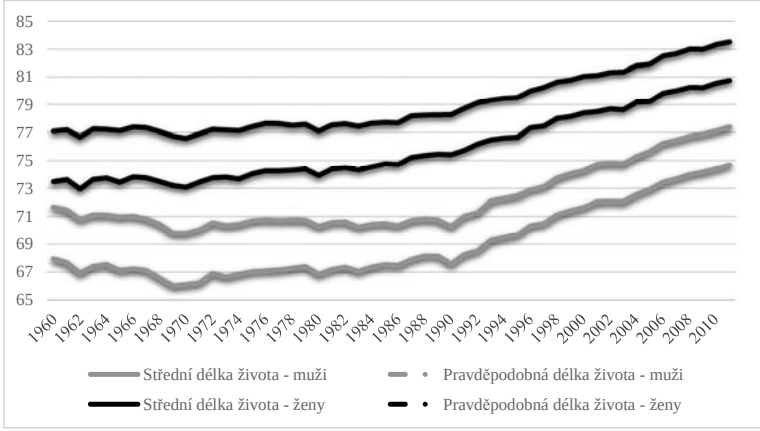
$$\tilde{e}(x) = x_D + \frac{l(x_D) - 0{,}5.l(x)}{l(x_D) - l(x_D + 1)} - x, \quad (4)$$

kde x_D je věk, ve kterém je ještě počet dožívajících se vyšší než polovina souboru osob dožívajících se přesného věku x, $l(x)$ je počet dožívajících přesného věku x z výchozího souboru živě narozených v tabulkové populaci, $l(x_D)$ je počet osob, které se dožili přesného věku x z výchozího souboru živě narozených v tabulkové populaci, $l(x_D+1)$ je počet dožívajících se přesného věku x_D+1. Protože se (podobně jako v případě střední délky života osoby v přesném věku x) jedná o délku zbývajícího života, je nutno (na rozdíl od výpočtu pro novorozence) ještě odečíst věk x (tj. již prožitou dobu).

3. Analýza vývoje úmrtnosti pomocí pravděpodobné a střední délky života

V tomto příspěvku bude nejprve porovnán vývoj pravděpodobné a střední délky života u novorozenců v České republice od roku 1960 do roku 2011. Je všeobecně známou skutečností, že pro novorozence je střední délka života o několik let nižší než pravděpodobná délka života.

Proto se v další části se zaměříme na zkoumání, v jakém věku poprvé překročí střední délka života *x*-letých pravděpodobnou délku života *x*-letých, a dále budeme zkoumat, k jakým změnám hodnoty tohoto věku dochází v průběhu sledovaného období. Získané výsledky budou prezentovány v grafické podobě a budou publikovány odděleně pro muže a pro ženy.

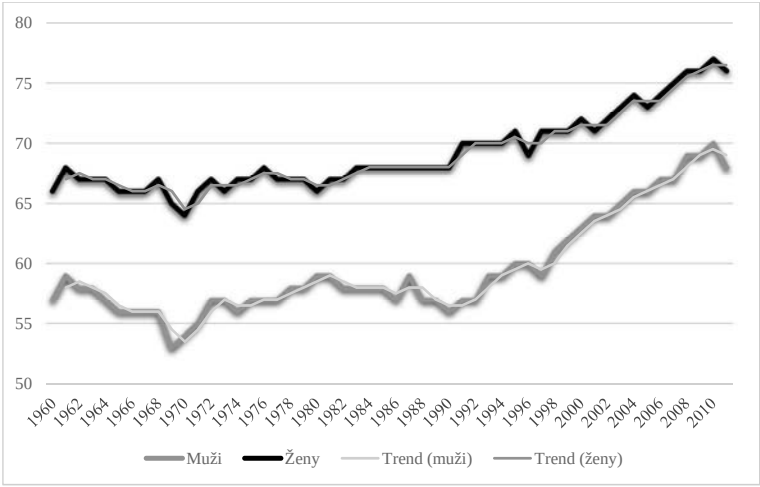


Graf 1 – Vývoj pravděpodobné a střední délky života pro novorozence v České republice od roku 1960 do roku 2011 – muži, ženy

Zdroj: data Eurostat, vlastní výpočty

Na grafu 1 je znázorněn vývoj střední a pravděpodobné délky života u novorozenců (chlapců a dívek) v České republice od roku 1960 do roku 2011. Je zřejmé, že u mužů do roku 1990 žádný nárůst nebyl. Během 60. let naopak došlo k poklesu, v 80. a 90. letech se objevil jen mírný růst střední délky života a stagnace pravděpodobné. Je zřejmé, že během celého sledovaného období docházelo u obou ukazatelů k nárůstu. Důležité je, že po celé sledované období dosahovala pravděpodobná délka života mužů vyšších hodnot než střední délka života. Tento rozdíl v roce 1960 činil téměř 4 roky, do roku 1990 se postupně snížil na 2,5 roku, v posledních 20 letech se pohyboval kolem těchto hodnot.

Pokud budeme zkoumat vývoj střední a pravděpodobné délky života u narozených dívek, tak vidíme, že (na rozdíl od mužů) nedocházelo k poklesu v roce 1969. I zde je však třeba zmínit, že do roku 1980 byl nárůst nepatrný, pravděpodobná délka spíše stagnovala, teprve po roce 1980 a zejména po roce 1990 je nárůst výrazný. Pokud budeme porovnávat rozdíly mezi pravděpodobnou a střední délkou života, tak zjistíme, že na počátku období byla hodnota rozdílu přibližně 3,7 roků. Později došlo ke snížení této hodnoty na necelé 3 roky. Příčinou by mohl být rychlejší nárůst hodnoty střední délky života při narození.



Graf 2 – Vývoj věku, ve kterém poprvé překročila střední délka života pravděpodobnou - Česká republika – muži, ženy

Zdroj: data Eurostat, vlastní výpočty

Na grafu 2 je znázorněn vývoj věku, ve kterém poprvé překročila střední délka života pravděpodonou pro muže a ženy v České republice. Trend vývoje byl v podstatě podobný jako u střední a pravděpodobné délky života. Odchylky hodnot v jednotlivých letech od základního trendu vývoje jsou však vyšší než u střední či pravděpodobné délky života. Zhruba do roku 1990 hodnoty v podstatě stagnovaly (s výrazným poklesem kolem r. 1969), po roce 1990 dochází k poměrně výraznému růstu hodnot. K výraznějšímu poklesu potom dochází v roce 2010. Hlavní příčinou tohoto vývoje je vyšší meziroční nárůst hodnoty pravděpodobné délky života (v porovnání se střední délkou života).

U žen byl trend vývoje také podobný jako u střední a pravděpodobné délky života. Odchylky hodnot v jednotlivých letech od základního trendu vývoje jsou však vyšší než u zmíněných ukazatelů. Zhruba do roku 1990 hodnoty v podstatě stagnovaly (s výrazným poklesem kolem r. 1970), po roce 1990 dochází k poměrně výraznému růstu hodnot věku.

Roustoucí vývoj věku, ve kterém poprvé překročila střední délka života pravděpodobnou také naznačuje, že v populaci přibývá starých osob. Můžeme tak soudit z vlastnosti průměru a mediánu, ze kterých plyne: pokud je průměr vyšší než medián, tak v souboru převažují vyšší hodnoty (tzn. osoby ve vyšším věku).

4. Závěr

V první části se příspěvek zabýval analýzou úmrtnosti české populace od roku 1960 do roku 2011 pomocí střední a pravděpodobné délky života při narození. Z vývoje zkoumaných ukazatelů můžeme usuzovat, že v průběhu sledovaného období docházelo k růstu hodnot obou charakteristik. Uvedený vývoj naznačuje, že jak u českých mužů, tak u českých žen docházelo po roce 1990 k výraznému zlepšování úmrtnostních poměrů.

Při zkoumání rozdílů mezi střední a pravděpodobnou délkou života zjišťujeme, že nedochází k nějakému výraznému sbližování zkoumaných charakteristik. Po roce 1990 k výraznému zlepšování. Během sledovaného období došlo jen k nepatrnému snížení rozdílů mezi pravděpodobnou a střední délkou života pro narozené (muži: snížení o 0,5 roku a ženy:

snížení o 0,7 roku). Zároveň tedy můžeme říci, že oba ukazatele vykazovaly podobný vývojový trend.

V poslední části se příspěvek zabýval zkoumáním, v jakém věku poprvé převýší střední délka života pravděpodobnou délku života. Ze získaných výsledků je zřejmé, že zkoumaná charakteristika vykazuje podobný trend vývoje jako střední a pravděpodobná délka života (platí jak pro muže, tak pro ženy). Ale i přesto můžeme vyslovit obecný závěr, že v průběhu sledovaného období se tento věk od roku 1990 se zvýšil více, než vzrostla délka života. Je to tedy výrazný nárůst.

Poděkování

Tento příspěvek vznikl za podpory projektu VŠE IGA 24/2013 "Úmrtnost a stárnutí populace České republiky".

Literatura

BOLESLAWSKI, L. – TABEAU, E. Comparing Theoretical Age Patterns of Mortality Beyond the Age of 80. *In: TABEAU, E. – VAN DEN BERG JETHS, A. and HEATHCOTE, CH. (eds.)* 2001. *Forecasting Mortality in Developed Countries: Insights from a Statistical, Demographic and Epidemiological Perspective*, 2001, p. 127 – 155.

BURCIN, B. – TESÁRKOVÁ, K. – ŠÍDLO, L. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. Demografie **52**, 2010: 77 – 89.

BURCIN, B. – HULÍKOVÁ-TESÁRKOVÁ, K. – KOMÁNEK, D. *DeRaS: software tool for modelling mortality intensities and life table construction*. Charles University in Prague, 2012, Prague. http://deras.natur.cuni.cz.

CIPRA, T. Matematické metody demografie a pojištění. 1990. Praha, STNL

EUROSTAT. Dostupný z WWW: http://epp.eurostat.ec.europa.eu/.

FIALA, T. *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha: Vysoká škola ekonomická v Praze, 2005.

GOMPERTZ, B.: On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* **115** (1825): 513–585.

GAVRILOV, L.A., GAVRILOVA, N.S.: Mortality measurement at advanced ages: a study of social security administration death master file. *North American actuarial journal* **15** (3) (2011): 432 – 447.

KOSCHIN, F. *Aktuárská demografie (úmrtnost a životní pojištění)*. Praha: Vysoká škola ekonomická v Praze, 2000.

MAKEHAM, W.M. On the Law of Mortality and the Construction of Annuity Tables. The Assurance Magazine, and Journal of the Institute of Actuaries 8 (1860): 301–310.

Adresy autorů:

Petra Dotlačilová, Ing.
Katedra demografie,
Fakulta informatiky a statistiky,
Vysoká škola ekonomická
Nám. W. Churchilla 4, 130 67, Praha 3
Petra.Dotlacilova@vse.cz

Jitka Langhamrová, doc., Ing., CSc.
Katedra demografie,
Fakulta informatiky a statistiky,
Vysoká škola ekonomická
Nám. W. Churchilla 4, 130 67, Praha 3
langhamj@vse.cz

Zahraniční migrace v ČR a SR v období 1991–2012
International migration in the Czech Republic and Slovakia in 1991–2012

Tomáš Fiala, Jitka Langhamrová

Abstract: The paper contains overview of the development of basic indicators of international migration in the Czech Republic and Slovakia in the period 1991–2012. Migration between the Czech Republic and Slovakia is distinguished from the migration with other countries. Besides registered emigration also the number of non-registered emigrants (estimated by the results of census) is taken into account. In Slovakia is this number higher than registered number of emigrants. In the first half of the 90th of the previous century a substantial part of the total amount of international migration of Czech Republic and Slovakia is the migration between these two countries, later the proportion of migration with other countries is increasing. Since 2001 the amount of international migration in the Czech Republic is relatively considerably increasing while in Slovakia it remains at the level of previous decade. Taking into account the unregistered migration the net international migration in Slovakia has negative value while in the Czech Republic is positive.

Abstrakt: Příspěvek obsahuje přehled vývoje základních charakteristik zahraniční migrace ČR a SR v období 1991–2012. Je rozlišena migrace mezi ČR a SR a migrace s ostatními zeměmi. Kromě registrované migrace je brán v úvahu též odhad (na základě výsledků sčítání lidu) počtu neregistrovaně vystěhovalých. Přitom na Slovensku je po roce 2000 tento počet vyšší než registrovaný počet emigrantů. V první polovině 90. let minulého století tvořila podstatnou část celkového objemu zahraniční migrace ČR a SR migrace mezi těmito zeměmi, později roste podíl migrace s jinými zeměmi. Od roku 2001 se poměrně výrazně zvyšuje objem zahraniční migrace v ČR, zatímco na Slovensku zůstává na úrovni poslední dekády minulého století. Po zohlednění neregistrované emigrace zůstává migrační saldo v ČR kladné, na Slovensku je však záporné.

Key words: Czech Republic, Slovakia, international migration, non-registered emigration.

Klíčová slova: Česká republika, Slovensko, zahraniční migrace, neregistrovaná emigrace.
JEL: J11, J15

1. Úvod

Po politických změnách, ke kterým došlo v Československu v roce 1989, následovaly postupně poměrně rychle změny společenské a ekonomické. Ta měly za následek i (někdy poměrně nečekané) změny demografického chování. Zatímco zrychlení růstu délky života mužů i žen nebylo příliš překvapivé, málokdo asi očekával tak prudký pokles plodnosti žen, kterého jsme byli svědky ve druhé polovině devadesátých let.

Nastaly rovněž předpoklady ke zvýšení zahraniční migrace. Byly zrušeny výjezdní doložky i vízová povinnost pro řadu zemí. To přineslo možnost dlouhodobých cest občanů Československa do zahraničí za účelem studia či práce. Na druhou stranu přibývalo zahraničních pracovníků i studentů v ČR i SR.

Obsahem článku je přehledná vývoje základních charakteristik zahraniční migrace v ČR a SR od roku 1991. Je rozlišena migrace mezi ČR a Slovenskem a migrace s ostatními zeměmi. Migrace mezi ČR a Slovenskem je považována za zahraniční migraci i v letech 1991 a 1992, i když se v té době se jednalo o migraci vnitřní mezi dvěma zeměmi federativního státu. Je rovněž zohledněna nehlášená emigrace odhadnutá na základě nedopočtů při sčítáni lidu v letech 2001 a 2011.

2. Metodologické poznámky

Statistická data týkající se migrace jsou mnohem méně přesná a hůře srovnatelná v čase než data týkající se přirozeného pohybu obyvatelstva. Údaje o zahraniční migraci nesleduje statistický úřad přímo, ale přebírá je od ministerstva vnitra, mohou se proto vyskytnout určité metodologické rozdíly.

Výrazný vliv měly i změny vymezení obyvatelstva ČR a s ní související změny vymezení zahraniční migrace. Zatímco do konce roku 2000 bylo do zahraničního stěhování ČR zahrnuto pouze stěhování osob s trvalým pobytem v ČR, od roku 2001 se do tohoto stěhování zahrnuje i stěhování cizinců s vízem nad 90 dnů i cizinců s přiznaným azylem. Od roku 2004 (vstup ČR a SR do EU) se údaje týkají též občanů zemí EU s přechodným pobytem a občanů třetích zemí s dlouhodobým pobytem. Změna zákonů týkající se pobytu cizinců v ČR na přelomu tisíciletí byla hlavní příčinou záporného migračního salda v roce 2001.

Od roku 2005 navíc nepublikuje ČSÚ počty migrantů tříděné podle zemí, ale pouze podle státního občanství. Protože však v letech 2002–2004 tvořily více než 96 % osob přistěhovalých ze SR do ČR osoby se slovenským státním občanstvím a přes 99 % osob vystěhovalých z ČR do SR byli občané Slovenska, byly od roku 2005 počty přistěhovalých ze Slovenska do ČR odhadnuty jako počty všech osob se Slovenským státním občanstvím přistěhovalých do ČR v daném roce, jako odhady počtu vystěhovalých z ČR do SR pak byly použity počty všech osob se slovenským státním občanstvím vystěhovalých z ČR.

Řada lidí v ČR i SR dlouhodobě (nebo trvale) pobývá v zahraničí, aniž by tuto skutečnost hlásila statistickému úřadu. Tito lidé by měli být považování rovněž za vystěhovalé. V článku se provádí (alespoň částečný) odhad jejich počtu na základě tzv. nedopočtu při sčítání lidu (rozdílu mezi počtem osob zjištěným při sčítání lidu a počtem osob zjištěným na základě každoročních bilancí). Počty těchto osob považujeme za počet neregistrovaně vystěhovalých v období mezi sčítáními. Protože nemáme další informace, předpokládáme pro jednoduchost, že neregistrované vystěhovávání probíhalo rovnoměrně po celé období, tj. že každý rok se neregistrovaně vystěhovalo 10 % celkového počtu neregistrovaně vystěhovalých za celé období mezi sčítáními.

3. Zahraniční migrace v ČR

V letech 1991–1993 tvořila většinu objemu zahraniční migrace ČR migrace se Slovenskem (viz Tab. 1). Jednou z možných příčin bylo rozdělení Československa na konci roku 1992. Počet přistěhovalých ze Slovenska se v roce 1992 přiblížil 12 tisícům a i v dalších letech se pohyboval v řádu několika tisíc ročně. Roční průměr za celé období byl o málo vyšší než 5 tisíc osob. Naproti tomu počet vystěhovalých na Slovensko prudce poklesl od roku 1994 z původních několika tisíc na několik set osob ročně. Průměrné roční migrační saldo činilo necelých 2 800 osob ve prospěch ČR ročně.

Počet přistěhovalých do ČR z jiných zemí byl poměrně vyrovnaný, dosahoval průměrně 6 800 osob ročně. Počet registrovaných vystěhovalých do jiných zemí sice činil v roce 1991 téměř 4 tisíce osob, v dalších letech se však pohyboval v řádu pouze několika set osob. Nedopočet při sčítání lidu v roce 2001 svědčí o tom, že počet neregistrovaně vystěhovalých činil průměrně ročně téměř 3,5 tisíce osob.

Za celé poslední desetiletí minulého století tvořili téměř 43 % všech přistěhovalých do ČR přistěhovalí ze Slovenska. Pokud zohledníme neregistrovanou emigraci, tvořila 52 % všech vystěhovalých, téměř 35 % vystěhovalo na Slovensko a jen 13 % do ostatních zemí. Neregistrovaná emigrace však mohla směřovat převážně do jiných zemí než na Slovensko, je tedy těžké odhadnout skutečný podíl vystěhovalých z ČR na Slovensko a do ostatních zemí.

Registrované migrační saldo v průměrné výši necelých 9 tisíc osob ročně tvořila téměř z 32 % migrace se Slovenskem. Po zahrnutí neregistrované emigrace činila průměrná roční hodnota migračního salda ČR (včetně migrace se Slovenskem) více než 5 tisíc osob,

Tab. 1: Zahraniční migrace z/do České republiky v letech 1991–2010

Charakteristika	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	průměr	
Přistěhovalí ze Slovenska	8 334	11 740	7 276	4 076	3 845	3 450	3 088	2 887	3 235	2 826	5 076	42,7%
Přistěhovalí z ostatních zemí	5 762	7 332	5 624	6 131	6 695	7 407	9 792	7 842	6 675	4 976	6 824	57,3%
Vystěhovalí na Slovensko	7 324	6 823	7 232	56	140	213	260	356	336	413	2 315	34,9%
Vystěhovalí do ostatních zemí	3 896	468	192	209	401	515	545	885	800	850	876	13,2%
Vystěhovalí neregistrovaní	3 452	3 452	3 452	3 452	3 452	3 452	3 452	3 452	3 452	3 452	3 452	52,0%
Saldo se Slovenskem	1 010	4 917	44	4 020	3 705	3 237	2 828	2 531	2 899	2 413	2 760	31,7%
Saldo ostatní země	1 866	6 864	5 432	5 922	6 294	6 892	9 247	6 957	5 875	4 126	5 948	68,3%
Saldo celkem (včetně nereg. migr.)	-576	8 329	2 024	6 490	6 547	6 677	8 623	6 036	5 322	3 087	5 256	x

pokračování

Charakteristika	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	průměr	
Přistěhovalí ze Slovenska	3 078	13 326	24 385	15 788	10 107	6 781	13 931	7 592	5 609	5 086	10 568	19,1%
Přistěhovalí z oststních zemí	9 840	31 353	35 630	37 665	50 187	61 402	90 514	70 225	34 364	25 429	44 661	80,9%
Vystěhovalí na Slovensko	8 711	14 455	18 262	21 152	1 946	629	802	585	4 167	6 424	7 713	27,6%
Vystěhovalí do oststních zemí	12 758	17 934	15 964	13 666	22 119	32 834	19 698	5 442	7 462	8 443	15 632	55,9%
Vystěhovalí neregistrovaní	4 604	4 604	4 604	4 604	4 604	4 604	4 604	4 604	4 604	4 604	4 604	16,5%
Saldo se Slovenskem	-5 633	-1 129	6 123	-5 364	8 161	6 152	13 129	7 007	1 442	-1 338	2 855	9,0%
Saldo ostatní země	-2 918	13 419	19 666	23 999	28 068	28 568	70 816	64 783	26 902	16 986	29 029	91,0%
Saldo celkem (včetně nereg. migr.)	-13 155	7 686	21 185	14 031	31 625	30 116	79 341	67 186	23 740	11 044	27 280	x

Zdroj dat: Český statistický úřad (ČSÚ)

V první dekádě současného století je na první pohled patrné výrazné zvýšení objemu zahraniční migrace (Obr. 1). Do značné míry však může být způsobeno změnou metodologie, kdy mezi zahraniční migraci je (na rozdíl od předchozích let) zahrnuto i stěhování cizinců s vízem nad 90 dnů i cizinců s přiznaným azylem. Od roku 2004 (vstup ČR do EU) se údaje týkají též občanů zemí EU s přechodným pobytem a občanů třetích zemí s dlouhodobým pobytem.

Tato změna se týká i migrace se Slovenskem. Roční počet přistěhovalých ze Slovenska je v průměru (10 tisíc osob ročně) dvojnásobný než v předchozím desetiletí. Jedná se však zřejmě převážně o dočasnou migraci – výrazně vzrostl i roční počet vystěhovalých z ČR na Slovensko, takže výsledné průměrné saldo je jen o málo vyšší než za období 1991–2000.

Mnohem více však vzrostl počet přistěhovalých do ČR z ostatních zemí. Dosáhl v průměru téměř 45 tisíc osob ročně, v roce 2007 se do ČR přistěhovalo více než 90 tisíc osob. Počet vystěhovalých rovněž několikanásobně vzrostl, sčítání lidu v roce 2011 potvrdilo i určité zvýšení neregistrované emigrace.

Vzhledem k výraznému zvýšení migrace z ostatních zemí tvořili přistěhovalí ze Slovenska v období 2001–2010 méně než 20 % všech přistěhovalých do ČR. Neregistrovaná migrace, tvořila pouze 16,5 % všech vystěhovalých, 27,6 % se vystěhovalo na Slovensko a téměř 60 % do ostatních zemí. Je tedy zřejmé, že na rozdíl od poslední dekády minulého století se v tomto století výrazně snížil podíl migrace se Slovenskem na celkové zahraniční migraci ČR.

Průměrné roční saldo registrované migrace ČR činilo téměř 32 tisíc osob ročně, tedy zhruba pětkrát více než v předchozí dekádě. Podíl migrace se Slovenskem dosahoval pouze 9 %. Po zohlednění neregistrované migrace bylo průměrné roční saldo zahraniční migrace ČR (včetně Slovenska) vyšší než 27 tisíc osob (Obr. 2).

V posledních letech je však patrný poměrně velmi výrazný pokles zahraniční migrace, do značné míry zřejmě způsobený pokračující ekonomickou krizí. V roce 2012 činilo saldo registrované migrace jen něco málo přes 10 tisíc osob (z toho se Slovenskem 4 tisíce), po zohlednění neregistrované emigrace by saldo bylo ještě nižší.

4. Zahraniční migrace v SR

Poznamenejme úvodem, že data pocházejí ze Slovenského statistického úřadu a že vzhledem k rozdělení Československa nedocházelo od roku 1994 k vzájemné výměně dat o migraci mezi ČR a SR. Z tohoto důvodu se slovenské údaje o migraci s ČR liší (někdy dost výrazně) od údajů ČSÚ o migraci se Slovenskem.

I na Slovensku tvořila na počátku sledovaného období většinu objemu zahraničního stěhování migrace s ČR (viz Tab. 2). V letech 1991–1994 byl počet přistěhovaných z ČR několikanásobně vyšší než počet přistěhovalých z ostatních zemí. I přes pozdější pokles byl roční průměr za celé období o něco vyšší než 3 tisíce osob. Počet vystěhovalých do ČR prudce poklesl od roku 1994 z původních několika tisíc na několik set osob ročně. Vzhledem k velmi vysokým hodnotám na začátku dekády je však roční průměr vyšší než 2,8 tisíce osob. Průměrné roční migrační saldo je proto jen o něco vyšší než 200 osob (ve prospěch Slovenska).

Tab. 2: Zahraniční migrace z/do Slovenské republiky v letech 1991–2010

Charakteristika	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	průměr	
Přistěhovalí z ČR	7 324	6 823	7 232	3 144	1 497	993	867	777	856	1 268	3 078	66,5%
Přistěhovalí z ostatních zemí	1 752	2 106	1 874	1 778	1 558	1 484	1 436	1 275	1 216	1 006	1 549	33,5%
Vystěhovalí do ČR	8 334	11 740	7 276	95	108	89	212	251	208	310	2 862	51,9%
Vystěhovalí do ostatních zemí	527	128	79	59	105	133	360	495	410	501	280	5,1%
Vystěhovalí neregistrovaní	2 376	2 376	2 376	2 376	2 376	2 376	2 376	2 376	2 376	2 376	2 376	43,1%
Saldo s ČR	-1 010	-4 917	-44	3 049	1 389	904	655	526	648	958	216	14,5%
Saldo ostatní země	1 225	1 978	1 795	1 719	1 453	1 351	1 076	780	806	505	1 269	85,5%
Saldo celkem (včetně nereg. migr.)	-2 161	-5 315	-625	2 392	466	-121	-645	-1 070	-922	-913	-892	x

pokračování

Charakteristika	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	průměr	
Přistěhovalí z ČR	990	749	650	987	1 144	1 164	1 178	1 405	1 440	1 160	1 087	21,2%
Přistěhovalí z ostatních zemí	1 033	1 563	1 953	3 473	4 132	4 425	7 446	7 360	4 906	4 112	4 040	78,8%
Vystěhovalí do ČR	398	449	448	662	734	706	775	638	605	629	604	10,2%
Vystěhovalí do ostatních zemí	613	962	746	924	1 139	1 029	1 056	1 067	1 374	1 260	1 017	17,2%
Vystěhovalí neregistrovaní	4 283	4 283	4 283	4 283	4 283	4 283	4 283	4 283	4 283	4 283	4 283	72,5%
Saldo s ČR	592	300	202	325	410	458	403	767	835	531	482	13,8%
Saldo ostatní země	420	601	1 207	2 549	2 993	3 396	6 390	6 293	3 532	2 852	3 023	86,2%
Saldo celkem (včetně nereg. migr.)	-3 271	-3 382	-2 874	-1 409	-880	-429	2 510	2 777	84	-900	-777	x

Zdroj dat: Štatistický úrad Slovenskej republiky (ŠÚSR)

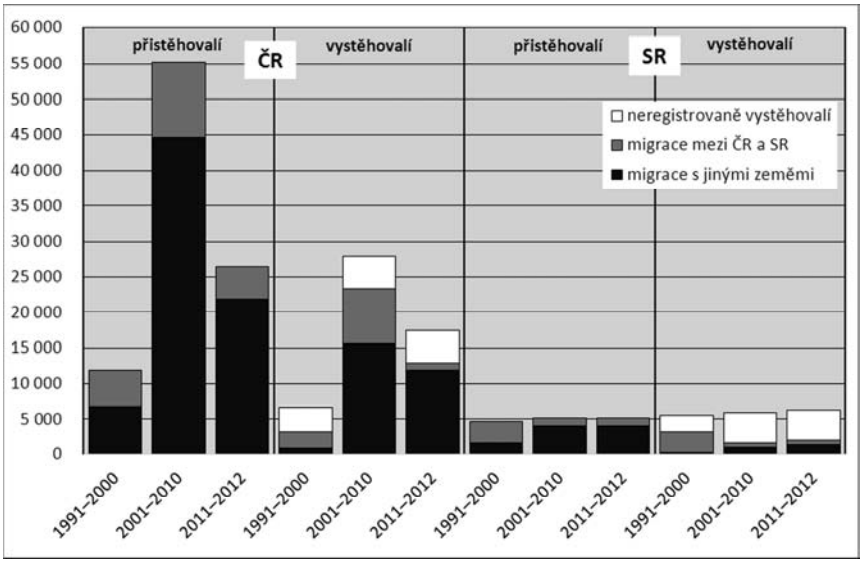
Počet přistěhovalých na Slovensko z jiných zemí byl poměrně vyrovnaný, pohyboval se kolem 1 500 osob ročně. Počet registrovaných vystěhovalých do ostatních zemí činil v průměru pouze 280 osob ročně. Migrační saldo tak dosahuje necelých 1 300 osob.

Nedopočet při sčítání lidu v roce 2001 však svědčí o poměrně vysokém počtu neregistrovaně vystěhovalých (průměrně ročně téměř 2 400 osob). Po zohlednění této neregistrované emigrace je celkové migrační saldo zahraniční migrace Slovenska (včetně migrace s ČR) záporné (průměrný roční úbytek téměř 900 osob).

Za celé desetiletí tvořili přistěhovalí z ČR téměř 2/3 všech přistěhovalých na Slovensko. Co týče vystěhovalých (včetně neregistrované emigrace), směřovalo jich téměř 52 % do ČR. Neregistrovaná emigrace tvoří 43 %, registrovaná migrace do ostatních zemí jen 5 % celkového počtu vystěhovalých.

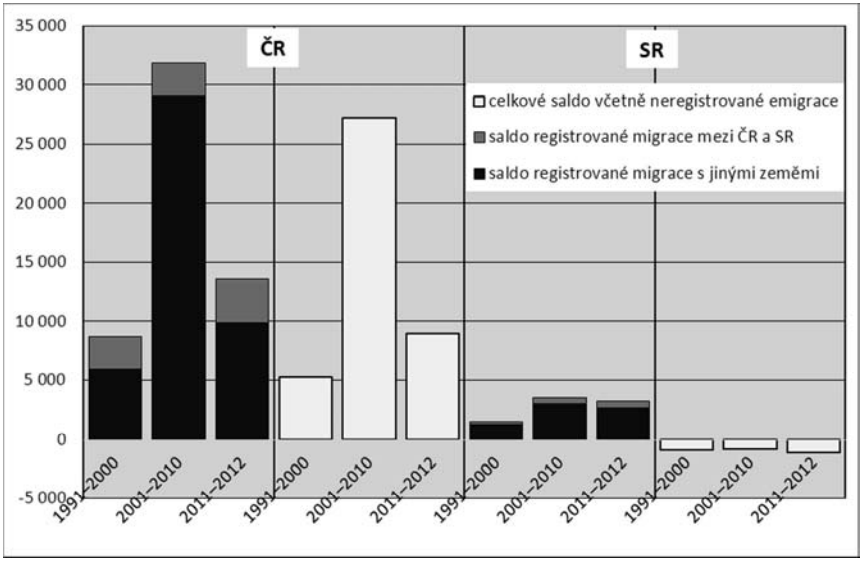
Na rozdíl od ČR nedochází na Slovensku v první dekádě současného století ke zvýšení objemu zahraniční migrace. Snižuje se však objem migrace s ČR, roste naopak migrace s jinými zeměmi a neregistrovaná emigrace.

Roční počet přistěhovalých z ČR je v průměru (necelých 1 100 osob ročně) třikrát menší než v předchozím desetiletí. Výrazně však klesá i roční počet vystěhovalých do ČR na Slovensko (v průměru 600 osob ročně), takže výsledné průměrné saldo (téměř 500 osob) je více než dvojnásobné v porovnání s předchozí dekádou.



Zdroj: Data ČSÚ a ŠÚSR, výpočet a zpracování vlastní

Obr. 18: Zahraniční migrace ČR a SR
průměrné roční počty přistěhovalých a vystěhovalých v uvedeném období



Zdroj: Data ČSÚ a ŠÚSR, výpočet a zpracování vlastní

Obr. 2: Zahraniční migrace ČR a SR
průměrné roční saldo registrované migrace a migrace včetně neregistrovaných vystěhování

Došlo k výraznému zvýšení objemu migrace s ostatními zeměmi. Průměrný roční počet přistěhovalých překročil 4 tisíce osob, počet registrovaných vystěhovalých do ostatních zemí dosahoval v průměru 1 tisíc osob ročně, průměrné migrační saldo vzrostlo na 3 tisíce osob ročně.

Nedopočet zjištěný při sčítání lidu v roce 2011 byl však téměř dvojnásobný v porovnání s předchozí dekádou (průměrně ročně téměř 4 300 osob). Po zohlednění této neregistrované emigrace je celkové migrační saldo zahraniční migrace Slovenska (včetně migrace s ČR) opět záporné, i když o něco nižší než v předchozí dekádě. Průměrný roční úbytek činí necelých 800 osob.

Podíl přistěhovalých z ČR činil v tomto období pouze 21 % všech přistěhovalých. Výrazně vzrostla neregistrovaná emigrace, činí více než 72 % odhadovaného počtu všech vystěhovalých. Podíl emigrace do ČR poklesl na 10 %, registrovaná migrace do ostatních zemí činí 17 % celkového počtu vystěhovalých.

5. Závěry

Na počátku 90. let vzhledem k rozdělení ČSFR tvořila výrazně vysoký podíl zahraniční migrace v obou zemích migrace mezi ČR a SR. V pozdějších letech postupně narůstal objem migrace s jinými zeměmi. Zatímco v ČR došlo po roce 2001 k výraznému zvýšení objemu zahraniční migrace (na více než čtyřnásobek), v SR zůstal se počet přistěhovalých i vystěhovalých zvýšil proti úrovni 90. let minulého století zhruba jen o 10 %. Odhadovaný počet neregistrovaně vystěhovalých ze Slovenska v období 2001–2010 je však téměř tak vysoký jako v ČR, podstatně vyšší než evidovaný počet vystěhovalých. Po přihlédnutí k této skutečnosti je průměrné saldo zahraniční migrace Slovenska (na rozdíl od ČR) po celé sledované období záporné

Článek vznikl za podpory Interní grantové agentury Vysoké školy ekonomické v Praze F4/24/2013 *Úmrtnost a stárnutí obyvatelstva ČR*.

Literatura

ROUBÍČEK, V. 1997. Úvod do demografie. 1. vyd. Praha. Codex Bohemia.

Český statistický úřad: Demografická ročenka České republiky 2012–xx: Dostupné z: http://www.czso.cz/csu/2013edicniplan.nsf/p/4019-13.

Štatistický úrad Slovenskej republiky: Pohyb obyvatelstva v Slovenskej republike v roku 2012–xx. Dostupné z: http://portal.statistics.sk/showdoc.do?docid=6674.

Adresa autorů:

Tomáš Fiala, RNDr., CSc.
Katedra demografie FIS VŠE
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
fiala@vse.cz

Jitka Langhamrová, doc., Ing., CSc.
Katedra demografie FIS VŠE
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
langhamj@vse.cz

Vzťah počtu nehospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja
Regional dependence of one day surgery healthcare young outpatients number

Beáta Gavurová, Samuel Korony

Abstrakt: Príspevok uvádza výsledky korešpondenčnej analýzy závislosti počtu nehospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja za obdobie 2009 - 2011. Agregovaný kraj má najväčšie podiely v odboroch Chirurgia a ORL. Agregovaný odbor má najväčšie podiely v krajoch Bratislava, Košice, Prešov a Žilina.

Abstract: The paper deals with correspondence analysis results of one day surgery young outpatients number dependence to Slovak regions during 2009 – 2011. Aggregate region has got the largest proportions in the sections of surgery and otolaryngology. Agregate section has got the largest proportions in the regions of Bratislava, Košice, Prešov and Žilina.

Kľúčové slová: jednodňová zdravotná starostlivosť, korešpondenčná analýza

Keywords: One day healthcare, Correspondence Analysis

JEL classification: C25, I12

1. Úvod

Jednodňová zdravotná starostlivosť alebo jednodňová (ambulantná) chirurgia (ďalej JZS) je definovaná ako operácia alebo procedúra, počas ktorej je pacient prijatý alebo prepustený z chirurgickej pohotovosti v ten istý deň. V súčasnosti je JZS stále viac považovaná za štandardnú plánovanú procedúru, ktorá môže byť výhodná nielen pre samotného pacienta a jeho rodinu, ale aj pre poskytovateľa zdravotnej starostlivosti (Gavurová et al. 2013).

Rozvoj kvalitných služieb JZS v európskych krajinách je aj prioritou ich vlád v oblasti zdravotnej starostlivosti. Jednodňová chirurgia na Slovensku, rovnako ako aj v zahraničí využíva pozitívne výsledky viacerých výskumov a medicínskej praxe deklarujúce fakt, že v domácom prostredí najlepšie prebieha liečba a rekonvalescencia pacientov po operačných výkonoch. Nedávny prieskum uskutočnený v devätnástich krajinách (Toftgaard – Parmentier, 2006) poukázal na signifikantné rozdiely v podiele realizovaných výkonov JZS, od menej než 10 % (napr. Poľsko), až po viac než 80 % (USA, Kanada). Rozdiely boli zaznamenané nielen v rámci jednotlivých krajín, medzi nemocnicami v danej krajine, jej oddeleniami, ale aj medzi špecialistami v rovnakej nemocnici. Implementácia systému JZS na Slovensku nie je zdravotnými poisťovňami ekonomicky vyhodnocovaná, interpretovaná a podporovaná s prihliadnutím na regionálne špecifiká, typ a vlastníctvo zdravotníckeho zariadenia, ako aj na ostatné sociálno-ekonomické podmienky nevyhnutných na jej rozvoj. Podchytenie všetkých signifikantných determinantov rozvoja a využívania systému JZS na Slovensku umožní identifikáciu relatívne presných úspor v systéme zdravotníctva, dosiahnutých zavedením JZS do praxe. Uvedené úspory je možné vyčísliť aj pre ostatné časti národného hospodárstva.

V našom príspevku sme na základe údajov poskytnutých Národným centrom zdravotníckych informácií chceli zistiť, či a ako je podiel počtu nehospitalizovaných pacientov JZS vo veku do 18 rokov („juniorov“) ovplyvnený geografickou polohou zdravotníckeho zariadenia (krajom), v ktorom bol pacient operovaný.

2. Vymedzenie materiálu skúmania a použitých metód

Podkladom pre naše analýzy boli údaje poskytnuté Národným centrom zdravotníckych informácií z Ročného výkazu J (MZ SR) 1-01 o jednodňovej starostlivosti za roky 2009 až 2011 s počtami pacientov, ktorým bol uskutočnený výkon daného typu podľa kódu číselníka výkonov JZS z Vestníka Ministerstva zdravotníctva SR zo dňa 1.3.2006, čiastka 9-16, časť 23 – „Odborné usmernenie MZ SR o výkonoch jednodňovej zdravotnej starostlivosti“ (Gavurová et al., 2013). Podľa uvedeného usmernenia je sedem špecializačných odborov JZS (Chirurgia, ortopédia, úrazová chirurgia a plastická chirurgia (ďalej „Chirurgia“), Gynekológia a pôrodníctvo (ďalej „Gynekológia“), Oftalmológia, Otorinolaryngológia (ďalej „ORL“), Urológia, Zubné lekárstvo a Gastroenterologická chirurgia a Gastroenterológia). Výkony JZS sa za posledné dva odbory vykazujú v minimálnej miere a preto sme ich nezahrnuli do ďalších analýz. Predpokladáme pritom, že zastúpenie jednotlivých typov výkonov JZS v každom odbore je zhruba rovnaké pre kraje. Pri relatívne konsolidovanom vývoji ľudskej populácie (bez veľkých regionálnych prírodných alebo ekologických katastrof) je to rozumný predpoklad.

Na analýzu vzťahu podielu počtu nehospitalizovaných pacientov juniorov JZS a kraja sme použili korešpondenčnú analýzu implementovanú v štatistickom systéme SPSS verzia 19. Korešpondenčná analýza je exploračná metóda pre analýzu vzťahu riadkových a stĺpcových podielov kontingenčných tabuliek. Najjednoduchší prístup k jej pochopeniu je považovať ju za analýzu hlavných komponentov kategorických dát (Jobson 1991).

3. Výsledky korešpondenčnej analýzy vzťahu počtu nehospitalizovaných juniorov JZS a kraja

Budeme vychádzať z tabuľky počtu nehospitalizovaných juniorov po krajoch (riadky) a špecializačných odboroch JZS (stĺpce) za všetky tri roky 2009 - 2011. V prvom stĺpci tabuľky 1 je rozdelenie po krajoch. V druhom až piatom stĺpci sú počty nehospitalizovaných juniorov podľa odboru výkonu JZS. Tabuľka obsahuje aj riadkové a stĺpcové úhrny. To je východisková situácia v korešpondenčnej analýze. Je vidieť, že v prípade viac ako dvoch riadkov alebo stĺpcov sa v tabuľke absolútnych počtov ťažko hľadajú nejaké súvislosti. A práve cieľom korešpondenčnej analýzy je zistiť, ktoré riadky a stĺpce sú navzájom podobné z hľadiska ich štruktúry (podielov). Ďalším krokom je urobiť tabuľky riadkových (tabuľka 2) a stĺpcových (tabuľka 3) podielov.

Tab. 1: Kraj verzus odbor JZS – počty nehospitalizovaných juniorov

Kraj \ Odbor	Chir	Gyn	Oftal	ORL	Urol	Spolu
B. Bystrica	553	35	18	602	13	1 221
Bratislava	888	20	36	2 600	27	3 571
Košice	992	66	83	2 401	2 880	6 422
Nitra	180	115	9	905	107	1 316
Prešov	334	173	52	3 919	32	4 510
Trnava	605	20	1	354	423	1 403
Trenčín	534	30	2	770	80	1 416
Žilina	1 027	177	120	2 309	417	4 050
Spolu	5 113	636	321	13 860	3 979	23 909

Tab. 2: Kraj verzus odbor JZS – riadkové podiely nehospitalizovaných juniorov

Kraj \ Odbor	Chir	Gyn	Oftal	ORL	Urol	Total
B. Bystrica	**,453**	,029	,015	**,493**	,011	1,000
Bratislava	**,249**	,006	,010	**,728**	,008	1,000
Košice	,154	,010	,013	**,374**	**,448**	1,000
Nitra	,137	,087	,007	**,688**	,081	1,000
Prešov	,074	,038	,012	**,869**	,007	1,000
Trnava	**,431**	,014	,001	**,252**	**,301**	1,000
Trenčín	**,377**	,021	,001	**,544**	,056	1,000
Žilina	**,254**	,044	,030	**,570**	,103	1,000
Mass	**,214**	,027	,013	**,580**	,166	

V tabuľke 2 sú riadkové profily (podiely v percentách) v 5 stĺpcoch. Ak by všetky kraje mali rovnomerne rozdelené podiely nehospitalizovaných juniorov v odboroch, potom by boli ich podiely 0,2. Korešpondenčná analýza skúma rozdiely medzi jednotlivými riadkovými profilmi a priemerným riadkovým profilom (Mass) v priestore menšej dimenzie.

Z tabuľky vidíme, že kraj B. Bystrica má prevládajúce podiely nehospitalizovaných juniorov v odboroch Chirurgia (45,3%) a ORL (49,3%). Podobne je na tom kraj Bratislava - Chirurgia (24,9%) s ešte väčším podielom na odbore ORL (72,8%). Kraj Košice je odlišný - ma najväčšie podiely v odboroch Urológia (44,8%) a ORL (37,4%). V krajoch Nitra a Prešov jednoznačne prevláda odbor ORL (68,8%, resp. 86,9%). Osobitné postavenie má kraj Trnava s podielmi 43,1 %, 25,2% a 30,1% v odboroch Chirurgia, ORL a Urológia. Posledné dva kraje – Trenčín a Žilina majú podobné profily ako kraje B. Bystrica a Bratislava. Priemerný kraj (riadok Mass) má najväčšie podiely v odboroch Chirurgia a ORL, čo je pochopiteľné vzhľadom na ich výskyt v štyroch krajoch.

Pre pohľad z druhej strany je vhodné pozrieť sa aj na stĺpcové podiely (tabuľka 3).

Tab. 3: Kraj verzus odbor JZS – stĺpcové podiely nehospitalizovaných juniorov

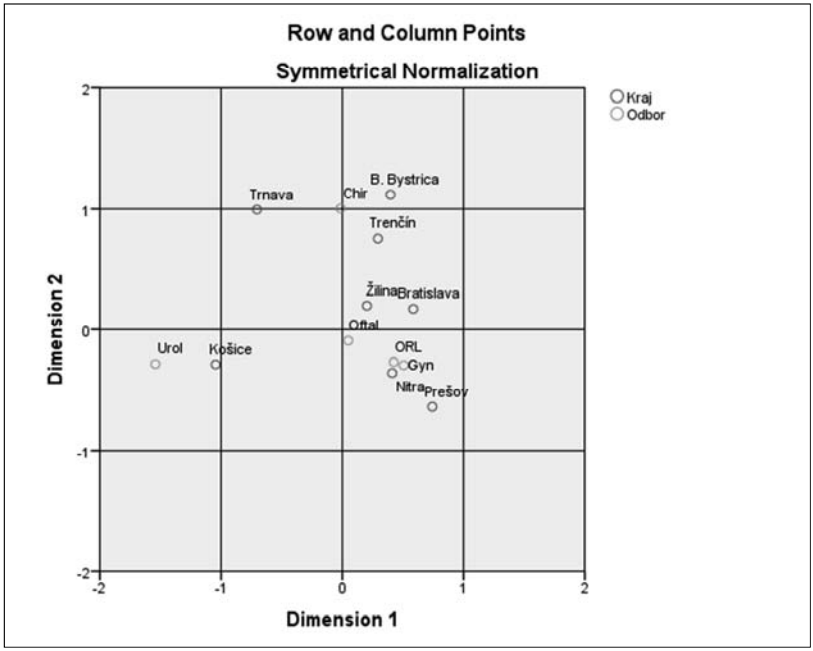
Kraj / Odbor	Chir	Gyn	Oftal	ORL	Urol	Mass
B. Bystrica	,108	,055	,056	,043	,003	,051
Bratislava	**,174**	,031	,112	**,188**	,007	**,149**
Košice	**,194**	,104	**,259**	**,173**	**,724**	**,269**
Nitra	,035	**,181**	,028	,065	,027	,055
Prešov	,065	**,272**	**,162**	**,283**	,008	**,189**
Trnava	,118	,031	,003	,026	,106	,059
Trenčín	,104	,047	,006	,056	,020	,059
Žilina	**,201**	**,278**	**,374**	**,167**	,105	**,169**
Total	1,000	1,000	1,000	1,000	1,000	

Korešpondenčná analýza v tomto prípade skúma rozdiely medzi jednotlivými stĺpcovými profilmi a priemerným stĺpcovým profilom (Mass). Ak by všetky odbory mali rovnomerne rozdelené podiely nehospitalizovaných juniorov po krajoch, potom by boli ich podiely 0,125. Maximálne podiely krajov v jednotlivých odboroch sú: Chirurgia – kraj Bratislava, Košice

a Žilina; Gynekológia – kraj Nitra, Prešov a Žilina; Oftalmológia – kraj Košice, Prešov a Žilina; ORL – Bratislava Košice, Prešov a Žilina; Urológia – Košice. Priemerný odbor (stĺpec Mass) má najväčšie podiely v krajoch Bratislava, Košice, Prešov, Žilina, čo sa dalo očakávať.

Ďalšou časťou výstupu je tabuľka aproximácie rozkladu kontingenčnej tabuľky na singulárne čísla (pre ušetrenie miesta ju neuvádzame). Prvý rozmer vysvetľuje 73,4 %. Druhý rozmer 21,6 %. Spolu je to 94,9 %, čo je výborné. Projekciou z pôvodne päťrozmernej kontingenčnej tabuľky do dvojrozmernej sme stratili len 5 %. Máme istotu, že vzťahy a detaily v rámci pôvodnej päťrozmernej štruktúry budú s dostatočnou presnosťou prenesené na plochu. Poslednou textovou časťou výstupu sú tabuľky kvality zobrazenia riadkov (kraje) a stĺpcov (odbory) (aj tie pre ušetrenie miesta neuvádzame).

Hlavným grafickým výstupom korešpondenčnej analýzy je korešpondenčný graf (mapa). Pri súčasnom zobrazení riadkov a stĺpcov kontingenčnej tabuľky sa nazýva biplot.



Obr. 1: Korešpondenčný biplot krajov a odborov nehospitalizovaných juniorov

Na korešpondenčnom biplote (obrázok 1) sú zobrazené kraje a odbory z pohľadu ich podielu (štruktúry) nehospitalizovaných juniorov. Pri jeho interpretácii musíme zobrať do úvahy, že sú na ňom nepresne zobrazené kraje Nitra a Žilina a odbory Gynekológia a Oftalmológia. Z krajov sú najbližšie k sebe B. Bystrica a Trenčín. A sú spolu s Trnavou aj blízko odboru Chirurgia, lebo majú zo všetkých krajov v ňom najväčší podiel nehospitalizovaných juniorov. Najodľahlejší je kraj Košice a je aj blízko odboru Urológia, pretože má v ňom najväčší podiel

nehospitalizovaných. Bratislavský kraj je vedľa odboru ORL s veľkým podielom nehospitalizovaných, podobne aj Prešovský kraj.

4. Záver

Ďalší vývoj JZS v nasledujúcich rokoch bude závisieť od mnohých vplyvov. Na prvom mieste je finančný vplyv, ktorý bude závisieť od prístupu poisťovní k systému JZS, jej hlavných aktérov, ako aj vládnej podpory. Poisťovne by mali obmedziť limitovanie počtu výkonov JZS a stanoviť jednotkovú cenu za výkon aspoň na úroveň hospitalizovaného pacienta. Tiež bude záležať aj na ďalšom rozvoji chirurgických metód a anestéziologickej starostlivosti a ich vplyvu na miniinvazívnu chirurgiu a pooperačné komplikácie a úmrtnosť. Nemenej dôležitým determinantom je aj sociálny faktor, ktorý vplýva na dĺžku pobytu v nemocnici po operácii, ako aj na voľbu výkonu formou JZS. Dôležitá je aj spokojnosť pacientov s realizáciou výkonu JZS, lekárov a lekárskeho personálu s podmienkami na výkon JZS, ako aj možnosti a prostriedky efektívnej komunikácie lekárskeho personálu s pacientmi.

Odborníci na danú problematiku zostávajú v otázkach ďalšieho rozvoja JZS skeptickí. Dôvodom sú pretrvávajúce problémy týkajúce sa úhrad poisťovní za chirurgické výkony, ako aj ich mesačné finančné limity, ktoré spôsobujú tvorbu čakacích listín aj v zariadeniach JZS. Nevyhnutné je také nastavenie systému zdravotnej starostlivosti, pri ktorom sa za jednoduché výkony dosiahnu nižšie platby, za zložitejšie vyššie, inak nemôže dôjsť k radikálnym zmenám v rozvoji JZS. Pokiaľ budú nemocnice spravodlivo platené za náročné výkony, možno sa ochotne zbavia jednoduchších, prípadne si vytvoria centrá jednodňovej chirurgie.

Vzhľadom na rozsiahlosť výstupov a obmedzenie veľkosti príspevku prezentujeme z našich rozsiahlych analýz len parciálne výsledky s cieľom odhaľovať kritické oblasti systému JZS na Slovensku a ich dopadov na jeho ďalší rozvoj. Našou ambíciou je tiež poukázať aj na nevyhnutnosť riešenia problematiky súvisiacej s údajovou základňou výkonov JZS – výkazníctvom a s tým súvisiace koncepčné a metodologické problémy, ktoré avizujú aj medzinárodné organizácie, ako sú OECD, WHO a Eurostat. Na elimináciu uvedených problémov zrealizovali tieto organizácie aktivity s cieľom vytvorenia jednotného medzinárodného dotazníka na získavanie konzistentných a porovnateľných medzinárodných údajov o chirurgických výkonoch. Tak sa získa hodnotná platforma v procese národného a medzinárodného benchmarkingu, ako aj ďalšieho rozvoja JZS na ceste k zvyšovaniu efektívnosti zdravotníckych systémov jednotlivých krajín.

Výstupy našich analýz poskytujú hodnotné informácie nielen pre zdravotnícke zariadenia, zdravotné poisťovne a rôzne iné zainteresované inštitúcie, ale aj pre pedagogický proces, nakoľko absentujú publikácie riešiace problematiku JZS na Slovensku, ktoré by pomohli študentom medicíny na univerzitách pochopiť aj ekonomickú stránku systému JZS, ktorá je neodlučiteľná od medicínskej.

5. Literatúra

GAVUROVÁ, B. – ŠOLTÉS, V. – KAFKOVÁ, K. – ČERNÝ, Ľ. 2013. Vybrané aspekty efektívnosti slovenského zdravotníctva. Jednodňová zdravotná starostlivosť a jej rozvoj v podmienkach Slovenskej republiky. Košice: Technická univerzita, 2013. 275 s. ISBN 978-80-553-1438-9.

JOBSON, J.D.:Applied Multivariate Data Analysis. Vol. II: Categorical and Multivariate Methods. New York: Springer - Verlag, 1992. 731 p. ISBN 0387978046

KORÓNY, S. – GAVUROVÁ, B.: Analýza vzťahu podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 93 - 98. ISSN 1336-7420

KORÓNY, S. – GAVUROVÁ, B.: Analýza vzťahu podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od odboru. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 99 - 104. ISSN 1336-7420

KORÓNY, S. – GAVUROVÁ, B.: Analýza vývoja podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 105 -111. ISSN 1336-7420

TOFTGAARD, C. – PARMENTER, G. 2006. International terminology in ambulatory surgery and its worldwide practice. In: Lemos P. at al., eds. Day Surgery Development and Practice. London, UK: International Association for Ambulatory Surgery (IAAS) 2006: 35-59.

Príspevok uvádza predbežné výsledky výskumu v súlade s podporeným projektom VEGA č. 1/1050/12 „Návrh systému merania výkonnosti v zdravotníckych zariadeniach na Slovensku a implementácia metrík výkonnosti.“

Adresa autorov:

Beáta Gavurová, doc., Ing., PhD., MBA.
Ekonomická fakulta
Technická univerzita v Košiciach
Němcovej 32
040 01 Košice
Email: beata.gavurova@tuke.sk

Samuel Koróny, RNDr., PhD.
Inštitút ekonomických vied
Ekonomická fakulta UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
Email: samuel.korony@umb.sk

Vzťah počtu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja

Regional dependence of one day surgery healthcare young inpatients number

Beáta Gavurová, Samuel Korónyi

Abstrakt: Príspevok uvádza výsledky korešpondenčnej analýzy závislosti hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja za obdobie 2009 - 2011. Agregovaný kraj má najväčšie podiely v odboroch Chirurgia a ORL. Agregovaný odbor má najväčšie podiely v krajoch Banská Bystrica, Košice a Prešov.

Abstract: The paper deals with correspondence analysis results of one day surgery young inpatients dependence to Slovak regions during 2009 – 2011. Aggregate region has got the largest proportions in the sections of surgery and otolaryngology. Agregate section has got the largest proportions in the regions of Banská Bystrica, Košice a Prešov.

Kľúčové slová: jednodňová zdravotná starostlivosť, korešpondenčná analýza

Keywords: One day healthcare, Correspondence Analysis

JEL classification: C25, I12

1. Úvod

V období transformácie systému verejného zdravotného poistenia je problematika odhaľovania rezerv v procese zvyšovania efektívnosti, ako aj optimalizácie liečebných a s nimi súvisiacich ekonomických procesov v zdravotníckych zariadeniach nesmierne zložitá a je predmetom neustálych rokovaní na rôznych úrovniach zdravotníckeho systému na Slovensku. Jednou z možností úspor finančných prostriedkov zdravotných poisťovní je zavedenie jednodňovej zdravotnej starostlivosti (JZS), výhodnej aj pre pacientov, ktorá funguje vo svete už viac ako tri desiatky rokov. Jej maximálny podiel na celkových chirurgických výkonoch dosahuje v niektorých krajinách takmer 90 % (USA), kým na Slovensku je to okolo 7 %. U nás JZS našlo podporu u zdravotných poisťovní, ako aj MZ SR, kde je evidentná podpora Vládneho programu MZ SR, žiaľ za 15 rokov sa ju nepodarilo dostatočne rozvinúť. Existuje mnoho dôvodov, ktoré bránia jej širšiemu zavádzaniu a využívaniu, čím by sa mohli ušetriť značné finančné zdroje zdravotníckeho systému, ktoré by bolo možné využiť v urgentných oblastiach (Gavurová, 2013).

V našom príspevku sme na základe údajov poskytnutých Národným centrom zdravotníckych informácií chceli zistiť, či a ako je počet hospitalizovaných pacientov JZS vo veku do 18 rokov („juniorov“) ovplyvnený geografickou polohou zdravotníckeho zariadenia (krajom), v ktorom bol pacient operovaný.

Zistené fakty nám pomôžu hlbšie získavať hodnotnú analytickú platformu v problematike riešenia oceňovania výkonov JZS, nastavenia rizikových indexov, ako aj odhaľovania socio-ekonomických trendov súvisiacich s rozvojom JZS na Slovensku.

2. Vymedzenie materiálu skúmania a použitých metód

Podkladom pre naše analýzy boli údaje poskytnuté Národným centrom zdravotníckych informácií z Ročného výkazu J (MZ SR) 1-01 o jednodňovej starostlivosti za roky 2009 až 2011 s počtami pacientov, ktorým bol uskutočnený výkon daného typu podľa kódu číselníka výkonov JZS z Vestníka Ministerstva zdravotníctva SR zo dňa 1.3.2006, čiastka 9-16, časť 23 – „Odborné usmernenie MZ SR o výkonoch jednodňovej zdravotnej

starostlivosti“ (Gavurová et al., 2013). Podľa uvedeného usmernenia je sedem špecializačných odborov JZS (Chirurgia, ortopédia, úrazová chirurgia a plastická chirurgia (ďalej „Chirurgia“), Gynekológia a pôrodníctvo (ďalej „Gynekológia“), Oftalmológia, Otorinolaryngológia (ďalej „ORL“), Urológia, Zubné lekárstvo a Gastroenterologická chirurgia a Gastroenterológia). Výkony JZS sa za posledné dva odbory vykazujú v minimálnej miere a preto sme ich nezahrnuli do ďalších analýz. Predpokladáme pritom, že zastúpenie jednotlivých typov výkonov JZS v každom odbore je zhruba rovnaké pre kraje.

Na analýzu vzťahu podielu počtu hospitalizovaných pacientov juniorov JZS a kraja sme použili korešpondenčnú analýzu implementovanú v štatistickom systéme SPSS verzia 19. Korešpondenčná analýza je exploračná metóda pre analýzu vzťahu riadkových a stĺpcových podielov kontingenčných tabuliek. Najjednoduchší prístup k jej pochopeniu je považovať ju za analýzu hlavných komponentov kategorických dát (Jobson 1991).

3. Výsledky korešpondenčnej analýzy vzťahu počtu hospitalizovaných juniorov JZS a kraja

Máme k dispozícii tabuľku počtu hospitalizovaných juniorov po krajoch (riadky) a špecializačných odboroch JZS (stĺpce) za roky 2009 – 2011, ktorá obsahuje aj riadkové a stĺpcové úhrny. V porovnaní so skupinou nehospitalizovaných juniorov je situácia zložitejšia pre nulové počty vo viacerých políčkach kontingenčnej tabuľky. Algoritmicky sa problém rieši tak, že miesto nuly sa dosadí vhodné malé kladné číslo. Najčastejší výskyt nulových početností je v odbore Urológia, Oftalmológia a Gynekológia. Za kraje je to Nitriansky a Žilinský kraj. Cieľom korešpondenčnej analýzy je zistiť, ktoré riadky a stĺpce sú navzájom podobné z hľadiska ich štruktúry (podielov). V ďalšom kroku urobíme tabuľky riadkových (tabuľka 2) a stĺpcových (tabuľka 3) podielov.

Tab. 1: Kraj verzus odbor JZS – počty hospitalizovaných juniorov

Kraj / Odbor	Chir	Gyn	Oftal	ORL	Urol	Total
B. Bystrica	6	0	2	1 005	0	1 013
Bratislava	5	2	6	213	0	226
Košice	73	98	25	173	0	369
Nitra	61	0	0	0	0	61
Prešov	268	0	0	91	0	359
Trnava	50	9	0	28	3	90
Trenčín	68	0	0	80	0	148
Žilina	68	0	0	6	0	74
Total	599	109	33	1 596	3	2 340

Tab. 2: Kraj verzus odbor JZS – riadkové podiely hospitalizovaných juniorov

Kraj / Odbor	Chir	Gyn	Oftal	ORL	Urol	Total
B. Bystrica	,006	,000	,002	**,992**	,000	1,000
Bratislava	,022	,009	,027	**,942**	,000	1,000
Košice	,198	**,266**	,068	**,469**	,000	1,000
Nitra	**1,000**	,000	,000	,000	,000	1,000
Prešov	**,747**	,000	,000	**,253**	,000	1,000
Trnava	**,556**	,100	,000	**,311**	,033	1,000
Trenčín	**,459**	,000	,000	**,541**	,000	1,000
Žilina	**,919**	,000	,000	,081	,000	1,000
Mass	**,256**	,047	,014	**,682**	,001	

V tabuľke 2 sú riadkové profily (podiely v percentách) v 5 stĺpcoch. Priemerný podiel hospitalizovaných juniorov v odboroch za kraj je 0,2. Korešpondenčná analýza skúma rozdiely medzi jednotlivými riadkovými profilmi a priemerným riadkovým profilom (Mass) v priestore menšej dimenzie.

Prevládajúce podiely hospitalizovaných juniorov za odbory v jednotlivých krajoch sú:

B. Bystrica a Bratislava má prakticky všetko v odbore ORL;

Košice ako jediný kraj majú rozložené podiely medzi Chirurgiu, Gynekológiu a ORL;

Nitra (Žilina) má (takmer) všetky hospitalizácie v odbore Chirurgia;

Prešov, Trnava a Trenčín majú najväčšie podiely v odboroch Chirurgia a ORL.

Priemerný kraj (riadok Mass) má najväčšie podiely v odboroch Chirurgia a ORL.

Pre pohľad z druhej strany je vhodné pozrieť sa aj na stĺpcové podiely (tabuľka 3).

Tab. 3: Kraj verzus odbor JZS – stĺpcové podiely hospitalizovaných juniorov

Kraj / Odbor	Chir	Gyn	Oftal	ORL	Urol	Mass
B. Bystrica	,010	,000	,061	**,630**	,000	**,433**
Bratislava	,008	,018	**,182**	**,133**	,000	,097
Košice	,122	**,899**	**,758**	,108	,000	**,158**
Nitra	,102	,000	,000	,000	,000	,026
Prešov	**,447**	,000	,000	,057	,000	**,153**
Trnava	,083	,083	,000	,018	**1,000**	,038
Trenčín	,114	,000	,000	,050	,000	,063
Žilina	,114	,000	,000	,004	,000	,032
Total	1,000	1,000	1,000	1,000	1,000	

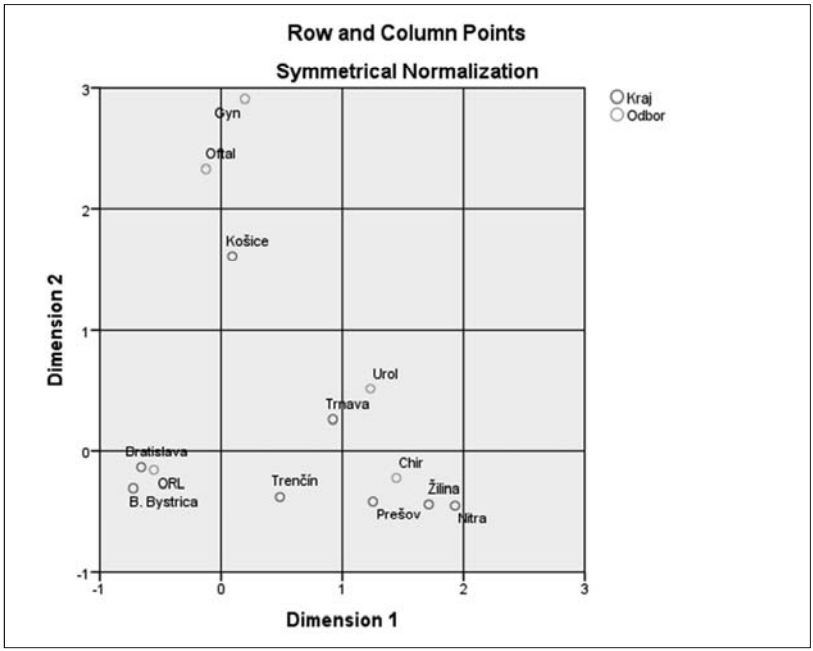
Korešpondenčná analýza v tomto prípade skúma rozdiely medzi jednotlivými stĺpcovými profilmi a priemerným stĺpcovým profilom (Mass). Priemerný podiel hospitalizovaných juniorov po krajoch za odbor je 0,125.

Nadpriemerné podiely krajov v jednotlivých odboroch sú:

Chirurgia – Prešov; Gynekológia – Košice; Oftalmológia – Bratislava a Košice; ORL – B. Bystrica a Bratislava; Urológia – Trnava. Priemerný odbor (stĺpec Mass) má najväčšie podiely v krajoch B. Bystrica, Košice a Prešov.

Ďalšou časťou výstupu je tabuľka aproximácie rozkladu kontingenčnej tabuľky na singulárne čísla (pre ušetrenie miesta ju neuvádzame). Prvý rozmer vysvetľuje 66,1 %. Druhý rozmer 29,9 %. Spolu je to 95,9 %, čo je ešte lepšie ako v prípade nehospitalizovaných juniorov. Poslednou textovou časťou výstupu sú tabuľky kvality zobrazenia riadkov (kraje) a stĺpcov (odbory) (aj tie pre ušetrenie miesta neuvádzame).

Hlavným grafickým výstupom korešpondenčnej analýzy je korešpondenčný biplot pri súčasnom zobrazení riadkov a stĺpcov kontingenčnej tabuľky.



Obr. 1: Korešpondenčný biplot krajov a odborov hospitalizovaných juniorov JZS

Na korešpondenčnom biplote (obrázok 1) sú zobrazené kraje aj odbory z pohľadu ich podielu (štruktúry) hospitalizovaných juniorov. Pri jeho interpretácii musíme zobrať do úvahy, že je na ňom nepresne zobrazený kraj Trnavský a odbor Urológia. Z krajov sú najbližšie k sebe dvojice B. Bystrica a Bratislava, Žilina a Nitra. Kraje B. Bystrica a Bratislava sú aj blízko odboru ORL s najväčším podielom hospitalizovaných juniorov. Kraje Žilina, Nitra a Prešov kraj sú zas blízko odboru Chirurgia s veľkým podielom hospitalizovaných. Najbližšie odbory v podiele hospitalizovaných sú Gynekológia a Oftalmológia. Kraj Košický je blízko odborov Oftalmológia a Gynekológia.

4. Záver

Prínosom by bola aj hlbšia analýza systému JZS na Slovensku zameraná na aspekty finančné, organizačné, spokojnosti pacientov (osobný prístup personálu, pohodlie, informovanosť), aspekty rizikovosti vyplývajúce z prostredia realizácie výkonov (napr. nozokomiálne nákazy), ako aj rizikovosti vyplývajúcej z predispozícií pacienta, komorbidít, miesta realizácie zdravotníckeho výkonu a pod. Opodstatnenosť riešenia tejto problematiky odôvodňuje aj fakt, že na Slovensku absentujú výskumné štúdie, ktoré by sa zameriavali na rozvoj JZS, jej efektívnosť, rizikovosť, cenové stratégie ZS, ako aj funkčnosť. Bez týchto analýz nie je možné odhaľovať rezervy v zvyšovaní efektívnosti zdravotníckeho systému, hľadanie možností úspor, efektívnu alokáciu zdrojov, ako aj zabezpečenie spokojnosti všetkých aktérov systému zdravotníctva. Analýza systému JZS poskytuje cenné východisko aj k realizácii strategického benchmarkingu.

Literatúra

GAVUROVÁ, B. – ŠOLTÉS, V. – KAFKOVÁ, K. – ČERNÝ, Ľ. 2013. Vybrané aspekty efektívnosti slovenského zdravotníctva. Jednodňová zdravotná starostlivosť a jej rozvoj v podmienkach Slovenskej republiky. Košice: Technická univerzita, 2013. 275 s. ISBN 978-80-553-1438-9.

JOBSON, J.D.:Applied Multivariate Data Analysis. Vol. II: Categorical and Multivariate Methods. New York: Springer - Verlag, 1992. 731 p. ISBN 0387978046

KORÓNY, S. – GAVUROVÁ, B.: Analýza vzťahu podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od kraja. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 93 - 98. ISSN 1336-7420

KORÓNY, S. – GAVUROVÁ, B.: Analýza vzťahu podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti od odboru. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 99 - 104. ISSN 1336-7420

KORÓNY, S. – GAVUROVÁ, B.: Analýza vývoja podielu hospitalizovaných detských pacientov jednodňovej zdravotnej starostlivosti. In: Forum Statisticum Slovacum. Roč.9, č. 6, 2013, s. 105 -111. ISSN 1336-7420

Príspevok uvádza predbežné výsledky výskumu v súlade s podporeným projektom VEGA č. 1/1050/12 „Návrh systému merania výkonnosti v zdravotníckych zariadeniach na Slovensku a implementácia metrík výkonnosti.“

Adresy autorov:

Beáta Gavurová, doc., Ing., PhD., MBA.
Ekonomická fakulta
Technická univerzita v Košiciach
Němcovej 32
040 01 Košice
Email: beata.gavurova@tuke.sk

Samuel Koróny, RNDr., PhD.
Inštitút ekonomických vied
Ekonomická fakulta UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
Email: samuel.korony@umb.sk

Koncentrácia tržieb v divízii Počítačové programovanie v roku 2010[1]
Concentration turnover in the division Computer Programming in 2010

Jozef Chajdiak

Abstract: The paper is based on data from 2010 analyzed the economic results of division Computer Programming. The above table is selected aggregates volume of economic indicators, table of selected financial ratios and the example of individual data on earnings of companies in the graph is the Lorenz curve and Gini coefficient is calculated concentrations. That is the procedure for calculating the coefficients in

Abstrakt: V príspevku sa na základe údajov z roku 2010 analyzujú ekonomické výsledky divízie Počítačové programovavie. Uvedená je tabuľka úhrnov výbraných objemových ekonomických ukazovateľov, tabuľka vybraných pomerových ukazovateľov a na príklade individuálnych údajov firiem o tržbách je uvedený graf Lorenzovej krivky a je vypočítaný Giniho koeficient koncentrácie. Uvedený je aj postup výpočtu koeficienta v Exceli. Výsledky ukazujú na veľmi vysokú koncentráciu tržieb v divízii Počítačové programovanie.

Key words: Turnover, values of economic indicators, Gini coefficient of concentration, Lorenz curve.

Kľúčové slová: tržby, hodnoty ekonomických ukazovateľov, Giniho koeficient koncentrácie, Lorenzová krivka.

JEL classification: C00

1. Úvod

Inovácie sú dôležitou súčasťou procesu výroby. V jednej z neformálnych diskusií na tému možností zvyšovania efektívnosti prihraničnej spolupráce malých a stredných podnikov SR sa konštatovalo, že výstupy produkcie firiem počítačového programovania sú nositeľmi inovácií. Výsledkom bolo riešenie otázky možností zapojenia podnikov divízie 62 SK NACE Počítačové programovanie do tejto úlohy. Účastníkov zaujímala ekonomická situácia podnikov v tejto divízii tak v oblasti veľkosti firiem ako aj ich stupňa koncentrácie meranej Giniho koeficientom koncentrácie a tiež úroveň hodnôt vybraných relatívnych ukazovateľov.

2. Popis súboru údajov

Anonymizované údaje sú za rok 2010 v eurách, z výkazu Súvaha a Výkaz ziskov a strát, za firmy účtujúce v sústave podvojného účtovníctva, ktoré spolu s daňovým priznaním odovzdali aj uvedené dva účtovné výkazy. Do analýzy boli zahrnuté len údaje s nenulovou veľkosťou osobných nákladov. Potrebnú podmnožinu údajov za firmy z divízie 62 SK NACE Počítačové programovanie autorovi poskytol SCB-Slovak Credit Bureau, s.r.o

3. Hodnoty ekonomických ukazovateľov

V tab. 1 sú úhrny objemových ekonomických ukazovateľov (v eurách) a v tab.2 sú uvedené hodnoty pomerových ukazovateľov.

[1] Príspevok bol spracovaný v rámci riešenia úlohy VEGA č. 1/1164/12 "Možnosti uplatnenia informačných a komunikačných technológií na zvyšovanie efektívnosti medzinárodnej spolupráce malých a stredných podnikov SR v oblasti inovácií" a úlohy VEGA č. 1/0335/13: "Štatistická analýza vybraných ukazovateľov konkurencieschopnosti na súbore podvojne účtujúcich podnikov SR.

.

Tab. 2: Rok 2010, Divízia 62SK NACE Počítačové programovanie – hodnoty ukazovateľov

Ukazovateľ	hodnota	Ukazovateľ	hodnota
rozsah (n)	1 573	zisk (po zdanení) (Zpo)	203 195 380
majetok (MAJ)	899 999 920	zisk (pred zdanením) Zpred	161 992 417
neobežný majetok (NM)	185 684 492	pridaná hodnota (PH)	608 934 548
obežný majetok (OM)	678 054 502	tržby (01) (Q01)	261 050 783
zásoby (ZAS)	34 227 390	tržby (05) (Q05)	1 148 004 325
krátkodobé pohľadávky (KP)	24 482 752	výrobná spotreba (VS)	598 030 209
finančné účty (FU)	372 679 167	spotreba materiálu (MAT)	49 122 930
vlastné imanie (VI)	246 665 194	služby (SLUZ)	548 907 279
základné imanie (ZI)	388 388 812	osobné náklady (ON)	362 046 489
krátkodobé záväzky (KZ)	60 141 529	odpisy (18+20) (ODP)	43 692 679
krátkodobá finančná výpomoc (KFV)	380 672 259	spotreba viazaných produktívnych faktorov SVPF	405 739 168
Bankové úvery (BU)	5 713 629		

Tab. 2: Rok 2010, Divízia 62SK NACE Počítačové programovanie – hodnoty ukazovateľov

Ukazovateľ	Ukazovateľ	hodnota
(Q01+Q05)/ON	Finančná produktivita práce meraná tržbami	3,89
PH/ON	Finančná produktivita práce meraná pridanou hodnotou	1,69
PH/SVPF	Účinnosť SVPF meraná pridanou hodnotou	1,50
(Q01+Q05)/SVPF	Účinnosť SVPF meraná tržbami	3,47
100*Zpo/VI	Rentabilita vlastného imania	41,7
100*PH/(Q01+Q05)	Podiel pridanej hodnoty na tržbách	43,2
100*Z/(Q01+Q05)	Ziskovosť tržieb	14,4
100*ON/(Q01+Q05)	Podiel osobných nákladov na tržbách	25,7
100*ODP/(Q01+Q05)	Podiel odpisov na tržbách	3,1
100*Z/PH	Rentabilita pridanej hodnoty	33,4
100*VI/MAJ	Miera samofinancovania	43,15

Údaje v tab. 1 a tab.2 dávajú čitateľovi konkrétnu numerickú predstavu o úrovni hospodárenia v divízii Počítačové programovanie.

4. Koncentrácia podľa tržieb

Koncentrácia a centralizácia je tradičnou témou ekonomických analýz. Na jednej strane vyššia koncentrácia znamená väčšie možnosti riešenia úloh, na druhej strane je to tendencia k monopolizácii s negatívnymi dôsledkami monopolizmu.

Koncentráciu meriame Giniho koeficientom koncentrácie GKK:

$GKK = 1 - 2\sum_{i=1}^{n} p_i \frac{(kpx_{i-1} + kpx_i)}{2}$, pričom $kpx_0=0$, $kpx_i = \sum_{j=1}^{i} px_j$ a $px_j = \frac{x_j}{\sum_{i=1}^{n} x_i}$.

Postup výpočtu GKK v Exceli:

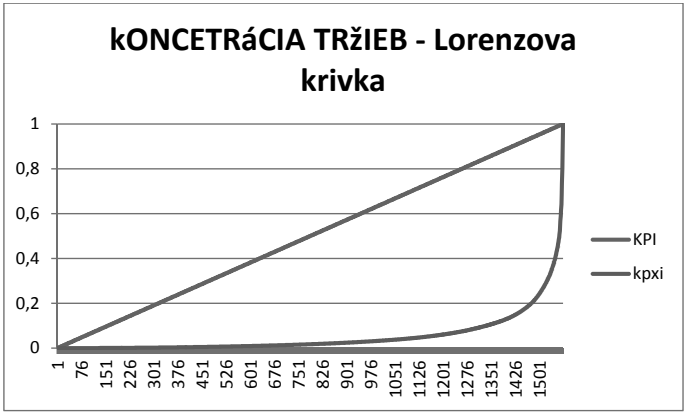
Koncentráciu v súbore jednotiek rozsahu n jednotiek hodnotíme podľa meritórnej vlastnosti X (tu tržby). Východiskové údaje uložíme na čistý hárok.

a) Do políčka B1 uložíme názov premennej „X“.
b) Do bloku B2.Bn+1 uložíme hodnoty premennej X (i=1,2,...,n).
c) Hodnoty x1 až xn v políčku B1 (názov) a B2 až Bn+1 usporiadame podľa veľkosti od najmenšej po najväčšiu.
d) Do políčka Bn+3uložíme súčet hodnôt z B2 po Bn+1.
e) Do políčka A1 uložíme text „I“ a do políčok A2 až An+1 postupnosť prirodzených čísiel 1, 2, ...,n.
f) Medzi 1 a 2. riadok vložíme nový prázdny riadok.
g) Do políčka C1 uložíme „ pxi“ a do C3 až Cn+2 uložíme pxi relatívne početnosti premennej X ($px_j = \frac{x_j}{\sum_{i=1}^{n} x_i}$).
h) Do políčka D1 uložíme „kpx“ a do D3 až Dn+2 zložíme kpxi kumulatívne relatívne početnosti premennej X $kpx_i = \sum_{j=1}^{i} px_j$.
i) Do políčkaE1 uložíme „Gkpxi“ , do políčka E2 uložíme 0 a do E3 až En+2 p(kpxi-1 + kpxi).
j) Do políčka En+2+1 vložíme súčet E2 až En+1.
k) V políčku En+2+2 uložíme text „GKK=“ a v En+2+2 vložíme príkaz 1-En+2+1

Výsledná hodnota Giniho koeficienta koncentrácie pre tržby v roku 2010 v divízii počítačové programovanie je 0,90, čo svedčí o veľmi vysokej koncentrácii podľa tržieb (veľa firiem ma malé tržby a niekoľko firiem veľmi vysoké tržby). Dokumentuje to aj obr. 1, na ktorom je zobrazená Lorenzova krivka (čím sa viac blíži krivka k uhlopriečke, tým je koncentrácia nižšia a naopak, čím sa viac blíži k x-ovej os, tým je koncentrácia väčšia).

5. Záver

Z pohľadu tržieb môžeme v divízii 62 SK NACE Počítačové programovanie konštatovať vysokú úroveň koncentrácie (prakticky to znamená monopolné postavenie časti firiem v súbore firiem Počítačového programovania). Z toho vyplýva dominantné postavenie veľkých firiem v organizácii úloh cezhraničnej spolupráce a servisná činnosť malých firiem pre veľké firmy.



Obr. 19: Lorenzova krivka pre tržby v divízii Počítačové programovanie v roku 2010

Literatúra

CHAJDIAK, J. 2013. *Štatistika jednoducho v Exceli*. Bratislava Statis.

Adresa autora (-ov):

Jozef Chajdiak, Doc., Ing. CSc.
Ústav manažmentu STU
Vazovova 5, Bratislava
chajdiak@statis.biz

Vybrané faktory predlženosti podnikov v podmienkach SR
Factors of over-indebtedness: The case of Slovakia

Štefan Kováč

Abstract: This paper deals with the analysis of over-indebtedness of Slovak enterprises using quantile regression. It verifies the assumption that the various degrees of over-indebtedness of enterprises is in relation to various factors, also relation of localization and industry in which it operates with condition of over-indebtedness.

Abstrakt: Príspevok sa venuje analýze predlženosti slovenských podnikov pomocou kvantilovej regresie. Overuje predpoklady, že s rôznymi stupňami predlženosti podnikov majú súvis rôzne mikrofaktory. Venuje sa tiež analýze dosahu vybraného odvetvia na stav predlženosti podnikov a významnosti faktora kraj v skúmaní.

Key words: quantile regression, over-indebtedness, high-risk enterprise

Kľúčové slová: kvantilová regresia, predlženie, rizikový podnik

JEL classification: C51, L25

1. Úvod

Fenomén predlženosti podnikateľských subjektov je nemenej dôležitý ako platobná neschopnosť. Toto tvrdenie je okrem iného podložené aj predpisom č. 7/2005 Z.z. - Zákon o konkurze a reštrukturalizácii a o zmene a doplnení neskorších zákonov. Na základe tohto zákona ma každý podnik, ktorý je v úpadku, povinnosť podať návrh na vyhlásenie konkurzu do 30 dní od zistenia tohto stavu. Podnik je v úpadku, ak je platobne neschopný alebo predlžený.

Stav platobnej neschopnosti je možné skúmať na mikroúrovni každého podniku individuálne. Nie je možné ho vyčítať z verejne dostupných údajov - z účtovných závierok (Zalai, 2008). Stav predlženia je možné zistiť aj zo statických výkazov ako sú účtovné závierky spoločností a táto skutočnosť ponúka možnosť analyzovať stav predlženosti podnikov v rámci celej krajiny, regionálne a odvetvovo porovnať podniky z pohľadu stavu predlženosti a tiež nájsť mikrofaktory pôsobiace na stav predlženosti (Kováč, 2013).

Odsek (3) §3 Zákona o konkurze a reštrukturalizácii hovorí, že *predlžený je ten, kto je povinný viesť účtovníctvo podľa osobitného predpisu, má viac ako jedného veriteľa a hodnota jeho záväzkov presahuje hodnotu jeho majetku.* Zákon nestanovuje pásma predlženosti a podniky silno aj slabo predlžené zaraďuje do jednej kategórie. Stav predlženosti sa určuje z rozdielu majetku a záväzkov spoločnosti, pričom záporný výsledok ho potvrdzuje. V praxi je z pohľadu zdravia podniku rozdiel medzi podnikmi slabo a silno predlženými.

Z uvedených informácií ako aj so skutkovej podstaty stavu predlženia sa dá vyvodiť súvis medzi týmto stavom a zvýšeným rizikom podniku, teda z pohľadu odbornej literatúry sa nejedná o „zdravý podnik“ (Kováč, 2013; Zalai, 2008). Táto skutočnosť a unikátnosť dostupných údajov[1] poskytuje možnosť odlišného a celoplošného pohľadu na skúmanú problematiku.

V príspevku sa budeme venovať overeniu nasledovných hypotéz:

H_1: Rast/pokles predlženosti podnikov na rôznych úrovniach má súvis s rôznymi mikrofaktormi.

[1] V príspevku pracujeme s anonymnými individuálnymi údajmi z účtovných závierok podnikateľských subjektov účtujúcich v podvojnom účtovníctve za rok 2010. Spolu ide o 140 493 účtovných závierok. Údaje boli spracované v spolupráci s Finančnou správou Slovenskej republiky.

H_2: Stav predlženosti podnikov súvisí s ich lokalizáciu v rámci krajiny.

H_3: Stav predlženosti podnikov súvisí s ich odvetvovou klasifikáciou.

2. Model a použité metódy

Pri návrhu modelu považujeme za východiskový údaj pre vysvetľovanú premennú rozdiel majetku a záväzkov, pričom záporná hodnota ukazovateľa vypovedá o stave predlženia firmy. Z dôvodu vysokej variability premennej sme sa rozhodli ju normovať nasledovným spôsobom aj napriek použitiu metódy kvantilovej regresie:

- *OI = (majetok spolu - záväzky) / záväzky*

Týmto podielom sme dosiahli nižšiu variabilitu vysvetľovanej premennej pri zachovaní ostatných dôležitých charakteristík. Pri voľbe vysvetľujúcich premenných sme sa inšpirovali najčastejšie používanými pomerovými ukazovateľmi hodnotenia bonity firiem a doterajšími zisteniami v tejto oblasti (Jones, 1987; Kováč, 2012; Nunes et al., 2010):

- *WCTA = finančné účty/ majetok spolu*
- *EBIT = (VH po zdanení + daň z BČ a MČ + nákladové úroky)/ majetok spolu*
- *TDTA = cudzie zdroje/ pasíva*
- *NINW = VH po zdanení/ vlastné imanie*
- *IT = zásoby/ (tržby z predaja tovaru a služieb)*360*
- *QR = finančné účty/ krátkodobé záväzky*
- *STA = tržby z predaja tovaru + výroba/ majetok spolu*
- *STRCA = obežný majetok/ majetok spolu*
- *STRSE = tržby z predaja tovaru + výroba/ náklady na tovar + výrobná spotreba*

Pri testovaní multikolinearity sme neodhalili žiadnu vzájomnú koreláciu medzi zvolenými premennými. V špecifikácii modelu sa vychádza z charakteru dostupných údajov z individuálnych účtovných závierok slovenských podnikov. V príspevku sú použité individuálne údaje z výkazov účtovných závierok podnikateľov účtujúcich v sústave podvojného účtovníctva za účtovné obdobie 2010. Podniky sú rozdelené podľa jednotlivých sekcií štatistickej klasifikácie ekonomických činností SK-NACE a krajov (teritoriálna úroveň NUTS 3), v ktorých majú sídlo. Pri aplikácii údajov na model sme vykonali transformáciu odvetvovej klasifikácie zlúčením niektorých sekcií do spoločných skupín na základe spoločných charakteristík týchto odvetví, a to nasledovným spôsobom:

- sekcie A B D E do skupiny A_E
- sekcie H I do skupiny HI
- sekcie J K L M do skupiny J_M
- sekcie N O P Q R S T U do skupiny N_U
- sekcie C F G sme ponechali samostatne.

Vzhľadom na skutočnosť, že vysvetľovaná premenná vykazuje vysoký stupeň variability, na identifikovanie mikrofaktorov predlženosti nie je vhodné použiť klasickú regresiu. Ako alternatíva sa odporúča kvantilová regresia (Koenker & Hallock, 2011; Koenker & Basser 1978; Kováč & Želinský, 2013; Nunes et al., 2010). Pre náhodnú premennú Y s distribučnou funkciou $F(y) = P(Y \leq y)$ je kvantil τ premennej Y definovaný ako inverzná funkcia $Q(\tau) = \inf\{y : F(y) \geq \tau\}$, pričom $0 < \tau < 1$. Predpokladajme, že kvantil τ podmieneného rozdelenia závislej premenej (Y_i) je lineárnou funkciou vektora nezávislých premenných (X_i). Kvantilovú podmienenú regresiu môžeme potom zapísať:

$$y_i = \alpha + \boldsymbol{\beta}_\tau \mathbf{x}_{i\cdot} + z_{\tau i} \tag{1}$$

a

$$Q_\tau(y_i \mid \mathbf{x}_{i\cdot}) \equiv \inf\{y_i : F_i(y_i \mid \mathbf{x}_{i\cdot}) \geq \tau\} = \alpha + \boldsymbol{\beta}_\tau \mathbf{x}_{i\cdot} \tag{2}$$

s nasledovným obmedzením:

$$Q_\tau(z_{\tau i} \mid \mathbf{x}_{i\cdot}) = 0 \tag{3}$$

kde:

y_i je *i*-tá zložka vektora závislej premennej **y**,

$\mathbf{x}_{i\cdot}$ je *i*-tý riadok matice **X** nezávislých premenných,

$\alpha, \boldsymbol{\beta}_\tau$ sú odhadované parametre pre rôzne hodnoty $\tau \in (0;1)$,

$z_{\square i}$ je náhodná zložka,

n je počet pozorovaní,

k je počet nezávislých premenných.

V príspevku sa zameriavame na predlženosť podnikov a jej determinanty pre 10-ty, 25-ty, 50-ty, 75-ty a 90-ty percentil rozdelenia predlženosti slovenských podnikov. Na diagnostiku odhadnutého modelu kvantilovej regresie sme použili mieru R^1 (analógia tradičného koeficientu determinácie) a zodpovedajúci test vierohodnostného pomeru. Obe techniky boli navrhnuté Koenkerom a Machadom (1999). Koeficient R^1 je lokálnou mierou vhodnosti modelu na určitom kvantile. Úplný model (t. j. s vysvetľujúcimi premennými) je lepší na kvantile τ ako model bez vysvetľujúcich premenných (teda len s interceptom), ak je τ-tá podmienená kvantilová funkcia významne presiahnutá vplyvom vysvetľujúcich premenných. Obe diagnostické techniky sú založené na hodnotách objektívnych funkcií. Koenker a Machado (1999) analyzovali správanie uvedených mier s použitím simulovaných údajov. Dospeli k záverom, že R^1 nevykazuje žiaduce správanie (má príliš malé hodnoty), ak je na príslušnom kvantile príliš vysoká variabilita.

Odhad modelu kvantilovej regresie bol uskutočnený v softvéri SAS použitím procedúry quantreg.

3. Výsledky a diskusia

V spracovanom prehľade uvádzame hodnotu regresných koeficientov jednotlivých premenných a faktorov vplývajúcich na vysvetľovanú premennú podľa uvedených kvantilov. Prvý kvantil (0,1) zahŕňa podniky silno predlžené, druhý kvantil označuje podniky stredne až slabo predlžené a podniky v tzv. šedej zóne (hodnota predlženosti blízka 0 z hora aj z dola). Tretí kvantil (0,5) označuje podniky, ktoré sú menej rizikové a nedosahujú stav predlženia. Kvantily (0,75) a (0,9) predstavujú silno zdravé podniky z pohľadu stavu majetku a záväzkov spoločnosti.

Analýza dosahu vybraných kvantitatívnych premenných na stav predlženia

Výstupom tejto analýzy je overenie predpokladu rôznorodosti súvisu jednotlivých ukazovateľov s rôznymi stupňami predlženosti podnikov. Ukazovateľ, ktorý potvrdil svoju významnosť v takmer celom spektre vysvetľovanej premennej bol pomer WCTA. Jeho významnosť však pri nepredlžených podnikoch klesala. Premenné TDTA, QR a STA preukázali svoju významnosť pri rizikových podnikoch, teda podnikoch, ktorých hodnoty predlženosti sa nachádzali v prvých dvoch až troch kvantiloch. Naopak, ukazovateľ STRSE

preukázal svoju významnosť pri podnikoch patriacich do posledných dvoch kvantilov avšak s minimálnym vplyvom.

Parameter	kvantil	0,1	0,25	0,5	0,75	0,9
WCTA	*Odhad*	-0,0071	0,0884	-1,0379	-0,4985	-0,0159
	P-hodnota	<,0001	<,0001	<,0001	<,0001	0,1376
EBIT	*Odhad*	0,0000	0,0000	0,0000	0,0000	0,0000
	P-hodnota	<,0001	<,0001	<,0001	<,0001	0,0141
TDTA	*Odhad*	-0,0001	-0,0001	0,0000	0,0000	0,0000
	P-hodnota	<,0001	<,0001	0,3504	0,4824	0,7142
NINW	*Odhad*	0,0000	0,0000	0,0000	0,0000	0,0000
	P-hodnota	0,6720	0,6308	0,8187	0,8673	0,8367
IT	*Odhad*	0,0006	0,0001	-0,0002	-0,0006	-0,0017
	P-hodnota	0,6731	0,8487	0,6916	0,7674	0,8345
QR	*Odhad*	0,0369	0,0446	0,9997	1,0182	1,1836
	P-hodnota	<,0001	<,0001	<,0001	<,0001	<,0001
STA	*Odhad*	-0,0011	-0,0006	0,0000	-0,0001	-0,0001
	P-hodnota	<,0001	<,0001	0,0005	0,0134	0,6823
STRCA	*Odhad*	0,0017	-0,0238	0,1408	0,1294	0,0003
	P-hodnota	0,1797	<,0001	<,0001	<,0001	0,9715
STRSE	Odhad	0,0000	0,0000	0,0000	0,0001	0,0003
	P-hodnota	0,4058	0,8570	0,5444	0,0002	0,0002
	Pseudo R	0,0922	0,0916	0,1032	0,1162	0,1228
	LR test	<,0001	<,0001	<,0001	<,0001	<,0001

Tab. 1: Výstup z testovania kvantitatívnych premenných

Premenné IT, NINW neprejavili žiadnu významnosť. Zaujímavým je aj ukazovateľ STRCA, ktorého významnosť sa v prvom a poslednom kvantile nepreukázala. Teda jeho hodnota nemá súvis so silne predlženými resp. silne nepredlžené podnikmi. Avšak existuje súvis s podnikmi, ktoré nadobúdajú mierne hodnoty predlženia resp. nepredlženia v porovnaní s ostatnými podnikateľskými subjektmi (podniky v tzv. šedej zóne). Ukazovateľ EBIT vykázal významnosť vo všetkých kvantiloch, no nadobúdnuté hodnoty regresných koeficientov boli veľmi blízke nulovej hodnote.

Analýza regionálneho dosahu vo vybraných regiónoch

V regionálnom porovnávaní podnikov bol ako referenčný kraj použitý Žilinský kraj. Z analýzy regresných koeficientov v jednotlivých kvantiloch možno pozorovať niekoľko zaujímavých zistení. Košický kraj v porovnaní s referenčným krajom dosahuje v poslednom kvantile (kam patria silne zdravé a nepredlžené podniky) kladnú hodnotu regresného koeficientu, čo znamená, že v porovnaní so Žilinským krajom pozícia bezrizikových podnikov v Košickom kraji v rámci Slovenska silnejšia ako v Žilinskom. Naopak, záporné hodnoty regresných koeficientov pri vysoko rizikových podnikoch napovedajú slabšiu pozíciu predlžených podnikov. Regionálne rozdiely je taktiež možné badať v Nitrianskom a Trenčianskom kraji. Zatiaľ čo pri podnikoch, ktoré sú z pohľadu rizikovosti a predlženosti zahrnuté v prvých dvoch kvantiloch majú tieto kraje hodnoty regresných koeficientov kladné a teda je možné predpokladať silnejšie postavenie predlžených podnikov v regionálnom porovnávaní oproti referenčnému kraju, podniky bezrizikové majú silnejšie postavenie práve v Žilinskom regióne. Štatistická významnosť regionálneho umiestnenia podniku v Trnavskom

kraji na základe zistených *p-hodnôt* nebola potvrdená. Tým pádom faktor umiestnenia podniku v tomto kraji model nemá význam ani v jednom kvantile. Pri ostatných krajoch je ich významnosť rozmanitá.

Parameter	kvantil	0,1	0,25	0,5	0,75	0,9
kraj BA	*Odhad*	-0,0843	-0,0570	-0,0476	-0,1437	-0,3099
	P-hodnota	<.0001	<.0001	<.0001	<.0001	0,0003
kraj BB	*Odhad*	-0,0073	0,0098	0,0012	-0,0464	-0,1857
	P-hodnota	0,6732	0,2474	0,8724	0,0639	0,0790
kraj KE	*Odhad*	-0,0678	-0,0093	-0,0241	-0,0604	0,0235
	P-hodnota	<.0001	0,2534	0,0006	0,0127	0,818
kraj NT	*Odhad*	0,0651	0,0278	0,0016	-0,0701	-0,2207
	P-hodnota	0,0001	0,0006	0,8244	0,0037	0,0302
kraj PO	*Odhad*	0,0189	0,0237	0,0052	-0,0069	0,0824
	P-hodnota	0,282	0,0053	0,4815	0,7841	0,4393
kraj TN	*Odhad*	0,0150	0,0140	-0,0023	-0,0070	-0,0277
	P-hodnota	0,4002	0,1047	0,7561	0,7842	0,7976
kraj TT	*Odhad*	0,0260	0,0192	0,0069	-0,0125	0,0060
	P-hodnota	0,1450	0,0262	0,3523	0,6263	0,9560

Tab.2: Výstup z testovania premenných „kraj“

Pôsobenie podniku v Bratislavskom regióne ovplyvňuje rizikovosť podniku vo všetkých kvantiloch.

Analýza dosahu vybraného odvetvia na stav predlženosti podnikov

Spojený sektor odvetví H a I vykazuje významnosť pri všetkých kvantiloch a zároveň jeho regresné koeficienty nadobúdajú záporné hodnoty. To znamená, že podniky pôsobiace v týchto odvetviach sú oproti referenčnému odvetviu (najrozšírenejšie odvetvie G) viac rizikové a majú tendenciu sa rýchlejšie ocitnúť v stave predlženia.

Parameter	kvantil	0,1	0,25	0,5	0,75	0,9
odv A_E	*Odhad*	0,2038	0,1261	0,1802	0,9368	2,5637
	P-hodnota	<.0001	<.0001	<.0001	<.0001	<.0001
odv C	*Odhad*	0,0188	0,0093	0,0608	0,1624	0,1309
	P-hodnota	0,2205	0,2104	<.0001	<.0001	0,1607
odv F	*Odhad*	0,0675	0,0235	-0,0013	-0,0761	-0,3408
	P-hodnota	<.0001	0,0013	0,8355	0,0004	0,0002
odv HI	*Odhad*	-0,2308	-0,2381	-0,0658	-0,1794	-0,4705
	P-hodnota	<.0001	<.0001	<.0001	<.0001	<.0001
odv J_M	*Odhad*	0,1518	0,0876	0,0847	0,2395	0,6111
	P-hodnota	<.0001	<.0001	<.0001	<.0001	<.0001
odv N_U	*Odhad*	0,0210	0,0618	0,0644	0,1392	0,1217
	P-hodnota	0,1394	<.0001	<.0001	<.0001	0,1585

Tab. 3: Výstup z testovania premenných „odvetvie“

Vplyv odvetvia stavebníctva na rizikovosť podnikov je významná v prvých dvoch a posledných dvoch kvantiloch. Zatiaľ čo pri vysokých hodnotách predlženosti sú hodnoty regresných koeficientov kladné, pri hodnotách označujúcich zdravé podniky sú koeficienty záporné. Túto skutočnosť môžeme interpretovať nasledovne: pri silne predlžených podnikoch

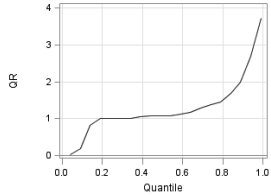
je riziko v odvetví stavebníctva oproti odvetviu maloobchodu nižšie a naopak - pri zdravých podnikoch je dominantnejšia pozícia maloobchodného a velkoobchodného odvetvia. Ostatné odvetvia dosahujú stabilne kladné hodnoty regresných koeficientov, čo vo všeobecnosti možno považovať za zistenie, ktoré hovorí o vyššej rizikovosti ostatných odvetví oproti vybranému referenčnému odvetviu G. Významnosť koeficientov je stabilná vo všetkých skúmaných odvetviach okrem odvetvia C a N až U. V týchto dvoch skupinách sa významnosť pri extrémnych hodnotách (prvý a posledný kvantil) nevyskytuje.

4. Záver

Z vykonanej analýzy podnikateľského prostredia z pohľadu predlženosti podnikov vyplýva niekoľko záverov.

Hypotéza H_1, ktorá predpokladala, že na firmy v rôznych stupňoch predlženosti majú vplyv iné faktory, sa potvrdila v prípade ukazovateľa STRCA, ktorý predstravuje pomer obežného majetku k celkovému majetku. Tento ukazovateľ neprejavil významnosť v extrémnych hodnotách predlženosti (prvý a posledný kvantil) teda jeho súvis s extrémnou predlženosťou resp. „nepredlženosťou“ (zdravou štruktúrou majetku a záväzkov) nebol preukázaný. V ostatných kvantiloch, v ktorých boli zaradené podniky slabo predlžené, resp. slabo nepredlžené (tzv. šedá zóna) svoju významnosť potvrdil. To znamená, že pri hodnotení „zdravia podniku“ má zmysel brať ohľad na tento ukazovateľ pri podnikoch s uvedenými hodnotami rozdielu majetku a záväzkov, pri extrémnych hodnotách svoju významnosť nepotvrdzuje.

V prípade mikrofaktora QR, ktorý v našom skúmaní predstavuje ukazovateľ pohotovej likvidity je potvrdená významnosť v celom rozpätí vysvetľovanej premennej. Zároveň je pozorovateľný rast hodnoty regresného koeficientu s rastom hodnoty rozdielu majetku a záväzkov spoločnosti. Vývoj hodnoty v jednotlivých kvantiloch je uvedený na Obr.1:



Obr.1: Zmena hodnoty reg. koeficientu pozdĺž kvantilov

Hypotéza H_2, ktorá predpokladala rozdiely medzi predlženými podnikmi v rôznych krajoch z pohľadu špecifikovaného modelu, sa potvrdila v prípade porovnania Košického a Žilinského kraja. „Zdravé“ podniky v Košickom kraji sú v porovnaní s podnikmi zaradenými do rovnakého kvantilu v lepšej kondícii (z pohľadu nami uvažovaných mikrofaktorov – ukazovateľov) ako podniky v Žilinskom kraji. Naopak, predlžené (rizikové) podniky v Košickom kraji sú na tom z pohľadu hodnotenia finančného zdravia podnikov horšie ako rovnako predlžené podniky v Žilinskom kraji.

Hypotéza H_3 bola potvrdená pre spojenú sekciu H a I odvetvovej klasifikácie ekonomických činností SK-NACE (Viď. interpretáciu v časti *Analýza dosahu vybraného odvetvia na stav predlženosti podnikov*).

Literatúra

JONES, F.: Current techniques in bankruptcy prediction. In: *Journal of Accounting Literature* (6) 1987: s. 131-164.

KOENKER, R., HALLOCK, K. Quantile Regression: An Introduction. *Journal of Economic*

Perspectives. Vol 15 (2011), 143–156.

KOENKER, R., BASSETT, G. 1978. Regression Quantiles. *Econometrica* 46 (1): 33-50.

KOENKER, R., MACHADO, J. A. F. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression.

KOVÁČ, Š.: Ekonometrická analýza ziskovosti podnikov SR. In: Chajdiak, J., Luha, J. (eds.): *Výpočtová štatistika 2012*: s. 59-69.

KOVÁČ, Š.: Aplikácia a komparácia predikčných modelov rizika bankrotu podnikov v podmienkach Slovenskej republiky. Diplomová práca. Ekonomická fakulta TUKE: Košice 2013. 146 s.

KOVÁČ, Š., ŽELINSKÝ, T.: Determinants of the Slovak Enterprises Profitability: Quantile Regression Approach In*: Statistika : Statistics and Economy Journal*. ISSN 1804-8765. Vol. 93, no. 3 (2013), p. 41-55

NUNES, P.M., SERRASQUERO, Z.S., LEITAO, J.: Are there nonlinear relationships between the profitability of Portuguese service SME and its specific determinants? In: *The Service Industries Journal.* Vol 30(5), 2010: s. 1312-1341.

Predpis č. 7/2005 Z.z. - Zákon o konkurze a reštrukturalizácii a o zmene a doplnení neskorších zákonov.

STANKOVIČOVÁ, I., VOJTKOVÁ, M.: *Viacrozmerné štatistické metódy s aplikáciami*. Bratislava: Iura Edition, 2007.

ZALAI, K.: *Finančno-ekonomická analýza podniku*. Bratislava: SPRINT, 2008.

Adresa autora:

Štefan Kováč, Ing.
Ekonomická fakulta, TU Košice
Němcovej 32, 040 01 Košice
stefan.kovac@tuke.sk

Segmentace států EU27 do čtyř skupin a dynamika segmentů
Segmentation of EU27 into four groups and their dynamics

Nikolay Kulbakov

Abstract: This paper is a continuation of the series of articles about the segmentation of EU27 countries by cluster analysis and economic indices. The paper contains the output from the cluster analysis based on the time series 2001-2012 years of macroeconomic indices for the EU27. The indicators that are applied have intense character, and it allows to compare the countries by productivity, standards of living, fiscal discipline and investor confidence.

Abstrakt: Článek je pokračováním seriálu článků o segmentaci států EU pomocí shlukové analýzy a ekonomických indikátorů. Příspěvek obsahuje komentář k výstupu shlukové analýzy provedené na základě časové řady 2001-2012 makroekonomických indikátorů pro země EU27. Použité indikátory převážně intensivního charakteru, které umožňují porovnání dle produktivity, životní úrovně, fiskální disciplíny a důvěry investorů.

Key words: cluster analysis, EU27, macroeconomic segmentation, MATLAB.

Kľúčové slová: shluková analýza, EU27, makroekonomická segmentace, MATLAB.

JEL classification: E01, C38

1. Úvod

Tento článek je pokračováním seriálu článků o segmentaci státu EU pomocí shlukové analýzy a makroekonomických indikátorů, viz také Kulbakov (2013). Primarním cílem tohoto výzkumu je, za využití metod shlukové analýzy popsaných v Rezanková, H., & Snásel, V. (2009), Löster, T. (2011) a dalších, sestavit adekvátní představu o ekonomické síle a pozici států v EU. Sekundárním cílem výzkumu je vývoj řady veřejně přístupných nástrojů v prostředí MATLAB, které budou pomáhat autorovi a všem zájemcům o podobnou problematiku provádět snadno a rychle shlukovou analýzu územních celků dle ekonomických indikátorů.

2. Použitá data

Pro účely analýzy byly použita data z databáze Eurostat pro 27 zemí Evropské Unii. K datu napsání článku do EU už bylo přijato i Chorvatsko, ale z důvodů chybějících dat, byla tato země vyřazena z analýzy. Celkem bylo použito devět makroekonomických indikátorů, které zachycují výkonnost ekonomik nehledě na velikost zemí a počty obyvatel a vhodných pro účel srovnávání. Výkonnost ekonomik a životní úroveň zastupují indikátory *HDP na obyvatele* (GDP per capita in PPS) jako index s průměrem 100 pro EU27 a *Roční čisté příjmy* (Annual net earnings, Single person without children, 50% of AW by PPS). Trh práce je zastoupen ukazateli *Míry ekonomické aktivity* (Activity rate 15-64) a *Míry zaměstnanosti* (Employment rate 15-64). Efektivita je zachycena indexy *Reálné produktivity práce na odpracovanou hodinu* (Real labour productivity per hour worked) s průměrem 100 pro 2005 rok a *Produktivity materiálů* (Resouce Productivity PPS per kg). Fiskální disciplína, mezinárodní obchodní pozice a důvěra investorů jsou zachyceny pomocí ukazatelů *Vládního deficitu k HDP* (General government deficit percentage of GDP), *Platební bilance k HDP* (Balance of Payments and International Investment Position items as share of GDP) a *Výnosů vládních dluhopisů* (EMU convergence criterion bond yields). Veškeré indexy obsahují data za státy EU27 a roky 2001-2012. Data vstupující do analýzy jsou normalizovaná dle charakterů dat a následujících vzorců.

Pokud větší hodnota je lepší byl použit vzorec: Iij = Xij - Xmini /Xmaxi –Xmini

Pokud menší hodnota je lepší byl použit vzorec: Iij =1 - (Xij - Xmini /Xmaxi -Xmini)

3. Shluková analýza

Podrobný popis algoritmu a odkaz na použitý programový kód MATLAB je v Kulbakov (2013). Stručně: výzkum je prováděn pomocí komparací dvou metod shlukové analýzy *hierarchické* a *k-mean*. Klasterní rozhodnutí je přijímáno na základě shody výsledků dvou metod a to minimálně 70% zemí má být zařazeno pomocí obou metod do stejných shluků. Konečné rozložení zemí po shlucích je přebíráno z výsledků *hierarchické analýzy*, protože tato metoda poskytuje na rozdíl od *k-mean* stabilní výstupy.

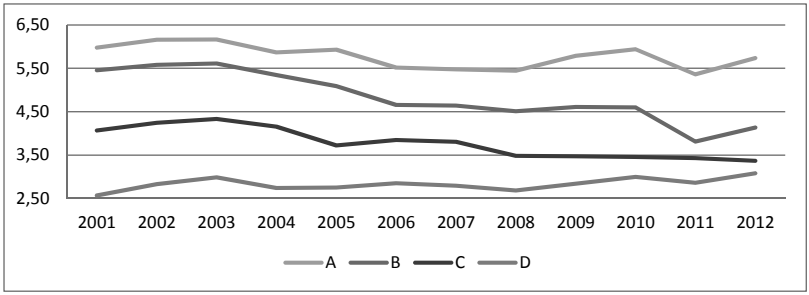
4. Výsledky

Ve výsledku analýzy autor dospěl k umístění 27 zemí EU do 4 shluků: A,B,C,D. Názvy shluků obsahují v sobě informaci. Hodnota A reprezentuje shluk, do kterého byly umístěny nejlepší státy z hlediska souhrnu normalizovaných indexů. Shluk D obsahuje nejslabší z hlediska sledovaných indikátorů státy pro příslušný rok. Distribuce do shluků se mění rok od roku a státy se pohybují mezi nimi. Výsledné roztřídění států se nachází v tabulce 1., kde vedle názvů států je časová řada ukazující na příslušný shluk v rocích 2001-2012, informace o tom jestli země do roku 2012 přijala euro a kreditní raiting S&P za květen 2012.

Tab. 3: Shlukové rozhodnutí pro EU27 a roky 2001-2012

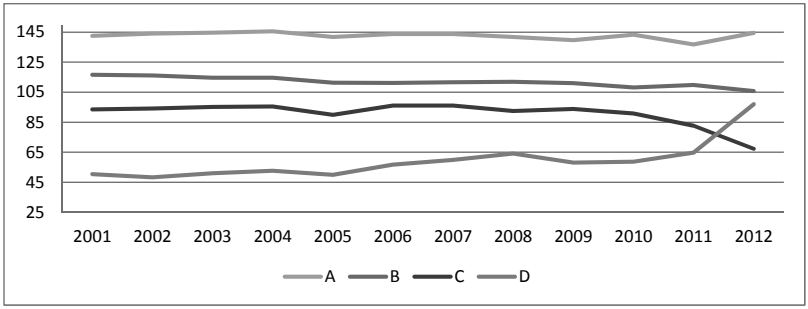
Country	01	02	03	04	05	06	07	08	09	10	11	2012	EA 2012	S&P 2012
Belgium	A	A	A	A	A	A	A	A	A	A	A	A	EA	AA
Bulgaria	D	D	D	D	D	D	D	D	D	D	D	C		BBB
Czech Republic	D	D	D	D	C	D	D	C	C	C	D	C		AA
Denmark	B	B	B	B	B	B	B	B	B	B	B	B		AAA
Germany	A	A	A	A	A	A	A	A	A	A	A	A	EA	AAA
Estonia	D	D	D	D	D	D	D	D	D	D	D	C	EA	AA-
Ireland	C	C	C	C	C	C	C	C	C	C	B	D	EA	BBB+
Greece	D	C	C	C	C	C	C	C	C	C	D	D	EA	B-
Spain	B	B	B	B	B	B	B	B	B	B	B	B	EA	BBB-
France	B	B	B	B	B	B	B	B	B	B	A	B	EA	AA+
Italy	B	B	B	B	B	B	B	B	B	B	B	B	EA	BBB+
Cyprus	C	C	C	C	C	C	C	D	C	B	B	B	EA	CCC
Latvia	D	D	D	D	D	D	D	D	D	D	D	C		BBB
Lithuania	D	D	D	D	D	D	D	D	D	D	D	C		BBB
Luxembourg	A	A	A	A	A	A	A	A	A	A	A	A	EA	AAA
Hungary	D	D	D	D	C	D	D	D	D	D	D	C		BB
Malta	C	C	C	C	C	C	C	C	C	C	C	D	EA	BBB+
Netherlands	A	A	A	A	A	A	A	A	A	A	A	A	EA	AAA
Austria	A	A	A	A	A	A	A	A	A	A	A	A	EA	AA+
Poland	D	D	D	D	D	D	D	D	D	D	D	C		A
Portugal	C	C	C	C	C	C	C	C	C	C	C	C	EA	BB
Romania	D	D	D	D	D	D	D	D	D	D	D	C		BB+
Slovenia	C	C	C	C	C	C	C	C	C	C	C	C	EA	A-
Slovakia	D	D	D	D	D	D	D	D	D	D	D	C	EA	A
Finland	B	B	B	B	B	B	B	B	B	B	A	B	EA	AAA
Sweden	B	B	B	B	A	A	A	A	A	A	A	A		AAA
United Kingdom	A	A	A	A	A	A	A	A	A	A	A	A		AAA

Shluk A zastupuje 7 až 9 nejsilnějších států EU. Shluk B obsahuje 5-6 států. Shluky A a B jsou stabilní, jenom Francie, Finsko a Švédsko se pohybují mezi A a B. Irsko a Kypr se pohybovaly mezi B a C. Shluk C má 3 až 11 států a společně se shlukem nejslabších států D jsou méně stabilní. Shluk D má 3 až 10 periferních států. Shluky A a D jsou protipóly pro EU27, viz obrázek 1.



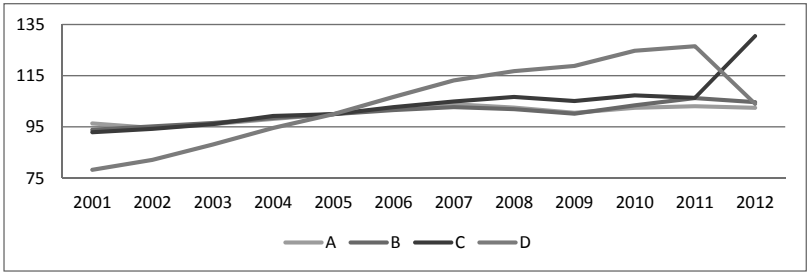
Obr. 20: žebříček shluků dle sumy normalizovaných průměrů rozhodovacích hodnot pro země EU27 rozdělených do čtyř shluků

Shluky A a B s hlediska HDP na osobu jsou vysoko nad průměrem EU27, protože průměrný HDP na osobu ve shluku D nedosahuje do roku 2012 ani půlky hodnoty ukazatele pro shluk lepších států A. Pro dynamiku vývoje HDP na osobu viz obrázek 2.



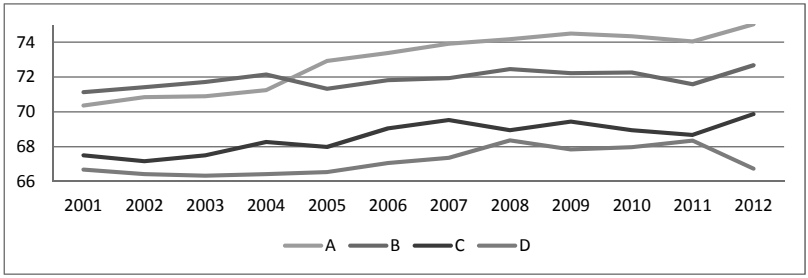
Obr. 2: HDP na obyvatele (index EU27 = 100, dle kupních sil) pro čtyři shluky EU27

Produktivita práce slabších států roste rychlejším tempem, než u silnějších. Ukazatel na obrázku 3. nezachycuje relativní produktivitu států navzájem, ale jenom tempo růstu reálné produktivity každého státu vůči bazickému roku 2005. Tempo růstu produktivity A a B stejné.



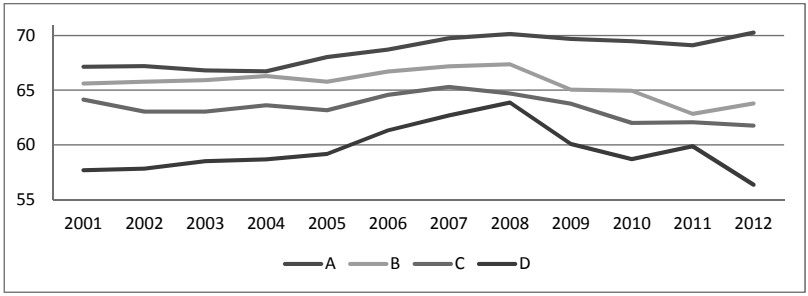
Obr. 3: Reálná produktivita práce na odprac. hod. (index 100=2005) pro čtyři shluky EU27

Míra ekonomické aktivity dospělého obyvatelstva efektivnějších států je oproti slabším státům vyšší o 2-8%. Vývoj ukazatele pro shluky A,B,C je shodný. U shluku A a B došlo ke křížení v roce 2004 vlivem přesunu Švédska, státu s vysokou mírou eko. aktivity, z A do B.



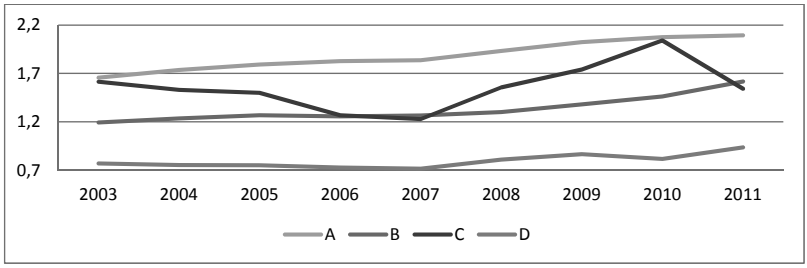
Obr. 4: Míra ekonomické aktivity (15 až 64 let) pro čtyři shluky EU27

Míra zaměstnanosti odpovídá rozložení na lepší a horší státy. Na obrázku 5. je vidět jak v předkrizovém roce se nůžky polárních shluků sevřely na minimum, ale po krizi se postupně otevřely na maximum. Zajímavý je fakt, že B a C se vyvíjí vedle sebe, a postupně propadají.



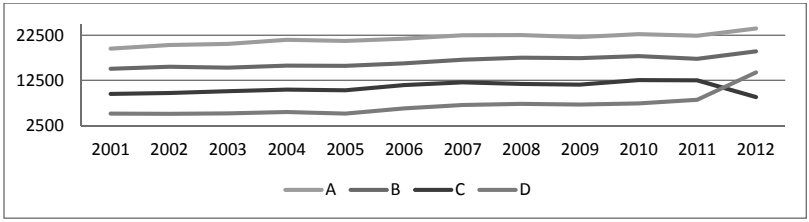
Obr. 5: Míra zaměstnanosti (15 až 64 let) pro čtyři shluky EU27

Graf produktivity materiálů viz 6., za výjimkou shluku C, má mírně rostoucí dynamiku a zachycuje informaci, že lepší shluk A má produktivitu zdrojů lepší o třikrát než horší státy. Shluk B má o dvakrát lepší produktivitu, než státy D. Shluk C kolísá kvůli své nestabilitě a vlivu Malty, která se vyznačuje největšími hodnotami. V roce 2012 statistika chybí.



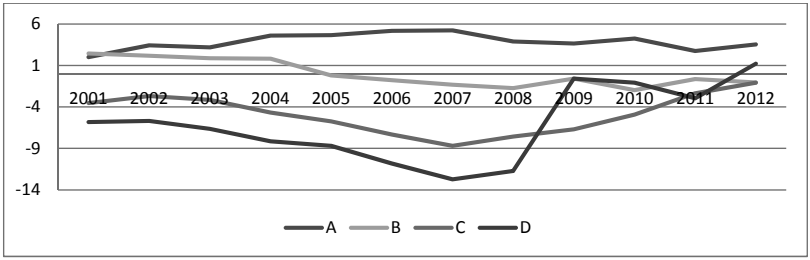
Obr. 6: Produktivita materiálů (€/kg) pro čtyři shluky EU27

Roční čisté příjmy na obrázku 7. odpovídají vývoji HPD na osobu. Ke křížení C a D došlo vlivem toho, že v roce 2011 shluky C a D obsahovaly 3 a 10 států a v roce 2012 naopak 11 a 3 států. Přičemž ve shluku D se v roce 2012 ocitlo Řecko, Irsko a Malta, kde průměrné příjmy v porovnání se státy z C jsou relativně větší.



Obr. 7: Roční čisté příjmy (€) pro čtyři shluky EU27

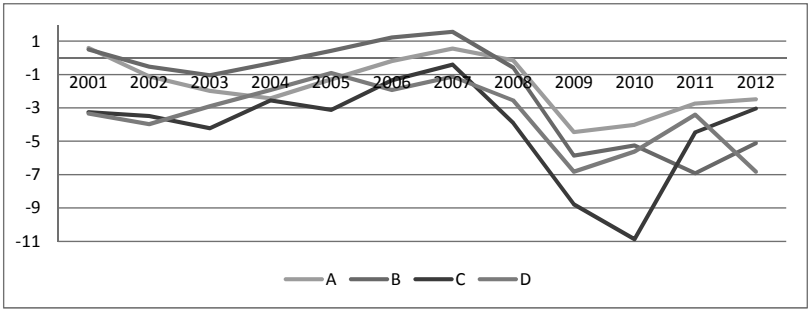
Shluk A se nachází v přebytku platební bilance, shluk B od roku 2004 mírně pod nulou. Shluky C a D v rocích 2001-2011 byly v deficitu platební bilance. Nůžky se otevřely nejvíc před krizí a nyní se uzavírají. D v roce 2012 měl v průměru vyrovnanou obchodní bilanci.



Obr. 8: Platební bilance k HDP (%) pro čtyři shluky EU27

Na obrázku 9. je vidět odraz světového hospodářského cyklu, během poslední krize se všechny státy EU ocitly v deficitu rozpočtu a značně prohloubily tuto mezeru. Fiskální disciplína se stala terčem pozornosti.

Na obrázku 10. je reakce investorů, shluk A půjčuje levně, shluk D draze, a vlivem Řecka se ocitl v 2012 daleko od ostatních shluků, protože i shluk C obsahuje několik států s velmi dobrým kreditním raitingem, příkladem jsou Česko, Litva, Slovensko a Slovinsko.

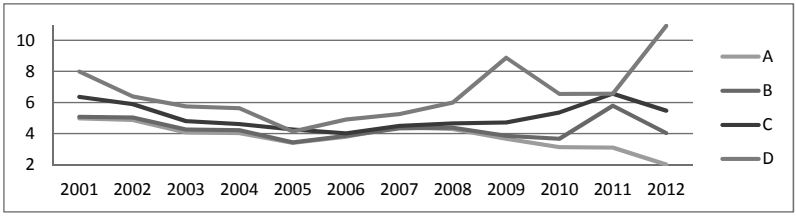


Obr. 9: Vládní deficit k HDP (%) pro čtyři shluky EU27

5. Závěr

Závěrem výzkumu jsou poznatky ohledně rozložení pozic států EU27 z hlediska významných makroekonomických indikátorů. Z výstupu plyne závěr, že poslední světová krize významně přemíchala ekonomické pozice půlky států EU mezi sebou. Kompletní analýza ukázala na místa pro budoucí vylepšení analýzy, například ukazatele *HDP na*

obyvatele a *Čistý roční příjem* jsou silně korelovány, konstrukce použitého ukazatele produktivity práce nevykazuje pozici zemí navzájem, ale pouze dynamiku přírůstků.



Obr. 8: Výnosy dluhopisů (%) pro čtyři shluky EU27

Acknowledgment

This article was created with the help of the Internal Grant Agency of University of Economics in Prague No. 6/2013 under the title „Evaluation of results of cluster analysis in Economic problems."

Literatura

CERNAKOVA, V., & HUDEC, O. (2012). Quality of Life: Typology of European Cities Based on Cluster Analysis. *E & M Ekonomie a Management*, *15*(4), 34–48.

EUROSTAT. (2013). *Eurostat research database*. Retrieved from http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

JAIN, A. K., MURTY, M. N., & FLYNN, P. J. (1999). Data clustering: A review. *Acm Computing Surveys*, *31*(3), 264–323. doi:10.1145/331499.331504

KULBAKOV, N.(2013) *Source for MATLAB cluster analysis, input and output data*. Retrieved from http://www.ilovecz.ru/research/matlab02.zip

KULBAKOV, N. (2013). *Segmentation of EU/EA Countries via Cluster Analysis of Macroeconomics Indicators*. VŠE. Retrieved from http://msed.vse.cz/files/2013/196-Kulbakov-Nikolay-paper.pdf

LÖSTER, T. (2011). *Hodnocení výsledkůmetod shlukové analýzy*. FIS VŠE.

MathWorks. (2013). Cluster Analysis. Retrieved from http://www.mathworks.com/help/stats/cluster-analysis.html

PAPAGEORGIOU, T., MICHAELIDES, P. G., & MILIOS, J. G. (2010). Business cycles synchronization and clustering in Europe (1960-2009). *Journal of Economics and Business*, *62*(5)

QUAH, D. T. (1996). Regional convergence clusters across Europe. *European Economic Review*, *40*(3-5), 951–958. doi:10.1016/0014-2921(95)00105-0

REZANKOVA, H., & LOSTER, T. (2013). *Shlukova analyza domacnosti charakterizovanych kategorialnimi ukazateli. E+M*. Ekonomie a Management, 16(3), 139-147. ISSN: 1212-3609

REZANKOVÁ, H., & SNÁSEL, V. (2009). *Shluková analýza dat*. Praha: Professional Publishing.

Adresa autora:
Nikolay Kulbakov
University of Economics, Prague
W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic
kulbakov@gmail.com

Divergence Decision Trees Used for D0 FNAL Particle Signal Separations
Divergenční rozhodovací stromy použité pro D0 FNAL separaci signálů

Václav Kůs, Michaela Sluková, Jan Kučera

Abstract. We deal with the new statistical classification method called Divergence Decision Tree (DDT), which links together the basic k–means classification and statistical ϕ-divergence measures as a decision criterion. K–means as a part of the DDT has several disadvantages, therefore other simple methods, e.g. fuzzy classifier, can be incorporated. Several versions of the DDT algorithm differing in the type of ϕ–divergences used are tested on data sets coming from single top quark decay channels measured within D0 experiment at Tevatron accelerator in FNAL.

Abstrakt: Zabýváme se novou statistickou klasifikační metodou nazývanou Divergenční rozhodovací strom (DDT), která spojuje základní separaci k-means se statistickými ϕ-divergenčními mírami ve formě rozhodovacího separačního kritéria. K-means jako část DDT má několik nevýhod, proto mohou být začleněny další jednoduché metody jako například fuzzy klasifikátor. Několik verzí DDT algoritmu lišících se v typu použité divergence je testováno na datových souborech ze single top kvark rozpadového kanálu měřených v rámci D0 experimentu na urychlovači Tevatron ve FNAL.

Key words: Data clustering, Separation tree, ϕ–divergence, K–means; Rényi divergence, Numerics

Klíčová slova: shlukování dat, separační strom, ϕ–divergence, k-means, Rényiho divergence, numerika

JEL classification: C61

1. Introduction and DDT algorithm

We were inspired by (Karakos et al., 2005) to work with Divergence Decision Trees (DDT). Thus in this paper we present how the DDT method works, we discuss the influence of different statistical ϕ–divergences used in DDT algorithm and we test the simple separation methods in node clustering. The divergence decision tree is an unsupervised classification method. The principle of this method is the following: the root of the tree is formed by all data set, which is transformed in each node of the tree (e.g. by means of Principal Component Analysis) and subsequently divided into two clusters in regard to some optimization criterion. If we achieve a convergence, which is predetermined, the cluster becomes a leaf and it is no longer being divided. Progress of the algorithm will end up exactly when all the clusters are marked as the leaves. DDT uses ϕ–divergences originated from statistics and information-theoretic field.

Let P,Q $\in \mathcal{P}(\mathcal{X},\mathcal{Y})$, where $\mathcal{P}$ is a set of probability measures on $(\mathcal{X},\mathcal{Y})$. Denote by $p = \mathrm{d}P/\mathrm{d}\mu$ and $q = \mathrm{d}Q/\mathrm{d}\mu$ the Radon-Nikodym derivatives of P and Q with respect to a measure μ. For a given divergence function $\phi\colon (0,\infty) \to \mathbb{R}$, convex on $(0,\infty)$ and strictly convex at 1, with $\phi(1) = 0$, we define ϕ–divergence of P and Q by

$$D_\phi(P,Q) = \int_{\mathcal{X}} q\,\phi\left(\frac{p}{q}\right)\mathrm{d}\mu, \quad P,Q \in \mathcal{P}(\mathcal{X},\mathcal{Y})\,.$$

We have used many divergences, e.g. the total variation $V(P,Q) = \int |p-q|\mathrm{d}\mu$, or the Hellinger squared distance $H^2(P,Q) = \int (\sqrt{p}-\sqrt{q})^2\mathrm{d}\mu$, or the Rényi decomposable divergences.

The splitting of the clusters in each node may be different in attributes. We used a Principal Component Analysis for transforming data into a set of lower dimension without

loss of mutual information. For each pair of new parameters of dataset we perform separation using k–means and then we compute the so called split-value for each separation:

$$S = \frac{N_1}{N} D_1 + \frac{N_2}{N} D_2 ,$$

where N means the total number of samples in dataset, N_k means the number of samples in k–th cluster (the left or the right node) and D_k is a divergence between empirical distribution functions of the k–th cluster and all dataset. We split the cluster according to that k–means result, whose split–value reaches the maximum. In fact this is a local maximum, so we say that the tree grows by greedy manner. Then we compute the contributions of both clusters $C_k = \frac{N_k}{N} D_k$. For the next splitting, the cluster with the maximum contribution is chosen. When the contribution of any cluster is lower than the predefined limits then that cluster becomes a leaf in the tree and we do not divide this cluster again. The aim of our work is to find clusters in the data set on the basis of observable characteristics (the unsupervised classification). One possible criterion is to maximize the distance between the clusters. This is just what this procedure guarantees.

From several applications of the DDT method it follows that k–means is not the best classification method for complicated datasets. Success of classification can be extremely affected by outliers and it also elongates the shape of clusters. We tried to solve the first problem by replacing k–means by k–means++, see Vassilvitskii and Arthur (2007). This method has the similar algorithm as k–means, however, it takes account of outliers. We also try replacing the k–means by Fuzzy Classification Method (FCM), where we can control the shapes of clusters by fuzzyfication factor and also by partial supervision. The FCM optimizes an objective function

$$Q = \sum_{i=1}^{C} \sum_{k=1}^{N} u_{ik}^m d_{ik}^2 + \alpha \sum_{i=1}^{C} \sum_{k=1}^{N} (u_{ik} - f_{ik} b_k)^2 d_{ik}^2 ,$$

where $U = [u_{ik}]\ i = 1,2, \dots c, k = 1, \dots . N$ is a partition matrix, u_{ik} is membership degree of the sample x_k to the cluster i, $d_{ik} = d(x_k, \mu_i)$ is the Euclidean distance between the sample x_k and the center of i-th cluster μ_i, m is the fuzzyfication factor ($m > 1$). Vector of tags $b = [b_1, \dots, b_N]^T$ gives the information about labeled samples (if sample x_k is labeled then $b_k = 1$, otherwise $b_k = 0$) and the matrix $F = [f_{ik}], i = 1,2, \dots c, k = 1, \dots . N$ contains the membership degrees assigned to the selected samples.

2. Rényi divergence and its robustness properties

Rényi decomposable divergences are the very new and promising concept used in statistical inference. They bring high robustness properties and relative practical feasibility. M-C simulation results for the Minimum Rényi Distance (MReD) estimates in the case of very sparse and scattered data (data with high variance) or small sample data sets were carried out and the effect of input parameter α to the robustness was explored. Heuristic approach is proposed for such MReD computations when the strict minimization leads to delta functions. Subsequently, these Rényi divergence based DDT signal separations were performed for the Single Top Quark decay channel, i.e. the data samples coming from high energy particle accelerator Tevatron in Fermilab obtained at the energy level 1.96TeV and the beam luminosity $9.5 fb^{-1}$.

Thus, let for some $\beta > 0$ it hold that $p^\beta, q^\beta, \ln p \in \mathrm{L}_1(Q)$ for distributions $P, Q \in \mathcal{P}$. Then for all α, $0 < \alpha \leq \beta$, and for $P, Q \in \mathcal{P}$ the expression

$$\Re_\alpha(P,Q) = \frac{1}{1+\alpha}\ln\left(\int p^\alpha \mathrm{d}P\right) + \frac{1}{\alpha(1+\alpha)}\ln\left(\int q^\alpha \mathrm{d}Q\right) - \frac{1}{\alpha}\ln\left(\int p^\alpha \mathrm{d}Q\right)$$

represents the Rényi family of pseudo-distances decomposable in the sense of

$$\Re_\alpha(P,Q) = \Re_\alpha^0(P) + \Re_\alpha^1(Q) - \frac{1}{\alpha}\ln\left(\int p^\alpha \mathrm{d}Q\right), \quad \text{where}$$

$$\Re_\alpha^0(P) = \frac{1}{1+\alpha}\ln\left(\int p^\alpha \mathrm{d}P\right), \quad \Re_\alpha^1(Q) = \frac{1}{\alpha(1+\alpha)}\ln\left(\int q^\alpha \mathrm{d}Q\right).$$

Further, for $\alpha \to 0$ it holds that $\Re_0(P,Q) = \lim_{\alpha\to 0}\Re_\alpha(P,Q) = \int (\ln q - \ln p)\mathrm{d}Q$. We use influence function (IF) to measure the robustness of the Rényi estimator. The IF summarizes the impact of single date on the estimator under consideration. There are more important characteristics derived from the influence function, but we are mainly interested in only two of them. First one is the *gross-error sensitivity* characterized by $\gamma^* = \sup_x|\mathrm{IF}(x; T_{\Re_\alpha}, \theta)|$. We require it to be finite, in other words, we want the influence function to be bounded. The second one is called the *rejection point* and it is defined by $\rho^* = \inf\{r > 0 \mid IF(x; T_{\Re_\alpha}, \theta) = 0 \text{ for } |x| > r\}$. It describes the values of samples, which will be treated as outliers and finally rejected.

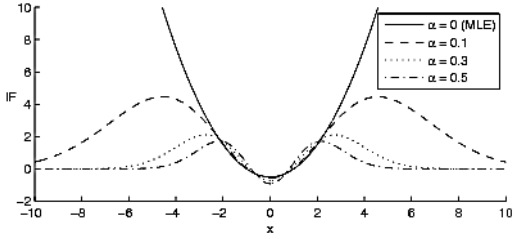


Figure 1: Influence function of Rényi estimator for Normal ($\mu = 0$ known, $\sigma = 1$ estimated)

As an example, we show the IF for Rényi divergence derived in Broniatowski, Toma and Vajda (2011) and Broniatowski and Vajda (2009) for parameter sigma of normal distribution, see Figure 1,

$$\mathrm{IF}(x; T_{\Re_\alpha}, \sigma) = \frac{(1+\alpha)^{5/2}\sigma}{2}\left[\left(\frac{x}{\sigma}\right)^2 - \frac{1}{1+\alpha}\right)\exp\left(-\frac{\alpha x^2}{2\sigma^2}\right)].$$

We can deduce from Figure 1 that Rényi estimators are robust for all $\alpha > 0$ in the sense that their influence functions are bounded. Also with higher α they are more robust against outliers since $\lim_{x\to\pm\infty}\mathrm{IF}(x; T_{\Re_\alpha}, \sigma) = 0$. Moreover, the convergence is faster with increasing α due to the term $e^{-\alpha x^2}$, see also Basu et al (1998).

2.1. Rényi divergence quality testing (the heuristic approach)

There is an inconvenience concerning the choice of "robust" α ($\alpha > 0.5$) for small data samples, or for very sparse or scattered data (data with high variance). This problem is illustrated in Figure 2, where we used a random set of 10 variables $\mathrm{X} = (x_1, \ldots, x_{10}) \sim \mathrm{N}(0,2)$ for the Rényi decomposable distance between the empirical distribution of observations X and normal distribution with parameters (μ, σ). We can see that there is a large shallow minimum

around the point $(0.3, 2.3)$, but closer examination shows that all of the eight bright points on the y axis ($\sigma = 0.001$) have steep minima with lower values of the distance function. These areas specify distributions corresponding to the Dirac δ −functions $\delta_{a_i}(x)$, where $a_i = x_i$ for $i \in \{1, \dots, 10\}$. Due to these extreme values of Rényi distance function, the Rényi estimator would prefer these singular estimates instead of the large shallow minimum around the point $(0.3, 2.3)$. Thus the estimator is so robust and the data are so sparse that the estimator always takes the single date as the only representative of the estimated distribution and all the other data are treated as outliers. To overcome this problematic behaviour, we devised a method inspired by image processing. The idea is to blur the resulting image so that there wouldn't be sharp and deep minima, which, from the viewpoint of image processing, are edges. This Blurring was created as a convolution of Rényi distance with averaging Gaussian mask. If we denote $d_{\mu,\sigma}$ the Rényi distance between the empirical distribution and $\mathrm{N}(\mu, \sigma)$, the distance after averaging is

$$\bar{d}_{\mu_i,\sigma_j} = \frac{1}{r^2} \sum_{k=i-r}^{i+r} \sum_{l=j-r}^{j+r} d_{\mu_k,\sigma_l} ,$$

where r is the radius of the averaging mask. Rényi distance after averaging is displayed in Figure 3.

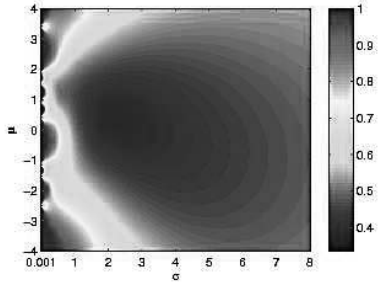


Figure 2: Rényi distance behavior

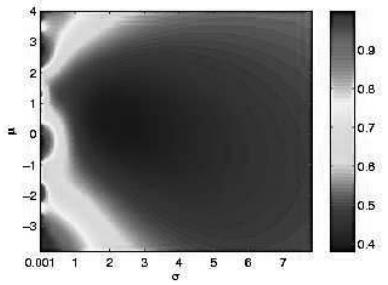


Figure 3: Rényi distance after averaging

We can see that the minima corresponding to the Dirac δ-functions are flattened after blurring and their values are higher than the minimum at the required point $(0.3, 2.3)$. The Blurring mechanism not only flattens the image, but it also slightly moves the minima. To overcome this setback and refine the result we used two-step algorithm:

1) Minimize the averaged distance $\bar{d}_{\mu,\sigma}$ (under Gaussian mask blurring),
2) Minimize the original Rényi distance near the local minima chosen by the step 1.

It means, we use the average Rényi distance $\bar{d}_{\mu,\sigma}$ for finding the overall local minimum that describes the whole bunch of data and then we find the exact final minimum from the original Rényi distance. Table 1 shows the differences between the three used algorithms.

Table 1: Rényi distance mean minimization for 100 repetitions, data $(\boldsymbol{x_1}, \dots, \boldsymbol{x_{10}}) \sim \boldsymbol{N(0, 2)}$

	$\hat{\mu}$	$\hat{\sigma}$
Rényi distance	0.1742	0.3448
Averaged Rényi distance	0.0058	2.8532
Two-step minimization	0.0346	2.3832

The first row corresponds to the unchanged original Rényi distance obtained directly from Theorem 1. We can see that the mean of the estimated parameter σ is very small even though

the generated values should have this parameter much bigger. In the second row we used only the averaging algorithm. In this case, the estimator prefers parametric estimator with the higher value of σ. The result of the proposed two-step algorithm is represented by the third row of Table 1. The corresponding results are much closer to the true values and they are apparently refined by the second step of the algorithm.

3. DDT application to D0 FNAL data sets

The DDT separation method has been applied also on a data sets acquired from the experiment D0 held in the collider accelerator Tevatron in Fermilab (USA). The data comes from the detection of proton–antiproton collisions, during which a large number of new particles arise. In our case we focus on the data from the Single Top Quark decay channels. D0 detectors record various parameters of the particles (railway, energy, momentum, angles, diffraction, etc.). According to these parameters it is possible to determine which particles are involved in. The number of parameters of the data (i.e. data dimension) is 39, which is a great complication for searching data structure, particularly for unsupervised methods that we applied to these data sets. The second problem regarding the separation (classification) consist in the number of samples measured, this is of the order of 10^5.

Table 2: Ratio of Signal (S) to Background (B) in each cluster in DDT separation technique

Classification of dataset from D0 FNAL *(dividing method: k–means, ϕ–divergence: Hellinger distance)*							
50 000 samples							
Cluster 1		Cluster 2		Cluster 3		Cluster 4	
19069		15890		7897		6604	
S	**B**	**S**	**B**	**S**	**B**	**S**	**B**
2685	16924	2524	13366	1407	6490	1270	5334
0.7 %	86.3 %	15.9 %	84.1 %	17.8 %	82.2 %	19.2 %	80.8 %

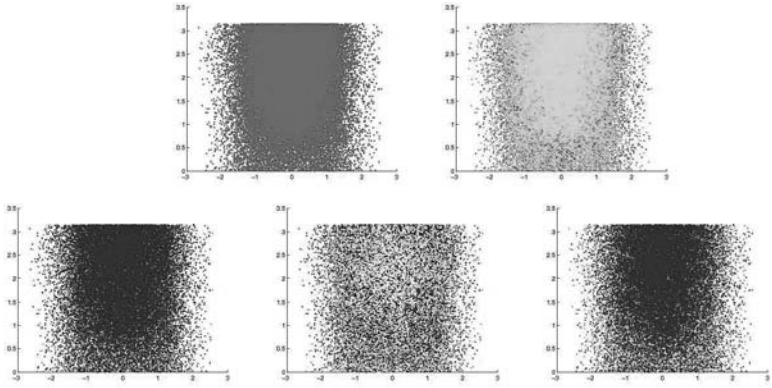


Figure 4: DDT using k–means and the total variation as a ϕ–divergence applied on dataset measured in Fermilab during experiment D0.

The goal of the method DDT was to find clusters that they either contain only signal or only background. Due to the fact that the number of data designated as the background is much larger than the number of signals, the calculation of the successes of the method was complicated. The results of the D0 data set classification can be found in Table 2. The results

are not as good as we have expected, however, we achieve an increase of the ratio of signal to background in 3 clusters of overall 4 clusters (initial value of about 15,8%, which means 7886 samples of signals (S) and 42114 samples of backgrounds (B)). The Figure 4 shows how complicated the structure of data is. The data set contains 39 parameters (separation attributes) for samples and in any pair of the parameters we cannot find any indication of clusters.

For the future work it is necessary to engage some supervised methods in order to improve the DDT method. A supervision could be very convenient for the data sets coming from D0 experiment, because training data sets generated by Monte Carlo are available in the experiment. Especially the Method of Distribution Mixtures (the so called MBC clustering) or Support Vector Machines (SVM) could be used in the DDT method in the tree nodes. For example, the SVM possesses considerable variability resulting from the possibility of selecting the suitable kernel transform function during the nonlinear SVM classification.

Acknowledgement. This work was supported by the grant MSMT INGO-II LG12020.

References

KARAKOS, D. et al. 2005. Unsupervised classification via decision trees: An information-theoretic perspective. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005.

VASSILVITSKII, S., ARTHUR, D. 2007. K-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, p. 1027-1035.

BASU, A., HARRIS, I. R., HJORT, N. L., JONES, M. C. 1998. Robust and efficient estimation by minimising a density power divergence. In *Biometrika*, 85, 549-559.

BRONIATOWSKI, M., TOMA, A., VAJDA, I. 2011. Decomposable pseudo-distances and Applications in Statistical Estimation. arXiv:1104.1541v1.

BRONIATOWSKI, M., VAJDA, I. 2009. *Several Applications of Divergence Criteria in Continuous Families*. Research report No 2257 September 2009, UTIA AV CR, Prague.

Authors addresses:

Václav Kůs, Ing, PhD.
Katedra matematiky
FJFI ČVUT
Trojanova 13, 120 00 Praha 2
vaclav.kus@fjfi.cvut.cz

Michaela Sluková, Bc.
Katedra matematiky
FJFI ČVUT
Trojanova 13, 120 00 Praha 2
slukomis@fjfi.cvut.cz

Jan Kučera, Bc.
Katedra matematiky
FJFI ČVUT
Trojanova 13, 120 00 Praha 2
kucerjan@fjfi.cvut.cz

Miery generalizovanej entropie
Generalized entropy measures

Viera Labudová

Abstract: The Lorenz curve and the Gini coefficient are the most fundamental tools used to measure income inequality. Theil's index is part of a special class of inequality measures known as Generalised Entropy measures. An important property of Theil's index is the additive decomposability characteristic, which implies that the aggregate inequality measure can be decomposed into inequality within and between any arbitrarily defined population subgroups.

Abstrakt: Lorenzova krivka a Giniho koeficient patria k základným mieram, ktoré sa používajú na meranie príjmovej nerovnosti. Theilov index patrí do skupiny mier generalizovanej entropie. Dôležitou vlastnosťou tohto indexu je aditívnosť rozkladu, ktorá umožňuje rozložiť nerovnosť meranú na celej populácii na nerovnosti merané na jej podmnožinách.

Key words: Generalised entropy measures, entropy, Theil index, inequality.

Kľúčové slová: Miery generalizovanej entropie, entropia, Theilov index, nerovnosť.

JEL classification: C44

1. Úvod

Meranie nerovnosti v rozdeľovaní príjmov alebo výdavkov je dôležitou súčasťou sociálno-ekonomických analýz. Spôsob merania závisí od použitého konceptu nerovnosti.

Podľa Sena (1997) možno rozdeliť miery nerovnosti na objektívne miery a normatívne miery. Hlavným znakom objektívnych mier nerovnosti je to, že využívajú štatistické a matematické nástroje. Sen medzi ne zaradil Lorenzovu krivka, Giniho index, miery generalizovanej entropie, pomer kvantilov, rozptyl príjmov, rozptyl logaritmu príjmov a variačný koeficient. Normatívne miery sa zvyčajne zaoberajú nerovnosťou z hľadiska jej vplyvu na spoločenský blahobyt. Najznámejšou mierou, ktorá je založená na spoločenskej funkciu blahobytu je Atkinsonov index (Charles-Coll, 2011).

V článku sa venujeme mieram generalizovanej entropie[1]. Východiskom pre túto skupinu mier (indexov) je koncept entropie, ktorý vychádza z informačnej teórie.

2. Shannonova definícia entropie[2]

Pôvodný koncept entropie pochádza od Ludwiga Boltzmanna (1877 v Frenken, 2007, p. 544). V teórii informácie bola táto teória rozpracovaná s využitím pravdepodobnosti (Shannon ,1948 v Frenken, 2007, p. 544). V roku 1960 vyvinul Henri Theile niekoľko aplikácií Shannonovej teórie informácie v ekonómii, ktoré publikoval v prácach *Economics and Information Theory* (1967 v Frenken, 2007, p. 544) and *Statistical Decomposition Analysis*(1972 v Frenken, 2007, p. 544).

Entropia je fyzikálna veličina, ktorá meria neusporiadanosť (náhodnosť, neporiadok,...) systému. V teórii informácie je definovaná ako miera neurčitosti t.j. protiklad k pojmu

[1] Anglický ekvivalent je *generalized entropy measure of inequality*. V slovenskej literatúre sme nenašli adekvátny preklad tohto pojmu. Používame preto ekvivalent prekladu, s ktorým sa môžeme stretnúť napr. v Netrdová – Nosek (2009), Novotný – Nosek (2006).

[2] Spracované podľa Palúch (2007). Všetky definície a vzťahy sú prevzaté bez dôkazov.

informácia. Pre definovanie entropie, ktoré vychádza z teórie informácie, je potrebné zaviesť pojmy informácia a pokus.

Informácia je definovaná ako reálna funkcia $I : A \to R$ (R je množina reálnych čísel), ktorá každému prvku z množiny A priradí nezáporné reálne číslo – množstvo informácie. Množinou A je σ-algebra[3] merateľných podmnožín základného priestoru Ω (množina elementárnych javov).

Funkcia $I : A \to R$ je definovaná tak, aby hodnota $I(A)$ pre $A \in A$ vyjadrovala množstvo informácie, ktorú dostaneme v správe, že nastal jav *A*. Toto je vyjadrené Shannonovou-Hartleyovou formulou[4]

$$I(A) = -k \cdot \log_2 (P(A)) \tag{1}$$

kde $P(A)$ je pravdepodobnosť javu *A*. Koeficient *k* vo vzťahu (1) závisí od jednotky merania. Ak zvolíme za jednotku merania jeden bit, potom *k*=1 a vzťah (1) má tvar

$$I(A) = -\log_2 (P(A)) \tag{2}$$

Ak teda dostaneme správu, že nastal jav $A \in A$, s pravdepodobnosťou $P(A)$, dostaneme s ňou informáciu $I(A) = -\log_2 (P(A))$ bitov.

Nech (Ω, A, P) je pravdepodobnostný priestor. Konečný merateľný rozklad istého javu je konečná množina javov (t.j. podmnožín Ω) $\{A_1, A_2, \ldots A_n\}$ taká, že $A_i \in A$ pre $i = 1, 2, \ldots n$, $\bigcup_{i=1}^{n} A_i = \Omega$ a $A_i \cap A_j = \emptyset$ pre $i \neq j$. Konečný merateľný rozklad $\mathbf{P} = \{A_1, A_2, \ldots A_n\}$ istého javu Ω nazývame pokusom.

Nech je (Ω, A, P) pravdepodobnostný priestor, na ktorom je daná informácia $I(A) = -\log_2 P(A)$. Nech $\mathbf{P} = \{A_1, A_2, \ldots A_n\}$ je pokus.

Entropia $H(\mathbf{P})$ pokusu $\mathbf{P}$ je stredná hodnota diskrétnej náhodnej premennej *X*, ktorá nadobúda na množine A_i hodnotu $I(A_i)$, t.j.

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i) P(A_i) = -\sum_{i=1}^{n} P(A_i) \log_2 P(A_i)^{5} \tag{3}$$

alebo

---

[3] Nech Ω je množina elementárnych javov (nazývaná aj základný priestor). σ-algebrou základného priestoru Ω nazývame taký systém A podmnožín množiny Ω, pre ktorý platí

1. $\Omega \in A$
2. Ak $A \in A$, potom aj $A^C \in (\Omega - A) \in A$
3. $A_n \in A$ pre $n = 1, 2, \ldots \infty$, potom aj $\bigcup_{n=1}^{\infty} A_n \in A$.

[4] Vzhľadom na to, že výskyt javov s nízkou pravdepodobnosť zvyšuje informáciu, pre "zosilnenie účinku" nízkych hodnôt pravdepodobnosti bola zvolená logaritmická funkcia.

[5] Ak sa v pokuse $\mathbf{P} = \{A_1, A_2, \ldots A_n\}$ objaví množina A_i s nulovou pravdepodobnosťou $P(A_i) = 0$, predpokladá sa, že výraz $-P(A_i) \log_2 P(A_i)$ je definovaný a platí $-0 \cdot \log_2 0 = 0$.

$$H(\mathbf{P}) = \sum_{i=1}^{n} P(A_i) \log_2 \frac{1}{P(A_i)} \tag{4}$$

Entropia nadobúda nezáporné hodnoty. V prípade, že pravdepodobnosť niektorého javu A_i je rovná 1, entropia sa rovná nule

$$H_{\min} = 1 \cdot \log_2 \left(\frac{1}{1} \right) = 0 \tag{5}$$

Maximálnu hodnotu nadobudne entropia v situácii, kedy je pravdepodobnosť všetkých javov A_i rovnaká $P(A_i) = \frac{1}{n}$ pre $i = 1, 2, \ldots n$

$$H_{\max} = \sum_{i=1}^{n} \frac{1}{n} \log_2(n) = n \frac{1}{n} \log_2(n) = \log_2(n) \tag{6}$$

Entropia môže byť považované za mieru neurčitosti (neistoty, premenlivosti). Čím je väčšia neistota (pochybnosť) pred správou, že nejaký jav nastal, tým väčšiu informáciu prináša správa o priemere. Theil (1972, v Frenken, 2007, p. 545) hľadá v tomto kontexte analógiu medzi entropiou a variabilitou.

3. Theilove indexy

Shannonova entropiu využil Theil na vytvorenie miery príjmovej nerovnosti. Ak nahradíme pravdepodobnosti $P(A_i)$, podielmi jednotlivcov na celkovom príjme y_i, potom vzťah (4) má tvar

$$H(y) = \sum_{i=1}^{n} y_i \log_2 \frac{1}{y_i} \tag{7}$$

Mieru nerovnosti vyjadril ako rozdiel medzi maximálnou hodnotou (6), ktorú entropia nadobúda v prípade, že každý jednotlivec má rovnaký podiel na celkovom príjme ($y_i = \frac{1}{n}$) a entropiou, ktorá odpovedá empirickému rozdeleniu príjmov

$$\log_2 n - H(y) = \log_2 n - \sum_{i=1}^{n} y_i \log_2 \left(\frac{1}{y_i} \right) = \log_2 n + \sum_{i=1}^{n} y_i \log_2 (y_i) =$$

$$= \log_2 n + \sum_{i=1}^{n} y_i \log_2 (y_i)$$

ak využijeme vzťah $\sum_{i=1}^{n} y_i = 1$, potom

$$\log_2 n + \sum_{i=1}^{n} y_i \log_2 (y_i) = \sum_{i=1}^{n} y_i \log_2 n + \sum_{i=1}^{n} y_i \log_2 (y_i) = \sum_{i=1}^{n} y_i \log_2 \left(\frac{y_i}{\frac{1}{n}} \right)$$

kde y_i vyjadruje podiel príjmu jednotlivca na celkovom príjme.

Pri prechode od relatívneho konceptu k absolútnemu konceptu má takto vyjadrená miera nerovnosti tvar[6]

$$\sum_{i=1}^{n} \frac{y_i}{Y} \log\left(\frac{y_i}{Y} \Big/ \frac{1}{N} \right) \tag{8}$$

kde y_i je príjem i-teho jednotlivca a $Y = \sum_{i=1}^{n} y_i$ je celkový príjem.

Odvodený Theilov index je indexom zo skupiny generalizovanej entropie. Miery zo skupiny generalizovanej entropie majú všeobecný tvar

$$GE(\alpha) = \frac{1}{n(\alpha^2 - \alpha)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right] \tag{9}$$

kde y_i je hodnota premennej (príjmu) pre *i*-tý objekt, $(i = 1, 2, \ldots n)$, $\bar{y}$ je jej priemerná hodnota. Špecifický tvar mier, ktoré patria do tejto skupiny je určený veľkosťou konštanty α, od ktorej závisí to, akým spôsobom bude miera reagovať na zmeny v jednotlivých spektrách príjmového rozdelenia.

Parameter α môže teoreticky nadobudnúť akúkoľvek hodnotu z intervalu $(-\infty; +\infty)$, v praktických realizáciách sa uvažujú len jeho nezáporné hodnoty $\alpha \geq 0$. Pre vyššie nezáporné hodnoty α index $GE(\alpha)$ reaguje citlivo na zmeny, ku ktorým dôjde v hornej časti príjmového rozdelenia. Ak je $\alpha \geq 0$ a nadobúda nízke hodnoty, miera $GE(\alpha)$ je senzitívnejšia v dolnej časti príjmového rozdelenia. V praxi sa najčastejšie využívajú tri tvary indexov z tejto skupiny a to pre $\alpha = 0$, $\alpha = 1$ a $\alpha = 2$[7].

Pre hodnotou $\alpha = 0$ dostávame *Theilov L index* (*mean logarithmic deviation*)

$$Theil\ L = GE(0) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{\bar{y}}{y_i} = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{y_i}{\bar{y}} \tag{10}$$

Pre hodnotu $\alpha = 1$ dostávame *Theilov T index*

$$Theil\ T = GE(1) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \log \frac{y_i}{\bar{y}} \tag{11}$$

Pre hodnotu $\alpha = 2$ má miera $GE(\alpha)$ tvar druhej odmocniny variačného koeficienta

---

[6] Uvádzame tvar, ktorý použil Theil.

[7] Pre hodnoty konštanty $\alpha = 0$ a α=1 je menovateľ vo vzťahu (3.22) rovný nule. Pri odvodzovaní obidvoch tvarov indexu generalizovanej entropie bolo použité L'Hospitalovo pravidlo.

$$\lim_{\alpha \to 0} \frac{1}{n(\alpha^2 - \alpha)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right] = \lim_{\alpha \to 0} \frac{\sum_{i=1}^{n} \left(\frac{y_i}{\bar{y}} \right)^{\alpha} \log \left(\frac{y_i}{\bar{y}} \right)}{n(2\alpha - 1)}$$. Podobne

$$\lim_{\alpha \to 1} \frac{1}{n(\alpha^2 - \alpha)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right] = \lim_{\alpha \to 0} \frac{\sum_{i=1}^{n} \left(\frac{y_i}{\bar{y}} \right)^{\alpha} \log \left(\frac{y_i}{\bar{y}} \right)}{n(2\alpha - 1)}$$

$$CV = GE(2) = \frac{1}{\bar{y}}\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{\frac{1}{2}} \tag{12}$$

$GE(\alpha)$ nadobúda minimálnu hodnotu 0 v prípade príjmovej rovnosti, s rastom jej hodnôt sa zvyšuje príjmová nerovnosť. Maximálna hodnota závisí od konštanty α .

Vlastnosti absolútnych mier generalizovanej entropie:

1. Minimálna hodnota pre indexy $GE(\alpha)$ je 0. Indexy ju nadobúdajú v prípade, že všetky príjmy sú rovnaké.[8]
2. Maximálna hodnota pre $GE(0)$ nie je definovaná, pre index $GE(1)$ je maximálnou hodnotou log *n*. U indexov s parametrom $\alpha > 1$ závisí maximálna hodnota $\frac{n^{\alpha} - n}{n(\alpha^2 - \alpha)}$ od konkrétnej hodnoty parametra.
3. Pre všetky hodnoty parametra $\alpha \geq 0$ indexy citlivo reagujú na prenos príjmov. Pre $GE(0)$ a $GE(1)$, závisí zmena veľkosti miery od veľkosti populácie a od úrovne jednotlivých príjmov, ktoré sú redistribuované[9]. Pre $\alpha > 0$, $\alpha \neq 1$ je zmena ovplyvnená okrem veľkosti populácie a hodnoty príjmu, ktorý je redistribuovaný aj od konkrétnej hodnoty parametra α .

4. Záver

V príspevku boli predstavené miery generalizovanej entropie, založené na koncepte entropie, ktorý vychádza z informačnej teórie. Stručne boli vyhodnotené ich vlastnosti vzhľadom na axiomatický koncept príjmovej nerovnosti. Ich výhodou, v porovnaní s ďalším veľmi často používaným Giniho indexom, je možnosť aditívneho rozkladu, ktorý umožňuje kvantifikovať vplyv nerovnosti meranej na populačných podskupinách na nerovnomernosť na celej populačnej množine.

Literatúra

FRENKEN, K. 2007. Entropy statistics and information theory. In *Elgar Companion to Neo-Schumpeterian Economics* [online]. Cheltenham, UK: Edward Elgar, 2007, p. 544-555. [08.10. 2012]. Dostupné na internete: <http://digamo.free.fr/elgarneoschump.pdf>.

CHARLES-COLL, J. A. 2011**.** Understanding Income Inequality: Concept, Causes and Measurement. In *Management Journals. International Journal of Economics and Management Sciences* [online]. 2011,vol. 1, no. 3, p. 17-28 [02.10. 2012]. Dostupné na internete: < http://www.managementjournals.org/ijems/3/IJEMS-11-1303.pdf>. ISSN 2162-6359.

LABUDOVÁ, V. 2010. Miery príjmovej nerovnosti. In *Forum statisticum Slovacum : vedecký časopis Slovenskej štatistickej a demografickej spoločnosti*. Bratislava: Slovenská štatistická a demografická spoločnosť, 2010. ISSN 1336-7420. Roč. 6, č. 5, 2010, s. 127-131.

---

[8] Ak sú všetky príjmy rovnaké $y_1 = y_2 = ... = y_n = \bar{y}$, potom $\frac{y_i}{\bar{y}} = 1$ a $\log\frac{y_i}{\bar{y}} = 0$. Vtedy $GE(0) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{y_i}{\bar{y}} = 0$ a $GE(1) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\bar{y}}\log\frac{y_i}{\bar{y}} = 0.$

[9] $\frac{d}{dy_i}GE(0) = -\frac{1}{n}\frac{1}{y_i}$; $\frac{d}{dy_i}GE(1) = -\frac{1}{n}\left[\frac{1}{\bar{y}}\left(1 + \log\frac{y_i}{\bar{y}}\right)\right]$

LABUDOVÁ, V. 2012. Miery príjmovej nerovnosti. In *Nerovnosť a chudoba v Európskej únii a na Slovensku: zborník statí z vedeckej konferencie, Herľany, 26. septembra 2012.* Košice: Ekonomická fakulta, Technická univerzita v Košiciach, 2012. ISBN 978-80-553-1225-5. s. 107-112.

LABUDOVÁ, V. 2013. Income inequality in Slovakia.In *Ekonomika a informatika: vedecký časopis FHI EU v Bratislave a SSHI*. Bratislava: Fakulta hospodárskej informatiky : Slovenská spoločnosť pre hospodársku informatiku, 2013. - ISSN 1336-3514. - Roč. 11, č. 1 2013, s. 94-104.

LABUDOVÁ, V. 2013. *Meranie príjmovej nerovnosti : habilitačná práca*. Bratislava, 2013. 131 s.

NETRDOVÁ, P. – NOSEK, V. 2009. Přístupy k měření významu geografického rozměru společenských nerovnoměrností. In *Geografie: Sborník české geografické společnosti.* Roč. 114, č. 1, s. 52-65.

NOVOTNÝ, J. – NOSEK, V. 2006. Regionální dimenze sociálně-ekonomických nerovností v Česku: pojetí, měření, empirie [online]. In *Sborník příspěvků z XXI. sjezdu České geografické společnosti. České Budějovice 30. 8. – 2. 9. 2006.* [15.10. 2012]. Dostupné na internete: <http://web.natur.cuni.cz/~nosek6/admin/volne/CB_conf.pdf>.

PALÚCH, S. 2007. *Teória informácie* [online]. Žilina: Žilinská univerzita, 2007. 135 s. [12.10. 2012]. Dostupné na internete: <http://frcatel.fri.uniza.sk/users/paluch/ti_vlna.pdf>.

SEN, A. K. 1997. *On Economic Inequality*. Oxford: Oxford University Press, 1997. 260 p. ISBN 0-19828193-5.

SIPKOVÁ, Ľ. 2004. Prehľad teoretických východísk merania príjmovej nerovnosti. In *Slovenská štatistika a demografia.* ISSN 1210-1095, 2004, roč. 14, č. 3, s. 36-49.

Adresa autora:

Viera Labudová, PhD.
Fakulta hospodárskej informatiky
Ekonomická univerzita v Bratislave
Dolnozemská cesta 1
852 35 Bratislava
viera.labudova@euba.sk

Normální délka života, pravděpodobná délka života a pravděpodobný věk úmrtí v České republice v letech 1920 – 2011
Modal age at death, median length of life and probable age at death in the Czech Republic in 1920 – 2011

Jana Langhamrová

Abstract: A very often used characteristic to evaluate the overall indicators of mortality is life expectancy. It is a characteristic of average; it is the average age of those dead in a stationary population. There are also other indicators characterizing the position like mode or median. For the overall characteristics of mortality is used so-called modal age at death (normal age at death), the age at which most people die. And median length of life (probable length of life), i.e. age, which would live just a half of people in age x years be given mortality. The life tables were calculated based on data on the number of deaths by age and sex for the years 1920 to 2011. The paper shows the development of a modal age at death and median length of life for the Czech Republic.

Abstrakt: Velmi často používanou charakteristikou k hodnocení souhrnných ukazatelů úmrtnosti patří střední délka života. Jde o charakteristiku typu průměr, je to průměrný věk zemřelých ve stacionární populaci. Existují také další ukazatele charakterizující polohu typu modus a medián. Pro souhrnné charakteristiky úmrtnosti se užívá tzv. normální délka života, věk, ve kterém lidé nejčastěji umírají (modus věku zemřelých) a pravděpodobná délka života (charakteristika typu medián), tj. věk, kterého by se při dané úmrtnosti dožila právě polovina x-letých. Z údajů o počtu zemřelých podle věku a pohlaví byly vypočteny úmrtnostní tabulky pro roky 1920-2011. V příspěvku je ukázán vývoj normální a pravděpodobné délky života a pravděpodobného věku úmrtí pro Českou republiku.

Key words: modal age at death, median length of life, probable age at death, life expectancy, Czech Republic, Gompertz-Makeham function

Klíčová slova: normální délka života, pravděpodobná délka života, pravděpodobný věk úmrtí, střední délka života, Česká republika, Gompertzova-Makehamova funkce

JEL classification: J110, J140, C020

1. Úvod

Jednou z nejčastěji používaných souhrnných charakteristik úmrtnosti dané země je střední délka života. Ovšem tato charakteristika v sobě zahrnuje vlastnosti průměru, se svými výhodami i nevýhodami. Ze statistiky víme, že existují i další míry polohy. Tyto jiné charakteristiky délek života nemají nedostatky průměru. Lze určit tzv. normální délku života, což je věk, ve kterém lidé nejčastěji umírají. Je to modus věku zemřelých ve stacionárním obyvatelstvu. Velmi často se normální délka života považuje za charakteristiku dlouhověkosti. Další využívanou charakteristikou délek života je tzv. pravděpodobná délka života osoby přesně *x*-leté. Pravděpodobná délka života je dána dobou, za kterou zemře právě polovina *x*-letých. Je to medián věku zemřelých starších *x*-let zmenšený o *x*. Pravděpodobná délka života jako ukazatel typu medián není závislá na extrémech.

2. Výpočet normální délky života

Normální délku života lze odhadnout jako hrubý odhad (dokončený věk, kdy je počet zemřelých maximální). Hledáme tedy věk, kdy je počet zemřelých v úmrtnostních tabulkách maximální.

Ovšem jak již bylo uvedeno, jde pouze o hrubý odhad a přesnějšího výsledku dosáhneme, pokud využijeme parametry Gompertzovy-Makehamovy funkce. Výpočet normální délky života byl proveden podle Fiala (2005).

Jako zdrojová data pro tento výpočet je nutné znát zemřelé podle věku ($M_{t,x}$) a střední stavy obyvatel podle věku ($\overline{S}_{t,x}$) nebo počáteční ($S_{t,x}$) či koncové stavy ($S_{t+1,x}$) počtu obyvatel.

Základními charakteristikami úmrtnosti, jsou specifické míry úmrtnosti. V případě jejich výpočtu za jeden kalendářní rok je vypočteme podle vzorce

$$m_{t,x} = \frac{M_{t,x}}{\overline{S}_{t,x}}. \tag{1}$$

Pokud neznáme střední, ale počáteční stavy, vypočteme podle vzorce

$$m_{t,x} = \frac{M_{t,x}}{\frac{S_{t,x} + S_{t+1,x}}{2}}, \tag{2}$$

kde $M_{t,x}$ je počet zemřelých v dokončeném věku x let roce t, $S_{t,x}$ pak počáteční stav x-letých osob v roce t. Vyrovnání měr úmrtnosti pro věk 60 a více let lze provést za pomocí Gompertzovy-Makehamovy funkce, která má následující tvar

$$\tilde{m}_x^{(GM)} = a + b \cdot c^{x+\frac{1}{2}}. \tag{3}$$

Zvolíme počátek prvního intervalu x_0 = 60 a délku intervalů k = 8. Vypočítáme součty empirických specifických měr úmrtností v jednotlivých intervalech a označíme je G_1, G_2, G_3

$$G_1 = \sum_{x=60}^{67} m_x, \tag{4}$$

$$G_2 = \sum_{x=68}^{75} m_x, \tag{5}$$

$$G_3 = \sum_{x=76}^{83} m_x. \tag{6}$$

Nyní již můžeme vypočítat hodnotu parametru c z Gompertzovy-Makehamovy funkce, jehož 8. Mocninu lze vyjádřit za pomocí součtu empirických specifických měr úmrtností v jednotlivých intervalech

$$c^8 = \frac{G_3 - G_2}{G_2 - G_1}. \tag{7}$$

Dále je nezbytné vypočítat hodnotu pomocného výrazu, díky níž jsme schopni v následujícím kroku vyjádřit zbylé dva parametry funkce

$$K_c = c^{60,5} \cdot (1 + c + ... + c^7) = c^{60,5} \cdot \frac{c^8 - 1}{c - 1}. \tag{8}$$

Parametry b a a je možné vypočítat na základě následujících výrazů

$$b = \frac{G_2 - G_1}{K_c \cdot (c^8 - 1)}, \tag{9}$$

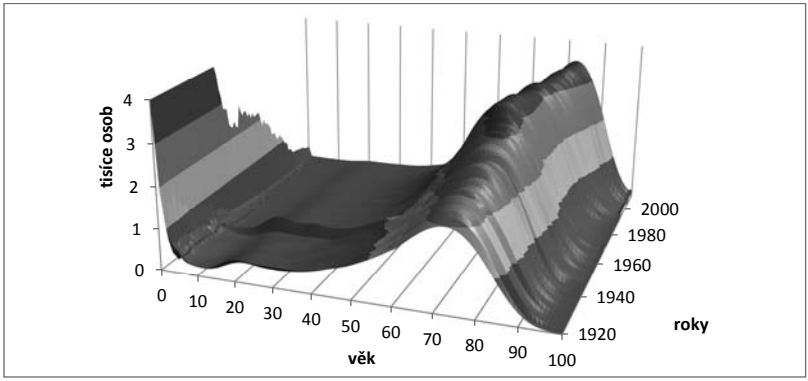
$$a = \frac{G_1 - b \cdot K_c}{8}. \tag{10}$$

A na základě těchto parametrů Gompertzovy.Makehamovy funkce můžeme nyní přesněji vypočítat normální délku života

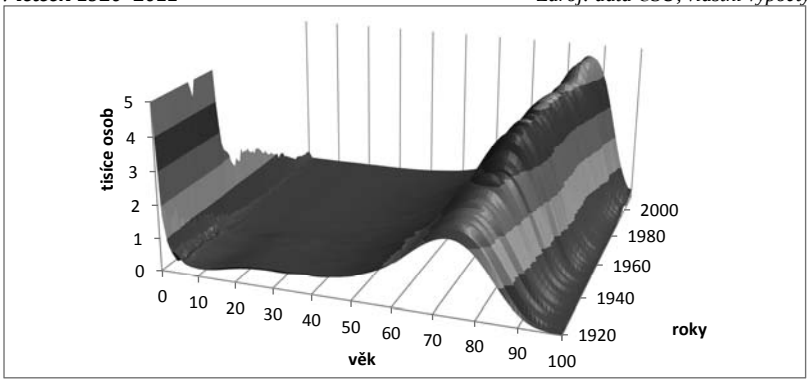
$$\hat{y} = \frac{\ln \dfrac{\ln c - 2a + \sqrt{(\ln c - 4a) \cdot \ln c}}{2b}}{\ln c}. \qquad (11)$$

3. Normální a pravděpodobná délka života v České republice

Na obrázcích č. 1 a 2 je vidět, vývoj počtu tabulkově zemřelých mužů a žen v dokončeném věku *x* let v České republice pro období 1920-2011. V roce 1920 byl ještě vysoký počet zemřelých na počátku života. V dobách, kdy byla vysoká úmrtnost dětí do jednoho roku, vykazovaly řady počtu tabulkových zemřelých dva mody. Normální délka života se uvažuje pro starší věk, tedy je to věk, ve kterém lidé nejčastěji umírají, a nepřihlížíme přitom k nízkému věku. Všechny grafy byly vytvořeny pomocí Excelu verze 2010.



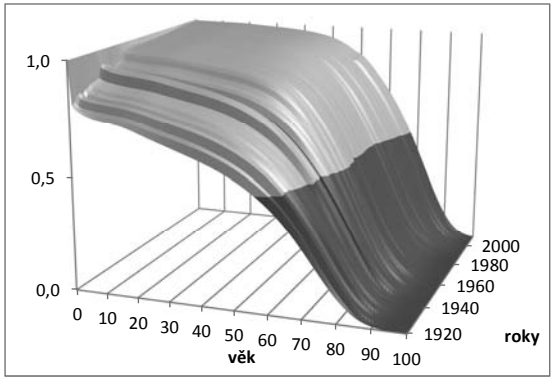
Obr. 7: Vývoj počtu tabulkově zemřelých mužů v dokončeném věku x let v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*



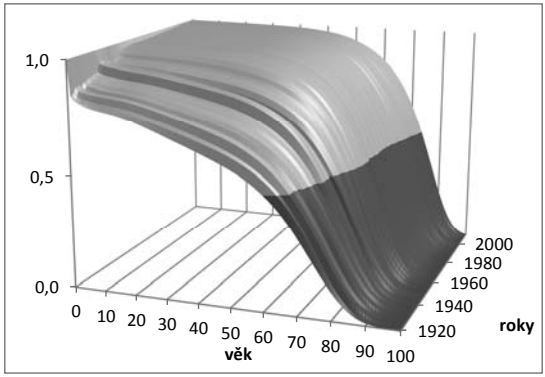
Obr. 8: Vývoj počtu tabulkově zemřelých žen v dokončeném věku x let v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*

V populacích s nízkou úmrtností je modus věku zemřelých jednoznačně ve vysokém věku. Normální délka v čase jak u mužů, tak u žen roste, ale pohybuje se v hodnotách od 72 do 87 let v celé časové řadě. Na rozdíl od střední délky života při narození nejsou tedy rozdíly v hodnotách mezi roky 1920 a 2011 tak velké.

Další využívanou charakteristikou délek života je tzv. pravděpodobná délka života osoby přesně *x*-leté. Pravděpodobná délka života je dána dobou, za kterou zemře právě polovina *x*-letých. Je to medián věku zemřelých starších *x*-let zmenšený o *x*. Pravděpodobná délka života je také často chápána jako poločas života souboru x-letých, doba, kterou potřebuje při dané úmrtnosti populace dožívajících se věku *x*-let k tomu, aby se tato populace snížila na polovinu. Pravděpodobná délka života jako ukazatel typu medián není závislá na extrémech.



Obr. 9: Vývoj počtu mužů dožívajících se přesného věku x let v úmrtnostních tabulkách v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*



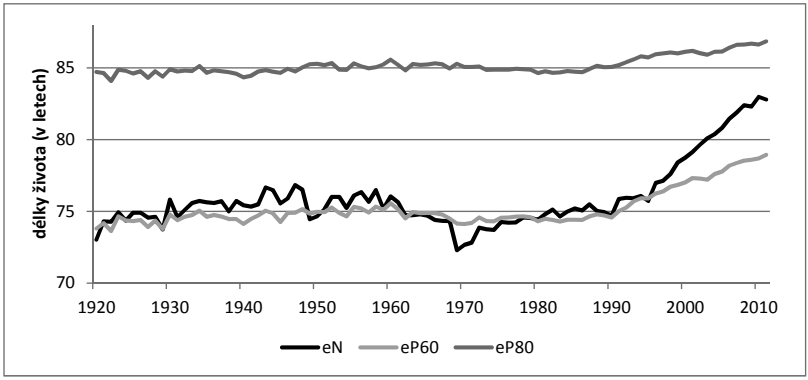
Obr. 10: Vývoj počtu žen dožívajících se přesného věku x let v úmrtnostních tabulkách v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*

Vývoj počtu mužů a žen dožívajících se přesného věku *x* let v úmrtnostních tabulkách České republiky v letech 1920–2011 je znázorněn v obrázcích 3 a 4. V těchto grafech je barevně odlišena polovina osob (medián) v tabulkovém souboru. V roce 1920 byl tedy věk 55,5 let mediánem počtu mužů dožívajících se přesného věku *x* let. V průběhu sledovaného období dochází ke zvyšování mediánu až na hodnotu 77,5 let v roce 2011. U žen byl tento věk v roce 1920 59,0 let a v roce 2011 již 83,5 let.

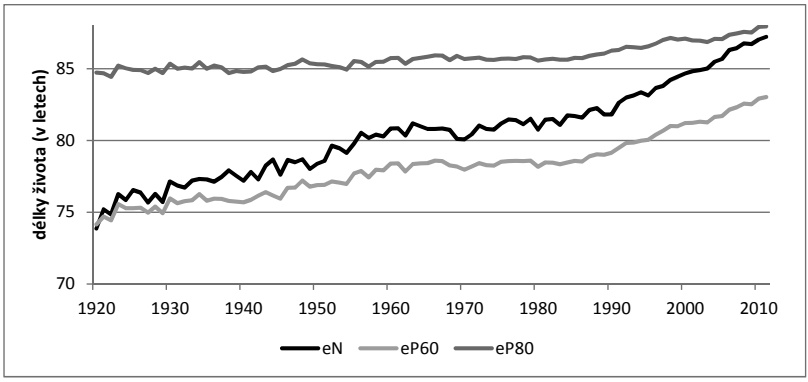
V obrázku č. 5 a 6 je záměrně porovnána normální délka života s pravděpodobným věkem úmrtí osoby v přesném věku 60 a 80 let v letech 1920–2011. Pravděpodobný věk úmrtí je možné vypočítat ze střední délky života v přesném věku *x* let po přičtení již prožitých let. Věk

60 let byl zvolen na ukázku, zda má při stávající úmrtnosti osoba právě 60 letá šanci dožít se normální délky života. Doplněn byl také pravděpodobný věk úmrtí 80 leté osoby. Normální délka života je v grafu značena eN, pravděpodobný věk úmrtí 60 letých eP60 a pravděpodobný věk úmrtí 80 letých eP80.

Střední délka života 60 letého muže za sledované období sice vzrostla, ale ne tak významně, jako střední délka života při narození. Pokud se v roce 1920 muž dožil 60 let, jeho pravděpodobný věk úmrtí byl obdobný jako normální délka života muže v témže roce. Je vidět, že především po roce 1996 se zvyšuje rozdíl mezi normální délkou života a pravděpodobným věkem úmrtí. Je to způsobeno změnou úmrtnostních poměrů ve vyšších věcích a změnami v intenzitě úmrtnosti podle věku. Dnešní muž se musí dožít alespoň věku 72 let, aby jeho pravděpodobný věku úmrtí byl shodný či větší než normální délka života.



Obr. 11: Vývoj normální délky života a pravděpodobného věku úmrtí 60 a 80 letých mužů v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*



Obr. 12: Vývoj normální délky života a pravděpodobného věku úmrtí 60 a 80 letých žen v České republice v letech 1920–2011 *Zdroj: data ČSÚ, vlastní výpočty*

Analogický graf pro ženy (obrázek č. 6) ukazuje, že pokud v roce 1920 měla žena, pokud se dožila 60 let šanci se dožít normální délky života, v následujícím období již to tak není. Pokud se dožije 80 let, tak v posledních letech (2005–2011) již má šanci se normální délky dožít. Normální délka života je v grafu značena eN, pravděpodobný věk úmrtí 60 letých eP60

a pravděpodobný věk úmrtí 80 letých eP80. Dnešní žena se musí dožít alespoň věku 79 let, aby jeho pravděpodobný věku úmrtí byl shodný či větší než normální délka života.

4. Závěr

Díky zlepšujícím se úmrtnostním poměrům populace České republiky zde přibývá starších osob. Dochází tedy ke stárnutí populace. Lze předpokládat, že úmrtnost bude i nadále klesat a přibližovat se úrovni známé z vyspělejších zemí. Do budoucna by mohlo být tedy vhodné kromě tradičních ukazatelů jako je délka života v přesném věku *x* let sledovat také například jaký je trend ve vývoji normální délky života, která je považována za charakteristiku dlouhověkosti či pravděpodobné délky života, která určuje tzv. životní poločas. I když z pohledu statistiky je leckdy vhodnější charakteristika typu medián, tedy v našem případě pravděpodobná délka života, která je velmi jednoduše interpretovatelná a není ovlivněna extrémními hodnotami, průměr lze lehce spočítat a určit medián je složitější. To lze chápat v době, kdy k výpočtům je používána kalkulačka. Dnes samozřejmě nepočítáme ukazatele na kalkulačce, neměl by být problém s výpočtem i ostatních délek života. To, že se v délkách života upřednostňuje právě střední délka života je dáno tradicí i tím, že v mezinárodním srovnání i srovnání v čase se do střední délky života promítá i vysoká či nízká kojenecká úmrtnost. Je tedy otázkou času, kdy se ukazatel střední délky života doplní či nahradí dalšími vhodnými charakteristikami úmrtnosti.

Poděkování

Tento příspěvek vznikl za podpory projektu VŠE IGA 24/2013 "Úmrtnost a stárnutí populace České republiky".

Literatura

Arltová, M., Langhamrová, Jitka, Langhamrová, Jana. *Development of life expectancy in the Czech Republic in years 1920-2010 with an outlook to 2050.* Prague Economic Papers, 2013. roč. 22, č. 1, s. 125–143. ISSN 1210-0455

Český statistický úřad. Dostupný z WWW: <http://czso.cz/>

Dotlačilová, P., Langhamrová, J., Šimpach, O. *Vybrané logistické modely používané pro vyrovnávání a extrapolaci křivky úmrtnosti a jejich aplikace na populace vybraných zemí Evropské unie.* Forum Statisticum Slovacum [online], 2012. roč. 8, č. 7, s. 21–25. ISSN 1336-7420.

Fiala, T.: *Výpočty aktuárské demografie v tabulkovém procesoru.* 1. vyd. Praha: Oeconomica, 2005. 177 s. ISBN 80-245-0821-4.

Koschin, F.: *Jak vysoká je intenzita úmrtnosti na konci lidského života?* Demografie, 1999. roč. 41, č. 2. s. 105-119.

Adresa autora:

Jana Langhamrová, Ing.
Katedra statistiky a pravděpodobnosti,
Fakulta informatiky a statistiky,
Vysoká škola ekonomická
Nám. W. Churchilla 4, 130 67, Praha 3
Xlanj18@vse.cz

Vliv částečných úvazků na flexibilitu trhu práce
Effect of part-time jobs on labour market flexibility

Jana Langhamrová

Abstract: The issue of labour market is a complicated mechanism that can have a serious impact on the whole economic process. If the labour market is flexible, it is undoubtedly able to effectively respond to any fluctuations in the economy and increase its competitiveness. The experience of Western Europe shows that the more flexible labour market has lower unemployment rate and higher labour productivity. With a variety of alternative forms of work we have new opportunities for exploitation of workers. Flexible working hours may reduce labour costs, etc. The paper examined the unemployment rate, percentage of the employed by part-time in various European countries and also reasons to work part-time in 2012.

Abstrakt: Problematika trhu práce je složitý mechanismus, který může mít vážný dopad na celý hospodářský proces. Pokud je trh práce pružný, je bezpochyby schopný efektivněji reagovat na případné výkyvy ekonomiky a zvyšuje svou konkurenceschopnost. Zkušenosti ze západní Evropy ukazují, že čím pružnější je trh práce, tím nižší by měla být nezaměstnanost a vyšší produktivita práce. S různými alternativními formami pracovních úvazků se před námi objevují nové možnosti pro využívání pracovníků, uplatnění flexibilní pracovní doby, možná snížení pracovních nákladů, atd. V příspěvku bude zkoumána míra nezaměstnanosti, podíly zaměstnaných osob na částečný úvazek v jednotlivých evropských státech a také důvody práce na částečný úvazek pro rok 2012.

Key words: flexibility of labour market, part-time job, unemployment rate, Czech Republic, European countries

Klíčová slova: flexibilita trhu práce, částečný úvazek, míra nezaměstnanosti, Česká republika, evropské země

JEL classification: J210, J220

1. Úvod

Na pracovní trh je možné nahlížet z mnoha různých úhlů pohledu. Stejně tak je možné hodnotit jeho flexibilitu na základě různých kritérií. Problematika trhu práce je složitý mechanismus, který může mít vážný dopad na celý hospodářský proces. V důsledku rozdílů mezi nabídkou práce a poptávkou po ní může docházet na pracovním trhu k určité nerovnováze. Pokud je trh práce pružný, je bezpochyby schopný efektivněji reagovat na případné výkyvy ekonomiky a zvyšuje svou konkurenceschopnost.

Zkušenosti ze západní Evropy ukazují, že čím pružnější je trh práce, tím nižší by měla být nezaměstnanost a vyšší produktivita práce. S různými alternativními formami pracovních úvazků se před námi objevují nové možnosti pro využívání pracovníků, uplatnění flexibilní pracovní doby, možná snížení pracovních nákladů, atd. Mezi méně obvyklé formy zaměstnávání můžeme zařadit například práci na částečný úvazek, zkrácenou pracovní dobu, pružnou pracovní dobu, práci z domova nebo sdílení jednoho pracovního místa více pracovníky. Pro osoby dobrovolně zaměstnané některou z výše zmíněných forem úvazku tato situace přináší nové možnosti souladu pracovního a soukromého života, možnosti určitého jiného životního stylu. Mezi tyto skupiny osob bychom nejspíše mohli zařadit studenty a absolventy, matky s malými dětmi, osoby se zdravotními omezeními či dlouhodobě nezaměstnané. Pro ně jsou flexibilnější formy zaměstnání bezpochyby velmi vítané. Hovoříme zde o určitém efektivnějším využívání pracovního potenciálu, který tyto skupiny osob bezpochyby pro pracovní trh mají.

2. Kde získat potřebná data

Podrobné informace o českém pracovním trhu se dlouhodobě získávají z Výběrového šetření pracovních sil (VŠPS), které zajišťuje Český statistický úřad. Toto šetření probíhá již od konce roku 1992 a provádí se ve všech okresech České republiky. Zjišťování probíhá nepřetržitě v průběhu celého roku a hlavním cílem VŠPS je získání pravidelných informací o situaci na trhu práce.

V roce 2002 došlo ke sjednocení formy i obsahu dotazování se standardy Evropské unie. Výsledky VŠPS jsou tedy srovnatelné s výsledky z ostatních zemí Evropy. Mezinárodní označení pro toto zjišťování zní Labour Force Survey (LFS).

Výběrové šetření se provádí za byty, které byly zvoleny pomocí dvoustupňového výběru. Pro jeden konkrétní byt jsou zjišťovány jeho základní identifikační údaje a také údaje o domácnostech, které v daném bytě hospodaří. Dále jsou zkoumány vazby mezi jednotlivými členy domácností a demografické údaje. Podrobněji se šetření věnuje osobám 15 letým a starším, které můžeme označit za obvykle bydlící v daném bytě. Pro tyto osoby se dále zjišťuje ekonomické postavení, vzdělávání, charakteristiky zaměstnání, předchozí pracovní zkušenosti, hledání zaměstnání, atp.

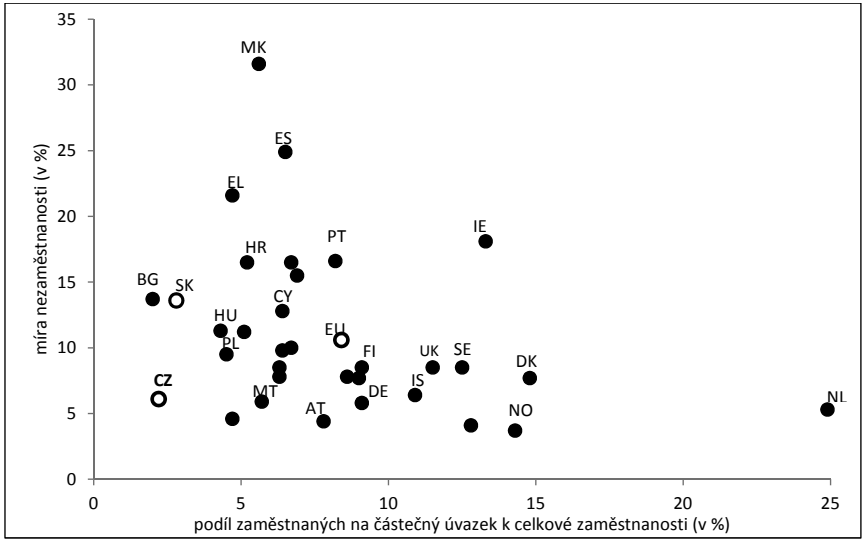
3. Částečné úvazky v evropských zemích

Od 80. let minulého století se většina evropských zemí snaží o postupné zkracování celkové pracovní doby za účelem řešení problému vzrůstající nezaměstnanosti. Jak jsou na tom ostatní evropské země a v kontextu s nimi i Česká republika si ukážeme dále. Do srovnání bude zařazeno 28 členských států Evropské unie, Island, Makedonie, Norsko, Švýcarsko a Turecko.

Nemělo by zřejmě příliš velký smysl zde uvádět počty částečných úvazků v tisících pro jednotlivé země bez toho, aby tato informace byla dána do kontextu s dalšími znalostmi o trhu práce dané země. Proto bude dále zkoumán podíl částečných úvazků na celkové zaměstnanosti mužů a žen ve věku 15-64 let v evropských zemích v roce 2012.

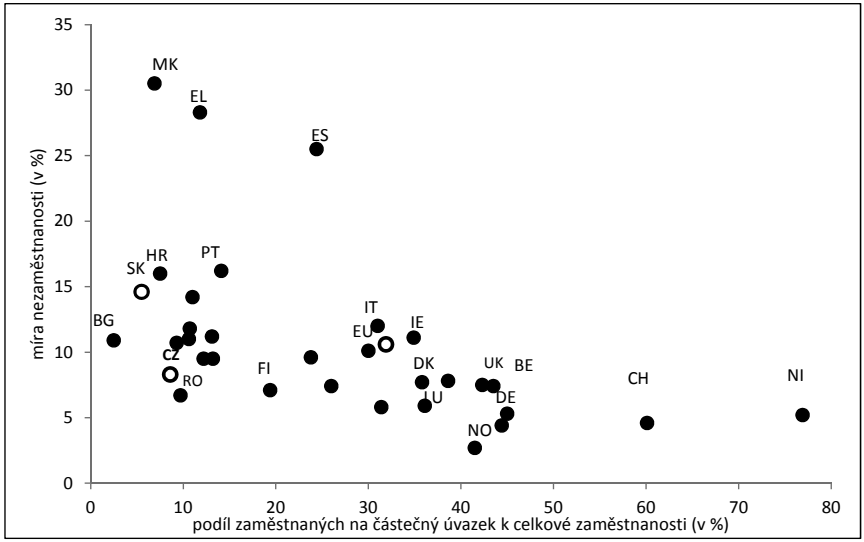
Jak již bylo dříve zmíněno, země s vyšším podílem osob zaměstnaných na zkrácený úvazek, mívají nižší míru nezaměstnanosti. Na obrázcích č. 1 a 2 je názorně zobrazeno rozmístění jednotlivých zemí vzhledem k hodnotám míry nezaměstnanosti a podílu zaměstnaných na částečný úvazek k celkové zaměstnanosti pro muže a ženy, jakých dosahovaly tyto země v roce 2012. Popisek pro danou zemi je uvedeny vždy nad značkou zobrazující hodnoty dané země či pomocí šipky. V obrázcích nejsou uvedeny popisky pro všechny země z důvodu přehlednosti a dále Česká republika, Slovensko a Evropská unie mají odlišné typy značek pro jejich snadnější identifikaci. Země, které se nacházejí blízko pravého dolního rohu, se obecně vyznačují nízkou mírou nezaměstnanosti a naopak vyšším podílem osob zaměstnaných na částečný úvazek k celkovému počtu zaměstnaných osob ve věku 15-64 let. Naopak země nacházející se v levém horním rohu se vyznačují vysokou mírou nezaměstnanosti a nízkým podílem osob zaměstnaných na částečný úvazek.

U mužů (viz obrázek č. 1) je nejvíce vpravo Nizozemsko, které se výrazně vzdálilo od ostatních zemí. Jedná se tedy o zemi, kde je v současnosti relativně nízká míra nezaměstnanosti mužů ve věku 15-64 let (5,3 %) a zároveň vysoký podíl mužů zaměstnaných na částečný úvazek ve stejné věkové kategorii (téměř 25 %). Dalšími zeměmi s nízkou hodnotou míry nezaměstnanosti mužů (do 5 %) a vyšším podílem pracujících na částečný úvazek (nad 12 %) je Norsko a Švýcarsko. Hodnoty zkoumaných ukazatelů za muže v Evropské unii jsou pro míru nezaměstnanosti 8,4 % a pro podíl zaměstnaných na částečný úvazek 10,6 %. Česká republika se nachází mezi zeměmi s nejnižším podílem mužů pracujících na částečný úvazek (2,2 %) a zároveň je, v porovnání s ostatními státy Evropy, zemí s vcelku nízkou mírou nezaměstnanosti mužů (6,1 %).



Obr. 13*: *Země Evropy zobrazené podle podílu zaměstnaných na částečný úvazek k celkové zaměstnanosti mužů a míry nezaměstnanosti mužů ve věku 15-64 let v roce 2012

Zdroj: data LFS, Eurostat



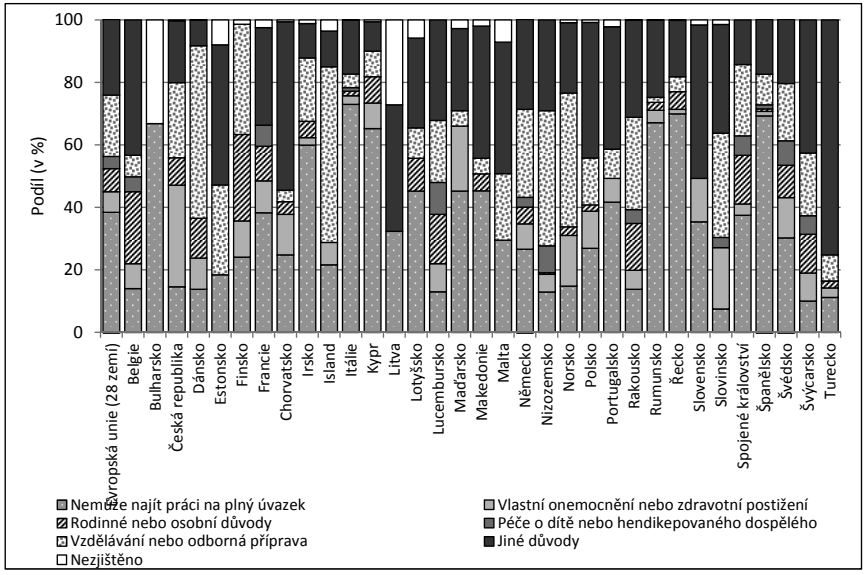
Obr. 14: Země Evropy zobrazené podle podílu zaměstnaných na částečný úvazek k celkové zaměstnanosti žen a míry nezaměstnanosti žen ve věku 15-64 let v roce 2012

Zdroj: data LFS, Eurostat

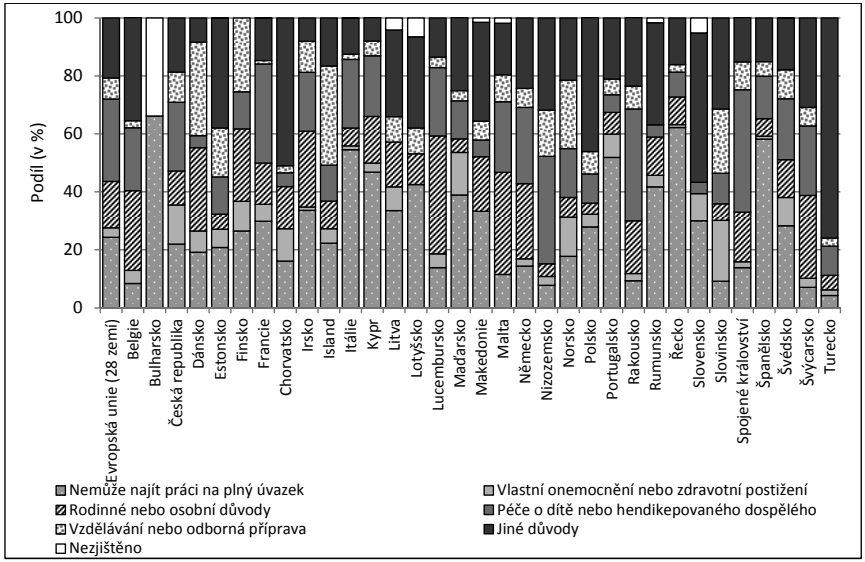
Pro ženy ve věku 15-64 let (viz obrázek č. 2) platí stejně jako pro muže ve stejném věkovém intervalu, že je zde v současnosti relativně nízká míra nezaměstnanosti mužů ve věku 15-64 let (5,2 %) a zároveň vysoký podíl žen zaměstnaných na částečný úvazek ve stejné věkové kategorii (téměř 77,0 %). Další zemí s vysokým podílem žen zaměstnaných na částečný úvazek a nízkou mírou nezaměstnanosti bylo v roce 2012 Švýcarsko. Evropská unie je pro ženy charakterizována hodnotami 10,6 % míry nezaměstnanosti a téměř 32,0 % podílem žen pracujících na částečný úvazek. Ženy v České republice mají oproti mužům vyšší podíl zaměstnaných na částečný úvazek (8,6 %), ale také míru nezaměstnanosti (8,3 %). I v tomto případě platí, že se Česká republika nachází mezi zeměmi s nejnižším podílem žen pracujících na částečný úvazek a zároveň je, v porovnání s ostatními státy Evropy.

Obecně tedy můžeme hodnotit Nizozemsko, Švýcarsko, Norsko, Dánsko, Irsko, Německo a Spojené království jako země s vysokým podílem osob zaměstnaných na částečný úvazek. Tyto země mají oproti ostatním zemím v Evropě flexibilnější pracovní trh z pohledu částečných úvazků.

Nyní se zaměříme na důvody zaměstnání mužů a žen na částečný úvazek v zemích Evropy. Bohužel nebyly zjištěny odpovědi pro všechny evropské země. Na obrázku č. 3 jsou zobrazeny odpovědi mužů ve věku 15-64 let zaměstnaných na částečný úvazek v roce 2012. Podíl mužů v EU, kteří pracují na částečný úvazek, protože nemohou najít práci na celý, tvořil téměř 40 %. Podíl mužů, kteří uváděli, jako důvod svůj zdravotní stav byl 6,6 %. Podíl mužů, kteří chtějí částečný úvazek skloubit se vzděláváním, a odbornou přípravou reprezentoval 20 % odpovědí. Odpověď rodinné a osobní důvody zvolilo téměř 7,5 % mužů v Evropské unii. V Bulharsku, Itálii, Rumunsku, Řecku, Španělsku či na Kypru uvedlo více než 65 % mužů, že hlavním důvodem jejich práce na částečný úvazek je nenalezení práce na úvazek plný. V Dánsku a na Islandu tvoří více než polovinu odpovědí vzdělávání nebo odborná příprava.



Obr. 15: Hlavní důvod zaměstnaní na částečný úvazek u mužů ve věku 15-64 let v zemích Evropy v roce 2012 *Zdroj: data LFS, Eurostat*



Obr. 16: Hlavní důvod zaměstnaní na částečný úvazek u žen ve věku 15-64 let v zemích Evropy v roce 2012 *Zdroj: data LFS, Eurostat*

Hlavní důvody práce na částečný úvazek žen ve věku 15-64 let jsou zobrazeny na obrázku č. 4. V EU byl podíl žen, které pracují na částečný úvazek, protože nenašly práci na plný 24,3 %. Podíl odpovědí, kde hlavním důvodem byla péče o jinou osobu, tvořil 28,4 %. Rodinné nebo osobní důvody byly u žen v EU zastoupeny v 16 % na celkovém podílu odpovědí. V Bulharsku a Řecku bylo hlavním důvodem z více než 60 % nenalezení práce na plný úvazek. V České republice byla zastoupena tato odpověď ve 22 % a tvoří tak druhý nejčastější důvod. Na prvním místě byla pro ženy v ČR v roce 2012 péče o jiné osoby (23,7 %). Rodinné a osobní důvody tvořily necelých 12 %, což je pod průměrem Evropské unie v tomto roce.

4. Závěr

Mezi jednotlivými zeměmi Evropy jsou značné rozdíly v podílu osob zaměstnaných na částečný úvazek. U žen je tento podíl výrazně vyšší než u mužů. Česká republika se nachází výrazně pod průměrem evropských států. Na českém pracovním trhu existují veliké rezervy v možném pracovním potenciálu hlavně u mladých, matek s dětmi a osob v předdůchodovém a důchodovém věku. Pokud by se v České republice podařilo nabídnout mladým matkám odpovídající částečné úvazky tak, aby mohly skloubit pracovní i rodinný život dle svých očekávání a potřeb, mohla by tato forma zaměstnávání mladých matek napomoci i mírnému zvýšení plodnosti. Přínosem pro trh práce mohou být již uvedené částečné pracovní úvazky pro studenty, kteří tak mohou skloubit práci v daném oboru a studium, což následně zvyšuje jejich uplatnitelnost na trhu práce a přispívá ke zvyšování ekonomického bohatství státu. Vzhledem k tomu, že se také prodlužuje střední délka života[1] a roste tzv. zdravá délka života[2],

[1] Také se jí říká naděje dožití a vyjadřuje počet roků, které v průměru ještě prožije osoba právě x-letá za předpokladu, že po celou dobu jejího dalšího života se nezmění řád vymírání, zjištěný úmrtnostními tabulkami. Jedná se tedy o hypotetický údaj, který říká, kolika let by se člověk určitého věku dožil, pokud by úroveň a struktura úmrtnosti zůstala stejná jako v daném roce

lze předpokládat, že lidé v předdůchodovém a důchodovém věku budou v budoucnu i v České republice více nabízet práci na částečný úvazek ve snaze se déle uplatnit na trhu práce. Politika zaměstnanosti, pracovně-právní legislativa, postoj vlády i to, jak bude vypadat sociální a daňový systém v České republice může mít vliv na fakt, zda zaměstnavatelé budou ochotni vytvořit více pracovních pozic na částečný pracovní úvazek.

Poděkování

Tento příspěvek vznikl za podpory projektu VŠE IGA 19/2012 " Flexibilita trhu práce České republiky".

Literatura

BLACKWELL, J. Social Security and Social Change, Harvester, Wheatsheaf. *Changing Work Patterns and their Implications for Social Protection*. 1994.
GOUDSWAARD, A. a M. DE NANTEUIL. LUXEMBOURG: OFFICE FOR OFFICIAL PUBLICATIONS OF THE EUROPEAN COMMUNITIES. *Flexibility and Working Conditions A Qualitative and Comparative Study in Seven EU Member States* [online]. 2000 [cit. 2013-11-07]. ISBN 92-828-9767-2. Dostupné z: http://www.eurofound.europa.eu/pubdocs/2000/07/en/1/ef0007en.pdf
Analýza flexibilních forem zaměstnávání a organizace pracovní doby v České republice. In: *Výzkumný ústav práce a sociálních věd* [online]. 2004 [cit. 2013-11-07]. Dostupné z: http://www.equalcr.cz/files/clanky/910/analyza_flexibilni_formy_zamestnavani_v_CR.pdf
Databáze Eurostat. *Eurostat* [online]. 2013 [cit. 2013-11-07]. Dostupné z: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database
Kulatý stůl: Slaďování pracovního a rodinného života. ČSÚ. *Český statistický úřad* [online]. 2013 [cit. 2013-11-07]. Dostupné z: http://www.czso.cz/csu/tz.nsf/i/kulaty_stul_sladovani_pracovniho_a_rodinneho_zivota20120522
Nové šance a rizika: flexibilita práce, marginalizace a soukromý život u vybraných povolání a sociálních skupin. 1. vyd. Editor Radka Dudová. Praha: Sociologický ústav Akademie věd ČR, 2008, 308 s. ISBN 978-807-3301-385.
Recese zhoršuje postavení mladých na trhu práce. *Český statistický úřad* [online]. 2011 [cit. 2013-11-07]. Dostupné z: http://www.czso.cz/csu/tz.nsf/i/recese_zhorsuje_postaveni_mladych_na_trhu_prace20111123
Veřejná databáze ČSÚ. ČSÚ. *Český statistický úřad* [online]. 2013 [cit. 2013-11-07]. Dostupné z: http://vdb.czso.cz/vdbvo
Výběrové šetření pracovních sil (VŠPS). *Český statistický úřad* [online]. 2013 [cit. 2013-11-07]. Dostupné z: http://www.czso.cz/vykazy/vykazy.nsf/i/vyberove_setreni_pracovnich_sil

Adresa autora:

Jana Langhamrová, Ing.
Katedra statistiky a pravděpodobnosti,
Fakulta informatiky a statistiky,
Vysoká škola ekonomická
Nám. W. Churchilla 4, 130 67, Praha 3
Xlanj18@vse.cz

[2] Uvádí průměrný počet let, které osoby určitého věku ještě prožijí bez zdravotního omezení

Chování studentů v procesu přijímacího řízení na vysoké školy v ČR
Behavior of the Students in the Admission Process to Universities in CR

Bohdan Linda, Jana Kubanová

Abstract: The system of the university education has fundamentally changed after 1989. The number of universities students has grown greatly. This fact influences the behavior of these students. This paper maps behavior of the students during the admission process to the bachelor studies.

Abstrakt: Po roce 1989 se zásadně změnil systém vysokého školství. Počet studentů na vysokých školách mnohonásobně vzrostl. Tato skutečnost ovlivňuje i chování těchto studenů. Příspěvek mapuje chování studentů při přijímacím řízení do bakalářského studia.

Key words: Graduates, admission process, the number of applications submitted, the number of applicants, the number of accepted, the number of registered.

Klíčová slova: Absolventi, přijímací řízení, počet podaných přihlášek, počet uchazečů, počet přijatých osob, počet zapsaných osob.

JEL classification: I21

1. Úvod

Změny, které nastaly po roce 1989, se dotkly všech oblastí našeho života, školství nevyjímaje. Některé změny vedly ke zlepšení, některé ke zhoršení situace v dané oblasti. Co se týče školství, lze jednoznačně říci, že kvalita absolventů vysokoškolského studia se rapidně zhoršila a i v současnosti se soustavně zhoršuje. Zásadní podíl na tomto stavu měl a v současnosti ještě stále má neprincipiální vznik nových vysokých škol a fakult. Ustupuje se tlaku uchazečů směrem k méně náročným studiím, jako jsou především některá humanitní studia, na úkor technických oborů. Neúměrně se zvyšují počty studentů na těchto oborech, což způsobuje vzhledem ke klesajícímu trendu populace nezájem o obory technické. Nekontrolovaný nárůst ekonomických vysokých škol a fakult vyvolal zápas o studenty, který bohužel nespočívá v soutěži o poskytování kvalitního vzdělání, ale naopak v nabídce snadného zisku příslušného titulu. Zastánci „nekontrolovaného“ vysokoškolského systému tvrdili a tvrdí, že praxe vyselektuje kvalitní vysoké školy a přesměruje zájemce z řad studentů na tyto školy. Bohužel ani po více než dvaceti letech se tak nestalo. O žalostném stavu vysokého školství svědčí i povzdech předsedy svazu průmyslu při svém televizním vystoupení, ve kterém si stěžoval, že v ČR je nedostatek kvalitních inženýrů (a nejen inženýrů, ale i technických pracovníků s výučním listem). O kvalitě absolventů nejen technických vysokých škol ví každý vysokoškolský učitel, vyučující před rokem 1989, své.

V tomto příspěvku se zabýváme zkoumáním chování uchazečů o vysokoškolské studium.

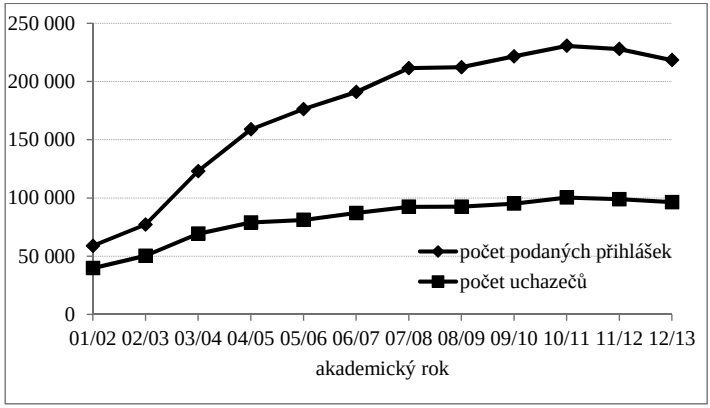
2. Změny v systému vysokého školství v České republice

Systém vysokého školství po roce 1989 doznal hlubokých změn. Vysoké školy a fakulty už nevychovávají studenty zaměřené pouze na oblasti, uváděné v jejich názvech, ale nabízejí i různé studijní programy, které nemusí mít s původním zaměřením školy či fakulty mnoho společného. Vzniklo mnoho soukromých vysokých škol, především ekonomického zaměření. Zavedl se systém třístupňového vysokoškolského vzdělání. V důsledku těchto změn se několikanásobně zvýšil počet studentů. Tyto změny a v důsledku těchto změn několikanásobně zvýšený počet studentů zásadním způsobem ovlivnily i chování studentů, a to jak v průběhu přijímacího řízení, tak i v průběhu samotného studia. Zatímco např. před rokem 1989 se mohl uchazeč o studium na vysoké škole v prvním kole zúčastnit pouze jednoho přijímacího řízení, v současnosti se může student v prvním kole zúčastnit přijímacího

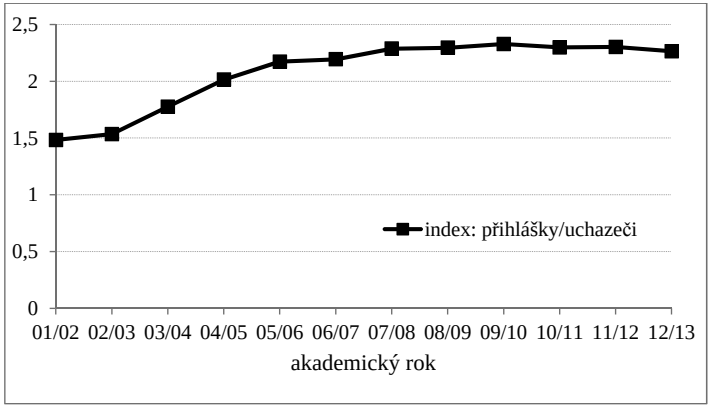
řízení na libovolném počtu vysokých škol. Chování studentů v průběhu přijímacího řízení v období po roce 1989 je popsáno v následující části.

3. Chování studentů v průběhu přijímacího řízení

Možnost, že se student může přihlásit na libovolný počet škol vede k tomu, že počet přihlášek na vysoké školy daleko překračuje počet uchazečů. Vývoj absolutních ukazatelů těchto počtů je uveden na obrázku 1. Poměr těchto dvou ukazatelů na obrázku 2 ukazuje, že zatímco v roce 2001/02 na jednoho studenta připadalo v průměru 1,48 přihlášky, v roce 2009/10 to bylo již 2,33. Z obou obrázků lze vidět, že přibližně v tomto roce nastala jistá stabilizace v počtu podaných přihlášek i v počtu podaných přihlášek na jednoho studenta.



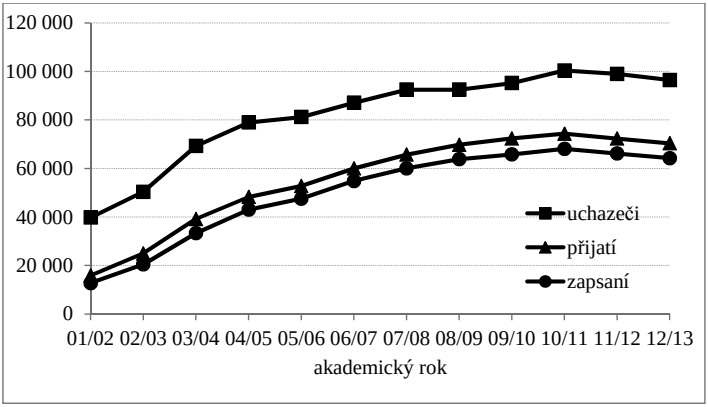
Obr.1: Počet podaných přihlášek do bakalářského studia - prezenční formy a počet uchazečů



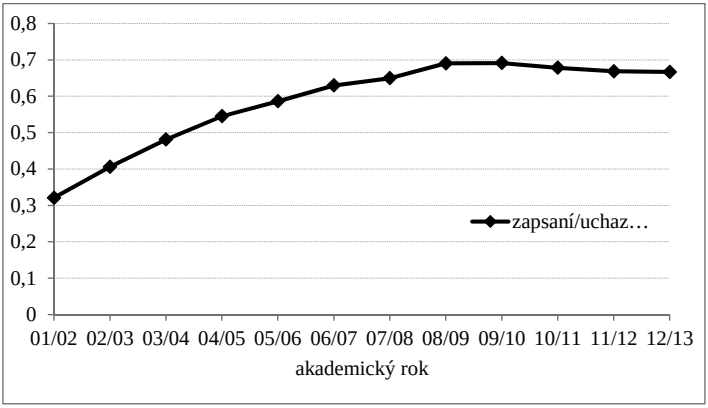
Obr.2: Poměr počtu podaných přihlášek do bakalářského studia - prezenční formy a počtu uchazečů

Zajímavou informaci nám dávají obrázky 3 a 4, které porovnávají počty uchazečů o vysokoškolské studium s počty přijatých a zapsaných studentů. Zatímco počet zapsaných

přibližně kopíruje počet přijatých, vidíme, že v absolutních počtech mezi uchazeči a přijatými je pořád přibližně stejný rozdíl. Podíváme-li se však na obrázek 4 vidíme, že poměr mezi počtem zapsaných a uchazečů narůstá poměrně rychlým tempem. Během 8 let se více než zdvojnásobil. Tento vysoký nárůst nelze vysvětlit zlepšením vědomostí, ale extenzivním zvyšováním počtu míst na vysokých školách a snižováním náročnosti přijímacího řízení. A bohužel právě nízké nároky na přijetí studenta zvyšují počet méně kvalitních vysokoškoláků, což v konečném důsledku vede ke snížení úrovně vysokoškolského studia.



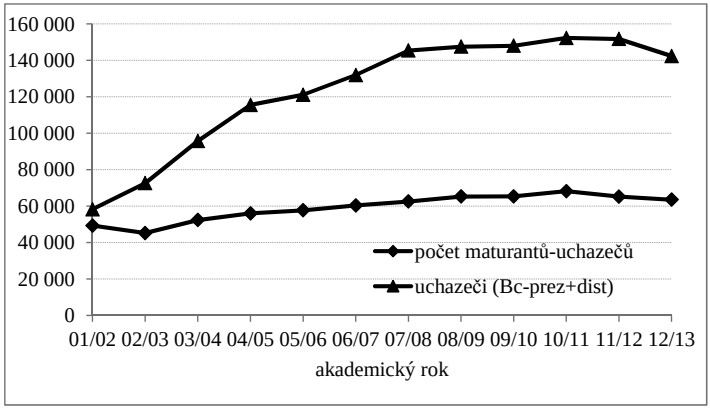
Obr. 3: Počet uchazečů, přijatých a zapsaných do bakalářského stupně studia - prezenční formy



Obr. 4: Poměr počtu zapsaných studentů a uchazečů do bakalářského stupně studia - prezenční formy

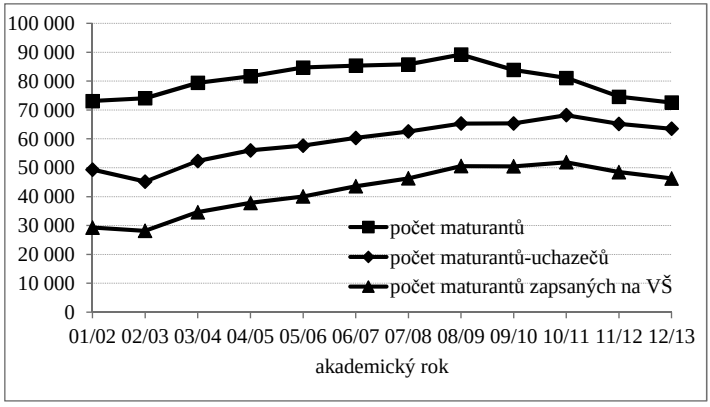
Obrázek 5 nás informuje o podílu čerstvých absolventů (studentů, kteří v roce jejich maturity podali přihlášku na vysokou školu) na uchazečích do vysokoškolského studia. Ze zmíněného obrázku plyne, že počet uchazečů rychle narůstá (z hodnoty 58 231v roce 2001/02

vzrostl na přibližně 2,5 násobek v roce 2007/08), zatímco počet čerstvých absolventů rostl velmi pomalu. Uvědomíme-li si, že v roce 1999 byla podepsána Boloňská dohoda, lze vysvětlit rychlý nárůst uchazečů tím, že mnoho zájemců o vysokoškolské studium dříve maturujících a netroufajících si na úplné vysokoškolské vzdělání před tímto rokem, se přihlásilo na studium bakalářské, které se oficiálně považuje za vysokoškolské.



Obr. 5: Počet uchazečů do bakalářského studia a počet čerstvých maturantů-uchazečů

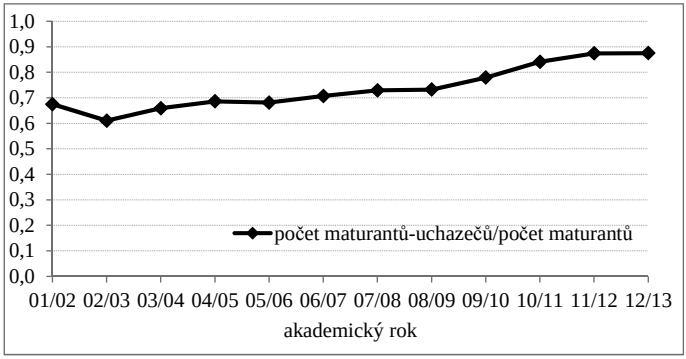
Poslední obrázek 6 dává vizuální představu o podílu čerstvých maturantů, ucházejících se o bakalářské studium a přijatých na vysokou školu. Z obrázku 6 vidíme, že od roku 2008/09 počet maturantů klesá, což je přirozený důsledek poklesu porodnosti po roce 1990.



Obr. 6: Počet čerstvých maturantů, maturantů - uchazečů a maturantů přijatých na VŠ

I když současně s tím klesá i absolutní počet maturantů-uchazečů, což je přirozený jev, obrázek 7 nás informuje, že soustavně narůstá procento maturantů – uchazečů (jakoby se po roce 1990 přestali rodit horší středoškoláci). Vzhledem k tomu, že střední školství v uvedených letech nezaznamenalo žádný pozitivní kvalitativní posun (viz nadnárodní

hodnocení kvality středoškoláků, ve kterém se neustále propadáme), znamená to pouze nezdravý nárůst sebevědomí absolventů středních škol.



Obr. 7: Počet maturantů-uchazečů/ počet maturantů

4. Závěr

Cílem příspěvku bylo popsat chování studentů se středoškolským vzděláním v přijímacím řízení na vysoké školy do bakalářského studia po podpisu Boloňských dohod (vznik bakalářského studia). Z uvedeného chování lze dedukovat, že mnoho studentů nemá zájem o studium konkrétního studijního oboru, ale jejich cílem je pouze získání jakéhokoli vysokoškolského diplomu. Jinak by nemohlo při současné úrovni přijímacího řízení více než 30% studentů podávat přihlášky na 3 a více vysokých škol. Dokonce jsou i tací studenti, kteří podávají přihlášky na 15 vysokých škol. Na druhé straně však nelze z tohoto stavu vinit jenom studenty. Ti se chovají pouze tak, jak jim pravidla povolují. Závěr je tedy ten, že české školství (a to nejen vysoké) se po roce 1989 vydalo špatným směrem. Stále více lze dedukovat snahu o komercializaci školství (neúměrný nárůst soukromých škol, neúměrný nárůst počtu studentů, špatná proporcionalita mezi studijními programy a obory atd.), která ze vzdělání dělá obchod. Je tedy nejvyšší čas, aby kompetentní lidé vyslyšeli hlasy seriózních pedagogů na všech stupních vzdělání a přikročili k nápravě. Aby předseda svazu průmyslu nemohl říci, že v národě, majícím 30, případně více procent vysokoškolsky vzdělaných občanů nejsou inženýři.

Literatura

http://www.msmt.cz/vzdelavani/skolstvi-v-cr/statistika-skolstvi/prijimaci-rizeni-ke-studiu-na-vysoke-a-vyssi-odborne-skole-1

Adresa autorů:

Bohdan Linda, doc., RNDr.,CSc.
Univerzita Pardubice,
Fakulta ekonomicko-správní
Studentská 95, 532 10 Pardubice
bohdan.linda@upce.cz

Jana Kubanová, doc., PaedDr., CSc.
Univerzita Pardubice,
Fakulta ekonomicko-správní
Studentská 95, 532 10 Pardubice
jana.kubanova@upce.cz

Hodnocení výsledků shlukování v ekonomických úlohách
Evaluation of Clustering in Economics problems

Tomáš Löster, Tomáš Pavelka

Abstract: Cluster analysis is a popular multivariate statistical method, which aims is to classify objects into clusters. Objects are characterized by different variables. The basic requirement is that the objects within the clusters are the most similar and objects from different clusters are the least similar. For clustering of objects we can use different methods and different metrics of distances (similarities). Choice of methods and metrics plays a key role. There are no rules which strictly define using of methods and metrics under specific conditions. Another important role is determine the number of clusters. There are many coefficients for evaluating of clustering. The aim of this article is to provide a brief overview how to proceed in evaluating of clustering, respectively, determine the number of clusters, including examples of clustering EU countries. We used economic data from EUROSTAT.

Abstrakt: Shluková analýza je oblíbená vícerozměrná statistická metoda, jejímž cílem je klasifikace objektů do shluků. Objekty jsou charakterizovány pomocí různých proměnných. Základním požadavkem je, aby objekty uvnitř shluků byly co nejpodobnější a objekty z rozdílných shluků co nejméně podobné. Ke shlukování objektů lze použít různé metody a různé metriky vzdáleností (podobností) objektů. Volba metody a metriky hraje klíčovou roli. Zároveň však neexistují striktní pravidla, která definovala, kterou metodu za jakých podmínek použít. Neméně důležitou roli hraje stanovení počtu shluků. Pro hodnocení shlukování existuje řada koeficientů. Cílem tohoto článku je poskytnout stručný přehled, jak postupovat při hodnocení shlukování, resp. stanovení počtu shluků, včetně příkladu shlukování zemí EU s využitím dat z trhu práce pocházející databáze EUROSTAT.

Key words: Cluster analysis, clustering methods, validity measure, labour market

Klíčová slova: Shluková analýza, metody shlukování, míry platnosti, trh práce

JEL classification: C 30, C 38, E20

1. Úvod

Shluková analýza je oblíbená, velmi často používaná vícerozměrná statistická metoda, jejíž cílem je klasifikace objektů do skupin, tzv. shluků. Svědčí o tom řada článků a studií, viz například Megyesiová (2005, 2006, 2011), Řezanková (2009, 2010), Stankovičová (2007) atd. Objekty, které jsou shlukovány, mohou představovat firmy, podniky, zákazníky, kraje, okresy či země EU. Základním požadavkem je, aby si objekty uvnitř jednotlivých shluků byly co nejvíce podobné a objekty ze dvou rozdílných shluků co nejméně podobné. Ke shlukování objektů lze použít různé metody, přičemž výběr konkrétní metody je na výzkumníkovi. Při volbě metody výzkumník musí také zvolit metriku, s jejíž pomocí budou měřeny vzdálenosti resp. podobnosti objektů. Volba metody a příslušné metriky hraje klíčovou roli při shlukování. Velkou roli v této oblasti hrají zkušenosti příslušného výzkumníka, protože neexistují striktní pravidla, která by jasně definovala, kterou metodu za jakých podmínek má výzkumník použít. Neméně důležitou roli hraje stanovení (optimálního) počtu shluků, do kterých budou objekty klasifikovány. Pro stanovení počtu shluků, stejně tak jako pro hodnocení jednotlivých metod shlukování a jejích výsledků existuje řada koeficientů (kritérií). Cílem tohoto článku je poskytnout stručný přehled, jak postupovat a rozhodovat se při hodnocení shlukování, včetně praktického příkladu, který se týká shlukování zemí EU s využitím dat o trhu práce. Data pocházejí z databáze EUROSTAT.

2. Metody shlukové analýzy

V současné vědecké literatuře jsou uváděny různé způsoby klasifikace metod shlukové analýzy. Mezi nejčastěji používané členění „tradičních" metod, které je uváděné ve většině zdrojů, je členění na *hierarchické* a *nehierarchické* metody shlukování.

Hierarchické shlukování představuje takové způsoby shlukování, jejichž postup směřuje k k vytváření stromovité struktury shluků. Jejich výstupem je mimo jiné tzv. *dendrogram*, který představuje což je grafické znázornění procesu shlukování v závislosti na zvolených metrikách. Důležitou vlastností hierarchických metod shlukování je skutečnost, že výsledky předešlého kroku jsou vždy přiřazeny k získaným výsledkům v následujícím kroku a je tak vytvářena stromová struktura. Výhodou hierarchických metod je, že není nutné dopředu znát počet shluků, což je považováno za jejich hlavní výhodu oproti nehierarchickým metodám shlukování. Jsou relativně rychlé, avšak nejsou vhodné pro rozsáhlé datové soubory.

Nehierarchické shlukování se nesoustřeďují na tvorbu dendrogramu, ale soustřeďují se na zařazování objektů do předem známého počtu shluků. Nejprve je potřeba stanovit počáteční rozklad objektů do shluků a pak pomocí iteračních postupů a metod původní rozklad zlepšovat. U této skupiny metod při postupném zlepšování rozkladu objektů může dojít k přeřazení objektu z jednoho shluku do druhého. Kvalita těchto metod závisí zejména na schopnosti uživatele vybrat počáteční rozklady.

Aplikace různých metod shlukování na stejné objekty popsané identickými vlastnostmi mohou přinášet různé výsledky. Jak se uvádí v Gan (2007), Halkidi (2002), „Nelze apriori říci, která z metod je nejlepší pro daný problém. Obyčejně platí, že metoda nejbližšího ouseda je nejméně vhodná a metoda průměrné vzdálenosti, resp. Wardova metoda vyhovují v mnoha případech nejlépe.". Vždy však platí, že je třeba využít také praktické zkušenosti výzkumníka s daným typem úlohy.

Mezi metody hierarchického shlukování lze zařadit například Metodu nejbližšího souseda, metodu nejvzdálenějšího souseda, metodu průměrné vzdálenosti, centroidní metodu.

Metoda nejbližšího souseda byla poprvé popsána, v roce 1957 P. H. A. Sneathem. Jedná se o nejstarší a nejjednodušší metodu. Při této metodě se hledají dva objekty, mezi kterými je nejkratší vzdálenost a spojí se do shluku. Další shluk je vytvořen připojením třetího nejbližšího objektu. Vzdálenost mezi dvěma shluky je definována jako nejkratší vzdálenost libovolného bodu ve shluku vůči libovolnému bodu v jiném shluku, viz Gan (2007). Jako zásadní nevýhoda této metody je uváděno, že dochází k tzv. *řetězení*, kdy do jednoho shluku mohou být zařazeny dva objekty, které jsou sice nejbližší, nicméně vzhledem k většině ostatních objektů nejbližší objekty nejsou.

Metoda nejvzdálenějšího souseda je založena na opačném principu, než metoda nejbližšího souseda. Jejím autorem je Sörensen. Je založena na spojování těch shluků, jejichž vzdálenost mezi nejvzdálenějšími objekty je minimální. Výhodou této metody je, že vytváří malé, kompaktní a dobře oddělené shluky. Oproti metodě nejbližšího souseda zde nevzniká problém s řetězením shluků.

Při **metodě průměrné vzdálenosti** kritérium pro vznik shluků představuje průměrnou vzdáleností všech objektů v jednom shluku ke všem objektům v druhém shluku. Výsledky této metody nejsou ovlivněny extrémními hodnotami, jako u metody nejbližšího a nejvzdálenějšího souseda. Vznik shluku je zde závislý na všech objektech. Dva shluky se spojí do nového shluku, pokud je mezi nimi minimální průměrná vzdálenost.

Centroidní metodu poprvé použili Sokal a Michener. pod názvem „weighted group method". Pro vyjádření nepodobnosti shluků se používá euklidova vzdálenost jejich těžišť (centroidů). Tato metoda nepoužívá mezishlukové vzdálenosti objektů. Do nového shluku se spojují takové dva shluky, mezi kterými je minimální vzdálenost jejich centroidů, přičemž jako

centroid je chápán jako průměr proměnných v jednotlivých shlucích. Výhodou této metody je, že není tak významně ovlivňována odlehlými objekty U této metody se mohou objevit také tzv. *zmatečné shluky*, což znamená, že vzdálenost mezi těžišti jednoho páru je menší, než vzdálenost mezi těžišti jiného páru utvořeného v předešlém kroku.

Mediánová metoda byla poprvé uvedena Gowerem pod názvem „unweighted group method". Cílem této metody je snaha odstranit nedostatek centroidní metody, viz výše. Gower konstatoval, že „... rozdílné počty objektů shluků způsobí rozdílnou váhu prvních dvou složek rekurzivního předpisu centroidní metody a tak se stává, že vlastnosti malých shluků se ve výsledném sjednocení ztrácejí". Mediánová metoda je obdobou centroidní metody a rozdíl spočívá v tom, že místo vzdáleností mezi centroidy shluků používá vzdálenost mezi mediány těchto shluků. Do jednoho shluku jsou spojeny ty dva shluky, mezi jejichž mediány je nejmenší vzdálenost. Výhoda této metody spočívá v odstranění rozdílné váhy, která je v centroidní metodě přiřazována různě velkým shlukům.

Wardova metoda řeší princip shlukování odlišným způsobem, než výše uvedené metody, které se zabývají optimalizací vzdáleností mezi jednotlivými shluky. Metoda se zabývá minimalizací heterogenity shluků, tj. shluky se vytváří pomocí maximalizace vnitroskupinové homogenty. Jako míra homogenity shluků je chápán vnitroskupinový součet čtverců odchylkek hodnot od průměru shluku a nazývá se *Wardovo kritérium*. Kritérium pro spojování shluků vychází z myšlenky, aby v každém kroku shlukování došlo k minimálnímu přírůstku Wardowa kritéria. Wardova metoda má tendenci odstraňovat malé shluky a tvořit shluky přibližně stejné velikosti.

Mezi nehierarchické metody shlukování lze zařadit například metodu *k*-průměrů. **Metoda *k*-průměrů** je vhodná v případě, že proměnné charakterizující objekty jsou pouze kvantitativní a je založena na přesouvání jednotlivých objektů mezi shluky. Jedná se o metodu, která patří do skupiny tzv. optimalizačních metod.

Kromě výše uvedených způsobů shlukování existuje také tzv. fuzzy shlukování. Tento způsob shlukování vychází z předpokladu, že je celkem *n* objektů a *k* shluků. Pro každý *i*-tý objekt a *h*-tý shluk je určena *míra příslušnosti* u_{ih}, která představuje pravděpodobnost, že daný objekt *i* je klasifikován do *h*-tého shluku. Fuzzy shlukování je proces, který narozdíl od výše uvedených postupů umožňuje zařazení jednoho objektu do více shluků, což je považováno za výhodu této metody.Výstupem této metody je tedy matice příslušností jednotlivých objektů do shluků.

Pro shlukování objektů lze také použít **dvoukrokovou shlukovou analýzu**. Tato metoda může být využita pro shlukování objektů, které jsou charakterizovány i samotnými nominálními proměnnými, případně proměnnými různých typů. Jako míra vzdálenosti, resp. nepodobnosti může být využita buď Euklidova míra (pouze pro případ kvantitativních proměnných) nebo věrohodnostní míra (pro proměnné různých typů). Tato metoda se skládá ze dvou fází. V první fázi se objekty shlukují do podshluků (malé shluky), jejichž počet je mnohem nižší, než je počet původních objektů. Přitom je použito tzv. inkrementální shlukování, kdy se objekty buď zařadí do některého z vytvořených shluků, nebo se vytvoří nový shluk. V obou fázích shlukování je použíta stejná míra nepodobnosti, tj. věrohodnostní míra, jak je tomu v systému SPSS.

Kromě samotných shlukovačích metod hrají významnou (klíčovou roli) také míry podobnosti. Podobnost je používána jako kritérium pro tvorbu shluků. Při měření podobnosti je potřeba rozlišit, jakými typy proměnných jsou charakterizovány vlastnosti jednotlivých objektů.

Nejvíce metod a postupů je použitelných pro situaci, kdy jednotlivé objekty jsou charakterizovány kvantitativními proměnnými. Měření podobnosti objektů v případě, že jsou charakterizovány kvantitativními proměnnými vychází ze *vzdáleností* objektů. Převod měr vzdáleností na míry podobnosti, resp. nepodobnosti se provádí podle jednoduchých pravidel. Při měření vzdálenosti se velmi často používají:

Euklidova vzdálenost (též geometrická metrika) představuje délku přepony pravoúhlého trojúhelníka. Výpočet této míry je založen na Pythagorově větě. Při použití tzv. Wardovy metody shlukování, se obvykle používá čtverec euklidovy vzdálenosti.

Hammingova vzdálenost není vhodná pro případ, kdy jednotlivé proměnné, které charakterizují objekty, jsou vzájemně korelované. Pokud by uvažované proměnné byly korelované, výsledné shluky by byly nesprávné.

Dále pak lze využít **Minkovského vzdálenost**, pro kterou platí, že stejně jako Hammingova vzdálenost, není vhodná pro případ, kdy proměnné charakterizující objekty jsou vzájemně korelované.

Případně lze použít **tětivovou vzdálenost, Mahalanobisovu vzdálenost**. Mahalanobisova vzdálenost odstraňuje problém, který vzniká při použití nestandardizovaných dat, které mohou způsobit rozdíly mezi shluky, a to v důsledku odlišností měrných jednotek. Tato míra je použitelná i v případě, že jsou proměnné charakterizující objekty vzájemně korelované.

3. Koeficienty pro hodnocení shlukování a stanovení počtu shluků

V této části textu se zaměříme na koeficienty, které slouží k hodnocení výsledků shlukování v případě, že uvažujeme shlukování objektů s pevným přiřazením do shluků. Pro hodnocení rozdělení množiny *n* objektů do k disjunktních shluků bylo navrženo mnoho koeficientů a to bez ohledu na způsob, jakým bylo rozdělování objektů uskutečněno, tedy nezáleží, zda-li shluky jsou výsledkem metod rozkladu a nebo výsledkem hierarchického shlukování]. Nebude zde popsán jejich výpočet, nýbrž pouze jejich přehled, ve kterém softwarovém produktu je možné tyto koeficienty nalézt a jakým způsobem se vyhodnocují. Důkladný popis těchto kritérií je uveden například v Löster (2011) či Řezanková (2009).

Tab. 4: Vybraná kritéria pro hodnocení výsledků disjunktního shlukování

Koeficient	**Hledaný extrém**	**Software**
Obrysový koeficient	maximum	S-PLUS, SPSS
CHF index (pseudo F)	maximum	SAS system LE, SYSTAT
PTS index (T-kvadrát)	minimum	SAS system LE, SYSTAT
RS (R-kvadrát, RSQ)	maximum	SAS system LE
SPRS (SPRSQ)	minimum	SAS system LE
BIC, AIC	minimum	SPSS
RMSSTD	minimum	SYSTAT
Daviesův-Bouldinův (DB)	minimum	SYSTAT
Dunnův separační index	maximum	SYSTAT

Zdroj: Vlastní zpracování

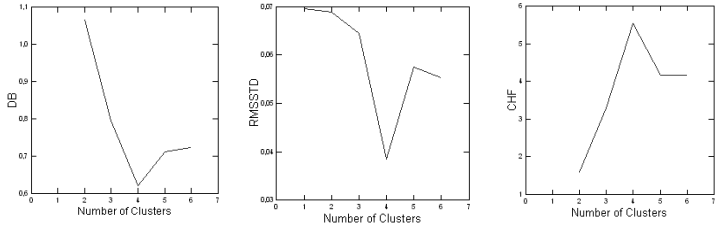
Při praktických úlohách je i v tomto případě vhodné stanovit hodnoty více koeficientů současně, protože neexistuje kritérium, které by s jistotou jednoznačně ohodnotilo výsledek

shlukování (metodu či počet shluků). Jak bylo uvedeno výše, v literatuře ani není jednoznačně vymezeno, za jakých podmínek je libovolný z koeficientů nejvhodnější pro hodnocení konkrétního shlukování. V případě, že se výsledky hodnocení shodují na základě hodnot více koeficientů současně, je možné tyto závěry považovat za „správné“. Jak uvádějí i samotní autoři koeficientů, viz Löster (2011) v některých případech je dokonce nutné hodnotit výsledky shlukování současně pomocí několika koeficientů.

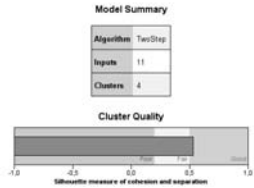
4. Shluky zemí EU s využitím dat, které se týkají trhu práce

V následující části si ukážeme, jak by se mělo postupovat při výsledném hodnocení shlukování (stanovení počtu shluků). Nebudou zde podrobně popisovány jednotlivé proměnné, které byly pro shlukování využity. Při shlukování je využito celkem 12 proměnných, které se týkají trhu práce. Jedná se například o míru nezaměstnanosti, míru dlouhodobé nezaměstnanosti, podíl osob starších 65 let na celkovém počtu nezaměstnaných atd.

Pro nalezení optimálního počtu shluků, do kolika rozdělit země EU, jsme využili dvě možnosti shlukování a to hierarchické shlukování a dvoukrokovou metodu shlukování. Použili jsme dva různé softwarové produkty, podle tabulky 1 a hledali optimální počet shluků. Na základě obou obrázků by jako optimální počet shluků byly stanoveny 4 shluky zemí EU.



Obr. 21: Kritéria pro stanovení optimálního počtu shluků



Obr. 2: Kritérium pro stanovení optimálního počtu shluků

5. Závěr

Shluková analýza je velmi populární vícerozměrná statistická metoda, jejíž cílem je vytváření skupin objektů, tzv. shluků. V současné literatuře je popsáno mnoho metod a koeficientů, které slouží ke shlukování. Zásadním problémem však je, že ve většině případů není popsané, jak se má výzkumník rozhodovat, aby získal co „nejlepší“ výsledky. Při výběru metod jsou částečně popsané možné problémy, například při korelovanosti proměnných, nicméně pravidlo, které by jednoznačně výzkumníkovi napomohlo, kterou metodu ve spojení s jakou metrikou (například vzdálenosti) použít, již neexistuje. Naším cílem je naznačit, jak

by se mohl výzkumník při shlukování orientovat. Stručně jsme popsali základní (tradiční) metody shlukování, jaké se často používají míry vzdáleností pro kvantitativní proměnné. Dále jsme v tabulce uvedli přehled běžně dostupných koeficientů, které nám pomohou vyhodnotit výsledky shlukování. Bohužel však, jak uvádí i samotní autoři koeficientů, i zde je nutné rozhodovat se současně na základě více hodnot, aby výsledky byly co nejobjektivnější a co nejlepší. Na příkladu shlukování zemí EU jsme si ukázali, jak postupovat a jak se rozhodovat. Využili jsme k tomu dva různé softwarové produkty a hledali jsme optimální počet shluků, do kolika bychom měli rozdělit země EU, které by pak dále mohly byly analyzovány. Na základě koeficientů jsme usoudili, že by bylo vhodné rozdělit země do čtyřech shluků.

Poděkování

Tento článek byl vytvořen s pomocí projektu Interní grantové agentury Vysoké školy ekonomické v Praze, č. 6/2013 pod názvem " Hodnocení výsledků metod shlukové analýzy v ekonomických úlohách."

Literatura

GAN, G., MA CH., WU J.: *Data Clustering Theory, Algorithms, and Applications*, ASA, Philadelphia, 2007.

HALKIDI, M., BATISTAKIS, Y., VAZIRGIANNIS, M.: Cluster Validity Methods: Part I., *SIGMOD Record*, No. 2, 2002, s. 40-45.

HALKIDI, M., BATISTAKIS, Y., VAZIRGIANNIS, M.: *Clustering algorithms and validity measures.* SSDBM, Athens, 2001.

LÖSTER, T.: *Hodnocení výsledků metod shlukové analýzy*. Disertační práce, VŠE v Praze, 2011, 137s.

MEGYESIOVÁ, S.: Softvérové riešenie úloh zhlukovej analýzy In *Forum Statisticum Slovacum*. ISSN 1336-7420. Roč. 1, č.2 (2005), s. 100-105.

MEGYESIOVÁ, S.: Význam analýzy hlavných komponentov pri riešení úloh viacrozmernej štatistiky. In *Forum Statisticum Slovacum*. ISSN 1336-7420. Roč. 2, č.1 (2006), s. 101-107.

MEGYESIOVA, S., LIESKOVSKA, V. (2012). *Are europeans living longer and healthier lives?*. In Loster Tomas, Pavelka Tomas (Eds.), *6th International Days of Statistics and Economics* (pp. 766-775). ISBN 978-80-86175-86-7.

ŘEZANKOVÁ, H., HÚSEK, D., LÖSTER, R.: *Clustering with Mixed Type Variables and Determination of Cluster Numbers*, CNAM and INRIA, Paříž, 2010, s. 1525-1532.

ŘEZANKOVÁ, H., HÚSEK, D., SNÁŠEL, V.: *Shluková analýza dat*, 2. vydání, Professional Publishing, Praha, 2009.

STANKOVIČOVÁ, I., VOJTKOVÁ, M.: *Viacrozmerné štatistické metódy s aplikáciami*, Ekonómia, Bratislava, 2007.

Adresa autorů:

Tomáš Löster, Ing., Ph.D.
Katedra statistiky a pravděpodobnosti
Vysoká škola ekonomická v Praze
Nám. W. Churchilla 4, 130 67, Praha 3
Tomas.Loster@vse.cz

Tomáš Pavelka, doc., Ing., Ph.D.
Katedra Mikroekonomie
Vysoká škola ekonomická v Praze
Nám. W. Churchilla 4, 130 67, Praha 3
Tomas.Pavelka@vse.cz

Analýza samovražednosti v Českej republike pomocou zhlukovej analýzy
Cluster analysis of suicidality in the Czech Republic

Elena Makhalova, Kornélia Cséfalvaiová, Jitka Langhamrová

Abstract: Number of suicides is increasing worldwide. Suicide attempts and completed suicides are very common in individuals with personality disorder. Tendency to end their own lives is more visible for men than for women, during the reporting period 2006-2010 five times more men committed suicide than women. Number of suicides and suicide methods are different for different age groups and gender. This paper tries to capture the relation between the number of suicides and the number of unemployed persons, as well as the relationship between the number of suicides and the average salary of employees in the Czech Republic. Using cluster analysis highlights the similarities and differences in different regions of the republic.

Abstrakt: Počet samovrážd rastie na celosvetovej úrovni. Pokusy o samovraždu a následná realizácia samovraždy sú častým prejavom správania osôb s poruchou osobnosti. Tendenciu ukončiť vlastný život majú výrazne vyššiu muži oproti ženám, za sledované päťročné obdobie 2006-2010 spáchalo samovraždu päťkrát viac mužov ako žien. Počet samovrážd a spôsob prevedenia samovraždy sú odlišné pre rôzne vekové skupiny a pohlavie. Predložená práca sa snaží zachytiť súvislosti medzi počtom samovrážd a počtom nezamestnaných osôb, rovnako ako vzťah počtu samovrážd a priemerného platu zamestnancov v Českej republike. Pomocou zhlukovej analýzy poukazuje na podobnosti a odlišnosti v jednotlivých krajoch republiky.

Key words: suicide, cluster analysis, correlation.

Kľúčové slová: samovražda, zhluková analýza, korelácia.

JEL classification: C38, J19

1. Úvod

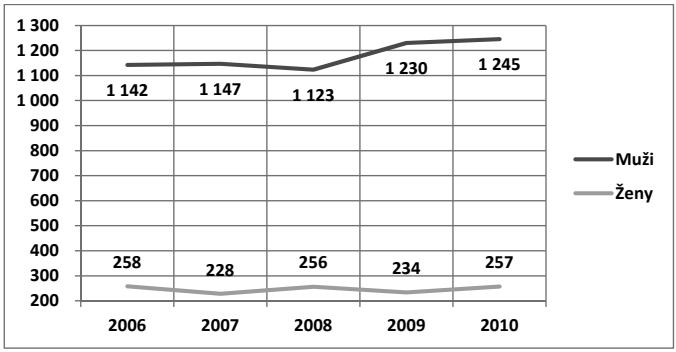
Samovražda je dobrovoľný a plánovaný úmysel danej osoby ukončiť svoj život, pričom toto rozhodnutie vedie k smrti (viď Český štatistický úrad). Dôvody samovrážd bývajú rôzne, vystupuje tu mnoho faktorov, medzi ktoré radíme napríklad nezamestnanosť, nespokojnosť so sociálnym postavením, nízke platové ohodnotenie alebo zlé medziľudské vzťahy či zdravotné problémy. Nepriaznivý ekonomický vývoj sa odzrkadľuje na raste počtu samovrážd v celej Európe. Dopady hospodárskej kríze sú pozorovateľné i v Českej republike. Počas recesie je zreteľný výskyt psychických problémov z dôvodu vysokej miery nezamestnanosti a nedostatku financií. Tento pocit napätia, stresu a frustrovania sa skôr či neskôr prejaví aj v rodinnom živote, a následne môže viesť k stavu beznádeje, kedy sa daný jedinec odhodlá k spáchaniu samovraždy. Samovražda neovplyvňuje iba samotného jedinca, ale i jeho najbližšie okolie, a preto skúmanie závislosti samovraždy na rôznych faktoroch je kľúčovým záujmom odborníkov z oblasti ekonómie, sociológie, psychológie, psychiatrie a sociálnej práce. Z tohto dôvodu sa i predložená práca zaoberá vývojom samovražednosti v Českej republike v súvislosti s počtom nezamestnaných osôb a priemerným platom zamestnancov v jednotlivých krajoch Českej republike za obdobie 2006-2010[1] (vstupné údaje sú prepočítané na 1000 obyvateľov stredného stavu a priemerná mzda sa týka podnikateľskej sféry). V práci sú použité dáta z Českého štatistického úradu a Ministerstva práce a sociálnych vecí.

[1] Český štatistický úrad pravidelne vydáva analýzu samovrážd vždy po uplynulom päťročnom období.

2. Analýza vývoja samovražednosti v Českej republike v období 2006-2010

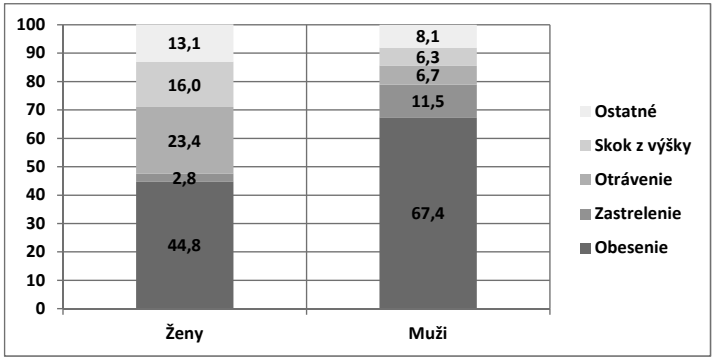
Z dostupných štatistík zisťujeme, že počet samovrážd v Českej republike má za sledované obdobie stúpajúcu tendenciu. V roku 2006 dobrovoľne odišlo z tohto sveta 1400 osôb, kým v roku 2010 sa tento počet zvýšil na 1502 osôb. K samovraždám dochádza častejšie pri mužoch ako pri ženách – počet samovrahov - mužov sa postupne zvýšil z hodnoty 1142 v roku 2006 na hodnotu 1245 v roku 2010. V Českej republike najvyšší počet samovrážd podľa pohlavia a veku je v skupine mužov vo veku 45-54 rokov. Naopak, v prípade žien sme nezaznamenali výraznejšie zmeny v počte samovrážd v období 2006-2010. V roku 2006 počet samovrahov - žien činil 258 a v roku 2010 tento počet predstavoval 257. Je zaujímavý nárast počtu samovrážd medzi rokmi 2008 a 2009, kedy sa prejavili známky hospodárskej krízy.

Na obrázku 1 je znázornený vývoj počtu samovrážd pre mužov a ženy medzi rokmi 2006 a 2010.



Obr. 22: Vývoj počtu samovrážd v Českej republike v období 2006-2010

Podľa spôsobu prevedenia samovraždy pozorujeme značné rozdiely v pohlaví (viď obrázok 2). Najčastejšie sa volí spôsob obesenia. V sledovanom období 2006-2010 týmto spôsobom prišlo o život 67,4 % mužov a 44,8 % žien. Muži sa ďalej v mnohých prípadoch rozhodnú i pre zastrelenie, otrávenie a skok z výšky. Ženy menej často siahnu po zbrani a častejšie sa rozhodnú pre otrávenie (23,4 %) či skok z výšky (16,0 %) v porovnaní s mužmi (6,7 % v prípade otrávenia a 6,3 % v prípade skoku z výšky).



Obr.2: Samovraždy podľa spôsobu prevedenia v Českej republike v období 2006-2010 (v %)

Ďalším krokom našej práce je pomocou korelačného koeficientu zistiť existenciu závislosti medzi počtom samovrážd na 1000 obyvateľov stredného stavu v jednotlivých krajoch Českej republiky a priemernou hrubou mesačnou mzdou, mierou nezamestnanosti, hrubým domácim produktom a podielom osôb s vysokoškolským vzdelaním v týchto krajoch. Pomocou korelačnej analýzy bolo zistené, že existuje nepriama závislosť medzi počtom samovrážd a priemernou hrubou mesačnou mzdou (čím vyššia priemerná mzda, tým nižší počet samovrážd) a taktiež bola zistená priama závislosť medzi počtom samovrážd a mierou nezamestnanosti. Ďalej z výsledkov vidíme nepriamu závislosť medzi počtom samovrážd a HDP a podielom osôb s vysokoškolským vzdelaním (viď tabuľka 1). Presnejšie výsledky by sme získali zahrnutím viacerých premenných do analýzy, ale v tejto práci sme uvažovali iba závislosť uvedených premenných. Pre záujemcov umožňuje databáza Českého štatistického úradu ďalšie, podrobnejšie kombinácie údajov.

	Miera samovražednosti	**Miera nezamestnanosti**	**Priemerná hrubá mesačná mzda**	**HDP v bežných cenách**	**Vysokoškolské vzdelanie**
Miera samovražednosti	1				
Miera nezamestnanosti	0,262	1			
Priemerná hrubá mesačná mzda	-0,210	-0,598	1		
HDP v bežných cenách	-0,228	-0,481	0,958	1	
Vysokoškolské vzdelanie	-0,351	-0,512	0,917	0,971	1

Tab. 1: Koeficient korelácie medzi premennými miera samovražednosti, miera nezamestnanosti, priemerná hrubá mesačná mzda, HDP v bežných cenách a vysokoškolské vzdelanie v Českej republike

Z tabuľky 1 vidíme, že existuje slabá závislosť medzi počtom samovrážd a mierou nezamestnanosti (0,262) a nepriama závislosť medzi počtom samovrážd a priemernou hrubou mesačnou mzdou (-0,210). Nepriama závislosť bola zistená i medzi počtom samovrážd a HDP (-0,228) a medzi počtom samovrážd a vysokoškolským vzdelaním (-0,351).

V nasledujúcej časti prevedieme zhlukovú analýzu, pomocou ktorej rozdelíme objekty do zhlukov. Objekty (kraje) patriace do jedného zhluku si budú vzájomne podobné (sú homogénne) a objekty (kraje) z rôznych zhlukov sa budú od seba odlišovať (sú heterogénne).

3. Zhluková analýza

V tejto časti sme sa zaoberali zhlukovou analýzou, účelom ktorej je zaradiť jednotlivé kraje Českej republiky do konkrétneho zhluku podľa podobnosti mzdy či miery nezamestnanosti. Dôležitou úlohou je určiť správny počet zhlukov.

Výpočty boli prevedené pomocou štatistického softvéru Matlab. Používali sme metódu K – means, v ktorej sme použili „Manhattanskú vzdialenosť“. Ďalej sme použili metódu inicializácie centroidov z náhodne vybraných bodov.

Boli preukázané nasledujúce výsledky (viď obrázok 3):

Obr.3: Zhluky jednotlivých krajov Českej republiky

Podľa obrázku 3 je vidieť, že lepším počtom zhlukov je 4, pretože hodnota „silhouette value" meria podobnosť krajov v rámci zhlukov. Pre lepšie pochopenie problematiky poslúži interpretácia jednotlivých hodnôt (viď tabuľka 2).

Kraj	Miera samovražednosti	Miera nezamestnanosti	Priemerná hrubá mesačná mzda	HDP v bežných cenách	Vysokoškolské vzdelanie
		Zhluk 1			
Juhočeský	0,133	0,035	19 240	193 369	0,092
Plzenský	0,147	0,037	20 429	178 660	0,091
Ústecký	0,146	0,068	19 633	239 721	0,063
Královohradecký	0,141	0,035	19 339	166 932	0,086
Pardubický	0,130	0,041	18 905	149 003	0,084
Vysočina	0,109	0,044	19 160	149 416	0,080
Olomoucký	0,162	0,050	18 884	169 727	0,096
Zlínsky	0,148	0,045	18 691	174 199	0,095
		Zhluk 2			
Stredočeský	0,139	0,033	20 931	392 496	0,103
Juhomoravský	0,118	0,049	20 278	376 971	0,128
Moravskosliezsky	0,138	0,059	20 101	370 225	0,093
		Zhluk 3			
Karlovarský	0,156	0,053	18 317	79 603	0,058
Liberecký	0,131	0,046	19 294	118 354	0,082
		Zhluk 4			
Praha	0,131	0,019	28 354	925 163	0,216

Tab. 2: Výsledky získané zhlukovaním

Tieto výsledky môžeme overiť, keď počítame celkový súčet „silhouette value" pre prípad 4 zhlukov (248 341) a pre prípad 5 zhlukov (148 306) a 6 zhlukov (93 510). Z toho je tiež vidieť, že celková hodnota je najväčšia v prípade 4 zhlukov.

Všetky objekty (kraje) sú rozdelené do 4 zhlukov. Je vidieť, hlavné mesto Praha je jednoznačne odlišná v porovnaní s ostatnými krajmi.

Priemerná mzda v Prahe sa výrazne líši od priemernej mzdy v ostatných krajoch a má najmenšiu mieru nezamestnanosti, vysoký hrubý domáci produkt a najväčší podiel osôb s vysokoškolským vzdelaním, preto sa objavuje v samostatnom zhluku. Do zhluku 3 patrí Karlovarský a Liberecký kraj, v ktorých sme zaznamenali malý podiel osôb s vysokoškolským vzdelaním.

Kraje nachádzajúce sa v prvom a druhom zhluku sa líšia podľa hodnoty HDP, ktorá je vyššia pre kraje v druhom zhluku. Do prvého zhluku sa priradili kraje s pomerne nízkym HDP a pomerne nízkou priemernou mesačnou mzdou. Výnimkou je Plzenský kraj, ktorý má najvyššiu priemernú mesačnú mzdu v tomto zhluku (20 429). Porovnaním miery samovražednosti v jednotlivých zhlukoch zisťujeme, že v priemere najvyššia miera samovražednosti je v treťom zhluku, teda v Karlovarskom a Libereckom kraji. Naopak, najnižšia miera samovražednosti je zaznamenaná v hlavnom meste Praha.

4. Záver

Rastúca tendencia samovražednosti je jav, ktorý nás núti na chvíľu sa pozastaviť a zamyslieť sa nad dôležitosťou a nenahraditeľnosťou ľudského života. Uvedomiť si váhu našej osoby a uvedomiť si naše okolie, ktoré nás obklopuje. Príčiny a okolnosti samovražednosti je potrebné skúmať v širších súvislostiach, okrem pocitu osamelosti a úzkosti predstavuje súčasná ekonomická recesia hrozbu pre rastúci počet samovrážd. Nepriaznivý hospodársky vývoj sa negatívne odzrkadľuje na duševnom stave mnohých obyvateľov. V práci sme preukázali rozdielnosť vývoja počtu samovrážd v Českej republike podľa pohlavia. Taktiež z dostupných údajov zisťujeme odlišný spôsob prevedenia samovraždy v prípade mužov a žien. Z nášho zistenia vyplýva ďalší záver – v Českej republike môžeme kraje podľa počtu samovrážd rozdeliť do 4 zhlukov a hlavné kritérium, ktoré ovplyvňuje počet samovrážd, je miera nezamestnanosti a vzdelanie.

Poďakovanie

Článok bol pripravený v spolupráci s Internou grantovou agentúrou Vysokej školy ekonomickej v Prahe, číslo 6/2013 pod názvom „Hodnotenie výsledkov zhlukovej analýzy pri ekonomických problémoch".

Literatúra

ČSÚ: *Sebevraždy v České republice*. [online] Český statistický úřad, Praha, 2011. http://www.czso.cz/csu/2011edicniplan.nsf/kapitola/4012-11-n_2011-16

ČSÚ: *Regionální časové řady*. [online] Český statistický úřad, Praha, 2013. http://www.czso.cz/csu/redakce.nsf/i/regionalni_casove_rady

MPSV: *Regionální statistika ceny práce.* [online] Ministerstvo práce a sociálních věcí, Praha. http://portal.mpsv.cz/sz/stat/vydelky

LANGHAMROVÁ, J. *Základy demografie (materiály ke cvičením)*. Praha: Oeconomica, 2013. ISBN 978-80-245-1956-2.

LÖSTER, T. LANGHAMROVÁ, J. *Disparities between regions of the Czech Republic for non-business aspects of labour market.* Prague 13.09.2012 – 15.09.2012. In: LÖSTER,

Tomáš, PAVELKA, Tomáš (ed.). International Days of Statistics and Economics at VŠE, Prague. Slaný : Melandrium, 2012, s. 689–702. ISBN 978-80-86175-86-7.

LÖSTER, T. *Hodnocení výsledků fuzzy shlukování*. In International collection of scientific work on the occasion of 60th anniversary of university education at faculty of Business Economy with seat in Košice of University of Economics in Bratislava. Praha: VŠE, 2012, s. 1--14. ISBN 978-80-86175-80-5.

LÖSTER, T. *Nerovnosti mezi regiony České republiky u podnikatelské sféry z hlediska trhu práce.* Herlany 26.09.2012. In:Nerovnosť a chudoba v Európskej únii a na Slovensku. [online] Košice : Ekonomická fakulta TU, 2012, s. 123–130. ISBN 978-80-553-1225-5.
URL: http://www3.ekf.tuke.sk/NaRE2012/subory/workshop/Herlany_Zbornik_web.pdf.

ŘEZANKOVA, H., & LÖSTER, T. (2013). *Shluková analýza domácností charakterizovaných kategoriálnimi ukazateli. E+M*. Ekonomie a Management, *16*(3), 139-147. ISSN: 1212-3609

ŘEZANKOVA, H., HÚSEK, D., SNÁŠE V., (2009). *Shluková analýza dat*, Professional Publishing, ISBN: 978-80-86946-81-8, EAN**:** 9788086946818.

Adresa autorov:

Mgr. Elena Makhalova
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 67 Praha 3
xmake900@vse.cz

Ing. Kornélia Cséfalvaiová
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 67 Praha 3
xcsek00@vse.cz

doc. Ing. Jitka Langhamrová, CSc.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 67 Praha 3
jitka.langhamrova@vse.cz

Statistický algoritmus výpočtu souřadnic vysílače
Statistical Algorithm for Determining Transmitters' Position

Michal Mandlik, Jaroslav Marek, Martin Svoboda

Abstract: In this article we use a new statistical procedure in a special regression model. The aim is estimation of the unknown plane coordinates of transmitter´s position. These coordinates are computed using the shift of signals from at least three receivers.

Abstrakt: V tomto článku používáme nový statistický algoritmus založený na regresním modelu. Cílem je nalezení odhadů neznámých souřadnic vysílače v rovině. Tyto souřadnice se počítají na základě měření posunu signálů z alespoň tří přijímačů.

Key words: regression model with a set of constrains, BLUE, linearization

Kľúčové slová: regresní model s podmínkou, nejlepší lineární nevychýlený odhad, linearizace

JEL classification: C13

1. Introduction

In this work, we will propose and subsequently examine a statistical model which will serve for finding out the most accurate estimators of the transmitter´s position coordinates. We will denote this position as P and its coordinates as γ_1 and γ_2. At our disposal are measured values of time differences $\Delta_{1,2}$, $\Delta_{1,3}$, $\Delta_{2,3}\ldots,\Delta_{n-1,n}$ and values of receivers´ plane coordinates labelled by $R_i=[x_i,y_i]$, $i\in\{1,2,..,n\}$. The situation is depicted for n = 3 in Figure 1.

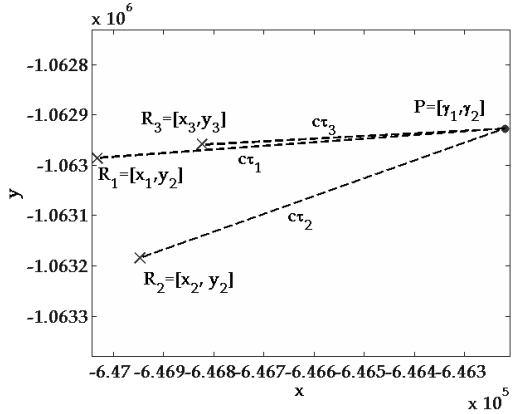


Fig. 23: The layout of the measurement

The distance between the transmitter P and the i^{th} receiver is given as $d(P,R_i)=c\cdot\tau_i$, where $c \doteq 3\cdot 10^8\, m/s$ is the speed of light. The time differences mentioned above are then $\Delta_{i,j}=\tau_i-\tau_j$.

Our algorithm is based on so-called model with a set of constrains of type II, cf. (Kubáčková, 1992; Kubáček, 1993; Kubáček, Kubáčková and Volaufová, 1995). In this algorithm we get

estimators $\hat{\hat{\boldsymbol{\beta}}}$, $\hat{\hat{\boldsymbol{\gamma}}}$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ is the vector of the unknown coordinates and $\boldsymbol{\beta} = (\Delta_{1,2}, \ldots, \Delta_{n-1,n}, x_1, y_2, \ldots, x_n, y_n)'$ denotes the vector of the measured data.

2. Theory

Definition 2.1 *The model of incomplete measurement with a set of constrains of type II is given in the form of*

$$\mathbf{Y} \sim \mathbf{N}[\mathbf{F}\boldsymbol{\beta}, \sigma^2\mathbf{V}], \quad \mathbf{b} + \mathbf{B}\boldsymbol{\beta} + \mathbf{G}\boldsymbol{\gamma} = \mathbf{0}, \tag{1}$$

where $\boldsymbol{\beta} \in R^{k_1}$, $\boldsymbol{\gamma} \in R^{k_2}$ *are the unknown parameters.*

If $h(\mathbf{F}_{(n,k)}) = k_1 < n \ \wedge \ h(\mathbf{B}_{(q,k_1)}, \mathbf{G}_{(q,k_2)}) = q < k_1 + k_2 \ \wedge \ h(\mathbf{G}_{(q,k_2)}) = k_2 < q$ *and* $\mathbf{V}$ *is a positively definite matrix then the model is regular.*

In the following text, we will consider only the regular model.

Theorem 2.1 BLUE *(Best Linear Unbiased Estimator) of the* $\left(\hat{\hat{\boldsymbol{\beta}}}, \hat{\hat{\boldsymbol{\gamma}}}\right)'$ *vector is given by*

$$\hat{\hat{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}'\left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\left(\mathbf{b} + \mathbf{B}\hat{\boldsymbol{\beta}}\right), \tag{2}$$

$$\hat{\hat{\boldsymbol{\gamma}}} = -\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\left(\mathbf{b} + \mathbf{B}\hat{\boldsymbol{\beta}}\right), \tag{3}$$

where

$$\mathbf{T} = \mathbf{B}\left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}' + \mathbf{G}\mathbf{G}', \tag{4}$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{V}^{-1}\mathbf{Y} \tag{5}$$

($\hat{\boldsymbol{\beta}}$ *is the estimator non-respecting the constraints between* $\boldsymbol{\beta}$ *and* $\boldsymbol{\gamma}$ *parameter).*

Proof: The derivation of the relations for the estimators is based on the least-squares method. For details see Kubáček, Kubáčková and Volaufová (1995) or Kubáček and Kubáčková (2000).

Theorem 2.2 *The covariance matrix of the* $\hat{\hat{\boldsymbol{\beta}}}$ *estimator is given by*

$$\begin{aligned} \operatorname{var}\left(\hat{\hat{\boldsymbol{\beta}}}\right) &= \sigma^2\left\{\mathbf{I} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}'\left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\mathbf{B}\right\} \times \\ &\times \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\left\{\mathbf{I} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}' \times \left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\mathbf{B}\right\}'. \end{aligned} \tag{6}$$

Proof: The proof is straightforward, i.e.

$$\begin{aligned} \operatorname{var}\left(\hat{\hat{\boldsymbol{\beta}}}\right) &= \sigma^2\left\{\mathbf{I} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}'\left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\mathbf{B}\right\}\hat{\boldsymbol{\beta}} = \\ &= \sigma^2\left\{\mathbf{I} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}'\left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\mathbf{B}\right\} \times \\ &\times \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\left\{\mathbf{I} - \left(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F}\right)^{-1}\mathbf{B}' \times \left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{T}^{-1}\right]\mathbf{B}\right\}'. \end{aligned}$$

Theorem 2.3 *The covariance matrix of the* $\hat{\hat{\boldsymbol{\gamma}}}$*, estimator is given by*

$$\operatorname{var}\left(\hat{\hat{\boldsymbol{\gamma}}}\right) = \sigma^2\left\{\left(\mathbf{G}'\mathbf{T}^{-1}\mathbf{G}\right)^{-1} - \mathbf{I}\right\} \tag{7}$$

3. Algorithm

Let us consider three positions $R_1=[x_1,y_1]$, $R_2=[x_2,y_2]$ and $R_3=[x_3,y_3]$. The uncertainty of these coordinates is described by covariance matrix $\mathrm{cov}(\mathbf{R})$.

Difference between received signals in these positions are $\Delta_{1,2}$, $\Delta_{1,3}$, $\Delta_{2,3}$. The accuracy of measurement of times is given by covariance matrix $\mathrm{cov}(\Delta)$.

The aim is to make the estimator of the position with unknown plane coordinates $P=[\gamma_1,\gamma_2]$. We consider model $\mathbf{Y} \sim \mathbf{N}[\boldsymbol{\beta}, \boldsymbol{\Sigma}]$, where

$$\boldsymbol{\beta} = (\Delta_{1,2},\ \Delta_{1,3},\ \Delta_{2,3},\ x_1,\ y_1,\ x_2,\ y_2,\ x_3,\ y_3)', \tag{8}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{cov}(\Delta) & 0 \\ 0 & \mathrm{cov}(\mathbf{R}) \end{pmatrix}. \tag{9}$$

The difference of distances $d(P,R_i)-d(P,R_j)$ is equal to $c\cdot\Delta_{i,j}$. Therefore, parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ have to fulfil the constraints in the form of

$$g_1(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sqrt{(x_1-\gamma_1)^2+(y_1-\gamma_2)^2} - \sqrt{(x_2-\gamma_1)^2+(y_2-\gamma_2)^2} - c\cdot\Delta_{1,2} = 0, \tag{10}$$

$$g_2(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sqrt{(x_1-\gamma_1)^2+(y_1-\gamma_2)^2} - \sqrt{(x_3-\gamma_1)^2+(y_3-\gamma_2)^2} - c\cdot\Delta_{1,3} = 0, \tag{11}$$

$$g_3(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sqrt{(x_2-\gamma_1)^2+(y_2-\gamma_2)^2} - \sqrt{(x_3-\gamma_1)^2+(y_3-\gamma_2)^2} - c\cdot\Delta_{2,3} = 0. \tag{12}$$

We need to linearize the nonlinear conditions, i.e.

$$\mathbf{g}(\boldsymbol{\beta},\boldsymbol{\gamma}) = (g_1(\boldsymbol{\beta},\boldsymbol{\gamma}),\ g_2(\boldsymbol{\beta},\boldsymbol{\gamma}),\ g_3(\boldsymbol{\beta},\boldsymbol{\gamma}))' = \mathbf{0} \tag{13}$$

by means of the Taylor series expansion in the linear form. These conditions are expressed by

$$\mathbf{b} + \mathbf{B}\,\delta\boldsymbol{\beta} + \mathbf{G}\,\delta\boldsymbol{\gamma} = \mathbf{0}, \tag{14}$$

where

$$\mathbf{B} = \frac{\partial \mathbf{g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\beta}'},\quad \mathbf{G} = \frac{\partial \mathbf{g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'},\quad \mathbf{b} = \mathbf{g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0),\ \delta\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0,\quad \delta\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0, \tag{15}$$

with $(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)$ being the initial solution. In our case, **B** and **G** matrices take the form of

$$\mathbf{B} = \frac{\partial \mathbf{g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\beta}'} = (B_{ij}) = \left(\frac{\partial g_i}{\partial \beta_j}\right)\Bigg|_{(\boldsymbol{\beta},\boldsymbol{\gamma})=(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)} \quad i\in\{1,2,3\},\quad j\in\{1,2,..,9\} \tag{16}$$

$$\mathbf{G} = \frac{\partial \mathbf{g}(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'} = (G_{kl}) = \left(\frac{\partial g_k}{\partial \gamma_l}\right)\Bigg|_{(\boldsymbol{\beta},\boldsymbol{\gamma})=(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)} \quad k\in\{1,2,3\},\quad l\in\{1,2,3\}\ . \tag{17}$$

The elements of these matrices can be easily computed through deriving functions g_i described in (10) – (12). After substitution of these expressions to (2) and (3), we get the estimators of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in our case.

4. Example

We have at our disposal measurements of plane coordinates

$$R_1 = [-647031960,\ -106298847] \tag{18}$$

$$R_2 = [-646822824,\ -106295875] \tag{19}$$

$$R_3 = [-646946970,\ -106318359] \tag{20}$$

with covariance matrix

$$\mathrm{cov}\,(\mathbf{R}) = 0.01\ \mathbf{I}\ [m^2]. \tag{21}$$

Measurements of differences $\Delta = (\Delta_{1,2},\ \Delta_{1,3},\ \Delta_{2,3})'$ between received signals in these three points lead to

$$c \cdot \Delta = c \cdot (0.00070131,\ -0.0055719,\ 0.002570374)'\ [m] \tag{22}$$

with covariance matrix

$$\mathrm{cov}(c \cdot \Delta) = c^2 (10^{-9})^2\ \mathbf{I} = 0.09\ \mathbf{I}\ [m^2]. \tag{23}$$

In our linearized model we will determine numerically from Theorem 2.1 the estimator of parameter $\hat{\hat{\boldsymbol{\gamma}}}$ and from Theorem 2.2 its covariance matrix. The results are:

$$\hat{\hat{\boldsymbol{\gamma}}} = \begin{pmatrix} -646220.24 \\ -10629265.69 \end{pmatrix} [m], \tag{24}$$

$$\mathrm{var}\left(\hat{\hat{\boldsymbol{\gamma}}}\right) = \begin{pmatrix} 0.0018 & -0.0018 \\ -0.0018 & 0.0053 \end{pmatrix} [m^2]. \tag{25}$$

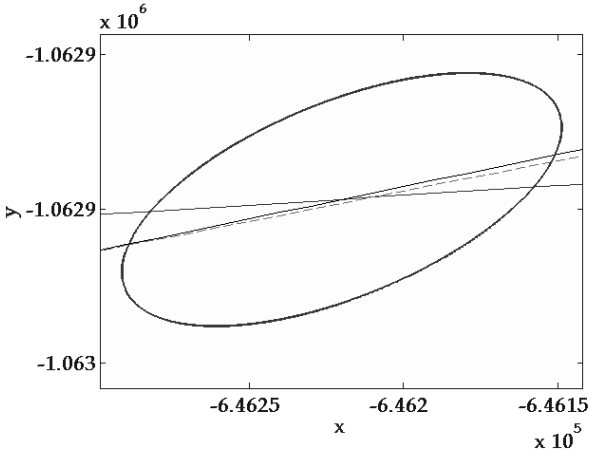


Fig. 2: The confidence domain

Figure 2 shows the confidence domain (cf. Kubáčková, 1992) for alpha = 5 %. Depicted are also three hyperbolic curves given by constrains (10), (11) and (12).

5. References

KUBÁČKOVÁ, L. 1992. Foundations of Experimental Data Analysis. CRCPress, Boca Raton Ann Arbor London Tokyo.

KUBÁČEK, L. 1993. *Two stage regression models with constraints*. Math. Slovaca (43), 643–658.

KUBÁČEK, L., KUBÁČKOVÁ, L., L., VOLAUFOVÁ, J. 1995. Statistical models with linear structures. Veda, Publishing House of the Slovak Academy of Sciences, Bratislava.

KUBÁČEK, L., KUBÁČKOVÁ L. 2000. Statistics and Metrology (in Czech). Publishing House of Palacký University, Olomouc.

KUBÁČEK, L. 2006. *Outliers in Models with Constraints*. Kybernetika 42 (6), 673–698.

KUBÁČEK, L., MAREK, J. 2004. *Partial optimum estimator in two stage regression model with constraints and a problem of equivalence*. Math. Slovaca 54.

SEBER, G. A. F., WILD, C. J. 2003. *Nonlinear Regression*, J. Wiley & Sons, New Jersey.

ZVÁRA, K. 1989. *Regression Analysis (in Czech).* Academia, Praha.

Authors:

Martin Svoboda, RNDr.
Univerzita Pardubice
Fakulta elektrotechniky a informatiky
náměstí Čs. legií 565, 530 02 Pardubice
martin.svoboda@upce.cz

Jaroslav Marek, Mgr., Ph.D.
Univerzita Pardubice
Fakulta elektrotechniky a informatiky
náměstí Čs. legií 565, 530 02 Pardubice
jaroslav.marek@upce.cz

Michal Mandlík, Ing.
Univerzita Pardubice
Fakulta elektrotechniky a informatiky
náměstí Čs. legií 565, 530 02 Pardubice
michal.mandlik@student.upce.cz

Vývoj regionálnych rozdielov priemernej doby pracovnej neschopnosti pre chorobu
Development of regional disparities of the average duration of sickness absence

Silvia Megyesiová, Vanda Lieskovská

Abstract: The economic crisis and its issues are in focus of attention not only of economists, but also sociologists, psychologists and physicians. It is clear nowadays that there exists relation between public health and economic crisis. Often stress and uncertainty that people survive in times of crisis affects their mental and overall health. In this paper we focus on the trend of the average duration of sickness absence. The average duration of sickness absence is the number of days of sickness absence per one newly reported case of sickness absence. The development of this indicator is recently very negative. In addition to a radical increase in the average duration of time required to cure the patient, dramatically has increased also the regional disparity of the selected indicator. Inhabitants of the regions with the highest unemployment rates are also burdened by the longest average duration of sickness absence.

Abstrakt: Ekonomická kríza a s ňou spojené problémy sú stredobodom pozornosti nielen samotných ekonómov, ale aj sociológov, psychológov a lekárov. Preukazuje sa súvis medzi verejným zdravím osôb a ekonomickou krízou. Často stres, ktorý osoby prežívajú v časoch krízy a neistoty vplýva na ich duševné a celkové zdravie. V príspevku sme sa zamerali na vývoj priemernej doby pracovnej neschopnosti v chorobe na novohlásený prípad pracovnej neschopnosti v chorobe. Vývoj v čase daného ukazovateľa je v poslednom období veľmi negatívny. Okrem radikálneho zvýšenia času potrebného na vyliečenie pacienta sa dramaticky zvýšila aj regionálna disparita tohto ukazovateľa. Práve obyvatelia regiónov s najvyššou mierou nezamestnanosti sú zároveň zaťažení najdlhšou priemernou dobou pracovnej neschopnosti v chorobe.

Key words: Sickness absence, regional disparities, Gini coefficient, coefficient of variation.

Kľúčové slová: Pracovná neschopnosť, regionálne rozdiely, Giniho koeficient, variačný koeficient.

JEL classification: A13, I15, I31

1. Úvod

Vplyv krízy na zdravie obyvateľstva je v súčasnosti veľmi diskutovanou a aktuálnou problematikou. Obavy obyvateľstva o ich existenciu, stres z možnej straty zamestnania a tým pádom aj neschopnosť úhrady niektorých svojich záväzkov, keď hlavne mladí ľudia čerpajú hypotéky a úvery na uspokojenie svojich potrieb, má vplyv na ich zdravie. Otázkami psychického zdravia v súvislosti s ekonomickou krízou sa zaoberali napríklad experti z Veľkej Británie[1]. Podľa expertov čoraz viac ľudí je nútených obrátiť sa na lekárov v oblasti duševného zdravia.

Duševné zdravie však môže vyvolať rôzne komplikácie, ktoré sa môžu prejaviť v zhoršení všeobecného zdravia jednotlivca. V príspevku sme sa preto zamerali na sledovanie vývoja vybraných údajov štatistiky pracovnej neschopnosti pre chorobu, nakoľko považujeme daný ukazovateľ za vhodný na meranie vývoja chorobnosti, ako aj na sledovanie regionálnych rozdielov v chorobnosti podľa krajov Slovenska.

[1] Royal College of Psychiatrists Mental Health Network, NHS Confederation & London School of Economics and Political Science: *Mental health and the economic downturn*, National priorities and NHS solutions

Štatistika pracovnej neschopnosti obsahuje široký rozsah informácií o pracovnej neschopnosti pre chorobu a úraz podľa štatistickej klasifikácie ekonomických činností, územia a iných kritérií.[2] Zamerali sme sa na analýzu vývoja priemernej doby pracovnej neschopnosti pre chorobu, ktorá sa vyjadrí ako počet kalendárnych dní pracovnej neschopnosti pripadajúci na jeden novohlásený prípad pracovnej neschopnosti. Regionálne rozdiely boli sledované na úrovni NUTS 3, táto klasifikácia vychádza z územnej klasifikácie Slovenska na osem krajov.

2. Regionálne disparity priemernej doby pracovnej neschopnosti pre chorobu

Existencia regionálnych disparít socio-ekonomických ukazovateľov je všeobecne známym faktom. Zamerali sme sa preto na sledovanie týchto disparít v rámci ukazovateľa, ktorý nám môže napovedať o existencii spojitosti medzi vyspelosťou regiónu, v zmysle napríklad vyššieho hrubého domáceho produktu na obyvateľa, nižšej miery nezamestnanosti, pričom predpokladáme, že vo vyspelejších regiónoch Slovenska (podrobnejšie pozri napr. Želinský‑Stankovičová, 2012) bude zároveň zdravie obyvateľstva merané priemernou dobou pracovnej neschopnosti pre chorobu na nižšej úrovni, než tomu bude v regiónoch zaostalejších.

Čím je priemerná doba pracovnej neschopnosti vyššia, tým viac súvisí s horším zdravím obyvateľov Slovenska a v prípade regionálneho porovnania, čím je hodnota tohto ukazovateľa vyššia v niektorom z krajov, tým viac signalizuje horšie zdravie v danom kraji. Priemerná doba pracovnej neschopnosti pre chorobu vykazuje už dlhšie obdobie rastúci trend. Kým ešte v roku 2001 pripadlo na jeden novohlásený prípad pracovnej neschopnosti pre chorobu 26,1 dní pracovnej neschopnosti, tak v roku 2008 bola priemerná dobra pracovnej neschopnosti už na úrovni 33 dní, čo znamenalo nárast približne o 7 kalendárnych dní v porovnaní s rokom 2001. Rapídne však priemerná doba pracovnej neschopnosti rástla práve od roku 2008. S týmto rokom sa spája vznik ekonomickej krízy, aj keď na Slovensku sa rozvinula až v roku 2009. V roku 2009 vzrástla priemerná doba pracovnej neschopnosti pre chorobu medziročne skoro až o 11 dní, čo predstavovalo maximálny medziročný nárast priemernej doby pracovnej neschopnosti pre chorobu. Tak prudký nárast tohto ukazovateľa môžeme dať aj do súvisu s príchodom ekonomickej krízy, pretože ako sme už spomenuli práve na prelome rokov 2008 a 2009 sa začína rozširovať kríza z amerického kontinentu na európsky kontinent.

V roku 2012 dosiahla priemerná doba pracovnej neschopnosti pre chorobu svoje maximum a to hodnotou 49,9 dňa. Z pohľadu rodovej rovnosti môžeme skonštatovať, že v sledovaných rokoch vykazovali ženy vyššie hodnoty daného ukazovateľa, pričom rozdiel v priebehu rokov rástol. Kým v roku 2001 bol rozdiel medzi pohlaviami len približne 2 dni, tak v roku 2012 dosiahol rozdiel priemernej doby pracovnej neschopnosti medzi pohlaviami rozdiel až 5,5 kalendárnych dní.

Tab. 5: Priemerná doba pracovnej neschopnosti pre chorobu

Územie, pohlavie / Rok	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Slovenská republika	**26,1**	**27,1**	**26,8**	**33,4**	**31,2**	**33,4**	**33,6**	**33,0**	**43,9**	**46,1**	**44,6**	**49,9**
Muži	25,2	26,2	25,7	32,8	31,3	32,9	..	30,4	42,7	44,9	44,6	47,0
Ženy	27,0	27,9	27,7	33,9	31,2	33,9	..	35,3	45,1	47,1	46,2	52,5

Ako miery regionálnych disparít priemernej doby pracovnej neschopnosti sme použili variačný koeficient a Giniho koeficient. Variačný koeficient sa vyjadrí ako podiel štandardnej odchýlky a priemeru, pričom ho môžeme vyjadriť aj v percentách (Löster‑Řezanková‑Langhamrová, 2009, Chajdiak, 2010) . Jeho výhodou je jednoduchý výpočet, ako aj to, že

[2] http://portal.statistics.sk/showdoc.do?docid=45

nám umožňuje porovnať variabilitu rôznych súborov (Sodomová a kol., 2000), v našom prípade variabilitu meranú v jednotlivých sledovaných obdobiach. Giniho koeficient sa používa hlavne ako nástroj na meranie dôchodkovej nerovnosti, pričom nadobúda hodnoty od 0 po 1. Jeho hodnota rovná nule, by charakterizovala absolútnu rovnosť a hodnota 1 absolútnu nerovnosť. Výpočet Giniho koeficienta je možné realizovať viacerými spôsobmi. V príspevku sme aplikovali nasledovný vzťah[3]:

$$Giniho\ index = \frac{1}{2n^2\bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| y_i - y_j \right| \qquad (1)$$

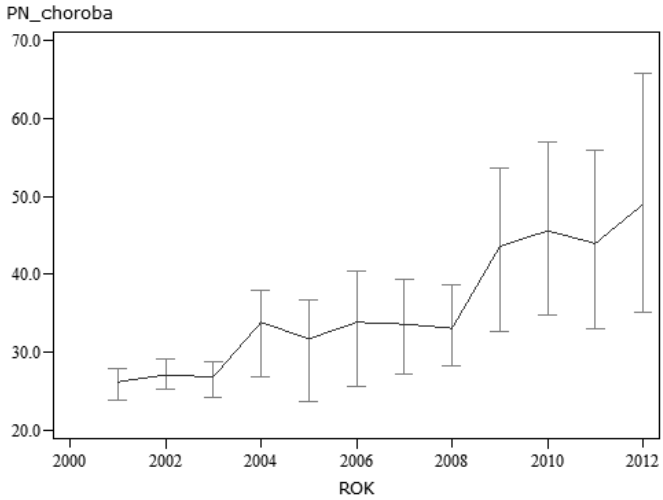
kde n - je celkový počet územných jendotiek,

y_i - je hodnota sledovaného ukazovateľa v i-tej územnej jdonotke,

y_j - je hodnota sledovaného ukazovateľa v j-tej územnej jdonotke,

$\bar{y}$ - je aritmetický priemer sledovaného ukazovateľa y.

Na nižšie uvedenom obrázku sú zobrazené hodnoty priemernej doby pracovnej neschopnosti pre chorobu od roku 2001 do roku 2012, pričom sú v grafe zachytené tak minimálne, mediánové, ako aj maximálne hodnoty tohoto ukazovateľa v jednotlivých krajoch SR. Údaje sú prevzaté z regionálnej databázy Štatistického úradu SR, Regstat[4]. V grafe je jednak zreteľne viditeľný nárast hodnôt priemernej doby pracovnej neschopnosti pre chorobu, ako aj nárast rozdielov medzi minimálnou a maximálnou hodnotou znaku.



Obr. 24: Minimálna, mediánová a maximálna doba priemerného počtu dní pracovnej neschopnosti pre chorobu krajov SR

[3] Matlovič, R., Matlovičová, K. (2005): Vývoj regionálnych disparít na Slovensku a problémy regionálneho rozvoja Prešovského kraja.

[4] http://px-web.statistics.sk/PXWebSlovak/

Minimálna hodnota priemernej pracovnej neschopnosti pre chorobu v roku 2001 bola dosiahnutá v Trenčianskom kraji (23,8) a maximálna hodnota v Prešovskom kraji (27,9). Absolútny rozdiel tak dosiahol hodnotu 4,1 kalendárneho dňa. Aj v nasledujúcom roku sa situácia nezmenila a najnižšia hodnota bola dosiahnutá v Trenčianskom kraji. Od roku 2003 bola minimálna priemerná hodnota pravidelne dosahovaná v Bratislavskom kraji. Najvyššia hodnota priemerného počtu dní pracovnej neschopnosti bola každoročne v sledovaných rokoch vykazovaná v Prešovskom kraji. Druhým najhorším krajom z pohľadu priemernej doby pracovnej neschopnosti je Košický kraj. Oba tieto kraje sú už dlhodobe na konci rebríčka regionálneho porovnania evidovanej miery nezamestnanosti. Nezamestnanosť, hlavne dlhodobá miera nezamestnanosti, je fenoménom obdobia krízy a spôsobuje značné tak ekonomické, ako aj sociálne, psychické problémy v krajinách (danou problematikou sa zaoberajú práce: Pavelka, 2011, Löster‑Langhamrová, 2011, Miskolczi ‑ Langhamrová ‑ Fiala, 2011).

V roku 2012 bola minimálna hodnota priemernej doby pracovnej neschopnosti pre chorobu na úrovni 35,2 dňa a maximálna hodnota 65,8 kalendárnych dní. Absolútny rozdiel sa tak vyšplhal až na 30,6 dňa, čo je v porovnaní s hodnotou z roku 2001 radikálny nárast absolútneho rozdielu. Najvyššie hodnoty sledovanej premennej boli dosahované v regiónoch s najvyššou mierou evidovanej nezamestnanosti a najnižšou mierou hrubého domáceho produktu na obyvateľa. Samozrejme nie je možné len na základe tohto jediného ukazovateľa usúdiť o tom, že obyvatelia zaostalejších regiónov sú na tom z pohľadu ich zdravia horšie než obyvatelia vyspelejších regiónov, avšak na základe porovnania priemernej doby pracovnej neschopnosti vidíme, že práve ťažko skúšané oblasti Slovenska sú zároveň zaťažené aj vysokým počtom kalendárnych dní práceneschopnosti pripadajúcich na jeden novohlásený prípad pracovnej neschopnosti. Hlavne choroby duševné si vyžadujú veľmi dlhé obdobie liečenia. Jednou z možností je, že vplyvom zhoršenej ekonomickej situácie v týchto regiónoch sa skutočne v značnej miere zhoršila zdravotná situácia daného obyvateľstva alebo je možné uvažovať aj o tom, že sa niektorí obyvatelia nechávajú z obavy o svoje pracovné miesta, ktoré sa majú napríklad rušiť, vypísať lekárom na dlhší čas. Na Slovensku už boli identifikované prípady brania úplatkov lekármi, ktorí odobrovali práceneschopnosti pacientov za odplatu.

Tab. 2: Ukazovatele regionálnej disparity priemernej doby pracovnej neschopnosti pre chorobu krajov na Slovensku

Ukazovateľ / Rok	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Variačný koeficient	5,0	5,0	6,1	11,8	12,5	12,8	11,8	10,5	14,3	14,2	15,7	18,9
Gini koeficient	0,026	0,026	0,032	0,060	0,061	0,064	0,062	0,054	0,073	0,073	0,080	0,096

Vychádzajúc z tabuľky 2 je zrejmé, že rozdiely medzi regiónmi rastú. Kým na začiatku sledovaného obdobia dosiahol variačný koeficient iba 5 %, postupne sa zvyšoval až na hodnotu 12,8 % v roku 2006. Obdobný trend mal aj vývoj Giniho koeficienta. V období rokov, keď sa ekonomike Slovenska darilo, sa koeficienty regionálnej disparity znížili a to konkrétne v rokoch 2007 a 2008. Ekonomická kríza mala na Slovenku vplyv na vývoj rôznych ekonomických ukazovateľov a charakteristík. Za takúto premennú môžeme považovať aj priemernú dobu pracovnej neschopnosti v chorobe, ktorá rapídne rástla od roku 2009. Okrem toho, že došlo k rastu hodnôt sledovanej premennej, zvyšovali sa aj regionálne rozdiely sledovaného ukazovateľa

V roku 2012 dosiahol variačný koeficient hodnotu 18,9 %, čo je najvyššia hodnota spomedzi všetkých hodnôt tohto ukazovateľa od roku 2001. Oba trendy, teda jednak zvyšovanie priemernej doby pracovnej neschopnosti v chorobe, ako aj zvyšovanie regionálnej disparity musíme hodnotiť veľmi negatívne.

S ekonomickou krízou úzko súvisí stav zdravia obyvateľstva, aj keď v tomto prípade sme sa zamerali na meranie zdravia iba jediným ukazovateľom. V prípade, že sa ekonomike nedarí, klesá zamestnanosť, zvyšuje sa tým pádom nezamestnanosť, klesá hrubý domáci produkt, klesá domáci dopyt, klesá dopyt po zdravých potravinách, ktoré sú často cenovo náročnejšie, a ako sme už poukázali odkazom na zahraničnú literatúru preukazuje sa vplyv stresu v čase krízy na duševné zdravie človeka. Všetky spomínané aspekty môžu spôsobovať zvyšovanie doby pracovnej neschopnosti pre chorobu, ktorá je potrebná pre prekonanie daného zdravotného problému. Musíme si uvedomiť, že hlavne psychické zdravotné problémy sú liečené dlhodobo.

3. Záver

Verejné zdravie je v súčasnosti v popredí záujmu tak samotných jedincov,m ako aj vlád krajín EÚ. To, že verejné zdravie je fenoménom súčasnosti, vyplýva jednak z najpálčivejšieho problému vyspelých krajín súčasnosti a to starnutím obyvateľstva a jednak z vplyvu ekonomickej krízy na zdravie obyvateľstva. Obidva faktory, tak starnutie, ako aj zhoršovanie zdravia vplyvom krízy vyvoláva tlaky na verejné rozpočty vlád. Je preto nevyhnutné sledovať akým spôsobom by sa dal eliminovať vplyv krízy na zdravie. Zatiaľ však vidíme hlavne negatívny vplyv ekonomickej krízy, s ktorou vo zvýšenej miere súvisí aj stres. Práve stres môže mať za následok zvýšenú chorobnosť a nárast prípadov chorôb s dlhšou dobou liečenia. Na Slovensku rastie priemerná doba pracovnej neschopnosti v chorobe rapídne rýchlo.

Kým v roku 2001 bola táto priemerná doba na úrovni 26,1 kalendárnych dní, tak v roku 2012 sa daný ukazovateľ vyšplhal až na 49,9 kalendárnych dní. Okrem tohto negatívneho trendu nárastu samotného ukazovateľa hlavne po roku 2008, je zreteľný aj rapídny nárast regionálnych disparít priemernej doby pracovnej neschopnosti pre chorobu. Najvyššie hodnoty sledovanej premennej boli v rokoch 2001 až 2012 zistené v Prešovskom kraji. Regionálne rozdiely merané variačným koeficientom vzrástli z 5,0 % v roku 2001 na úroveň 18,9 % v roku 2012.

Príspevok bol spracovaný v rámci projektu VEGA č. 1/0906/11.

Literatúra

CHAJDIAK, J. 2010. Štatistika jednoducho. Bratislava, STATIS 2010. ISBN 978-80-85659-60-3.

LÖSTER, T. – ŘEZANKOVÁ, H. – LANGHAMROVÁ, J. 2009. Statistické metody a demografie. 1. vydanie. VŠEM, Praha. s. 297. ISBN 978-80-86730-43-1.

LÖSTER, T. – LANGHAMROVÁ, J. 2011. Analysis of long-term unemployment in the Czech Republic. In *International Days of Statistics and Economics. Conference Proceedings.* ISBN 978-80-86175-77-5. pp. 307 – 316.

MATLOVIČ, R., MATLOVIČOVÁ, K. 2005. Vývoj regionálnych disparít na Slovensku a problémy regionálneho rozvoja Prešovského kraja. In: *Acta Facultatis Studiorum Humanitatis et Naturae Universitatis Prešoviensis*, Prírodné vedy, *Folia geographica*, 2005. XLIII, 8. ISSN 1336-6157. s. 66-88.

MISKOLCZI, M. – LANGHAMROVÁ, J. – FIALA, T. 2011. Unemployment and GDP. In *International Days of Statistics and Economics. Conference Proceedings.* ISBN 978-80-86175-77-5. pp. 407 – 415.

PAVELKA, T. 2011. Long term unemployment in the Czech Republic in comparison with the other countires of the European Union. In *International Days of Statistics and Economics. Conference Proceedings.* ISBN 978-80-86175-77-5. pp. 481 – 489.

Royal College of Psychiatrists Mental Health Network, NHS Confederation & London School of Economics and Political Science. 2009. *Mental health and the economic downturn, National priorities and NHS solutions*. Royal College of Psychiatrists reference: Occasional Paper OP70. 2009. Dostupné na: http://www.rcpsych.ac.uk/files/pdfversion/OP70.pdf

SODOMOVÁ, E. a kol.: *Štatistik: modul A*. Bratislava: Ekonóm. 2000. ISBN 80-225-1270-2

ŽELINSKÝ, T. – STANKOVIČOVÁ, I. 2012. Spatial aspects of poverty in Slovakia. In *The 6th International Days of Statistics and Economics. Conference Proceedings.* ISBN 978-80-86175-86-7. pp. 1228 – 1235.

http://portal.statistics.sk/showdoc.do?docid=45

Regionálna databáza, ŠÚ SR. Dostupné na: http://px-web.statistics.sk/PXWebSlovak/

Adresa autorov:

Silvia Megyesiová, Ing. PhD.
Podnikovohospodárska fakulta, EU
Tajovského 13, 041 30 Košice
silvia.megyesiova@euke.sk

Vanda Lieskovská, prof. Ing. PhD.
Podnikovohospodárska fakulta, EU
Tajovského 13, 041 30 Košice
lieskovska@euke.sk

Rozpoznávanie entít v texte
Entity recognition in text

Andrej Mihálik

Abstract: The paper is focused on the task of information extraction, specifically its subtask of entity recognition in text. At first, the field of practice for text mining is defined as well as groups of related technologies. After clarifying the term of entity and its importance for text analysis, main features used for entity extraction are presented as well as strategies of implementation of entity recognition into practice either by rule definition or use of statistical classifiers and their strengths and weaknesses. In the practical section of the paper rule-based entity extraction (using noun phrases) is demonstrated on product reviews.

Abstrakt: Príspevok sa zameriava na úlohu extrakcie informácii, a síce jej čiastkovú problematiku rozpoznávania entít v texte. Úvodom je definované pole pôsobnosti hĺbkovej analýzy textu, ako aj jeho členenie na jednotlivé skupiny súvisiacich technológii. Po objasnení pojmu entity a jej význame pre analýzu textu sú uvádzané hlavné znaky používané pre ich identifikáciu a spôsoby implementácie rozpoznávania entít spočívajúce buď v definovaní pravidiel alebo použití štatistického klasifikátora spolu s ich silnými a slabými stránkami. V praktickej časti je ďalej demonštrované použitie extrakcie entít založenom na pravidlách, konkrétne na identifikácii slovných spojení zložených z podstatných mien na dátach tvorených produktovými recenziami.

Kľúčové slová: hĺbková analýza textu, extrakcia informácii, spracovanie prirodzeného jazyka, extrakcia entít, extrakcia entít založená na pravidlách

Key words: text mining, information extraction, natural language processing, entity extraction, rule-based entity extraction

JEL classification: C19

1. Úvod

Matematika, štatistika a výpočtové technológie nám umožňujú objavovať stále nové postupy a algoritmy na spracovanie obrovského množstva dát produkovaných ľudskou spoločnosťou. Vďaka nim si dokážeme zachovať nad dátami určitý „nadhľad“ odlíšením podstatného od nepodstatného, a teda trendov od šumu. Matematické a štatistické modely však od dát väčšinou požadujú určitú pevne danú štruktúru. Štruktúrované dáta však predstavujú len zlomok dostupných dát, neštruktúrované dáta predovšetkým v podobe textu sú oveľa frekventovanejšie. Pre uplatnenie spomínaných postupov je v prípade týchto dát nevyhnutná ich transformácia na štruktúrované dáta a číselné ukazovatele. Dosiahnutie tejto transformácie je náplňou hĺbkovej analýzy textu (text miningu). Táto vedecká disciplína nachádzajúca prieniky so štatistikou, hĺbkovou analýzou dát (data miningom), strojovým učením, manažérskymi vedami a umelou inteligenciou predstavuje technológie na analýzu a spracovanie neštruktúrovaných dát (Miner, 2012). Hĺbková analýza textu nie je jednotnou metódou, skôr si ju môžeme predstaviť ako pojem zastrešujúci veľké množstvo navzájom si konkurujúcich postupov na rôznych stupňoch vývoja. Podľa Minera môžeme tieto postupy zaradiť do siedmych veľkých zoskupení po zohľadnení:

- granularity dát podľa toho, či nás zaujímajú samotné slová alebo celý dokument
- zamerania (špecifické informácie alebo celok)

- dostupných informácii (máme k dispozícii pevne dané triedy alebo ich vytvoríme podľa spoločných vlastností)
- preferencie syntaxe (štruktúry) alebo sémantiky (významu)
- zdroja textu (jednoduchý text alebo text s odkazmi v prípade webových zdrojov)

Na základe týchto kritérií členíme technológie spojené s hĺbkovou analýzou textu na extrakciu informácií, extrakciu konceptov, spracovanie prirodzeného jazyka, zber informácií, zhlukovanie, klasifikáciu dokumentov a hĺbkovú analýzu webu (web mining) (Miner, 2012). V závislosti od úlohy väčšinou kombinujeme viacero z týchto postupov na dosiahnutie najlepšieho možného výsledku. Veľmi často tak napríklad používame spracovanie prirodzeného jazyka na syntaktickú analýzu textu, ktorá potom môže slúžiť ako vstup napríklad pre klasifikáciu dokumentu. V našom príspevku sa zameriame na analýzu obsahu dokumentu (podľa prvého členenia pôjde teda o analýzu na úrovni slov), pričom sa pokúsime z neho extrahovať určité špecifické informácie. Cieľom extrakcie informácii je identifikovať v texte entity, odhaliť vzťahy medzi nimi a štrukturovať tak text.

2. Problém extrakcie entít

Extrakcia entít predstavuje čiastkovú úlohu v problematike extrakcie informácii. Vstupom pre túto úlohu je jeden alebo viacero dokumentov a výstup je zoznam entít prítomných v tomto texte. Entita je v tomto prípade väčšinou nejaký konkrétny objekt, a teda ľudia, miesta, organizácie alebo veci, niekedy sem zahŕňame aj čas a čísla. V prípade tvorby systémov, ktoré majú slúžiť na zodpovedanie užívateľských otázok ide teda o odpovede na otázky ako: Kto? Čo? Kde? Kedy? Koľko? Definícia entity je okrem toho závislá aj na konkrétnej aplikácii: iné entity sú zaujímavé v produktových recenziách (produktové atribúty) a iné v článkoch o známych osobnostiach (osoby). Z hľadiska vetnej stavby predstavujú entity podmet alebo predmet a asociujeme ich s podstatnými menami. Získaním entít a teda zodpovedaním už uvedených základných otázok môžeme ľahko zistiť o čom článok je, medzi ďalšie aplikácie rozpoznávania entít patria okrem iného:

- Ak nás zaujímajú názory zákazníkov pre daný produkt, môže byť vhodné na základe recenzií alebo článkov o konkrétnom produkte zistiť, akými atribútmi sa vyznačuje: v prípade auta môže ísť o výkon a typ motora, spotrebu, dizajn alebo objem kufra. Ak poznáme atribúty produktu, môžeme v zákazníckych recenziách každému z nich priradiť zákaznícky sentiment, čo nám pomôže odhaliť jeho silné a slabé stránky. Týmto typom rozpoznávania entít sa budeme zaoberať aj ďalej v krátkej prípadovej štúdii.
- Extrakcia entít nám môže slúžiť na automatické zistenie kľúčových slov (alebo tagov) článkov na našej webstránke.
- Veľmi vhodnou je táto technika aj pri konverzii neštruktúrovaného textu do štruktúrovanej podoby (napríklad tabuľky).

Cieľom extrakcie entít je teda rozhodnúť, či je dané slovo alebo skupina slov súčasťou určitej entity. Pri tomto rozhodnutí sa berú do úvahy väčšinou znaky slova samotného, ako aj znaky slov v blízkosti tohto slova.

Znaky, ktoré pri takejto klasifikácii zohľadňujeme, môžeme rozčleniť do troch hlavných kategórii:

- prítomnosť slova v určitom slovníku – pre uľahčenie rozpoznávania slov ako entít a určenie druhu entity nám môžu veľmi pomôcť slovníky obsahujúce najčastejšie sa vyskytujúce názvy entít. Môže pritom ísť napríklad o zoznam krstných mien alebo väčších miest. Problémom môže byť nejednoznačná kategorizácia slov vyskytujúcich sa vo viacerých slovníkoch (napr. Martin môže byť aj krstné meno, ale aj mesto), v tomto prípade musíme zobrať do úvahy ďalšie vlastnosti slova.
- tvar slova môžeme veľmi jednoducho implementovať v našom extraktore pomocou regulárnych výrazov. Ak slovo napríklad začína veľkým písmenom a nie je zároveň prvým

slovom vo vete, je veľmi pravdepodobné, že ide o vlastné meno, v niektorých jazykoch (napr. nemčina) tak môžeme pomerne jednoducho identifikovať podstatné mená vôbec. Značky a akronymy pre názvy organizácii taktiež vyhľadáme pomerne jednoducho podľa viacerých veľkých písmen nasledujúcich za sebou, pričom môžu byť oddelené bodkou. Prítomnosť čísla alebo pomlčky v slove môže odhaliť názov chemickej zlúčeniny, striedanie veľkých a malých písmen je často príznačné pre názvy produktov (napr. iPad).

- gramatické vlastnosti slova – na odhalenie týchto vlastností nám slúži spracovanie prirodzeného jazyka, ktorá na základe syntaktických vlastností vety poskytuje algoritmy na zaradenie slov do slovných druhov (POS tagging), vďaka čomu môžeme vo vete ďalej rozoznávať väčšie zoskupenia slov s určitými vlastnosťami (chunking alebo shallow parsing). Ako bolo spomenuté, entity sú skoro výlučne podstatné mená, ak teda úspešne identifikujeme podstatné meno, odhalili sme aj potenciálnu entitu. Dôležité je ale uvedomiť si, že entita nemusí zodpovedať jednému slovu. Pre odhalenie entít presahujúcich hranicu slova je preto potrebné skúmať zoskupenia slov s určitými vlastnosťami, veľmi často to bývajú slovné spojenia tvorené dvoma a viacerými podstatnými menami. Entita „Fakulta managementu Univerzity Komenského“ sa skladá zo štyroch po sebe nasledujúcich podstatných mien, v prípade, že by sme sa pozerali na slová jednotlivo, mohli by sme tak mylne identifikovať štyri nezmyselné entity.

Po úvodnej analýze textu, definícii a určení znakov, podľa ktorých sa pri extrakcií entít budeme riadiť je potrebné určiť spôsob implementácie extrakcie entít. Dominujú pritom dva základné prístupy (Ingersoll, 2013):

- extrakcia entít založená na pravidlách – je historicky prvým, dnes ale menej používaným prístupom. Na základe vyššie uvedených znakov si určíme pravidlá pre to, čo je a čo nie je entita, každému pravidlu priradíme váhu a vyhodnotíme pravdepodobnosť toho, že dané slovo je entitou. Ak táto pravdepodobnosť prekročí určitú hranicu, slovo klasifikujeme ako entitu a prípadne určíme aj jej druh. Pravidlá a váhy musíme potom veľmi často prehodnocovať v závislosti od toho, s akým textom pracujeme. Ak si ako pravidlo pre entitu určíme veľké začiatočné písmeno, toto pravidlo môže byť veľmi úspešné v odbornom texte, ale menej pri analýze chatu na internetovom portáli, kde niektorí užívatelia zvyknú používať zásadne malé písmená. Tento prístup teda nemusí byť dostatočne flexibilný. Jeho výhodou je, že už na základe niekoľko málo pravidiel môžeme odhaliť veľký podiel relevantných entít. Vždy však bude existovať pomerne veľké množstvo prípadov, ktoré môžu byť výnimkami a na zachytenie ktorých by bolo potrebné veľké množstvo veľmi komplexných pravidiel. Pri tomto prístupe tak pri pomerne malej námahe môžeme dosiahnuť uspokojivé výsledky, pre ďalšie zlepšenie úspešnosti však vyžadované úsilie rastie neúmerne. Ak chceme teda vytvoriť aplikáciu s výsledkami približujúcimi sa klasifikáciou ľuďmi, môže ísť o pomerne nákladnú metódu.
- použitie štatistických klasifikátorov – v tomto prípade ide o štandardnú klasifikáciu „pod dohľadom učiteľa“ so známymi triedami (nie je entita, je entita, prípadne aj jednotlivé druhy entít – osoba, miesto, organizácia) a definovanými znakmi, ktoré chceme použiť ako vstup pre klasifikáciu slova. Pravidlá pre klasifikáciu textu tak vznikajú automaticky na základe podmienených pravdepodobností príslušnosti ku konkrétnej triede za výskytu daných znakov. Na vytvorenie modelu je však nevyhnutné použitie dátovej množiny slúžiacej na trénovanie, čo v praxi znamená, že potrebujeme človekom klasifikovaný text, teda text, kde je pre každé slovo identifikovaná jeho trieda. Rozsah takto ručne klasifikovaného textu by mal pritom predstavovať aspoň 30 000 slov (Ingersoll, 2013). Hoci tento prístup vyžaduje vyvinutie väčšieho úsilia na začiatku, dosahuje v prípade úspešnej implementácie lepšie výsledky ako prístup založený na pravidlách. Okrem toho nám umožňuje kvantifikovať úspešnosť nášho modelu pomocou:
  - miery presnosti (precision):

$$presnosť = \frac{počet\ korektne\ identifikovaných\ entít}{počet\ identifikovaných\ entít} \quad (1)$$

- miery rozpoznania (recall):

$$rozpoznanie = \frac{počet\ korektne\ identifikovaných\ entít}{celkový\ počet\ entít} \quad (2)$$

- a F-skóre predstavujúceho harmonický priemer z mier presnosti a rozpoznania:

$$F = 2.\frac{presnosť.rozpoznanie}{presnosť+rozpoznanie} \quad (3)$$

3. Prípadová štúdia: extrakcia entít z recenzii o notebookoch

Zdrojom dát pre náš systém pre extrakciu entít boli recenzie notebookov nachádzajúce sa na http://reviews.cnet.com/laptops/. Použitím regulárnych výrazov sme získali zoznam URL s dokopy 70 recenziami. Zdrojový kód každej stránky s recenziou sme najprv museli očistiť, aby sme extrahovali čistý text použiteľný pre našu analýzu, použili sme pritom nasledovné regulárne výrazy:

```
text = re.findall(r'<div id="editorReview">(.*?)<div class="pageNav"
section="paginate">',html,re.DOTALL)[0]

sub_strings  = ['<p>', '</p>', '<div.*?>', '</div>', '<img.*?/>',
'<span class="image-credit.*?/span>', '<noscript>', '<!--.*?>',
'<table.*?/table>', '<style.*?/style>', '<a.*?>', '</a>', '<br/>',
'\n','<br>','<b>','</b>','<i>','</i>']

for sub_string in sub_strings:

  text = re.sub(sub_string,'',text,0,re.DOTALL)
```

Pre extrakciu recenzií sme sa rozhodli použiť aplikáciu pravidiel, pričom sa obmedzíme na jediný gramatický znak: príslušnosť slova k slovnému spojeniu tvorenému podstatnými menami. V tomto prípade teda môžeme stotožniť úlohu rozpoznávania entít s identifikáciou fráz zložených z podstatných mien. Na tento účel použijeme knižnicu jazyka Python nltk slúžiacu na spracovanie prirodzeného jazyka. Proces identifikácie entít potom pozostáva z nasledujúcich krokov:

1. Text rozčleníme na vety.

```
sentences = nltk.sent_tokenize(text)
```

2. Vety ďalej rozdelíme na jednotlivé slová.

```
sentences = [nltk.word_tokenize(s) for s in sentences]
```

3. Ku každému slovu priradíme slovný druh (POS tagging).

```
sentences = [nltk.pos_tag(s) for s in sentences]
```

4. Vyhľadáme vo vetách frázy zložené z podstatných mien. Najprv si musíme definovať vhodné gramatické pravidlo, v tomto prípade je to následnosť jedného alebo viacerých podstatných mien nezávisle od typu podstatného mena. Toto pravidlo následne aplikujeme na každú vetu každej recenzie.

```
grammar = "NP: {<NN.*>+}"
chunk_parser = nltk.RegexpParser(grammar)
for sentence in sentences:
  parsed_tree = chunk_parser.parse(sentence)
```

5. Každé jedinečné slovné spojenie spĺňajúce toto pravidlo si uložíme do textového súboru.

Výsledkom tohto procesu je zoznam zhruba 4000 slov, z ktorého uvádzam krátky zoznam:

```
b keys, back, back catalogs, back edge, back panel, back side
edges, back surface, back-breaker, background, backlight,
```

```
backlighting, backpack, backspace, backspace keys, backup, backup
discs, bag, bags, balance, bandwagon
```

Tento v tomto prípade pomerne rozsiahly zoznam entít (atribútov produktu ale aj podstatných mien vyskytujúcich sa vo frázach používaných v recenziách) môžeme v závislosti od aplikácie ďalej skrátiť napríklad:

- obsiahnutím slov, ktoré sa vyskytujú vo všetkých recenziách častejšie ako 5-krát
- vytvorením zoznamu entít pre recenziu týkajúcu sa úplne odlišného produktu (napríklad kozmetiky) a vylúčením týchto slov zo zoznamu, aby sme vylúčili frázy vyskytujúce sa v recenziách, ktoré nesúvisia s produktom nášho záujmu.

4. Záver

V príspevku sme prezentovali hlavné teoretické východiská spojené s problematikou extrakcie entít. Hĺbkovú analýzu dát (text mining) môžeme vnímať ako súbor technológii slúžiacich na štrukturovanie textu do podoby použiteľnej štatistickým, matematickým a výpočtovým aparátom. Podľa konkrétnych potrieb nášho vedeckého skúmania pritom rozlišujeme extrakciu informácií, extrakciu konceptov, spracovanie prirodzeného jazyka, zber informácii, zhlukovanie, klasifikáciu dokumentov a hĺbkovú analýzu webu. Obvykle sa tieto technológie používajú v kombinácii. Problematika extrakcie informácii je zameraná na hľadanie častí textu (slov a slovných spojení) s určitými vlastnosťami. Patrí sem aj rozpoznávanie entít alebo nezávislých objektov, ktoré v texte prinášajú odpovede na otázky: Kto? Čo? Kde? Kedy? Koľko? Pre odhalenie entity nám slúžia znaky slov: ich príslušnosť k preddefinovaným zoznamom entít, tvar slova a jeho gramatické vlastnosti. Pri tvorbe záverov o tom, či dané slovo je alebo nie je entita na základe uvedených znakov môžeme definovať pravidlá alebo použiť štatistický klasifikátor. V prvom prípade sa náš model môže skomplikovať existenciou veľkého množstva výnimiek a špecifikami konkrétneho textu, v druhom je potrebné mať k dispozícii dátový set na trénovanie modelu. V praktickej časti sme si prezentovali ako pomocou regulárnych výrazov môžeme získať text z internetových zdrojov a na základe gramatických vlastností slov vytvoriť veľmi jednoduchý systém na extrakciu entít.

Literatúra

BIRD, S. - KLEIN, E. - LOPER, E. 2009. Natural Language Processing with Python. Sebastopol, CA : O'Reilly, 2009. ISBN 978-0-596-51649-9.

INGERSOLL, G. - MORTON, T. - FARRIS, A. 2013. Taming Text: How to Find, Organize, and Manipulate It. Shelter Island, NY : Manning Publications Co. All, 2013. s. 320. ISBN 978-1933988382.

MINER, G. & al. 2012. Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications. Waltham, MA : Academic Press, 2012. s. 1000 s. ISBN 978-0123869791.

Mgr. Andrej Mihálik
Fakulta managementu UK
Odbojárov 10
820 05 Bratislava
Andrej.Mihalik@fm.uniba.sk

Extrémne príjmy a ich vplyv na miery príjmových nerovností
Extreme incomes and their influence on income inequality measures

Ivan Mojsej, Alena Tartaľová

Abstract: The presence of the extreme values on the both tails of the income distribution can affect the characteristics constructed from the data. Social indicators of poverty and inequality are known to be potentially sensitive to the occurrence of extreme incomes. The paper presents sensitivity analysis of indicators estimated from EU SILC which is the reference source for comparative statistics on income distribution and social exclusion in the EU. We considered simple data adjustment, trimming and winsorizing.

Abstrakt: Prítomnosť extrémnych hodnôt na oboch koncoch rozdelenia príjmov môže ovplyvniť vypočítané číslené charakterstiky z týchto údajov. Sociálne indikátory nerovnosti a chudoby su známe tým, že sú citlivé na výskyt extrémenych hodnôt. V práci prezentujeme analýzu citlivosti vybraných indikátorov odhadnutých na základe údajov z databázy EU SILC, ktorá predstavuje referenčný zdroj údajov, ktorý slúži na porovnávanie príjmového rozdelenia a sociálnej exklúzie v rámci EÚ. Uvažovali sme dve metódy úpravy údajov, „trimming" a „winsorizing".

Key words: inequality, extreme income, trimming, winsorizing, EU SILC

Kľúčové slová: nerovnosť, extrémne príjmy, trimming, winsorizing, EU SILC

JELclassification: C13**,** I30, I32

1. Úvod

Hlavným nástrojom v analýzach o príjmoch a životných podmienkach v EÚ je databáza EU SILC (EU Statistics on Income and Living Conditions), ide o zisťovanie, ktoré sa od roku 2003 nariadením EK č. 1177/2003 každoročne realizuje. Výberové zisťovanie EU SILC realizuje od roku 2005 na Slovensku Štatistický úrad Slovenskej republiky. Ide o harmonizované zisťovanie členských štátov EU, ktorého úlohou je zabezpečiť produkciu pravidelných, včasných a kvalitných údajov o príjmoch, chudobe a sociálnom vylúčení. Zisťovanie EU SILC sa stalo zdrojovou základňou pre analýzy životnej úrovne obyvateľstva, ako i pre koncepčné zámery a prijímanie opatrení smerujúcich k zvyšovaniu kvality života občanov SR. Jednotkami výberu v EU SILC sú hospodáriace domácnosti a jej súčasní členovia, preto tieto mikroúdaje umožňujú porovnanie na úrovni domácností rovnako aj na úrovni jednotlivca. Hlavným cieľom zisťovania je ponúknuť porovnateľné indikátory chudoby a nerovnosti v rámci EÚ. Problémom týchto indikátorov je však ich citlivosť na extrémne vysoké, ale tiež extrémne nízke príjmy. Navyše niektoré indexy (napríklad Atkinsonov index nerovnosti, Wattsov index chudoby, indexy založené na entropii a pod.) nie sú pre záporné príjmy, ktoré sa v databáze vyskytujú, definované. Podľa odporúčania Eurostatu (pozri Eurostat 2006), by sa mali záporné príjmy nahradiť nulou alebo úplne z databázy vylúčiť. Vo viacerých prácach autorov bolo ukázané, že odhad indexov chudoby a nerovnosti je citlivý na výskyt extrémnych príjmov a to rovnako v prípade extrémne vysokých ako aj nízkych príjmov. Pred samotným výpočtom určitého indikátora by sa teda mala určitá úprava vstupných údajov urobiť. Cieľom tohto príspevku je analýza vplyvu extrémnych hodnôt na vybrané indikátory a tiež metódy úpravy údajov a ich efekt na výslednú hodnotu indikátora. Ukážeme si dve metódy – metódu nazvanú „trimming" a metódu „winsorizing", ktoré neprekladáme, keďže zatiaľ nemajú slovenský ekvivalent. (pozri Eurostat 2007)

2. Úprava súboru s extrémnymi hodnotami a výsledky analýzy citlivosti

Na úpravu údajov sa používajú najmä dve metódy – trimming a winsorizing. Často krát sa tieto metódy zamieňajú, dokonca niekedy aj nesprávne používajú.

Trimming – predstavuje odseknutie určitého percenta údajov na pravom aj ľavom konci rozdelenia, teda vylúčenie najnižších a najvyšších. Môže sa pritom použiť aj to, že sa odsekne napríklad 10 najnižších a 10 najvyšších príjmov. Uvádzame aj kód v programe R na príklade odseknutia 5% hodnôt na oboch koncoch rozdelenia..

```
x_0.05 <- x[ x>quantile(x, .05) & x<quantile(x, .95)]
```

Winsorizing – tento spôsob je podobný predchádzajúcemu, no namiesto vylúčenia hodnôt, sú extrémne nízke a extrémne vysoké hodnoty nahradené hodnotou, ktoré predstavuje tzv. prah odseknutia. V programe R by úprava údajov vyzerala takto:

```
winsorize <- function(x, q=0.05) {
 extrema <- quantile(x, c(q, 1-q))
 x[x<extrema[1]] <- extrema[1]
 x[x>extrema[2]] <- extrema[2]
 x
}
```

Analýzu citlivosti vykonáme na vybraných indikátoroch a indexoch nerovnosti, ktoré sa dajú rozdeliť do troch skupín (pozri Van Kern, 2006):

- Indikátory založené na strednej hodnote – patrí sem *stredná hodnota*, teda aritmetický priemer hodnôt a *medián*. Očakávame, že stredná hodnota, bude viac ovplyvnená úpravou údajov ako medián.

- Indikátory nerovnosti:
  - *Podiel P80/P20* – veľmi jednoduchý, ale pritom efektívny spôsob ako vyjadriť príjmovú nerovnosť je porovnanie decilov. Pomer P80/P20 vyjadruje pomer príjmu, ktorý sa nachádza na 80-tom percentile, teda oddeľuje 20 % najvyšších príjmov, a príjmu, ktorý sa nachádza na 20-tom percentile a oddeľuje 20 % najnižších príjmov. Napríklad hodnota pomeru rovná 3 vyjadruje, že 20 % domácností s najvyššími príjmami má príjem 3-krát vyšší ako 20 % domácností s najnižšími príjmami. Všeobecne platí, že čím je tento podiel vyšší, tým väčšia je miera príjmovej nerovnosti.
  - *Podiel P90/P10* – podobne ako podiel P80/P20, vyjadruje pomer 10 % najvyšších a 10% najnižších príjmov.

- Indexy príjmovej nerovnosti
  - *Giniho koeficient* - Tento ukazovateľ príjmovej nerovnosti je určite najznámejším koeficientom resp. indexom, ktorý sa používa na hodnotenie príjmovej nerovnosti. Giniho koeficient je matematicky založený na Lorenzovej krivke. Giniho koeficient

potom počítame ako pomer plochy medzi nivelizovanou a skutočnou Lorenzovou krivkou k ploche pod nivelizovanou krivkou. (pozri Cowell, 2000)

- *Atkinsonov index* - Index je založený na výpočte tzv. spravodlivého primerného príjmu, ktorý je definovaný ako príjem skupiny, ktorý je rovnomerne rozdelený medzi príjemcov. (pozri Cowell, 2000)
- *Generalized Entropy (GE) index* – Ide o všeobecný index, ktorý spĺňa vlastnosti tzv. „ mean independence“, čo znamená, že ak sa všetky príjmy vynásobia určitou konštantou, miera nerovnosti sa nezmení. Ďalej je to „additive decomposability“ odkazujúca k možnosti dekomponovať nerovnosť na sumu vnútro-skupinových nerovnosti a nerovností medzi skupinami. Dôležitou charakteristikou je aj splnenie tzv. „transfer axiom“, ktorý vyžaduje, aby pri akomkoľvek transfere od bohatého jedinca k chudobnejšiemu došlo k zvýšeniu nerovnosti. (pozri Cowell, 2000)

V tabuľke 1 sú zhrnuté výsledky pre základné charakteristiky a indikátory nerovnosti. V prvom stĺpci je výpočet prevedený pre pôvodné údaje, bez úpravy hodnôt. Ďalších šesť stĺpcov predstavuje výpočet pre dva spôsoby úpravy údajov s rôznymi hranicami.

Tak, ako sme očakávali, medián nie je ovplyvnený žiadnou úpravou údajov. Toto pozorovanie podporuje metodiku využívania mediánového príjmu pri určovaní hranice chudoby. Stredná hodnota je viac ovplyvnená, najnižšia hodnota priemerného príjmu je pre prípad, ak odsekneme z údajov 5 % hodnôt z oboch koncoch rozdelenia. Ak údaje z oboch koncov rozdelenia neodsekneme, ale iba nahradíme príslušnými percentilmi, stredná hodnota sa zmení (zníži sa), ale zmena nie je taká výrazná ako pri prvej metóde. Podiel percentilov je pri metóde „winsorizing“ nezmenený, čo je pochopiteľné, keďže sme použili hranicu maximálne 5 %, pri ktorej sa uvažované percentily a teda ani ich podiel nezmenil. Iná situácia je už pri metóde „trimming“, kde sa podiel percentilov zmenil, nakoľko sa zmenil aj počet údajov.

Tab.6: Porovnanie pôvodných a upravených charakteristík príjmovej nerovnosti

	Pôvodné údaje	**Trimming**			**Winsorizing**		
		1%	**2,5%**	**5%**	**1%**	**2,5%**	**5%**
Počet hodnôt	49 286	48 288	46 804	44 345	49 286	49 286	49 286
Stredná hodnota	7117,864	6966,141	6907,701	6842,970	7017,161	6980,367	6936,827
Medián	6546,722	6546,722	6546,722	6546,722	6546,722	6546,722	6546,722
P90/P10	3,005	2,879	2,704	2,446	3,005	3,005	3,005
P80/P20	1,942	1,902	1,865	1,784	1,942	1,942	1,942

Zdroj: Vlastné spracovanie na základe údajov EU SILC

Analýzu citlivosti na úpravu údajov o extrémne nízke a vysoké hodnoty sme urobili aj pre vybrané indexy, ktoré bežne používame na meranie príjmovej nerovnosti. Giniho koeficient sa úpravou údajov zmenil najmenej. Index vypočítaný z pôvodných údajov mal hodnotu na úrovni 24, 6 %. Po úprave údajov metódou „trimming“ sa hodnoty indexy znížili a pohybujú sa od 22,1 % po 18,4 %. To znamená, že ak vynecháme 5 % najnižších a 5 % najvyšších príjmov, príjmová nerovnosť bude na úrovni 18, 4 %. Pri úprave údajov metódou „winsorizing“ nie je zmena Giniho koeficientu taká výrazná. Hodnoty indexu sa pohybujú od 21, 4 % po 23,4%. Opäť, najnižšia hodnota Giniho koeficientu je v prípade, ak sa 5 %

najvyšších a 5 % najnižších hodnôt nahradí 95. resp. 5 percentilom. O Giniho koeficiente je známe (pozri De Maio, 2007), že je najviac citlivý na zmenu príjmu v strede rozdelenia. Výraznejšia zmena po úprave údajov nastala pri Atkinsonovom indexe. Index vypočítaný z pôvodných údajov mal hodnotu 10,8 %, no po použití úpravy údajov metódou „trimming“ s hranicou 5 %, sa hodnota zmenšila o viac ako polovicu na 5, 2%. Najvýraznejšia zmena nastala po úprave údajov pri výpočete GE indexu. Index vypočítaný zo všetkých hodnôt je 11, 2 %. Ak odsekneme 5 % hodnôt na oboch koncoch rozdelenia, hodnota indexu je iba 5,2 %. Pri nahradení týchto hodnôt v metóde „winsorizing“ je hodnota indexu 7,2 %. Týmto sa nám potvrdilo to, čo pozorovali aj iní autori.

Tab.2: Porovnanie pôvodných a upravených indexov na meranie príjmovej nerovnosti

Koeficient	Pôvodné údaje	Trimming			Winsorizing		
		5%	2,5%	1%	5%	2,5%	1%
Giniho	0,246	0,184	0,204	0,221	0,214	0,227	0,234
Atkinsonov	0,108	0,052	0,066	0,080	0,072	0,083	0,092
GE index	0,112	0,052	0,066	0,079	0,072	0,082	0,090

Zdroj: Vlastné spracovanie na základe údajov EU SILC

3. Záver

Veľmi vysoké ako aj veľmi nízke príjmy predstavujú určitú kontamináciu údajov. Nájsť vhodnú metódu, ako na extrémne hodnoty v analýze prihliadať je veľmi zložité. Napriek tomu, je nutné kontrolovať vplyv týchto hodnôt s cieľom zlepšiť porovnateľnosť výsledkov z rôznych krajín. Metód na úpravu údajov existuje niekoľko, my sme v príspevku popísali dve základné metódy nazývané „trimming“ a „winsorizing“. Tieto metódy nie sú v podmienkach SR veľmi známe o čom svedčí aj fakt, že zatiaľ nemajú slovenský ekvivalent.

V príspevku sme sa venovali analýze indikátorov nerovnosti v prípade prítomnosti extrémnych príjmov. Táto situácia je pri analýze príjmov domácnosti častá, keďže príjmové rozdelenie je pravostranne zošikmené s extrémnymi hodnotami najmä na pravom konci rozdelenia. Ako zdroj údajov sme použili databázu EU SILC. Ukázali sme, že najmä indexy príjmovej nerovnosti, Giniho, Atkinsonov a GE index sú citlivé na akúkoľvek zmenu, či úpravu údajov. Najmenej ovplyvnený úpravou údajov je pritom Giniho index a naopak najviac ovplynený je GE (Generalized Entropy) index, čo potvrdzuje aj zistenia iných autorov. Možným riešením tohto problému by bolo napríklad samostatné modelovanie chvostov rozdelenia. Ďalším námetom je aj štúdium robustných metód a odhad parametrov rozdelenia pomocou tzv. OBRE (optional B-robust estimator) algoritmu, ktorý je popísaný v prácach autorov Victoria- Feser, 1996 a Victoria-Feser, 2000.

Príspevok bol vytvorený s podporou vedeckovýskumných projektov VEGA1/0127/11 Priestorová distribúcia chudoby v Európskej únii a VEGA1/0344/14 Matematické a štatistické metódy v ekonomickom rozhodovaní.

Literatúra

COWELL F.A. 2000. Measurement of Inequality. In Atkinson A.B., Bourguignon F. (Eds.) Handbook of Income Distribution . Amsterdam, Elsevier, Vol. 1, pp. 87-166.

R DEVELOPMENT CORE TEAM. 2012. R: A language and environment for statistical computing. Viedeň: R Foundation for Statistical Computing. ISBN 3-900051-07-0. URL http://www.R-project.org/.

De MAIO, F.G. 2007. Income Ineaquality Measures, J Epidemiol Community Health. 2007 October; 61(10): 849–852.

EUROSTAT. 2006, 'Some proposals on the treatment of negative incomes', EU-SILC Documents TFMC-15/06, European Commission, Eurostat.

EUROSTAT. 2007, An examination of outliers at upper end of income distribution. Report N. ISRI.12, Project EU-SILC (Community statistics on income and living conditions) 2005/S 116-114302 – Lot 1 (Methodological studies to estimate the impact on comparability of the national methods used).

Van KERN, P. 2006. Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC.

VICTORIA-FESER, M.P. AND ALAIZ M.P. 1996. Modelling Income Distribution in Spain: A Robust Parametric Approach. *DARP Discussion Paper 20*, London School of Economics

VICTORIA-FESER, M.P. 2000. Robust methods for the analysis of income distribution, inequality and poverty. *International Statistical Review 68* (3), 277-293

STANKOVIČOVÁ, I. 2009. Analýza monetárnej chudoby v domácnostiach Českej republiky, In: *Forum Statisticum Slovacum*, č.7, 2009, s. 151-156

ŠÚ SR. (2011). Zisťovanie o príjmoch a životných podmienkach EU SILC 2010 (UDB_31/08/11). [databáza s mikroúdajmi]. Bratislava: Štatistický úrad SR.

Adresa autorov:

RNDr. Ivan Mojsej, PhD.
Ústav matematických vied
Jesenná 5, 040 01 Košice
ivan.mojsej@upjs.sk

Mgr. Alena Tartaľová, PhD.
Katedra aplikovanej matematiky
a hospodárskej informatiky
Nemcovej 32, 040 01 Košice
alena.tartalova@tuke.sk

Aplikace modifikované metody CPM v rámci logistického řetězce
Application of modified CPM within the logistics chain

Lubor Možný, Vojtěch Ondryhal, Marek Sedlačík

Abstract: The contribution is focused on the functionality analysis of the logistics chains. The paper describes a possible modification of the Critical Path Method through portfolios of safety factors in the vertices and edges. The algorithm is implemented in Python environment and introduced in the final part of the paper.

Abstrakt: Příspěvek je zaměřen na analýzu a optimalizaci logistického řetězce. Je navržena modifikace metody hledání kritické cesty rozšířená o portfolio bezpečnostních kritérií hran i uzlů uvažovaného řetězce. Vytvořený algoritmus je zpracován v programovacím jazyce Python a je demonstrován na simulovaných datech.

Key words: Critical Path Method, logistics chain, security portfolio.

Klíčová slova: Metoda kritické cesty, logistický řetězec, bezpečnostní portfolio.

JEL classification: C44

1. Úvod

Příspěvek je zaměřen na problematiku logistických řetězců a zde často využívanou metodu hledání kritické cesty (CPM z anglického Critical Path Method). Úvod je věnován charakteristice pojmů, další část naznačuje uplatnění modifikované CPM metody. Navržený algoritmus byl implementován v programovém jazyce Python.

Metoda kritických cest je běžně používána pro procesy a projekty. Pro uplatnění metody je nutné splnit několik podmínek. Jednou z těchto podmínek je jasně definovaný počáteční a závěrečný bod. Model pro CPM musí dále obsahovat místa, kde se procesy setkávají a vytváří omezení pro další navazující činnosti. Hlavním úkolem je najít optimální cestu plnění procesů s minimální (kritickou) délkou trvání celého projektu. Obdobnou činnost a procesy lze identifikovat i v logistickém řetězci.

Díky paralele logistického řetězce a modelů pro použití CPM lze modifikovanou metodu použít pro optimalizaci pohybu v logistickém řetězci. Přístavy, letiště a nádraží mohou představovat v rámci modifikované metody CPM uzly logistického řetězce. Hranou mohou být spojnice uzlů (silnice, námořní cesty,…). Navržená modifikace běžně užívané metody spočívá v zapojení bezpečnostního portfolia (Foltin, Sedlačík, Ondryhal, 2013), které obsahuje všechna významná (případně stanovená) rizika pro daný uzel nebo hranu. Výstupem je analýza definující optimální kritickou cestu vzhledem ke všem stanoveným bezpečnostním faktorům, které mohou ovlivňovat hrany a uzly (NATO, 2008), (Zsidsin, 2008).

Bezpečnostní portfolio je tedy souborem rizik, která jsou identifikována pro uzly a hrany řetězce vycházející z hrozeb v rámci daného bezpečnostního prostředí. Jasné definování bezpečnostního portfolia umožňuje systému poskytnout optimálnější výstup pro uživatele logistického řetězce. Pouhá definice bezpečnostních rizik není dostatečným vstupem pro analýzu, z tohoto důvodu jsou jednotlivým rizikům přiřazena váhová ohodnocení jejich významu.

Dříve než přistoupíme k analýze bezpečnostního prostředí, zavedeme v souladu s (Jablonský, 2001) následující značení.

2. Bezpečnostní portfolio a kritéria logistického řetězce

Předpoklade grafu je uspořádaná dvojice $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ množina hran a uzlů. Přesněji řečeno uvažujeme konečný, ohodnocený a orientovaný multigraf (Jablonský, 2001). Dále budeme používat označení:

$$\boldsymbol{V} = \{V_1, \dots, V_n\} = V_i; i = 1, \dots, n\},$$

kde *n* je počet uzlů $\boldsymbol{G}$, V_1 je vstupní uzel a V_n je výstupní uzel logistického řetězce. Obdobně

$$\boldsymbol{E} = \{E(r, s_p)_{;}\, r, s \in \{1, \dots, n\},\ p = 1,2, \dots\},$$

kde $E(r, s_1)$,…,$E(r, s_p)$ jsou všechny exitující hrany mezi uzly V_r a V_s.

Jak bylo uvedeno, pro základní metodu CPM není nezbytně nutná analýza potencionálních rizik pro logistický řetězec. Pro další činnost algoritmu (systému) je metoda CPM obohacena právě o portfolio bezpečnostních rizik, jež je základem pro vytvoření nové kritické cesty zohledňující nepředvídané narušení logistického řetězce. Dle (Zsidsin, 2008) lze bezpečnostní kritéria pro daný uzel identifikovat následovně:

- kritérium času (X_1)
- kritérium prostoru (X_2)
- kritérium nákladů (X_3)
- kritérium informační (X_4)
- kritérium flexibility a pružnosti (X_5)

Bezpečnostní kritéria X_1,…, X_5 popisující bezpečnostní situaci v uzlech $V_1, \dots, V_n$ lze souhrnně vyjádřit pomocí *bezpečnostní matice uzlů* **X**:

$$\mathbf{X}_{5\times n} = (x_{ij})_{i=1,\dots,5\ j=1,\dots,n},$$

kde *i*-tý sloupec matice popisuje bezpečnostní kritérium uzlu V_i, $i = 1, \dots, n$. Individuální portfolio bezpečnostních kritérií pro uzel vyjadřuje relativní bezpečnost uvedeného prvku řetězce. Všechna bezpečnostní kritéria uzlů jsou považována za minimalizační a mohou nabývat hodnot z intervalu $\langle 0,1\rangle$. Hodnota vyjadřuje, jaký vliv má dané kritérium na celkovou bezpečnostní funkci uzlu. Při mezní hodnotě 0 je vliv nulový, opačně hodnota 1 charakterizuje úplný (zásadní) vliv kritéria na celkovou bezpečnostní funkci uzlu. Konkrétněji $X_i \in \langle 0,1\rangle$, $i = 1, \dots,5$.

Podobně jako u bezpečnostních kritérií pro uzly lze identifikovat bezpečnostní kritéria pro funkci hran viz (Zsidsin, 2008):

- kritérium času (Y_1)
- kritérium komunikace (Y_2)
- nákladové kritérium (Y_3)
- informační kritérium (Y_4)
- kritérium flexibility a pružnosti (Y_5)
- kritérium množství (Y_6)
- kritérium kvality (Y_7)

Bezpečnostní situaci hran $\boldsymbol{E} = \{E(r, s_{p_{;}}\, r, s \in \{1, \dots, n\}, p = 1,2, \dots\}$ lze obdobně vyjádřit pomocí popsaných kritérií $Y_1, \dots, Y_7$. *Bezpečnostní matice hran* **Y** je potom:

$$\mathbf{Y}_{7\times l} = (y_{ij})_{i=1,\dots,7\ j=1,\dots,l},$$

kde $l = |\boldsymbol{E}|$ a jednotlivé sloupce matice popisují bezpečnostní situaci všech hran z $\boldsymbol{E}$. Identifikovaná bezpečnostní kritéria hran logistického řetězce vyjádříme stejným způsobem jako u uzlů, tj. $Y_i \in \langle 0,1\rangle, i = 1, \dots, l$.

Konkrétní bezpečnostní kritéria uzlů a hran nemají totožný význam pro logistický řetězec respektive pro algoritmus. Z tohoto důvodu jsou již zavedená kritéria doplněna o váhy

popisující jejich bezpečnostní význam. Pro všechny uzly $V_1, \dots, V_n$ je *váha bezpečnostního kritéria* X_k stejná a je označena jako α_k, kde $\alpha_k \in \langle 0,1 \rangle$ a $k = 1, \dots, 5$. Co se týká hran, je *váha bezpečnostního kritéria* Y_k označena β_k, kde $\beta_k \in \langle 0,1 \rangle$ a $k = 1, \dots, 7$. Jinými slovy uvedené parametry definují významnost bezpečnostních kritérií uzlů a hran. Uvedené parametry musí být stanoveny před zahájením samotného algoritmu.

3. Navržený algoritmus

Při hledání optimální kritické cesty v daném čase t_j pro $j = 1, \dots, m$ a při současném hodnocení portfolia bezpečnostních kritérií pro jednotlivé uzly a hrany je nezbytné nalezení optimálního scénáře $\boldsymbol{G}_{t_j}^{opt}$ z množiny všech potencionálních scénářů U_{t_j}. V daném čase t_j jsou všechny uzly $V_1, \dots, V_n$ a všechny hrany $E_1, \dots, E_l$ z pohledu bezpečnosti ohodnoceny výše popsaným způsobem. Hledání optimální posloupnosti uzlů a hran v čase t_j, které reprezentuje optimální kritickou cestu od vstupního uzlu po výstupní uzel pro stanovené váhy bezpečnostních kritérií uzlů $\alpha_1, \dots, \alpha_5$ a pro stanovené váhy bezpečnostních kritérií pro hrany $\beta_1, \dots, \beta_7$ lze zapsat následovně:

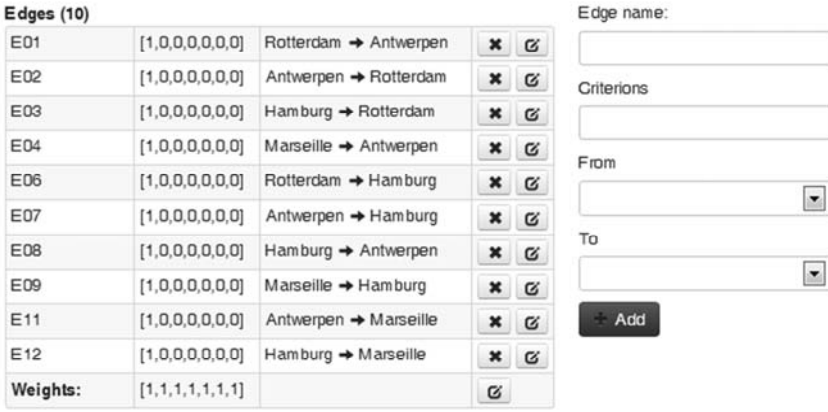
1. Vytvořením grafu $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ Jednotlivé hrany a jejich čísla musí být uvedena s ohledem na definované uzly. Z tohoto identifikujeme matici:
$$\mathbf{A}_{n \times n} = (a_{ij})_{i,j=1,\dots,n},$$
kde a_{ij} představuje počet hran mezi uzly V_i *a* V_j. Je evidentní, že matice **A** je symetrická.
2. Nalezení bezpečnostních kritérií $X_1, \dots, X_5$ pro všechny uzly V_i, kde $i = 1, \dots, n$:
$$\mathbf{X}_{5 \times n} = (x_{ij})_{i=1,\dots,5\ j=1,\dots,n}.$$
3. Nalezení bezpečnostních kritérií $Y_1, \dots, Y_7$ pro všechny hrany E_i, kde $i = 1, \dots, l$:
$$\mathbf{Y}_{7 \times l} = (y_{ij})_{i=1,\dots,7\ j=1,\dots,l}.$$
4. Pro $i = 1, \dots, n$ vyjádření bezpečnostního portfolia $\mathrm{V}_{t_j}^{V_i}$ pro všechny uzly V_i logistického řetězce:
$$\mathrm{V}_{t_j}^{V_i} = \sum_{i=1}^{5} \alpha_k x_{ki}$$
5. Pro $i = 1, \dots, l$ vyjádření bezpečnostního portfolia $\mathrm{E}_{t_j}^{E_i}$ pro všechny hrany E_i logistického řetězce:
$$\mathrm{E}_{t_j}^{E_i} = \sum_{i=1}^{7} \alpha_k y_{ki}$$
6. Definování a nalezení všech potencionálních scénářů $U_{t_j} = \{\boldsymbol{G}_s;\ s = 1, \dots\}$, kde $\boldsymbol{G}_s = (\boldsymbol{V}_s, \boldsymbol{E}_s)$ jsou podgrafy $\boldsymbol{G}$ mající stejný vstupní a výstupní uzel. Následně bezpečnostní ohodnocení $S_{t_j}^s$ scénáře G_s je:
$$S_{t_j}^s = \sum_{V_i \in \boldsymbol{V_s}} \mathrm{V}_{t_j}^{V_i} + \sum_{E_k \in \boldsymbol{E_s}} \mathrm{E}_{t_j}^{E_k}$$
7. Optimální scénář $\boldsymbol{G}_{t_j}^{opt}$ ze všech přípustných scénářů U_{t_j} nalezneme jako scénář s ohodnocením:
$$S_{t_j}^{opt} = \min_{S_{t_j}^s \in U_{t_j}} S_{t_j}^s$$
Jestliže je minima dosaženo pro více scénářů, je zvolen jeden z nich jako optimální.

4. Ilustrační příklad

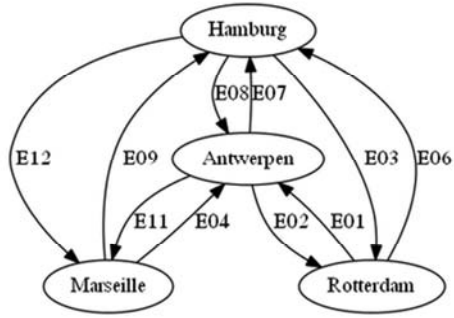
Na základě výše uvedeného algoritmu byl vytvořen program, který simuluje hledání optimální kritické cesty v daném čase. Algoritmus byl zpracován v programovacím jazyce Python (Harms, Mcdonald, 2003). Vstupními daty jsou uzly, hrany a hodnoty bezpečnostních kritérií uzlů a hran. Výstupem je pak optimální kritická cesta.

Následuje ukázka editoru uzlů, hran (obrázek 1 a 2) a výstup hledání optimální cesty a automaticky generované schéma (obrázek 3) na základě definovaných uzlů a hran. Jako příklad byly zvoleny čtyři evropské přístavy, hodnoty kritérií jsou nastaveny fiktivně.

Obr.1: Seznam uzlů a jejich editace



Obr.2: Seznam hran a možnost editace/přidání hrany



Obr.3: Automaticky generovaný graf z uzlů a hran

Program dovolí vybrat startovací a cílový uzel a nalezne všechny možné cesty. Cesty jsou ohodnoceny a seřazeny podle hodnoty skóre. Optimální cesta je ta s nejnižším skóre viz obrázek 4.

Obr.4: Nalezené cesty s uvedením skóre

5. Závěr

Příspěvek naznačuje, jakým způsobem lze navržený algoritmus a vytvořenou aplikaci uplatnit v rámci logistického řetězce. Pomocí nastíněné modifikace metody CPM lze zvyšovat zajištění logistických kanálů v měnícím se bezpečnostním prostředí. Lze předpokládat, že z důvodu stále rostoucího významu logistiky může daná aplikace najít významné uplatnění v civilním i vojenském prostředí.

Literatúra

FOLTIN, P., SEDLAČÍK, M., ONDRYHAL, V. *Bezpečnostní aspekty logistických řetězců. In: Manažment, teória, výučba a prax 2013: Zborník z príspevkov z medzinárodnej vedecko-odbornej konferencie.* Liptovský Mikuláš: Akadémia ozbrojených síl, 2013, s. 88-96. ISBN 978-80-8040-477-2.

HARMS, D., MCDONALD, K. *Začínáme programovat v jazyce Python*. Praha: Computer Press, 2003. ISBN ISBN: 80-722-6799-X.

JABLONSKÝ, J. 2001. *Operační výzkum*. Praha: VŠE, 2001. ISBN 80-2450162-7.

NATO Research and Technology Organisation. 2008. *Improving Common Security Risk Analysis.* Neuilly-sur-Seine: NATO, 2008. str. 3-23. ISBN 978-92-837-0045-6.

ZSIDSIN, G. A. 2008. *Supply chain risk: a handbook of assessment, management, and performance*. New York: Springer, 2008. ISBN 03-877-9934-6.

Bc. Lubor Možný (2.ročník-MN)
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
lubor.mozny@unob.cz

Ing. Vojtěch ONDRYHAL, Ph.D.
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
vojtech.ondryhal@unob.cz

RNDr. Marek SEDLAČÍK, Ph.D.
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
marek.sedlacik@unob.cz

Demografický vývoj ako významný determinant cien obytných nehnuteľností v SR
Demographics as an important determinant of house prices in SR

Miroslav Pánik

Abstract: The aim of the paper is an analysis of residential real estate prices using quantitative methods, construction of economic model which describes the relationship between house prices and demographics factors. Work objectives have been achieved using mathematical and statistical methods. The regression model that explains the formation of housing prices was formed using the regression analysis and the model is based on one key determinants: population aged from 25 to 44 years.

Abstrakt: Cieľom príspevku je analýza cien nehnuteľností určených na bývanie pomocou kvantitatívnych metód, konštrukcia ekonomického modelu popisujúceho vzťah medzi cenami nehnuteľností a demografickými faktormi. Cieľ práce bol dosiahnutý použitím matematicko-štatistických metód. Pomocou korelačnej a regresnej analýzy bol zostavený regresný model, ktorý vysvetľuje tvorbu cien bývania na základe determinantu obyvateľstvo vo veku od 25-44.

Key words: house prices, demographics factors, regression and correlation analysis

Kľúčové slová: ceny obytných nehnuteľností, demografické faktory, regresná a korelačná analýza

JEL classification: C10

1. Úvod

Vývoj cien nehnuteľností má značný vplyv na hospodárstvo ako celok, pričom hospodársky vývoj a vývoj na realitnom trhu sú vzájomne prepojené. Rozvoj realitného trhu je do veľkej miery závislý na stave ekonomiky v rámci hospodárskeho cyklu. Extrémne tlaky v ekonomike môžu viesť k vzniku krízy na trhu nehnuteľností. Výrazný pokles cien nehnuteľností môže do značnej miery destabilizovať bankový systém a spôsobiť tak rozsiahle ekonomické problémy.

Ceny nehnuteľností ovplyvňuje množstvo dopytových a ponukových faktorov pôsobiacich na realitnom trhu. Medzi kľúčové faktory patrí disponibilný dôchodok, hrubý domáci produkt, objem úverov na bývanie, objem stavebnej produkcie bytových budov, rast počtu obyvateľov a počtu domácností (Cár, 2009).

Špirková (2009) uvádza, že najdôležitejšie determinanty cien sú ekonomický rast, príjem domácností, úrokové miery a dostupnosť úverov, demografické faktory, dane, dotácie a subvencie štátu, výstavba nových bytov a špekulácie súvisiace s očakávaním rastu cien.

Hilbers a kol. (2008) uvádza, že náklady kupujúceho pri kúpe bývania významne ovplyvňujú príjmy resp. výška nájomného v prípade prenájmu nehnuteľnosti, dane a subvencie zo strany štátu. Dôležitým faktorom ktorý ovplyvňuje dopyt po nehnuteľnostiach je demografický vývoj, konkrétne populačný rast a vývoj počtu a veľkosť domácností, ktoré vplývajú na dopyt po nehnuteľnostiach. Ponuku nehnuteľností na bývanie ovplyvňuje dostupnosť a cena stavebného pozemku, stavebné náklady, legislatíva. Vo všeobecnosti sa ponuka prispôsobuje dopytu s oneskorením, ktoré spôsobuje získavanie stavebných povolení, návrh a realizácia výstavby.

Cár (2009, s. 2-8) zaradil demografické faktory medzi veľmi významné determinanty, ovplyvňujúce dopyt po nehnuteľnostiach. Medzi tieto faktory patria najmä:

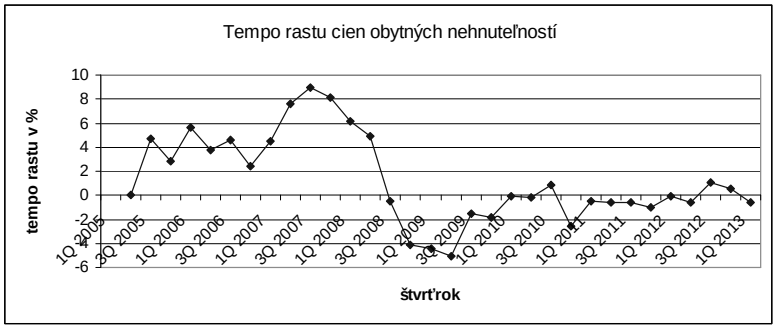
- počet obyvateľstva

- počet hospodáriacich domácností
- pôrodnosť
- úmrtnosť
- sobášnosť a rozvodovosť
- potreba bývania

a. Vývoj cien obytných nehnuteľností v SR

Ceny nehnuteľností na Slovensku mierne rástli od 1. štvrťroku 2005 do 3. štvrťroku 2007. Tento nárast sa vo všeobecnosti pripisuje kladným očakávaniam najmä zahraničných investorov, ktoré vyplynuli zo vstupu Slovenska do Európskej únie v roku 2004. Na konci roka 2007 však bolo možné pozorovať extrémny nárast cien, ktorý vyvrcholil v polovici roku 2008. Táto fáza sa označuje ako realitná bublina, ktorá je spojená s rôznymi deformáciami realitného trhu, ako napr. rovnaká a mnoho krát aj vyššia cena starších bytov ako novostavieb.

Koncom roku 2008 nastal zlom vo vývoji cien. V dôsledku globálnej hypotekárnej krízy a celkovej ekonomickej recesie došlo k prasknutiu realitnej bubliny a nasledoval prudký pokles cien nehnuteľností, ktorý sa spomalil až na konci roku 2009. V roku 2010 ceny stagnovali na úrovni z roku 2007. Od roku 2010 až doteraz je trend vývoja cien približne konštantný až mierne klesajúci. Prudký pokles tempa rastu cien nehnuteľností určených na bývanie je zobrazený na obrázku 1.



Obr. 25: Tempo rastu cien obytných nehnuteľností (zdroj: NBS)

Mnoho realitných odborníkov obdobie od roku 2010 nazýva ako obdobie kryštalizácie realitného trhu. Zmenilo sa správanie kupujúcich na strane dopytu, ale aj správanie realitných inštitúcií na strane ponuky. Klesala celková ponuka starých bytov, pretože obyvatelia odložili kúpu novej nehnuteľnosti na neskôr a radšej zotrvali vo svojom staršom byte. Na strane ponuky sa taktiež zmenilo správanie developerov, ktorí v období realitnej expanzie stavali veľkometrážne, často krát dispozične ťažko využiteľné byty. Odborníci sa zhodujú, že v roku 2010 bolo v SR približne 6000 nových voľných bytov, z ktorých 4500 bolo v Bratislave. Ponuka bytov v tomto období teda jednoznačne prevyšovala dopyt.

b. Demografický vývoj

Ceny obytných nehnuteľností významne ovplyvňuje aj demografický vývoj – celkový počet obyvateľov, celkový prírastok, pôrodnosť, úmrtnosť, migrácia a iné. Celkový počet obyvateľov v SR síce rástol, avšak prirodzený prírastok dlhodobo klesá a v súčasnosti osciluje okolo 1%.

Dôležitým demografickým faktorom je aj veľkosť a štruktúra cenzových domácností. Špirková (2009) uvádza, že na základe vypracovanej projekcie demografického vývoja do roku 2015 je zrejmé, že populačný vývoj charakterizujú výrazne sa znižujúce prírastky

obyvateľstva, ktoré budú mať za následok úbytok počtu obyvateľov v SR. Postupne sa znižuje percento spolunažívania cenzových domácností, čo ovplyvní nároky na počet bytov. V roku 1991 bola priemerná veľkosť cenzovej domácnosti 2,89 osôb, v roku 2015 sa predpokladá 2,71 osôb. V roku 1991 pripadalo 112,96 cenzových domácností na 100 bytov a v roku 2015 sa predpokladá 105,3 cenzových domácností na 100 bytov, uvádza Špirková (2009).

Podľa aktuálneho štatistického zisťovania ŠÚSR, v súčasnosti viac ako 56 % mladých ľudí vo veku 25-34 rokov žije v spoločnej domácnosti s rodičmi. V rámci zisťovania Eurostatu sa Slovensko umiestnilo na poslednom mieste spomedzi všetkých krajín EU. Zotrvávanie mladých dospelých v spoločnom byte so svojimi rodičmi má podľa mnohých odborníkov hlavne ekonomické dôvody. Významný vplyv však zohráva aj stagnujúca štátna bytová politika. Vysoké ceny nehnuteľností v pomere s relatívne nízkou mzdou obyvateľstva zapríčiňujú neschopnosť osamostatniť sa. Tento trend vedie k populačnému úpadku, starnutiu populácie a následnému brzdeniu ekonomického rastu.

2. Materiál a metódy

Cieľom príspevku bolo preukázať vzájomný vplyv demografických faktorov na ceny obytných nehnuteľností a zostrojiť lineárny regresný model. Na základe predchádzajúcich štúdií sme ako kľúčový demografický faktor zvolili počet obyvateľov vo veku 25-44 rokov. Práve v tomto veku je najväčší predpoklad, že si ľudia založia rodinu, resp. zaobstarávajú vlastné bývanie. Závislou premenou bola priemerná cena obytných nehnuteľností (€/m2).

Údajová základňa bola čerpaná z Národnej banky Slovenska (NBS) a zo Štatistického úradu SR (ŠÚSR). Dáta tvoria časové rady od 1. štvrťroka 2004 po 4. štvrťrok 2010. Výraznou komplikáciou pri analýze cien nehnuteľností je krátkosť dostupných časových radov cien nehnuteľností. Výpočty boli realizované pomocou programov MS Excel a Eviews.

Pri kvantitatívnej analýze bola na posúdenie vzťahu medzi cenami obytných nehnuteľností a demografickým faktorom použitá regresná a korelačná analýza.

2.1 Korelačná analýza

Korelačná analýza uplatňuje štatistické metódy a postupy na posúdenie intenzity štatistickej závislosti medzi kvantitatívnymi premennými (Pacáková, 2009)

Jednou z najznámejších korelačných charakteristík je Pearsonov koeficient korelácie ρ_{xy}, ktorý meria obojstrannú lineárnu závislosť dvoch premenných, *x* a *y*. Koeficient korelácie nadobúda hodnoty z intervalu <-1, 1>, pričom znamienko určuje smer závislosti:

- $\rho_{xy} = 0$ a ak premenné *x* a *y* majú dvojrozmerné normálne rozdelenie potom, premenné *x* a *y* nie sú lineárne závislé
- $\rho_{xy} > 0$, potom medzi premennými *x* a *y* je priamy lineárny vzťah
- $\rho_{xy} < 0$, potom medzi premennými *x* a *y* je nepriamy lineárny vzťah

Ak sa koeficient korelácia rovná 1 hovoríme o úplnej priamej lineárnej závislosti. Ak sa koeficient korelácia rovná -1 hovoríme o úplnej nepriamej lineárnej závislosti a premenné *x* a *y* sú vo vzťahu funkčnej závislosti (Pacáková, 2009).

Hodnotenie intenzity závislosti (v absolútnej hodnote) pri dostatočne veľkom rozsahu súboru je nasledovné:

- Hodnota ρ_{xy} od 0,8 do 1: veľmi silná závislosť
- Hodnota ρ_{xy} od 0,4 do 0,8: stredne silná závislosť
- Hodnota ρ_{xy} od 0,1 do 0,4: slabá závislosť

2.2 Regresná analýza

Regresný model je matematický predpis, ktorý zjednodušene charakterizuje vzťahy medzi premennými (XinZan, Xiaogang Su, 2009).

Premenné v modeli potom môžu byť

- vysvetľovaná (závislá) y – premenná, ktorej závislosť od iných skúmame.
- vysvetľujúca (nezávislá) $x_1, x_2 \ldots x_k$, - premenné, ktoré vyvolávajú zmeny závislej premennej

Vzťah medzi premennými môže byť

- jednostranný – ak zmeny vysvetľujúcej premennej spôsobujú zmeny vysvetľovanej premennej, ale zmeny vysvetľovanej premennej nevyvolávajú zmeny vysvetľujúcej premennej
- obojstranný – keď sa vysvetľujúca a vysvetľovaná premenná navzájom ovplyvňujú

Jednoduchý lineárny regresný model môžeme zapísať ako

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon, \; i = 1, 2 \ldots n \qquad (1)$$

y_i je i-tá pozorovaná hodnota vysvetľovanej premennej

β_0, β_1 sú neznáme parametre regresného modelu

x_i je i-tá hodnota vysvetľujúcej premennej

ε_i je náhodná chyba i-teho pozorovania

n je počet pozorovaní

Na odhad neznámych parametrov β_0, β_1 sa používa metóda najmenších štvorcov (MNŠ), ktorú bližšie popisuje Pacáková (2009, s.185-188).

Pri analýze časových radov musí byť časový rad reziduí lineárnej regresie stacionárny, aby odhadnutý vzťah prezentoval rovnováhu z dlhodobého hľadiska. V opačnom prípade by mohli byť výsledkom tzv. falošné regresie. Stacionaritu reziduí sme testovali pomocou rozšíreného Dickey-Fuller testu (ADF test). Nulová hypotéza H_0 je jednotkový koreň, ktorý predstavuje nestacionaritu. Premenné v modeli boli stacionárne po prvej diferencii (d=1). Viac o stacionarite časových radov a falošnej regresii píše Hamilton (1994).

3. Výsledky

Korelačná analýza potvrdila hypotézu, že vývoj počtu obyvateľstva vo veku 25-44 rokov významne vplýva na vývoj cien obytných nehnuteľností. Korelačný koeficient medzi týmito premennými je:

$$\rho_{xy} = 0{,}86$$

Medzi skúmanými premennými teda existuje silná štatistická závislosť. Výsledky ADF testu, ktorý testuje stacionaritu rezíduí, sú prezentované v tabuľke 1.

Tab. 7: Výsledky ADF testu(zdroj: autor)

Premenná	P hodnota ADF testu
Obyvateľstvo od 25 - 44 rokov	0,0048
Cena obytných nehnuteľností	0,0042

Podľa ADF testu sú reziduá časových radov obyvateľstvo od 25 - 44 rokov a cenami nehnuteľností určených na bývanie stacionárne. Teda lineárna závislosť medzi premennými nie je falošná.

Po korelačnej analýze bol odhadnutý v súlade s teóriou v kapitole 2.2 jednoduchý lineárny regresný model, ktorý je uvedený na obrázku 2.

Dependent Variable: CENY
Method: Least Squares
Date: 11/05/13 Time: 17:10
Sample: 2004Q1 2010Q4
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11558.19	1513.499	-7.636738	0.0000
OB2544	7649.298	909.4302	8.411089	0.0000

R-squared	0.731256	Mean dependent var	1170.497
Adjusted R-squared	0.720920	S.D. dependent var	231.3725
S.E. of regression	122.2295	Akaike info criterion	12.51843
Sum squared resid	388441.5	Schwarz criterion	12.61359
Log likelihood	-173.2580	F-statistic	70.74641
Durbin-Watson stat	0.221850	Prob(F-statistic)	0.000000

Obr. 2: Odhad regresného modelu

Regresný model má tvar:

$$\boldsymbol{CENY} = -11558.19 + 7649.298\boldsymbol{OB2544}$$

Analýza výstupu z programu Eviews na obrázku 2. ukazuje, že *p* hodnota determinantu obyvateľstvo od 25 do 44 rokov je štatisticky významná, $0{,}00 \leq 0{,}05$, čo potvrdzuje opodstatnenosť faktora zahrnutého do modelu. Koeficient determinácie R^2 je 0,73, teda model vysvetľuje realitu takmer na 73%.

4. Záver

Pomocou korelačnej analýzy bola potvrdená hypotéza, ktorá bola stanovená v úvode príspevku. Zistili sme významnú štatistickú závislosť medzi cenami obytných nehnuteľností a počtom obyvateľov vo veku 25-44 rokov. Môžeme konštatovať, že nárast obyvateľstva vo veku 25 až 44 rokov spôsobuje rast cien, pretože dopyt sa zo zvyšujúcou kúpyschopnou populáciou bude zvyšovať. Pomocou ADF testu sme vylúčili falošné regresie a pomocou regresnej analýzy zostavili jednoduchý lineárny regresný model pozostávajúci z dvoch premenných.

V ďalšom výskume by bolo možné skúmaný populačný interval rozdeliť na menšie časti a taktiež zahrnúť do modelu aj iné socio-ekonomické premenné.

Poďakovanie

Článok bol financovaný pomocou programu na podporu mladých výskumníkov, názov projektu: Ceny nehnuteľností ako determinujúci faktor demografického potenciálu a ekonomickej aktivity obyvateľstva

Literatúra

CÁR, M. 2009 b. Aktuálny a očakavaný vývoj cien nehnuteľností na bývanie na Slovensku. In *Biatec*. ISSN 1335-0900, 2009, roč 17, č.11, s. 2-6.

CÁR, M. 2009. Výber faktorov ovplyvňujucich ceny nehnuteľností nabývanie na Slovensku. In *Biatec*. ISSN 1335-0900, 2009, roč. 17, č.3, s. 2-8.

HAMILTON, J. D. 1994. *Time series analysis*. Priceton University Press, 1994, s. 799. ISBN 9780691042893

HILBERS a kol. 2008. *House Price Developments in Europe*. Working papers WP/08/211. 2008

PACÁKOVÁ, V. 2009. *Štatistické metódy pre ekonómov*. Bratislava: IURA Edition, 2009, s. 411, ISBN 978-80-8078-284-9

ŠPIRKOVÁ, D. - IVANIČKA, K. - FINKA, M. 2009. *Bývanie a bytová politika – vývoj, determinanty rozvoja bývania a nové prístupy v nájomnej bytovej politike na Slovensku*. Bratislava: Vydavateľstvo STU v Bratislave, 2009, s. 191, ISBN 978-80-227-3173-7.

XIN, Z. – XIAOGANG, S. 2009. *Linear regression analysis: theory and computing*. World scienfic, 2009, s. 328, ISBN 9789812834102

Adresa autora:

Miroslav Pánik, Ing., PhD.
Ústav manažmentu - STU
Vazovova 5, 812 43 Bratislava
miroslav_panik@stuba.sk

Neurónová sieť založená na Gumbelovej distribučnej funkcií s aplikáciou na vysoko rozmerný dátový súbor
Neural network based on Gumbel distribution function applied to high dimensional dataset

Lukáš Pastorek

Abstract: The aim of this paper is an application and brief description of a neural network that uses Gumbel distribution function as a tool for distribution of the weights in a vector space. This network is compared with Kohonen's self-organizing maps on an artificial dataset with the high dimension. The neural network possess lower mean square errors than Kohonen's maps throuhout neural learning.

Abstrakt: Cieľom tohto článku je programová aplikácia a stručné predstavenie neurónovej siete, ktorá používa Gumbelovu distribučnú funkciu ako nástroj na distribúciu váh vo vektorovom priestore. Táto sieť je porovnávaná s Kohonenovými samoorganizujúcimi sa mapami na umelom dátovom súbore s vysokým rozmerom. Neurónová sieť dosahuje nižšej chybovosti ako Kohonenova mapa.

Key words: Gumbel cummulative distribution function, Kohonen's self-organizing map, Gompertz curve, high dimensional space

Kľúčové slová: Gumbelova kumulatívna distribučná funkcia, Kohonenova samoorganizujúca sa mapa, Gompertzova krivka, vysoko rozmerný priestor

JEL classification: C45

1. Úvod

Umelé neurónové siete s učením bez učiteľa sa používajú ako analytický nástroj pri riešení aproximačných úloh vo viacrozmernom priestore. Ich podstatnou výhodou oproti klasickým optimalizačným nástrojom nelineárnej regresie, je ich nezávislosť na teoretických distribučných predpokladoch. Cieľom týchto metód je optimálne rozmiestnenie predvoleného počtu modelových vektorov v priestore vektorov dátového súboru. Hľadanie optimálnej pozície sa uskutočňuje na základe heuristického, iteratívneho učenia sa modelu a presúvania modelových vektorov v rámci viacrozmerného priestoru. Toto učenie môžeme označovať za mapovania priestoru vstupných vektorov prostredníctvom modelových vektorov. Po utlmení učenia a zastavení pohybu modelových vektorov sa priestor vstupných vektorov roztriešti na rovnaký počet regiónov, ako je počet modelových vektorov. Každý región je reprezentovaný jedným modelovým vektorom, ktorý tento priestor zároveň definuje. Ku každému modelovému vektoru pripadá vektorový priestor, ktorý je bližšie k danému modelovému vektoru ako ku ktorémukoľvek inému modelovému vektoru. Tieto regióny sa označujú ako tzv. Voronoive regióny (Řezanková, 2009). Regióny vstupných vektorov môžeme volať zhluky a proces učenia a následného delenia priestoru označovať za zhlukovanie.

Umelé neurónové siete s učením bez učiteľa (na rozdiel od metód s učiteľom), nemajú množinu klasifikovaných vektorov, ktoré by prostredníctvom penalizačnej funkcie korigovali priebeh učenia. Správnosť rozdelenia vstupných vektorov do prirodzených zhlukov zostáva väčšinou neznáme a je na expertnom úsudku analytika, či je dané rozdelenie vstupných vektorov vecne a logicky správne a či nedošlo k umelému rozdeleniu niektorých prirodzených zhlukov. Z dôvodu neexistencie jednoznačne správneho riešenia alebo univerzálneho ukazovateľa kvality zhlukovania, je vhodné, aby sme pri aplikácií na reálne dátové súbory porovnávali výsledky viacerých aproximačných metód medzi sebou.

V tomto článku prezentujeme stručný popis neurónovej siete, ktorá je odvodená od princípov Kohonenových samoorganizujúcich sa máp a je založená na ne-gausovskej distribúcií váh (energie) v modeli počas učenia. Hlbšiu teoretickú analýzu a popis iných, podobne navrhnutých algoritmov nájdeme u L. Pastoreka (Pastorek, 2013).

Článok je rozdelený do nasledujúcich častí. Prvá časť sa zaoberá teoretickým predstavením Kohonenových samoorganizujúcich sa sietí. Druhá časť pojednáva o Gompertzovskej samoorganizujúcej sa mape a Gompertzovej funkcii. Tretia časť obsahuje informácie o uskutočnenom experimente a testovacom súbore. Posledná záverečná časť sumarizuje výsledky testu.

2. Samoorganizujúce sa mapy

Vetva samoorganizujúcich sa máp reprezentuje neurónové siete s učením bez učiteľa, ktoré kombinujú pôsobenie tzv. učiacej funkcie a funkcie okolia (susedstva). Siete sa vyvíjajú prostredníctvom adaptačného predpisu, ktorý určuje budúcu pozíciu modelového vektoru v priestore. Táto pozícia je závislá na uvedených funkciách a pozícií mapovaného vstupného vektoru.

Kohonenove samoorganizujúce sa mapy (Kohonen, 2001) sú najznámejší reprezentant tejto vetvy, pričom distribúcia váh v modeli, a teda aj vyplývajúci pohyb modelových vektorov, je založený na existencií fixnej mriežky tzv. neurónov. Každý neurón je asociovaný práve s jedným modelovým vektorom, takže zložitá definícia susedstva vo viacrozmernom priestore je zjednodušená na dvojrozmernú mriežku neurónov. Mriežka presne a jednoznačne definuje susedné neuróny. Pri iteratívnom učení, modelové vektory, ktoré spolu susedia na mriežke, reagujú podobne aj v priestore vstupných vektorov.

Pri sekvenčnom učení modelu, v každom iteratívnom kroku náhodne vyberieme jeden vstupný vektor $\mathbf{x}_i$ z dátového súboru $X = \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbf{R}^d, i = \{1,\ldots,n\}\}$, kde n je dĺžka trénovacieho súboru a $\mathbf{R}^d$ je d-rozmerný vektorový priestor. Tento vstupný vektor následne predložíme populácií modelových vektorov $C = \{\mathbf{c}_j \mid \mathbf{c}_j \in \mathbf{R}^d, j = \{1,\ldots,k\}\}$, kde c_j je modelový vektor a k je počet modelových vektorov. S každým modelovým vektorom je asociovaný práve jeden neurón na dvojrozmernej mriežke. Po predstavení vstupného vektoru modelu, modelové vektory súťažia o najbližšiu pozíciu k danému vektoru

$$\|\mathbf{x} - \mathbf{c}_v\| = \arg\min_j \|\mathbf{x}_i - \mathbf{c}_j\|, \tag{1}$$

kde $\mathbf{c}_v$ je *víťazný modelový vektor*, ktorý reprezentuje Voronoi región (zhluk), do ktorého bude vstupný vektor prvotne patriť. Tento modelový vektor bude adaptovať svoju pozíciu najvýraznejšie a priblíži sa najbližšie k predloženému vstupnému vektoru. I ostatné modelové vektory adaptujú svoju pozíciu, avšak, nie v rovnakej miere. O sile, s akou sa modelový vektor priblíži, rozhoduje jeho poloha na 2-rozmernej mriežke. Ak sa neurón asociovaný s daným modelovým vektorom nachádza v tesnej blízkosti víťazného neurónu, modelový vektor bude meniť svoju pozíciu výrazne. Ak je naopak ďaleko od víťazného neurónu, bude sa meniť jeho poloha len minimálne. Tento spomínaný princíp je pretavený v tzv. funkcii susedstva.

Nielen funkcia susedstva rozhoduje o sile, s akou sa modelový vektor ku vstupnému vektoru pritiahne. Svoju úlohu zohráva i parameter učenia, ktorý riadi celkovú aktivitu modelu. Ten utlmuje v priebehu učenia aktivitu všetkých modelových vektorov bez rozdielu.

Adaptačný vzorec u Kohonenových máp vyzerá nasledovne

$$\mathbf{c}_j(t+1) = \mathbf{c}_j(t) + \alpha(t) h_{jv}(t) [\ \mathbf{x}(t) - \mathbf{c}_j(t)\ ], \tag{2}$$

kde $\mathbf{c}_j(t+1)$ je budúca pozícia j-teho modelového vektoru v nasledujúcom iteratívnom kroku $t+1$. Parameter $\alpha(t)$ je učiaci parameter v čase (iteratívnom kroku) t a môže byť definovaný ako ktorákoľvek klesajúca funkcia v čase t, napr. $\alpha(t) = \alpha_{IN}(0.005/\alpha_{IN})^{t/T}$, kde α_{IN} je hodnota parametra na začiatku učenia a T je počet iteratívnych krokov. Týmto učiacim parametrom sa násobia váhy všetkých modelových vektorov rovnakou mierou v danom iteratívnom kroku. Na konci učenia je tento parameter rovný alebo blízky nule. Funkcia $h_{jv}(t)$ je funkciou okolia (susedstva), pričom jej hodnota závisí na Euklidovskej vzdialenosti medzi skúmaným neurónom a víťazným neurónom na mriežke. Funkcia má najčastejšie tvar gaussovskej krivky $h_{jv}(t) = e^{-d_{jv}^2/2\sigma_t^2}$, kde d_{jv} je vzdialenosť na mriežke od skúmaného j-teho neurónu k víťaznému neurónu v, $d_{jv} = \|r_v - r_j\|$ a σ_t je fixná vzdialenosť v čase t.

3. Gompertzova funkcia a Gompertzovská samoorganizujúca sa mapa

V tejto časti sa venujeme bližšie samoorganizujúcej sa mape, ktorá využíva Gumbelovú kumulatívnu distribučnú funkciu, známejšiu pod pojmom Gompertzova krivka, ako nástroj pre distribúciu síl medzi neurónmi v modeli.

Gompertzovu krivku (Gompertz, 1825) zostrojil Benjamin Gompertz, ako prostriedok na odhad poistného. Týmto spôsobom, však, aj odhalil zákon úmrtnosti v ľudskej populácii. Využitie jeho zákona sa neobmedzilo len na oblasť poisťovníctva a demografie, ale aplikačne expandovalo i do ostatných oblastí. Pôvodný demografický význam Gompertzovej krivky sa pretransformoval na všeobecný rastový zákon, pričom sa modifikované verzie pôvodnej Gompertzovej funkcie stali jednými z najdôležitejších rastových modelov v prírodných vedách. Rigoróznejšiu analýzu vývoja a teoretickej konštrukcie nájdeme u L. Pastoreka (Pastorek, 2013).

Tento rastový model hovorí o exponenciálnom raste limitovanom iným exponenciálnym pôsobením, pričom následkom získame asymetrickú, asymptoticky zhora a zdola obmedzenú sigmoidnú krivku, ktorá je charakterizovaná troma fázami. Prvá fáza (úsek pred inflexným bodom) je charakterizovaná exponenciálny rastom. Druhá fáza (po inflexnom bode) je úsek, v ktorom sa stretávajú viaceré sily, rast sa spomaľuje, degraduje, odchyľuje sa čoraz výraznejšie od exponenciálneho rastu. Posledná tretia fáza má takmer lineárny priebeh, pričom rast je minimálny.

Gompertzovský rast môžeme interpretovať i z iného hľadiska. Na začiatku samoorganizujúce sa systémy rastú najviac. Vďaka interaktivite a silným väzbám medzi prvkami systému dochádza k obrovskému transferu energie a informácií o možnostiach rastu. Postupne sa priestor na rast vyčerpáva a expanzia systému je sprevádzaná čoraz výraznejším uvoľňovaním väzieb, lokálnou špecializáciou, rastúcou nezávislosťou aktivít a poklesom celkovej aktivity. Finálna fáza je útlm rastu.

Neurónové výpočtové učenie môžeme ľahko napojiť na tento koncept. Na začiatku, keď má model veľký priestor na zmapovanie, si neuróny vymieňajú informácie veľmi intenzívne, systém je veľmi aktívny. Postupne sa priestor vstupných vektorov vyčerpáva a model prestáva rásť. V poslednej fáze sa modelové vektory sústreďujú len na mapovanie vstupných vektorov vo svojom okolí bez akejkoľvek referencie na činnosť iných neurónov.

V súlade s Pastorekom (Pastorek, 2013) je adaptačný krok Gompertzovskej samoorganizujúcej sa mapy definovaný ako

$$\mathbf{c}_j(t+1) = \mathbf{c}_j(t) + E[\ \mathbf{x}(t) - \mathbf{c}_j(t)\ ], \tag{3}$$

kde E je vnútorná energia systému, ktorá je definovaná ako

$$E = \alpha(t) - b\left[\frac{\alpha(t)}{b}\right]^{1-c^{-\phi(t)}}, \tag{4}$$

kde symboly ba c sú zvolené konštanty, $\alpha(t)$ je parameter globálneho rastu definovaný ako lineárne klesajúca funkcia

$$\alpha(t) = \alpha_{IN}\left(1 - \frac{t}{T}\right) + b \tag{5}$$

a exponent $\phi(t)$ je parameter sily interakcie

$$\phi(t) = \left[\frac{c}{Q_{\lambda(t)}}\right] d_{jv}. \tag{6}$$

Symbol d_{jv} je euklidovská vzdialenosť na fixnej topologickej mriežke medzi j-tym neurónom a víťazným neurónom v. Gompertzovská samoorganizujúca sa mapa využíva rovnaký princíp fixnej mriežky ako Kohonenova mapa. $Q_{\lambda(t)}$ je λ -tý kvantil počítaný zo všetkých nenulových Euklidovských vzdialeností medzi víťazným neurónom v a inými neurónmi na mriežke v iteratívnom kroku t. Poradové číslo kvantilu λ získame z predpisu

$$\lambda(t) = \lambda_{IN}\left[\frac{\lambda_{FI}}{\lambda_{IN}}\right]^{t/T}, \tag{6}$$

kde význam hodnoty kvantilu λ je identický s významom funkcie okolia, kde λ_{FI} a λ_{IN} sú parametre funkcie okolia na začiatku učenia a na jeho konci.

4. Testovací dátový súbor a experiment

Kohonenovu a Gompertzovskú samoorganizujúcu sa mapu podrobíme testu, pri ktorom budeme sledovať vývoj ukazovateľa priemernej štvorcovej odchýlky (MSE) vektorov dátového súboru od ich najbližších modelových vektorov v priebehu učenia.

Testovací súbor bude umelo vygenerovaný podľa nasledujúcich pravidiel:

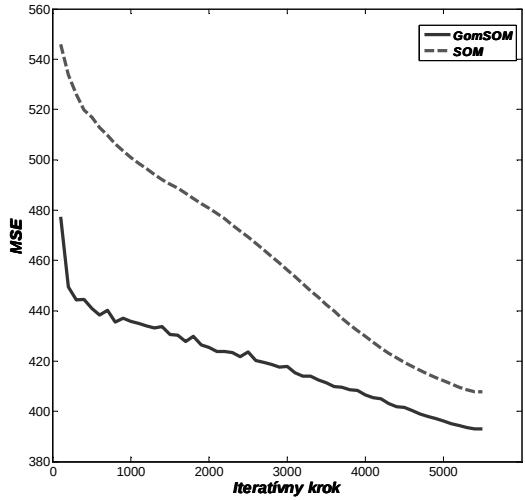
1. Súbor bude obsahovať 6 zhlukov, pričom zhluky budú obsahovať nasledujúci počet vektorov: n_1=400, n_2=70, n_3=70, n_4=400, n_5=70, n_6=70.
2. Každý vektor sa bude skladať z 1000 zložiek (rozmerov).
3. Hodnoty v každom zhluku budú generované normálnym rozdelením.
4. Stredné hodnoty rozdelení sa budú líšiť nielen medzi zhlukmi ale aj medzi dimenziami v rámci zhlukov. Stredné hodnoty pre jednotlivé dimenzie jednotlivých zhlukov budú pseudonáhodne vygenerované celé čísla na otvorenom intervalu (0,100).
5. Zhluky budú mať fixnú hodnotu smerodajnej odchýlky v každej dimenzií: σ_1=5, σ_2=5, σ_3=5, σ_4=20, σ_5=20, σ_6=20.

Týmto spôsobom zostrojíme jeden početný koncentrovaný zhluk, dva malé koncentrované zhluky, jeden početný rozptýlený zhluk a dva malé rozptýlené zhluky.

Parametre v metódach sú nastavené nasledujúcim spôsobom. Obe metódy majú spoločnú počiatočnú hodnotu učiaceho parametra, $\alpha_{IN} = 1$. Gompertzovská samoorganizujúca sa mapa (GomSOM) má parametre okolia $\lambda_{IN} = 0{,}05$ a $\lambda_{FI} = 0{,}001$ a Kohonenova mapa (SOM) má $\lambda_{IN} = 4$a $\lambda_{FI} = 0$. Konštanty $b = 0{,}005$ a $c = 2$.

Algoritmy sú spustené tridsaťkrát s použitím náhodného nastavenia hodnôt prvkov modelových vektorov. Hoci je priestorová inicializácia pri každom spustení odlišná, počiatočné vektorové hodnoty majú obe metódy vždy identické. Pri každom spustení absolvujú algoritmy 5 učiacich epoch, kedy sú modelom päťkrát predstavené vstupné vektory súboru. Hodnota ukazovateľa MSE je vypočítaná pri každom spustení algoritmov v každom stom iteratívnom kroku. Následne sú hodnoty priemerované pre každú stú hodnotu naprieč tridsiatimi behmi. Obe mapy majú nastavenú fixnú topológiu štvorcovej mriežky (10x10 neurónov). Algoritmy boli testované s použitím softvéru Matlab a balíka SOM Toolbox (Vesanto, 2000).

Pri bližšom pohľade na výsledky testu na Obr.1 vidíme výrazný náskok Gomperzovskej samoorganizujúcej sa mapy počas celej doby učenia. Najvýraznejší je náskok algoritmu GomSOM v prvej polovici učiaceho procesu, pričom sa metóda SOM postupne približuje k chybovým hodnotám metódy GomSOM. Porovnateľné hodnoty, však, nikdy nedosiahne.



Obr. 26: Vývoj ukazovateľa MSE v priebehu algoritmického učenia

5. Záver

V tomto príspevku bola predstavená jedna z Gomperzovských neurónových sietí s učením bez učiteľa – Gompertzovská samoorganizujúca sa mapa. Táto metóda používa Gumbelovu kumulatívnu distribučnú funkciu ako hlavný nástroj na distribúciu energie v rastúcom systéme, t.j. váh priradených k modelovým vektorom. Jej výkon bol porovnaný s klasickou Kohonenovou samoorganizujúcou sa mapou v podmienkach dátového súboru s veľmi vysokou dimenziou. Gompertzovská samoorganizujúca sa mapa vykazovala výrazne menšiu chybovosť počas celej dlžky algoritmického učenia, čím prekonala úspešnosť modelu SOM. Gompertzovská samoorganizujúca sa mapa sa prejavila byť vhodný analytický nástroj pri riešení úloh s vysokou dimenziou.

Poďakovanie: Táto práca bola spracovaná s finančnou podporou grantu IGA VSE F4/17/2013.

Literatúra

GOMPERTZ, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In: Philosophical transactions of the Royal Society of London, roč. 115, s. 513–583.

KOHONEN, T. 2001. Self-organizing maps (3rd ed.). Berlin : Springer.

PASTOREK, L. 2013. Gompertzian Fractal Dynamics Applied to Self-Organizing Networks. (v recenznom konaní – under review)

ŘEZANKOVÁ, H., HÚSEK, D., SNÁŠEL, V. 2009. Shluková analýza dat. Praha: Professional Publishing,

VESANTO, J., ALHONIEMI, E., HIMBERG, J., KIVILUOTO, K., PARVIAENEN, J. 2000. SOM Tool-box for Matlab 5. http://www.cis.hut.fi/somtoolbox/

Adresa autora (-ov):

Lukáš Pastorek, Mgr.
Katedra statistiky a pravdepodobnosti
Fakulta informatiky a statistiky VŠE v Praze
nám. W. Churchilla 4
130 67 Praha 3
lukas.pastorek@vse.cz

Flexibilní formy zaměstnanosti v některých zemích střední Evropy[1]
Flexible forms of employment in some countries of Central Europe

Tomáš Pavelka, Tomáš Löster

Abstract: The flexibility of the labour market may take various forms. Flexible forms of employment are one of them. Among flexible forms of employment can be included temporary work contracts, part–time jobs or working from home. Self-employment or in other words entrepreneurship can be perceived as one of the alternatives of flexible forms of employment. Article discusses how these forms of employment are common in a selected group of countries. The economies of these countries were recently, after several years of rapid economic growth, hit by the economic recession. Article, addition to short describing of the economic situation in the last decade, also examines whether the economic cycle had an impact on the incidence of flexible forms of employment in these countries.

Abstrakt: Flexibilita trhu práce může nabývat různých podob. Jednou z nich jsou flexibilní formy zaměstnávání. Mezi flexibilní formy zaměstnávání lze zařadit dočasné pracovní kontrakty, zkrácené pracovní úvazky nebo práce z domova. Jako jedna z alternativ těchto flexibilních forem zaměstnávání se jeví i sebezaměstnávání či jinými slovy podnikatelská činnost osob. Článek pojednává o tom, jak jsou uvedené formy zaměstnávání běžné ve vybrané skupině zemí. Ekonomiky těchto zemí byly nedávno po několika letech rychlého hospodářského růstu zasaženy hospodářskou recesí. Článek vedle popisu ekonomické situace v posledních deseti letech zkoumá také to, zda tento ekonomický cyklus měl dopad na výskyt flexibilních forem zaměstnání v těchto zemích.

Key words: Economic recession, Labour market flexibility, employment

Kľúčové slová: ekonomická recese, flexibilita trhu práce, zaměstnanost

1. Úvod

Flexibilitu trhu práce představuje rychlost a způsob, jak se trh práce přizpůsobuje výkyvům v ekonomice či obecně ve společnosti. Často je zmiňováno, že nízká flexibilita trhu práce vede k vyšší nezaměstnanosti a k prodlužování jejího trvání (Pavelka a kol. 2011). Flexibilita trhu práce je však často uváděná jako protiklad ochrany zaměstnanců. Flexibilita trhu práce může mít pozitivní, ale i negativní dopady jak na zaměstnance, tak také na zaměstnavatele. Je třeba nalézt určitý kompromis mezi flexibilitou a ochranou trhu práce. V této souvislosti se často zmiňuje pojem Flexikurita, která je podporovaná i v rámci Evropské unie.

Flexibilita trhu práce může být pojata různě široce. Nejčastěji se uvádí tyto typy flexibility trhu práce (např. Atkinson 1984 nebo Nekolová 2008):

- vnější numerická flexibilita,
- vnitřní numerická flexibilita,
- funkční flexibilita,
- finanční flexibilta,
- mzdová flexibilita.

V tomto článku zúžíme flexibilitu trhu práce na některé flexibilní formy zaměstnávání. Konkrétně bude pozornost věnována dočasným pracovním kontraktům, práci na zkrácené pracovní úvazky a práci z domova. Jako určitou alternativu těmto flexibilním formám zaměstnanosti lze uvést i sebezaměstnávání či jinými slovy podnikání osob. Do analýza výskytu flexibilních forem zaměstnávání budou zahrnuty některé státy střední Evropy: Česká

[1] Článek je součástí řešení projektu „Flexibilita trhu práce“, který je podpořen Interní grantovou Vysoké školy ekonomické v Praze a je veden pod číslem VŠE IGS 19/2012.

republika, Německo, Maďarsko, Polsko, Rakousko, Slovenko a navíc i Slovinsko. Ekonomiky těchto států byly v roce 2009 zasaženy hospodářskou recesí. Článek se pokouší zjistit, zda tento ekonomický propad měl dopad na výskyt flexibilních forem práce. Právě z tohoto důvodu je nejprve v článku stručně popsán vývoj reálného hrubého domácího produktu a míry nezaměstnanosti v těchto zemích.

2. Ekonomická recese a nezaměstnanost

V období před rokem 2008 dosahovaly sledované ekonomiky střední Evropy relativně vysoký růst reálného hrubého domácího produktu. Např. v roce 2007 dosahovaly nejvyššího tempa růstu Slovensko (o 10,5 %), Slovinsko (7,0 %), Polsko (6,8 %) a Česká republika (5,7 %). Naopak ekonomické problémy se projevily v Maďarsku, které ve stejný rok vzrostlo pouze o 0,1 %, což znamenalo meziroční zpomalení růstu o 3,8 p. b. Země původní evropské patnáctky rostly pomaleji, v roce 2007 vzrostl reálný hrubý domácí produktu v Německu o 3,3 % a v Rakousku o 3,7 %. Evropská ekonomika jako celek zaznamenala v roce 2007 růst reálného produktu o 3,2 %. V roce 2009 však nastal zlom, finanční krize, která měla kořeny ve Spojených státech, začala negativně ovlivňovat vývoj evropských ekonomik. V roce 2009 se již krize projevila plně a reálný hrubý domácí produkt klesl s výjimkou jedné ve všech sledovaných zemích. Největší pokles reálného hrubého domácího produktu byl zaznamenán ve Slovinsku (-7,9 %), následované Maďarskem (-6,8 %), Německem (-5,1 %), Slovenskem (-4,9 %), Českou republikou (-4,5 %) a Rakouskem (-3,8 %). Jedinou zemí, ve které rostl reálný hrubý domácí produkt i v roce 2009 bylo Polsko (1,6 %). Hlavní příčinou byl velký a relativně uzavřený vnitřní trh a také struktura mezinárodního obchodu Polska. Reálný hrubý domácí produkt celé Evropské unie ve stejném roce klesl o 4,5 %. V následujících dvou letech vzrostl reálný produkt ve všech sledovaných ekonomikách. Toto oživení bylo relativně slabé a svou roli samozřejmě hrál i statistický efekt nízké základny z roku 2009. V loňském roce některé sledované státy pokračovaly ve svém růstu: Polsko (1,9 %), Slovensko (1,8 %), Rakousko
(0,9 %) a Německo (0,7 %), ale některé se opětovně propadly do recese: Slovinsko (-2,5 %), Maďarsko (-1,7 %) a Česká republika (-1,0 %).

Ekonomická recese měla samozřejmě dopad také na trh práce, respektive na nezaměstnanost. Nejnižší míru nezaměstnanosti dosahovaly všechny sledované státy v roce 2008. V tomto roce již sice docházelo ke zpomalení růstu ekonomik, ale jak je všeobecně známo, nezaměstnanost reaguje s určitým časovým zpožděním. Nejnižší míru nezaměstnanosti dosahovalo v roce 2008 Rakousko (3,9 %) následované Českou republikou (4,4 %, Slovinskem (4,5 %), Polskem (7,2 %), Německem (7,6 %), Maďarskem (7,9 %) a Slovenskem (9,5 %). V celé Evropské unii dosahovala ve stejném roce míra nezaměstnanosti 7,1 %. Následná hospodářská recese vedle k velmi výraznému nárůstu míry nezaměstnanost. V roce 2009 vzrostla míra nezaměstnanosti meziročně ve všech sledovaných zemích, přičemž tento nárůst pokračoval ve většině zemí i v roce 2010. Mezi roky 2008-2010 vzrostla míra nezaměstnanosti v České republice o 3,0 p. b., Maďarsku o 3,3 p. b., Rakousku o 0,6 p. b., Polsku o 2,5 p. b., Slovinsku o 2,9 p. b. a Slovensku o 4,9 p. b. Jedinou zemí, ve které míra nezaměstnanosti mezi lety 2008 a 2010 klesla, bylo Německo. V Německu sice v roce 2009 vzrostla mírně míra nezaměstnanosti, ale v roce 2010 byla nakonec ve srovnání s rokem 2008 míra nezaměstnanosti nižší o 0,4 p. b. V Evropské unii jako celku se míra nezaměstnanosti mezi roky 2008 a 2010 zvýšila o 2,6 p. b. Ve zbývajících dvou letech klesla míra nezaměstnanosti v České republice, Německu, Maďarsku, Rakousku a Slovensku. Tento pokles byl však s výjimkou Německa velmi mírný. V Polsku a Slovinsku však míra nezaměstnanosti v roce 2012 přesahuje její hodnoty v roce 2010.

3. Flexibilní formy zaměstnávání

Ekonomická recese a nárůst nezaměstnanosti by měl mít spojitost i s vývojem flexibilních forem zaměstnávání. Jak již bylo uvedeno výše, vyšší zastoupení flexibilnějších forem zaměstnávání by mohlo přispět k rychlejšímu vymanění se z ekonomické recese. V této kapitole bude popsán vývoj flexibilních forem zaměstnanosti v období od roku 2005 do roku 2012, tedy v období před ekonomickou recesí, v období samotné ekonomické recese, ale i v období které ekonomickou recese následovalo. Pozornost bude věnována dočasným pracovním úvazkům, zkráceným pracovním úvazkům a práci z domova. Vedle těchto tradičně uváděných flexibilních forem však bude pozornost věnována i vývoji sebezaměstnanosti neb tato forma zaměstnanosti může, jak bylo uvedeno výše, představovat v některých zemích alternativu tradičním flexibilním formám zaměstnanosti.

Dočasné pracovní úvazky

Dočasné pracovní úvazky či také úvazky na dobu určitou představují formu zaměstnávání, která umožňuje firmám plynule reagovat na výkyvy v poptávce (např. sezónní), aniž by musely vynakládat náklady, které jsou spojeny se zaměstnáváním na dobu neurčitou. Dočasné pracovní úvazky mohou mít podobu vlastního zaměstnávání firmou či najmutí zaměstnanců přes agentury práce. Je zřejmé, že pokud dojde k ekonomické recesi, jsou právě zaměstnanci na dočasné pracovní úvazky postiženi jako první.

Tabulka č. 1 zachycuje procento zaměstnaných osob s dočasnými pracovními úvazky z celkového počtu zaměstnanců v období 2005 – 2012. Je třeba si uvědomit, že se nejedná o absolutní počty, ale o procentní podíl.

Nejvíce jsou dočasné pracovní úvazky využívané v Polsku, Slovinsku a také v Německu, naopak nejméně na Slovensku a také v České republice. Srovnáme-li období před krizí (2005-2008) s obdobím krize a po krizi (2009-2012) je zřejmé, že v zemích, které vykazují vysoký podíl dočasných kontraktů, došlo v po krizovém období ke stagnaci či mírnému snížení tohoto procentního podílu. Jistou výjimkou je Německo, kde podíl zaměstnanců s dočasnými úvazky v loňském roce meziročně klesl o 0,9 p. b., což by mohlo být vysvětleno pozitivním ekonomickým růstem, při kterém jsou zaměstnavatelé ochotni k zaměstnávání na dobu neurčitou. V České republice, Maďarsku a na Slovensku se tyto zkrácené úvazky stávají v posledních letech častějšími, mezi roky 2008 – 2012 vzrostl podíl zaměstnanců se zkrácenými úvazky na celkovém počtu zaměstnanců v České republice o 1,1 p. b., v Maďarsku o 1,6 p. b. a na Slovensku dokonce o 2,2 p. b.

Tab. 1: Zaměstnanci s dočasnými pracovními úvazky z celkového počtu zaměstnanců v %

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Česká republika	7,9	8,0	7,8	7,2	7,5	8,2	8,0	8,3
Německo	14,3	14,6	14,7	14,8	14,6	14,7	14,8	13,9
Maďarsko	7,0	6,7	7,3	7,8	8,4	9,6	8,9	9,4
Rakousko	9,1	9,0	8,9	9,0	9,1	9,3	9,6	9,3
Polsko	25,6	27,3	28,2	26,9	26,4	27,2	26,8	26,8
Slovinsko	17,2	17,1	18,4	17,3	16,2	17,1	18,0	17,0
Slovensko	4,9	5,0	5,0	4,5	4,3	5,6	6,5	6,7
EU28	**14,0**	**14,5**	**14,5**	**14,1**	**13,5**	**13,9**	**14,0**	**13,7**

Pramen: Eurostat

Zkrácené pracovní úvazky

Zkrácené úvazky jsou úvazky na kratší pracovní dobu než odpovídá právními normami stanovené pracovní době na plný úvazek. Zkrácené pracovní úvazky se výrazně rozšířily ve vyspělých letech v posledních dvou dekádách. Důvody existence zkrácených úvazků jsou

různé, nemusí jít vždy pouze o aktivitu ze strany zaměstnavatelů, ale sami zaměstnanci často požadují zkrácené pracovní úvazky (např. ženy na mateřské dovolené). Dochází však i k situacím, při kterých by zaměstnanci chtěli pracovat na plné úvazky, ale práci neseženou. Jsou tak nuceni vzít i úvazky na kratší dobu.

Tabulka 2 zachycuje procentní podíl zaměstnanosti na zkrácené úvazky z celkové zaměstnanosti v letech 2005 – 2012. Je důležité si uvědomit, že jde o zaměstnanost a ne o zaměstnance. Podrobnou analýzu srovnání zemí podle procentního podílu zaměstnanosti na zkrácené úvazky v rámci Evropské unie lze nalézt v článku *Labour Market Flexibility from the Perspective of Part-time Job* (Langhamrová 2013).

Nejvyšší podíl zaměstnanosti na zkrácené úvazky v průběhu celého sledovaného období vykazuje Německo a Rakousko, a jak je zřejmé z tabulky č. 2, tento podíl se navíc stále zvyšuje. Zkrácené úvazky jsou hojně využívané zejména ve státech západní Evropy, kde jsou časté i ve vysoce kvalifikovaných zaměstnáních. Naopak ve státech střední a východní Evropy byly zkrácené úvazky často nabízené u méně náročných pozic, s čímž je spojeno i relativně nízké finanční ohodnocení, které bránilo většímu rozšíření zkrácených úvazků. Situace se v posledních letech mění pouze velmi pozvolna. Jak je patrné z tabulky č. 2, v České republice se po krizi mírně zvýšilo procento zaměstnanosti na zkrácené úvazky. Obdobný vývoj je patrný také ve Slovinsku, Maďarsku či Slovensku. Naopak jedinou zemí, u které došlo v po krizovém období ve srovnání s obdobím před krizí k poklesu procentního podílu zaměstnanosti na zkrácené úvazky, je Polsko.

Tab. 2: Zaměstnanost na zkrácené úvazky z celkové zaměstnanosti v %

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Česká republika	4,4	4,4	4,4	4,3	4,8	5,1	4,7	5,0
Německo	23,4	25,2	25,4	25,1	25,3	25,5	25,7	25,7
Maďarsko	3,9	3,8	3,9	4,3	5,2	5,5	6,4	6,6
Rakousko	20,8	21,3	21,8	22,6	23,7	24,3	24,3	24,9
Polsko	9,8	8,9	8,5	7,7	7,7	7,7	7,3	7,2
Slovinsko	7,8	8,0	8,1	8,1	9,5	10,3	9,5	9,0
Slovensko	2,4	2,7	2,5	2,5	3,4	3,8	4,0	4,0
EU28	**17,2**	**17,5**	**17,5**	**17,5**	**18,0**	**18,5**	**18,8**	**19,2**

Pramen: Eurostat

Práce z domova

Práce z domova znamená, že zaměstnanec většinu pracovní doby provádí práci z domova a do prostor zaměstnavatele dochází výjimečně. Tato možnost flexibilního zaměstnávání je spojena zejména s rozvojem informačních technologií, díky kterým mohou zaměstnanci využívat doma výpočetní techniku k plnění svých pracovních úkolů. Práce doma však může mít i jiné podoby než jen práce s výpočetní technikou. Jako nejčastější výhoda práce z domova z pohledu zaměstnanců je uváděná větší možnost skloubení pracovního a osobního života. U zaměstnavatele jsou to pak nižší náklady spojené se samotným pracovním místem, např. náklady na provoz kanceláře.

Tabulka č. 3 zachycuje procento zaměstnanců, kteří pracují z domova z celkového počtu zaměstnanců v období 2005 – 2012.

Jak je patrné z tabulky č. 3, práce z domova je rozšířená zejména v Rakousku, přičemž její rozšíření se neustále zvyšuje. V loňském roce již 6 % z celkového počtu zaměstnanců v Rakousku pracovalo z domova. Vysoký podíl zaměstnanců pracujících z domova mělo také Slovinsko. V ostatních sledovaných státech dosahovalo procento zaměstnanců pracujících z domova nižší hodnoty než je celoevropský průměr. Je zajímavé, že v České republice,

Německu, Maďarsku a Slovensku vlivem ekonomické krize došlo spíše k poklesu podílu osob pracujících z domova na celkovém počtu zaměstnanců.

Tab. 3: Zaměstnanci pracující z domova z celkového počtu zaměstnanců v %

	2005	2006	2007	2008	2009	2010	2011	2012
Česká republika	1,1	1,2	1,0	0,9	0,6	0,8	0,7	0,7
Německo	1,8	1,7	1,5	1,9	1,4	1,3	1,6	1,5
Maďarsko	1,1	0,8	0,7	0,9	0,8	0,8	1,0	1,2
Rakousko	2,9	5,7	5,7	5,6	5,8	6,0	6,1	6,0
Polsko	1,3	1,3	1,3	1,3	1,4	1,5	1,6	1,5
Slovinsko	5,8	4,8	4,5	4,6	5,4	5,7	5,3	5,1
Slovensko	2,6	2,6	2,4	2,1	1,8	1,6	1,9	2,1
EU28	**2,4**	**2,2**	**2,3**	**2,6**	**2,5**	**2,7**	**3,0**	**3,0**

Pramen: Eurostat

Sebezaměstnaní

Jak je zřejmé z předcházejících částí textu, v některých zemích nejsou flexibilní formy zaměstnávání hojně využívané. Jedním z vysvětlení může být to, že v některých zemích tvoří k tradičně uváděným flexibilním formám zaměstnávání alternativu sebezaměstnávání. Sebezaměstnávání představuje podnikající osoby. Vedle zaměstnanců jsou podnikající osoby součástí celkové zaměstnanosti. Sebezaměstnané či podnikatele lze rozčlenit na ty, kteří jsou bez zaměstnanců a na ty, kteří mají zaměstnance (jsou zaměstnavatelé). Speciální skupinou jsou pak pomáhající rodinní příslušníci. Při hodnocení významu sebezaměstnáných jako jedné z flexibilních forem práce je třeba upozornit, že výše uvedený ukazatel zaměstnanosti na zkrácené úvazky může v sobě zahrnovat i sebezaměstnané, kteří pracují na zkrácený úvazek.

Tabulka č. 4 zachycuje procento sebezaměstnaných na celkové zaměstnanosti v období 2005 – 2012.

Z tabulky č. 4 je zřejmé, že nejvyšší procento sebezaměstnaných na celkové zaměstnanosti vykazuje dlouhodobě Polsko následované Českou republikou. Na rozdíl od Polska, kde podíl sebezaměstnaných po roce 2005 nejprve mírně klesl a poté stagnoval, v České republice byl podíl sebezaměstnaných do krize relativně stabilní, ale po vypuknutí krize postupně narůstá. V této souvislosti lze uvést, že náklady pro zaměstnavatele při zaměstnávání zaměstnance jsou ve srovnání s využívání dodávek práce od podnikatele výrazně vyšší. Také z pohledu zaměstnance nejsou daňové výhody práce na vlastní účet zanedbatelné v porovnání s pozicí zaměstnance. Postupný nárůst podílu sebezaměstnaných na celkové zaměstnanosti je patrný také na Slovensku. Naopak v Německu a Rakousku je tento podíl relativně stabilní.

Tab. 4: Sebezaměstnaní z celkové zaměstnanosti v %

	2005	2006	2007	2008	2009	2010	2011	2012
Česká republika	15,1	15,3	15,4	15,2	15,9	16,8	17,2	17,5
Německo	10,8	10,7	10,5	10,3	10,5	10,5	10,5	10,5
Maďarsko	13,1	12,1	11,8	11,6	11,9	11,7	11,4	10,9
Rakousko	11,6	11,7	11,7	11,1	10,9	11,3	11,3	11,0
Polsko	20,0	19,4	18,7	18,3	18,3	18,7	18,7	18,4
Slovinsko	9,3	10,4	10,0	9,3	10,1	11,6	11,9	11,6
Slovensko	12,5	12,5	12,8	13,6	15,5	15,8	15,8	15,3
EU28	**14,6**	**14,6**	**14,4**	**14,2**	**14,3**	**14,6**	**14,4**	**14,5**

Pramen: Eurostat, vlastní výpočet

4. Závěr

Flexibilní formy zaměstnávání nejsou v analyzované skupině zemí rozšířeny stejně. Česká republika vykazuje velmi nízké rozšíření dočasných pracovních kontraktů, částečných úvazků i práce z domova, naopak sebezaměstnání je velmi rozšířeno. Postupně však v České republice dochází ke zvyšování podílu těchto flexibilních forem práce. Obdobný podíl i vývoj flexibilní formy práce zaznamenávají i Slovensko a Maďarsko. Polsko vykazuje nejvyšší podíl dočasných kontraktů ze všech sledovaných zemí. Podíl zkrácených úvazků a sebezaměstnání však v Polsku v posledních deseti letech mírně klesá. Slovinsko vykazuje vysoký výskyt všech uvedených flexibilních forem zaměstnání. Německo v případě dočasných kontraktů dosahuje v podstatě průměrných hodnot Evropské unie a výrazně tento průměr převyšuje u zkrácených pracovních úvazků. V případě práce z domova však Německo dosahuje podprůměrných hodnot. V Rakousku, podobně jako v Německu, jsou značně rozšířené zkrácené pracovní úvazky a navíc i práce z domova.

Nedávná ekonomická recese neměla na výskyt flexibilní forem zaměstnávání ve většině sledovaných zemí v podstatě žádný vliv. Země, které měly vysoký podíl určité flexibilní formy zaměstnání, si tento podíl udržují i v po krizovém období. Země, u kterých byl podíl flexibilních forem zaměstnání nízký, se jejich výskyt postupně zvyšuje.

Literatura

ATKINSON, J. (1984) *Flexibility, Uncertainty and Manpower Management*, IMS Report No.89, Institute of Manpower Studies, Brighton.

EUROSTAT. Databáze zaměstnanosti a nezaměstnanosti (VŠPS). Online. Citováno: 12.11.2013.
http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/data/database.

LANGHAMROVÁ, J. 2013. Labour Market Flexibility from the Perspective of Part-time Job. In: PAVELKA, T., LÖSTER, T. (ed.). *International Days of Statistics and Economics*. Slaný: Melandrium, 2012, s. 903–911. ISBN 978-80-86175-86-7.

NEKOLOVÁ, M. 2008. Flexicurity- hledání rovnováhy mezi flexibilitou a ochranou trhu práce v České republice. VÚPSV, Praha 2008. ISBN 978-80-87007-89-1.

PAVELKA, T., LÖSTER, T., MAKOVSKÝ, P., LANGHAMROVÁ, J. 2011. *Dlouhodobá nezaměstnanost v České republice.* 1. vyd. Slaný: Melandrium, 2011. 116 s. ISBN 978-80-86175-76-8.

Adresa autora (-ov):

Tomáš Pavelka, doc., Ing., Ph.D.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 67 Praha 3
pavelkat@vse.cz

Tomáš Löster, Ing. Ph.D.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 67 Praha 3
losterto@vse.cz

Využití vícekriteriálních rozhodovacích metod v regionální analýze udržitelného rozvoje
Using multicriteria decision methods in the analysis of regional sustainable development

Ludmila Petkovová, Lenka Hudrlíková

Abstract: The aim of the paper is to apply multicriteria decision methods for aggregation in a composite indicator approach. This approach was used for ranking of Czech NUTS3 regions from point of view of sustainable development. The indicators in each pillar (environmental, economic and social) are merged by means of linear aggregation. The pillars are aggregated by methods derived from multicriteria decision theory (namely Borda and Condorcet). The methods differ in a level of compensability. Compensability means poor performance in some indicator can be compensated by sufficiently high values of others indicators and vice versa. Borda allows for compensability but Condocet does not. Indicators used for this analysis were provided by Czech Statistical Office. The final ranking is thoroughly discussed.

Abstrakt: Cílem tohoto příspěvku je ukázat možnosti využití vícekriteriálních rozhodovacích metod v úlohách agregace a tvorby kompozitních indikátorů. Metody zde byly použity pro regiony České republiky na úrovni NUTS 3 v oblasti udržitelného rozvoje. Pro agregaci uvnitř pilířů bylo využito lineární agregace a pro agregaci pilířů vícekriteriálních rozhodovacích metod v podobě Bordova a Condorcetova přístupu. Oba přístupy se liší v pojetí možností substituce/kompenzace vstupujících proměnných. Zatímco první přístup připouští možnost kompenzace slabých výsledků u některých ze sledovaných indikátorů dostatečně vysokými hodnotami jiných indikátorů, druhý nikoliv. Indikátory použité pro analýzu vychází z analýz udržitelného rozvoje na regionální úrovni provedených Českým statistickým úřadem. Závěry obsahují diskuzi nad výsledky vytvořeného pořadí regionů.

Key words: Multi-criteria decision analysis, Sustainable development, Composite indicators, Regional analysis

Kľúčové slová: Vícekriteriální rozhodovací metody, Kompozitní indikátory, Udržitelný rozvoj, Regionální analýza

JEL classification: R11

1. Úvod

Ačkoliv je Česká republika spíše malou zemí, je možné sledovat rozdílný vývoj jednotlivých regionů, který je způsobený především rozdílnými přírodními podmínkami a odlišným historickým vývojem. Česká republika prochází v posledních 30 letech velkými ekonomickými a sociálními změnami a značná pozornost je v tomto kontextu věnována i otázce vývoje jednotlivých jejích částí. Otevřenou otázkou stále zůstává zda-li tento přerod povede k sbližování jednotlivých region či naopak. Ekonomický rozkvět, který je s posledními desetiletími spjat sebou přináší nejen pozitiva, ale i hrozby v oblasti sociálních nerovností, korupce a devastace přírody. Na druhou stranu s sebou ale nese uvědomení a pocit odpovědnosti. Zhodnocení stavu a vývoje všech složek udržitelného rozvoje v regionech České republiky je ze všech těchto důvodů zcela nepostradatelné. Je to však velmi nelehká úloha vyžadující zhodnocení mnoha na první pohled nesourodých a někdy kvantitativně těžko zachytitelných jevů, a to vše při maximální jednoduchosti a výstižnosti výsledku. Indikátorové sady prezentující desítky až stovky údajů jsou v těchto ohledech nepřehledné a omezující. Trendem posledních let je proto ústup od prezentace takto náročně pojatých výsledků a hledání nových cest, například v podobě kompozitních indikátorů. Hlavní výhoda využití kompozitních indikátorů spočívá v možnosti shrnutí komplexních jevů. Díky

konstrukci kompozitního indikátoru lze následně provést snadné porovnání regionů. Často může platit i jen to, že poukázání na slabé regiony může přivést větší pozornost k určitému problému, což je první krok v cestě za zlepšením.

2. Data

Původní sada indikátorů udržitelného rozvoje byla vytvořena Českým statistickým úřadem v roce 2008 a v roce 2010 byla zrevidována. Od té doby již nebyla provedena další komparativní studie o udržitelném rozvoji českých regionů. Pro naši analýzu jsme vyšly ze souboru ukazatelů Českého statistického úřadu, který byl naplněn nejnovějšími dostupnými daty. Právě s ohledem na dostupnost dat jsme musely přistoupit k několika málo úpravám původní indikátorové sady. Indikátory jsou rozděleny do tří pilířů udržitelného rozvoje - ekonomický (12 ukazatelů), sociální (12) a environmentální (12). Kompletní seznam s detailním vysvětlením změn a úprav oproti původnímu souboru ukazatelů ČSÚ je dostupný v (Fischer a kol. 2013).

Pro srovnání regionů jsme zvolily úroveň NUTS 3 územních správních celků, kde do analýzy vstoupil soubor čtrnácti krajů České republiky. Pro nižší územní celky již nejsou dostupná vhodná data.

3. Metody

Na základě teorie o udržitelném rozvoji byla zvolena dvojstupňová agregace. První krok spočívá v agregaci v rámci pilíře, kdy je pro každý pilíř (ekonomický, environmentální, sociální) provedena samostatná agregace. Výsledné skóre pilířů byly dále agregovány do jednoho finálního skóre.

Ukazatele nejsou vyjádřeny ve stejných měrných jednotkách a je tedy třeba nejdříve provést normalizaci dat. Normalizace dat v neposlední řadě také pomáhá s nastavením směru závislostí. Indikátory jsme převedly do směru „vyšší hodnota, lepší pořadí“, a to pomocí z-skóre, kdy pro každou napozorovanou hodnotu x_{ij}^t, kde i ($i=1,...,Q$) je označení ukazatele, j ($j=1,...,N$) je označení jednotky a t rok pozorování, je vypočtena hodnota dle vzorce:

$$I_{ij}^t = \frac{x_{ij}^t - x_{ij=\bar{j}}^t}{\sigma_{ij=\bar{j}}^t} \quad (1)$$

kde $x_{ij=\bar{j}}^t$ je průměr mezi porovnávanými jednotkami a $\sigma_{ij=\bar{j}}^t$ směrodatná odchylka. Normalizovaná data mají nulovou střední hodnotu a rozptyl 1 [$N\sim(0,1)$]. Výhodou této metody je, že zajišťuje nezkreslení od průměru, sjednotí různé škály a rozptyl (variabilitu). Díky lineárnímu vztahu zůstanou po normalizaci zachovány relativní rozdíly mezi hodnotami. Metoda pouze částečně eliminuje vliv odlehlých pozorování a ukazatel s extrémními hodnotami má tak větší vliv na výsledný kompozitní indikátor. Protože není důvod předpokládat, že jeden z pilířů (či ukazatelů) je významnější než další, neaplikujeme žádné váhy, neboli používáme rovných vah. To samozřejmě zjednodušuje i interpretaci výsledných hodnot. Pro kontrolu tohoto předpokladu jsme provedly exploratorní analýzu dat, která zahrnovala korelační analýzu a metodu hlavních komponent pro poznání struktury dat. Jednoznačně se neprojevila potřeba aplikace různých vah. Navíc názor na důsledky korelace mezi jednotlivými ukazateli nemusí být jednotný (Saltelli, 2012). Na jedné straně silná korelace mezi ukazateli může být chápána jako problém, protože může ukazovat na dvojí zahrnutí stejného jevu. Na druhou stranu, silná korelace mezi ukazateli může být znakem měřeného komplexního jevu či ukazatele mohou reflektovat nezaměnitelné různé znaky sledovaného jevu a úpravy pak nejsou na místě. Korelace mezi ukazateli byla prozkoumána a bylo ověřeno, že není tvořena nadbytečností ukazatele či více ukazatelů.

Základní otázka u agregačních metod je substituce/kompenzace mezi ukazateli v agregovaném ukazateli. To znamená možnost, aby nízká hodnota jednoho ukazatele byla kompenzována dostatečně vysokou hodnotou jiného ukazatele. Metody agregace se od sebe liší právě stupněm kompenzace. U lineární agregace pomocí aritmetického průměru je možnost kompenzace konstantní. K nekompenzovatelné agregaci ukazatelů vedou některé nelineární metody odvozené z metod vícekriteriálního porovnání.

V rámci pilířů bylo k agregaci využito lineární metody agregace. Tato metoda je nejčastěji užívaná z důvodu snadné srozumitelnosti a jednoduché interpretace. Jelikož lineární agregační metody dovolují plnou kompenzaci, je možné aplikovat tuto metodu uvnitř pilířů, kde lze dovolit kompenzaci.

Pro agregaci pilířů je lineární metoda zcela nevhodná a zvolily jsme raději metody odvozené z teorie vícekriteriálního rozhodování. Dva hlavní směry představují Borda a Condorcet. (Vansnick, 1990) prokázal, že Condorcetova teorie volby je nekompenzační, zatímco Bordova je kompenzační právě ve smyslu možnosti nahrazení jedné nízké hodnoty ukazatele dostatečně velkou hodnotou jiného ukazatele. Condorcet přináší nelineární porovnání, což implikuje nemožnost substituce mezi ukazateli. V takovém případě váhy vyjadřují míru důležitosti daného dílčího ukazatele. Z tohoto důvodu je právě Condorcetův přístup či některá z jeho modifikací přijímána jako vhodný nástroj pro agregaci ukazatelů do jednoho kompozitního ukazatele.

Existuje obecná shoda (Moulin, 1988), (Truchon, 1995) nebo (Young, 1995)), že Bordova metoda je vhodná pro zvolení nejlepší jednotky, avšak Condorcetova metoda je vhodnější pro určení pořadí všech porovnávaných jednotek. Informace o intenzitě preference jednotky u metod odvozených z teorie vícekriteriálního rozhodování se vytrácí a tyto metody pracují jen s ordinálním pořadím jednotek u dílčích ukazatelů. Tato vlastnost je v tomto případě pozitivní, jelikož tyto metody jsou použity až při druhém stupni agregace, kdy agregujeme tři již agregované indikátory.

Bordova metoda je ve své podstatě skórovací pravidlo. Pro *N* jednotek platí, že pokud je jednotka hodnocena nejhůře, nedostane žádný bod, pokud je hodnocena jako druhá nejhorší dostane jeden bod. Takto proces pokračuje až do *N*-1 bodů, které obdrží nejlepší jednotka. Pro výpočet je vhodné využít matici četností, tzn. utříděnou informaci, kolikrát byla jednotka hodnocena na 1., 2. až *N*-tém místě. Matice četností pak přehledně udává počet bodů každé jednotky, dle toho v kolika indikátorech se umístila jako první až *N*-tá. Výsledné pořadí jednotek je samozřejmě takové, že jednotka s nejvyšším skóre je nejlepší atd. Z toho vyplývá, že Bordův přístup může zvolit jako nejlepší takovou jednotku, která není v nejvíce ukazatelích první. V postupu jsou využita všechna pořadí a finální hodnocení tak závisí na všech ukazatelích. Vynechání jednoho ukazatele či přidání nového ukazatele by mohlo zcela změnit výsledné hodnocení. Je zřejmé, že Bordovo pravidlo dovoluje kompenzaci mezi ukazateli a pořadí tak závisí na počtu porovnávaných jednotek.

Condorcetův přístup je založen na párových porovnáních mezi všemi uvažovanými jednotkami. Pro výpočet se používá matice sestavená z koeficientů konkordance reprezentujících počet ukazatelů, pro něž daná jednotka byla lepší než jiná jednotka při párovém porovnání. Vznikne čtvercová matice *E* o rozměrech *NxN*, kde pro každé *r≠s* (*r* značí *r*-tý řádek matice a *s* značí *s*-tý sloupec matice) pro prvky matice platí $e_{rs} + e_{sr} = N$, kde *N* je počet jednotek. Při aplikaci vah lze matici upravit tak, že váhy ukazatele, které jsou ve prospěch jednotky *a* (oproti jednotce *b*) při všech párových porovnání jsou sečteny do koeficientu konkordance místo počtu vítězství při párovém porovnání. Za předpokladu, že součet vah je 1 platí pro upravenou matici *E** vztah $e^*_{rs} + e^*_{sr} = 1$. Pro *N* jednotek vznikne *N×(N-1)* porovnání a tudíž stačí při aplikování Condorcetova pravidla pracovat pouze s páry ukazatelů, u kterých je koeficient větší než ½ počtu ukazatelů (v případě matice *E**, kde

koeficient je větší než 50 %). Za nejlepší je zvolena jednotka s největším počtem vítězství v párovém porovnání. Condorcet formuloval závěr, že pokud existuje jednotka, jež získá prostou většinu nad ostatními jednotkami v párovém porovnání, pak tato jednotka má být první.

Dalším možným postupem, jež vychází z matice párových porovnání a Condorcetova přístupu, je Copelandovo pravidlo, které vede k zabránění tvorby cyklů provázejících Condorcetův postup. Copelandova metoda je založena na rozdílu počtu jednotek z množiny *N*, nad kterými jednotka *a* zvítězila a počtu jednotek z množiny *N*, které naopak byly lepší než jednotka *a*. Jinými slovy jedná se o počet vítězství jednotky *a* nad jinými jednotkami snížený o počet proher jednotky *a* v porovnání s každou další jednotkou z množiny *N*. Výpočet lze nejsnadněji provést pomocí matice párových porovnání *E**. Všechny hodnoty větší než 0,5 v matici *E** jsou nahrazeny +1, hodnoty nižší než 0,5 nahrazeny -1 a hodnota 0,5 bude odpovídat 0. Celkové skóre je podle Copelanda dáno součtem hodnot pro každou jednotku. Na prvním místě je samozřejmě jednotka s nejvyšším skóre a dále jsou jednotky seřazeny dle klesajících hodnot skóre.

4. Výsledky

Výsledky hodnocení krajů v rámci jednotlivých pilířů udržitelného rozvoje České republiky jsou uvedeny v tabulce 1. Jsou zde patrné silné a naopak slabé stránky jednotlivých krajů.

Tab. 8: Výsledky hodnocení v prvním stupni agregace

	Ekonomický pilíř	Sociální pilíř	Environmentální pilíř
Hl. m. Praha	1	1	14
Středočeský kraj	3	7	13
Jihočeský kraj	12	6	5
Plzeňský kraj	7	3	3
Karlovarský kraj	13	11	2
Ústecký kraj	9	14	8
Liberecký kraj	6	9	1
Královéhradecký kraj	8	4	7
Pardubický kraj	10	8	11
Kraj Vysočina	14	5	9
Jihomoravský kraj	2	2	6
Olomoucký kraj	5	12	10
Zlínský kraj	11	10	4
Moravskoslezský kraj	4	13	12

Specifickým krajem je Hl. m. Praha. Tento kraj je vymezen hranicemi města a je centrem jak vládních, tak ale i mnoha vzdělávacích institucí a obchodních společností. Ekonomická síla a vysoce kvalifikované pracovní síly navázaní na centrální státní úřady či vědecko-výzkumná a vzdělávací pracoviště jsou hlavním důvodem předního umístění v ekonomickém a sociálním pilíři. Naopak absence ploch zemědělského a přírodního charakteru (toto přirozené zázemí města spadá administrativně již do Středočeského kraje) způsobuje umístění tohoto regionu na poslední příčce v pilíři environmentálním. V určité míře lze spatřovat podobnost v umístění v prvních dvou pilířích i u Jihomoravského kraje, jehož

správním centrem je město Brno. Byť v porovnání s Prahou je Brno centrem spíše regionálního významu. Podobnost pak rozhodně není u umístění ve třetím pilíři, neboť území tohoto kraje nezahrnuje pouze samotné město (jako je tomu v případě Prahy), ale též přírodně velmi rozmanité regiony jižní Moravy.

V Tabulce 2 jsou zobrazeny výsledná pořadí krajů České republiky po použití vícekriteriálních rozhodovacích metod v podobě Bordova a Condorcetova přístupu v druhém stupni agregace. Condorcetův přístup byl navíc doplněn o Copelandův postup výpočtu. Rozdíl v pořadích jednotlivých krajů není při použití jednotlivých metod příliš veliký, maximální odchylka je v posunu o dvě až tři místa (Hl. m. Praha). O jedno až dvě místa jsou pak posunuty Plzeňský kraj a Moravskoslezský kraj.

Tab. 2: Výsledky hodnocení v druhém stupni agregace

	Bordův přístup	Condorcetův přístup	Copelandova metoda
Hl. m. Praha	3-4	2-3	1
Středočeský kraj	6-7	6-7	6-7
Jihočeský kraj	6-7	6-7	6-7
Plzeňský kraj	2	2-3	3-4
Karlovarský kraj	9	9-10	9-10
Ústecký kraj	14	13	13
Liberecký kraj	3-4	4	3-4
Královéhradecký kraj	5	5	5
Pardubický kraj	12-13	11-12	11-12
Kraj Vysočina	10	9-10	9-10
Jihomoravský kraj	1	1	2
Olomoucký kraj	11	11-12	11-12
Zlínský kraj	8	8	8
Moravskoslezský kraj	12-13	14	14

Pro tři výše zmíněné kraje je charakteristické, že dosáhly velmi podobných výsledků (velmi dobrého umístění či naopak velmi špatného umístěni) ve dvou pilířích a zcela opačného výsledku v pilíři třetím. Kraj Hl. m. Praha již byl v tomto ohledu popsán výše. Moravskoslezský kraj dosáhl nadprůměrných výsledků v ekonomickém pilíři, jeho skóre ve zbývajících dvou oblastech je však jedno z nejhorších. Plzeňský kraj pak dosáhl dobrých výsledků v pilíři sociálním a environmentálním. V ekonomickém pilíři se však nalézal slabě pod průměrem. Tyto výsledky ho posunuly na přední příčky celého hodnocení spolu s Hl. m. Praha a Jihomoravským krajem.

Na předních místech se umístili kraje s největšími městy České republiky - Hl. m. Praha, Plzeňský a Jihomoravský kraj. Ve všech případech se jedná a důležitá centra ekonomiky, vědy a výzkumu a vzdělávání. V případě Prahy a v některých oblastech i v Jihomoravském kraje jsou zde soustředěny důležité státní instituce.

Na druhém konci hodnocení jsou pak především kraje Ústecký a Moravskoslezský. Oba kraje již dlouhodobě patří mezi problémové regiony České republiky. V oblasti ekonomické a sociální to je způsobeno výrazným útlumem těžkého a těžebního průmyslu po roce 1990 a s tím spojeným zhoršením sociální situace obyvatel. Z hlediska životního prostředí pak jsou v obou krajích rozsáhlá území, velmi silně poškozena právě těžbou uhlí a na ni navázanou výrobou elektrické energie a v případě Moravskoslezského kraje též hutnictvím a ocelářstvím.

5. Závěr

V příspěvku bylo využito metod vícekriteriálního rozhodování k vytvoření pořadí regionů České republiky na úrovni NUTS 3 v oblasti udržitelného rozvoje. Metody zde používané se liší svým přístupem ke kompenzaci agregovaných ukazatelů (v tomto případě pilířů udržitelného rozvoje). Výsledky ukazují pouze na malé rozdíly mezi uvedenými přístupy, což je způsobeno i malým počtem agregovaných proměnných. Nekompenzovatelnost Condorcetova přístupu způsobuje rozdílné výsledky ve srovnání s Bordový postupem u krajů, které se vyznačují vysoce nadprůměrným či naopak vysoce podprůměrným umístěním ve dvou pilířích a zcela opačným výsledkem u pilíře třetího. Došly jsme k závěrům, že vytvořená pořadí se příliš neliší a lze na jejich základě usuzovat na vedoucí či naopak zaostávající regiony. Výsledky také ukazují na důležitost ekonomické prosperity, která v případě České republiky jde často ruku v ruce se sociální stabilitou. Vzhledem ke stejným vahám, které byly přirazeny všem třem pilířům, je tak následkem propojení sociální a ekonomické oblasti mírné potlačení vlivu environmentálního pilíře. Tento fakt lze spatřovat jak na pozicích vedoucích regionů (např. Hl. m. Praha), tak ale i na zcela opačném konci žebříčku v Ústeckém kraji. Výrazněji jsou tyto výsledky viditelné u přístupů, které nevedou ke kompenzaci pilířů. Jak již bylo řečeno výše, tři nejlépe hodnocené regiony v obou metodách vynikají po ekonomické stránce, která sebou přináší i vysoké skóre v sociálním pilíři. Vyjma Hl. m. Prahy je ale nezanedbatelné i jejich skoré v pilíři environmentálním. Lze tedy usuzovat na správnost jejich umístění při porovnání. Otázkou do dalších obdobných prací v oblasti udržitelného rozvoje je možnost rozdělení regionů dle existence velkých sídel v regionu. Podle zde prezentovaných výsledků, se jedná o významné hledisko ovlivňující konečné pořadí.

Literatura

ČESKÝ STATISTICKÝ ÚŘAD. (2010). *Vybrané oblasti udržitelného rozvoje v krajích České republiky 2010*. Praha: Český statistický úřad

FISCHER J., PETKOVOVÁ L., HELMAN K., KRAMULOVÁ J., ZEMAN J. (2013). Sustainable development indicators at the regional level in the Czech Republic. *Statistika*, r. 50 č. 1

MOULIN, H. 1988. *Axioms of cooperative decision making*, Cambridge University Press.

SALTELLI, A. 2012. Composite Indicators: An introduction. Paper presented at the 10th JRC Annual Seminar on Composite Indicators.

TRUCHON, M. 1995. Voting games and acyclic collective choice rules. In: *Mathematical Social Sciences*, č. 29, s. 165-179.

YOUNG, P. 1995. Optimal voting rules. In: *The Journal of Economic Perspectives*, č. 9, s. 51-64.

VANSNICK, J.-C. 1990. Measurement theory and decision aid. Readings in multiple criteria decision aid. Springer.

Adresa autorov:

Ludmila Petkovová, Ing.
Katedra ekonomické statistiky
Fakulta informatiky a statistiky VŠE v Praze
Nám. W. Churchilla 4, 130 00 Praha 3
Ludmila.Petkovova@vse.cz

Lenka Hudrlíková, Ing.
Katedra ekonomické statistiky
Fakulta informatiky a statistiky VŠE v Praze
Nám. W. Churchilla 4, 130 00 Praha 3
Lenka.Hudrlikova@vse.cz

Tento příspěvek vznikl v rámci projektu Vysoké školy ekonomické v Praze č. 11/2012 Konstrukce a verifikace indikátorů udržitelného rozvoje ČR a jejích regionů.

Shluky zemí Evropské unie podle struktury státního rozpočtu
Clusters of European Union Countries by government budget structure

Tomáš Pivoňka, Tomáš Löster

Abstract: In this paper we are consider with government budged and its structure. We use the method of cluster analysis to investigate the impact of current economic crisis on the cluster composition and the impact of crisis on chosen variables. As an object of clustering we use the states of European Union except of Croatia. The states are divided into four groups with usage of twelve budget characteristics. The analysis compares the situation in 2000, 2006 and 2012. Nordic countries differ from the rest of European Union. These countries have higher social benefits and compensation of employees and higher taxes on the other hand. In 2012, we can see the group of countries which were connected with some kind of budget or financial issues in recent time. This cluster consists of Greece, Portugal, Ireland, Cyprus and other. The values of budget balance deteriorated during recent economic crisis.

Abstrakt: V tomto příspěvku se věnujeme problematice státního rozpočtu a jeho struktuře. Pomocí shlukové analýzy potom zkoumáme vliv krize na složení jednotlivých shluků a také vliv krize na vývoj jednotlivých ukazatelů. Jako objekty shlukování jsou zde použity státy Evropské unie s výjimkou Chorvatska, kde nebyla data dostupná. Země jsou rozděleny do shluků podle dvanácti položek státního rozpočtu, kde dvanáctou položkou je pak saldo rozpočtu. Severské státy se svou strukturou rozpočtu liší od zbytku Evropy ve všech sledovaných letech. V těch to zemích je štědrá sociální politika, vysoké daně a lepší hodnoty salda rozpočtu. V roce 2012 byl vytvořen shluk států, které byly v nedávné době ve velkých problémech, jako je Řecko, Španělsko, Portugalsko, Kypr a další. Obecně za všechny země Evropské unie došlo v době krize k prohloubení deficitu rozpočtu.

Key words: Budget balance, Economic crisis, cluster analysis

Kľúčové slová: Státní rozpočet, Ekonomická krize, shluková analýza

JEL classification: C38, G31

1. Úvod

Státní rozpočet je v poslední hodně diskutované téma vzhledem k problémům, které měly některé státy Evropské unie. Státní rozpočet můžeme rozdělit na dvě agregované položky. Jedná se o příjmy a výdaje. Příjmy jsou více svázané s vývojem výstupu ekonomiky. Jelikož se jedná především o daně, v době ekonomické recese dochází k nižším výběrům daní a státní rozpočet je tak více ohrožen deficitem. Což je případ současného stavu nejen evropských ekonomik. Obecně můžeme říci, že stát může financovat své výdaje pomocí výběru daní, tiskem peněz a financováním na dluh, tedy půjčkami.

Cílem tohoto příspěvku je posoudit podobnost zemí jednotlivých zemí podle struktury státního rozpočtu. Vybrali jsme proto 12 proměnných a aplikovali jsme metodu shlukové analýzy, která umožní vytvořit skupiny zemí, které jsou si navzájem podobné. Strukturu těchto shluků jsme porovnávali v letech 2000, 2006 a 2012. První dvě období se týkají situace před současnou ekonomickou krizí a rok 2012 by pak popisovat situaci během krize. Dalším cílem článku je tedy posoudit dopad krize jednak na strukturu shluků a jednak na vývoj jednotlivých charakteristik.

Jako objekty analýzy byly použity země Evropské unie s výjimkou Chorvatska. U Chorvatska nebyla dostupná starší data a nebylo by tedy možno porovnávat situace v jednotlivých letech. Data jsou čerpána z databáze Eurostat a jsou vždy vztažena k HDP, aby bylo možno mezinárodní srovnání.

Článek je rozložen následovně. V první části se zabýváme stručnou charakteristikou problematiky státního rozpočtu. Následuje obecný popis metody shlukové analýzy. V dalších částech pak popisujeme vybrané proměnné a prezentujeme vlastní výzkum. Poslední část je závěr obsahující shrnutí výsledků.

2. Problematika státního rozpočtu

Problematika státního rozpočtu se týká fungování a financování vládních aktivit. Na jedné straně to mohou být vládní nákupy statků služeb, dále potom sociální politika vlády (různé transfery ekonomickým subjektům) a na druhé straně potom výběr daní. Celkové vládní výdaje musí být z něčeho financovány. Existují obecně tři typy financování výdajů vlády a to z daní, na dluh nebo tiskem peněz. Všechny tyto způsoby jsou do jisté míry formou zdanění. Dluh se musí někdy v budoucnu splatit i s úroky. Tisk peněz, tedy půjčování si od centrální banky, v sobě nese riziko zvýšení inflace, což se projeví v poklesu kupní síly peněz.

Pokud bude stát financovat svoje výdaje pomocí půjček, a tyto půjčky budou mít dlouhodobý charakter, jedná se o mezi generační transfer, kdy dluh vytvořený v současnosti bude muset být splacen budoucí generací, tedy zdaněním našich potomků. Zde se dotýkáme otázky, zda má stát na to, aby si mohl na trhu půjčovat peníze. Nedávná situace v Řecku ukázala, co se může stát, pokud stát již nedosáhne na peníze z trhu a musí být dotován z půjček od mezinárodních organizací typu IMF (Mezinárodní měnový fond). V této situaci se již nenašel na trh subjekt, který by byl ochoten půjčit Řecku peníze. Toto je otázka rizikovosti, důvěryhodnosti a schopnosti splácet.

Vládní výdaje mohou být rozděleny na dvě větší skupiny a to na nákup statků a služeb a na sociální politiku. Vládní nákupy statků a služeb se týkají veřejných statků. Tyto statky jsou charakteristické tím, že nikoho nemůžeme vyloučit ze spotřeby. Mezi tyto statky bychom mohli řadit třeba obranu státu a právní systém. Tyto statky jsou poskytovány výhradně vládou. Můžeme zde zmínit i další příklady jako zdravotnictví, vzdělání a dopravu, kde kromě vládního sektoru může figurovat i soukromý subjekt. Sociální politika se týká kompenzacemi zaměstnancům, benefity rodinám a například starobním důchodem. K celkovým výdajům se potom musí připočíst ještě úroky z půjček.

Rozpočtové omezení vlády lze formalizovat do podoby, viz (WICKENS, 2011)

$$P_t g_t + P_t h_t + B_t^G = P_t^B B_{t+1}^G + \Delta M_{t+1} + P_t T_t, \qquad (1)$$

Kde g_t, h_t jsou vládní nákupy a transfery, T_t jsou celkové daně, M_{t+1} jsou peníze v oběhu nezatížené úrokem (vydávané centrální bankou) a B_t^G jsou potom vládní dluhopisy v čase t. P_t je cenová hladina v čase t P_t^B je cena vládního dluhopisu vyjádřená přes diskont R_t jako

$$P_t^B = {}^1/_{1+R_t'} \qquad (2)$$

Rovnici rozpočtového omezení můžeme interpretovat tak, že vláda financuje své výdaje z daní a penězi od centrální banky. Rozdíl, většinou tedy záporný, musí vyrovnat půjčkami, tedy vládními dluhopisy.

3. Metodika

Pro analýzu používáme metodu shlukové analýzy, která umožňuje seskupovat objekty, v tomto případě jsou to země EU do skupin. Objekty uvnitř skupiny vykazují homogenitu,

tedy jsou si navzájem podobné. Objekty v různých shlucích se naopak významněji liší. To nám umožňuje posoudit blízkost jednotlivých zemí EU 27 podle zvolených kritérií.

Na základě zkušeností s tímto typem úlohy, viz (PIVOŇKA, LÖSTER, 2013) jsme se rozhodli pro rozdělení zemí do čtyř shluků. Byla pro to použita Wardova metoda ve spojení se čtvercem Euklidovy vzdálenosti Více o postupech používaných při shlukové analýze lze nalézt například v (ŘEZANKOVÁ, 2009) a (GAN, G., MA, CH., WU, J, 2007).

4. Výběr proměnných

Pro účely shlukové analýzy bylo vybráno celkem 12 proměnných, které se týkají státního rozpočtu. Jedná se o položky jak z příjmové, tak z výdajové strany, které jsou následně doplněny o celkové saldo rozpočtu. Jmenovitě se pak jedná z výdajové stránky

a. Kompenzace zaměstnancům
b. Podpory, dotace
c. Důchody z vlastnictví
d. Úroky splatné
e. Sociální benefity a další sociální transfery
f. Ostatní běžné transfery
g. Kapitálové transfery

Ze strany příjmů se pak jedná o

h. Běžné daně
i. Příspěvky na sociální zabezpečení
j. Přijaté kapitálové transfery
k. Vládní výdaje ve formě tvorby hrubých fixních investic

Všechny zmíněné proměnné jsou vyjádřeny jako podíl k hrubému domácímu produktu z důvodu možného porovnání výsledků mezi sebou. Jako zdroj dat byla využita databáze Eurostat. Analýze obsahuje 27 zemí EU, teda bez Chorvatska jako nejčerstvějšího člena. Chorvatsko bylo z důvodu nedostatku dat vyřazeno z analýzy.

Je známo, že jsou v Evropské Unii různé státy s různým přístupem k sociální politice. Více sociálně zaměřené země jsou například země na jihu Evropy, Francie a Skandinávské země. Na severu Evropu jsou na druhou stranu vyšší daně, tudíž stát si na svou sociální politiku dokáže více vydělat.

Výdaje státního rozpočtu jsou jistě mnohem více rigidní v porovnání s příjmy. Jelikož příjmy jsou tvořeny především daněmi, jsou tedy více závislé na vytvořeném produktu. V době ekonomické recese se tedy vybere méně daní a státní rozpočty jsou tak ohroženy větší mírou deficitu. Můžeme tedy shrnout, že příjmy státního rozpočtu jsou procyklické.

5. Vlastní výzkum

Aplikací metody shlukové analýzy na země EU 27 podle vybraných charakteristik jsme vytvořili 4 shluky zemí EU. Porovnáváme výsledky shlukové analýzy ve třech letech, a sice v roce 2000, 2006 a 2012. Rok 2000 charakterizuje období dlouho před současnou ekonomickou krizí, 2006 pak situaci před vypuknutím krize a rok 2012 pak ukazuje situaci během současné krize. Cílem tedy je posoudit, jestli se současná ekonomická situace podepsala na změnu ve shlucích zemí EU podle kritérií struktury státního rozpočtu.

Tab. 9: Rozdělení zemí do shluků

2000		2006		2012	
1	Denmark	1	Denmark	1	Denmark
	Finland		Finland		Luxembourg
	Sweden		Sweden		Finland
2	Belgium	2	Belgium		Sweden
	Greece		Germany	2	Belgium
	Italy		Italy		CZE
	Hungary		Austria		Germany
3	Germany	3	Greece		France
	France		France		Netherlands
	Austria		Cyprus		Austria
	Slovenia		Hungary		Slovenia
	Slovakia		Malta		Slovakia
4	Bulgaria		Poland	3	Ireland
	CZE		Portugal		Greece
	Estonia		Slovenia		Spain
	Ireland		UK		Italy
	Spain	4	Bulgaria		Cyprus
	Cyprus		CZE		Hungary
	Latvia		Estonia		Malta
	Lithuania		Ireland		Portugal
	Luxembourg		Spain		UK
	Malta		Latvia	4	Bulgaria
	Netherlands		Lithuania		Estonia
	Poland		Luxembourg		Latvia
	Portugal		Netherlands		Lithuania
	Romania		Romania		Poland
	UK		Slovakia		Romania

Zdroj: Vlastní výzkum

Shluky jsou pojmenované čísly a čísla neznamenají pořadí od nejlepšího k nejhoršímu. Shluk číslo jedna má relativně stálou strukturu ve všech třech sledovaných letech. Severské státy jsou tedy oproti ostatním zemím rozlišné z hlediska struktury státního rozpočtu. V roce 2012 se k této skupině přidalo ještě Lucembursko. Další shluky v analýze již takovou v čase stabilitu nemají. V roce 2006 byl největší shluk co do počtu zemí číslo 4. V dalších letech se jednotlivé země začaly mezi shluky přelívat, kdy v roce 2012 vznikly 3 shluky s podobným množstvím zemí. Zajímavé je sledování Řecka a Španělska, které měly asi největší problémy v poslední době. V roce 2012 se do společnosti Řecka a Španělska dostalo i Portugalsko, Itálie, Irsko, Kypr, Maďarsko, Malta a UK. Až na UK se jedná o země, které byly spojovány s dluhovou krizí a označovány za méně bezpečné v regionu Evropy. Charakteristiky jednotlivých shluků uvádí následující tabulka.

Tab. 2: Charakteristiky shluků v jednotlivých letech

	2000				2006				2012			
	1	2	3	4	1	2	3	4	1	2	3	4
KOMPENZACE ZAMĚSTNANCŮM	15,1	10,8	10,5	10,3	15,2	10,0	12,2	9,0	13,9	10,0	11,8	9,3
PODPORY	1,8	1,1	2,1	1,2	1,7	1,8	1,0	1,1	1,8	1,8	0,9	0,6
DŮCHODY Z VLASTNICTVÍ	3,3	6,4	3,2	2,4	1,7	3,6	3,0	1,0	1,2	2,3	3,9	1,5
ÚROKY	3,3	6,4	3,2	2,4	1,7	3,6	3,0	1,0	1,2	2,3	3,9	1,5
BĚŽNÉ DANĚ	24,6	12,7	10,5	9,9	23,3	13,8	10,4	9,1	19,8	10,7	11,6	6,3
PŘÍSPĚVKY ZE SOC. ZABEZPEČENÍ	9,3	13,6	16,4	10,6	8,0	15,4	11,7	10,9	8,9	16,2	10,7	10,0
SOCIÁLNÍ BENEFITY	16,0	14,8	16,7	11,5	15,5	17,1	14,4	10,4	16,6	16,3	16,5	11,7
OSTATNÍ TRANSFERY	2,8	1,5	1,8	1,2	2,8	2,0	2,3	1,8	3,1	2,2	2,0	2,1
PŘIJATÉ KAP. TRANSFERY	0,4	1,1	0,2	0,4	0,4	0,4	1,0	0,6	0,4	0,5	1,4	1,6
VYPLACENÉ KAP. TRANSFERY	0,4	2,0	3,1	1,1	0,4	1,8	1,0	1,3	1,0	1,5	1,9	0,8
THFK	2,3	2,8	2,5	3,1	2,4	1,7	3,4	4,0	3,1	2,4	2,2	4,3
SALDO ROZPOČTU	4,3	-2,0	-3,6	-0,9	3,8	-1,6	-3,8	0,3	-1,9	-3,5	-6,1	-2,1

Zdroj: Vlastní výzkum

V tabulce jsou uvedeny střední hodnoty všech proměnných v jednotlivých shlucích. Hodnoty, které zaslouží větší pozornost, jsou označeny šedou barvou. Nejprve budeme popisovat rozdíly mezi shluky v jednotlivých letech. Shluk číslo jedna (tedy severské státy) se vyznačuje tím, že má vysoké kompenzace zaměstnancům, poměrně vysoké sociální benefity a přesto byl státní rozpočet v přebytku. Je to hlavně díky vysokým daním. Ostatní shluky měly v průměru deficitní státní rozpočet. Shluk číslo 3 měl vysoké příspěvky ze sociálního zabezpečení stejně tak i sociální benefity. Mezi tyto státy patří i například Francie. Shluk číslo 4 s největším počtem zemí se vyznačuje nižší úrovní zdanění.

Pokud se budeme pohybovat dále v čase, pak můžeme zopakovat tvrzení o shluku číslo jedna. Tento shluk má vysoké daně, vysoké kompenzace zaměstnancům a vysoké sociální benefity. V roce 2006 měl státní rozpočet stále v přebytku.

Rok 2012 znázorňuje situaci po vypuknutí hospodářské krize. Pokud budeme porovnávat situaci mezi jednotlivými léty, pak můžeme shrnout, že došlo k poklesu kompenzace zaměstnancům u shluku číslo jedna. I tak má tento shluk nejvyšší hodnoty u tohoto ukazatele. Snížily se i vybrané běžné daně.

Podíváme-li se blíže na shluk číslo 3 v roce 2012, vidíme, že platí nejvyšší úroky, vyplácí větší množství peněz na sociální dávky a rozpočet má nejvíce deficitní. Země v tomto shluku mají také poměrně nízké příspěvky na sociální zabezpečení v porovnání s tím, co vyplácejí.

Ostatní položky vykazují relativně stabilní průběh v jednotlivých letech a mají i podobné hodnoty v porovnání s ostatními shluky.

6. Závěr

Prezentovaný článek se zabýval problematikou státního rozpočtu a jeho struktury. Pro analýzu byla použita metoda shlukové analýzy, kdy jsme vytvářeli shluky zemí Evropské unie, s výjimkou Chorvatska, které jsou si navzájem podobné na základě vybraných položek státního rozpočtu. Bylo vybráno 12 rozpočtových položek, kdy poslední bylo saldo rozpočtu. Porovnávali jsme zde shluky a hodnoty proměnných ve třech letech, a sice v roce 2000, 2006 a 2012. Rok 2012 je reprezentantem situace v období za současné ekonomické krize. Jako cíl práce jsme si stanovali porovnat shluky zemí a hodnoty proměnných v jednotlivých letech a posoudit tak dopad ekonomické krize.

Zdrojem dat byla databáze Eurostat, kdy každá veličina byla vztažena k HDP v odpovídajícím roce.

Severské státy se ukázaly být odlišné od zbytku Evropy v každém roce pozorování. Ty to státy mají poměrně vysoké kompenzace zaměstnancům, vysoké sociální benefity a na druhé straně pak i vysoké daně. Celkem pak v letech 2000 a 2006 měly kladné saldo rozpočtu. V roce 2012 se pak saldo rozpočtu dostalo do deficitu, stále je však toto číslo nejvyšší v porovnání s ostatními shluky. Je tedy možné shrnout, že se krize podepsala i na tento relativně stabilní (z hlediska struktury rozpočtu) region.

Země jako Řecko, Španělsko, Kypr, Portugalsko byly spojeny v roce 2012 do jednoho shluku spolu s dalšími zeměmi, které měly problémy v poslední době. Celkem tento shluk čítal 9 zemí. Tyto země v roce 2012 platily nejvyšší úroky, vysoké sociální dávky, nižší příspěvky do sociálního zabezpečení a tedy i vyšší deficit rozpočtu.

Poděkování

Tento článek byl vytvořen s pomocí Interní grantové agentury Vysoké školy ekonomické v Praze č. 6/2013 pod názvem „Hodnocení výsledků shlukové analýzy v ekonomických problémech“

Literatura

Databáze EUROSTAT; http://epp.eurostat.ec.europa.eu

GAN, G., MA, CH., a WU, J. 2007. *Data Clustering Theory, Algorithms, and Applications*. Philadelphia: ASA-SIAM.

PIVOŇKA, T. a LÖSTER, T. 2013. Clustering of EU Countries Before and During Crisis, „The 7th International Days of Statistics and Economics“ Conference Proceedings. Prague, Czech Republic. S 1110 – 1120.

PIVOŇKA, T. a LÖSTER, T. 2013. The structure of labor market in the european union countries, In: 3rd International Scientific Conference “Whither Our Economies– 2013” Conference Proceedings. Vilnius. Lithuania. s. 111 – 119.

ŘEZANKOVÁ, H., HÚSEK, D., a SNÁŠEL, V.2009. *Shluková analýza dat*, Prague: Professional Publishing.

WICKENS, M.2011. *Macroeconomic Theory*, Princeton University Press

Adresa autorů

Tomáš Pivoňka, Ing.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 97 Praha 3
Pivonkat@gmail.com

Tomáš Löster , Ing, PhD.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4, 130 97 Praha 3
tomas.loster@vse.cz

Porovnanie inovačnej výkonnosti SR s kľúčovými krajinami EÚ a V4 v obdobiach 2008-2012
Comparison of innovation Performance of SR with key EU Countries and V4 in time periods 2008-2012

Milan Potančok[1]

Abstract: The paper deals with defining the European Innovation Scoreboard. Afterwards evaluates Slovak republic according to the Summary Innovation Index time series 2008-2012. The article also comparison of the performance scores per dimension between SR and countries V4.

Abstrakt: Príspevok sa zaoberá vymedzením Európskeho inovačného rebríčka. Ďalej hodnotí Slovenskú republiku na základe súhrnného inovačného indexu v časových radoch 2008-2012. Článok tiež porovnáva výkonnostné skóre podľa dimenzií medzi SR a krajinami V4.

Key words: European Innovation scoreboards, innovation, Summary Innovation Index

Kľúčové slová: Európsky inovačný rebríček, inovácia, sumárny inovačný index.

JEL classification: O30, B40, C18

1. Úvod

Inovácie predstavujú jeden z najvýznamnejších nástrojov ekonomického rastu. Ak chce Slovensko napredovať potrebuje byť tvorivejšie a inovatívnejšie. Európska politika inovácií sa zameriava na rýchlo rastúce inovatívne firmy prinášajúce nové produkty, resp. riešenia reagujúce na trhové požiadavky s cieľom zvýšenia konkurencieschopnosti subjektov. Investície do vedy a výskumu (VaV) sú kľúčovým prvkom podpory inovatívnych nápadov a následného ekonomického rastu. Z tohto dôvodu je zvýšenie investícií do VaV jedným z cieľov Európy 2020.

Príspevok sa zaoberá definovaním EIS, rozdelením krajín podľa inovačnej výkonnosti a obsahuje rebríček vybraných krajín EÚ-27 podľa sumárneho inovačného indexu v rozmedzí rokov 2008 až 2013 a porovnanie SII a tempa rastu SII vybraných krajín s priemerom EÚ. Príspevok sa taktiež venuje hodnoteniu inovačnej výkonnosti Slovenska a krajín V4 a ich porovnanie s priemerom EÚ.

2. Európsky inovačný rebríček (EIS)

Európsky inovačný rebríček sa zostavuje každoročne od roku 2001 za účelom sledovania a porovnávania relatívnej inovačnej výkonnosti členských krajín Európskej únie prostredníctvom viacerých ukazovateľov. Európsky inovačný rebríček (EIS – European Innovation scoreboards) bol ako nástroj vyvinutý na základe iniciatívy Európskej komisie v rámci Lisabonskej stratégie. Európsky inovačný rebríček členských krajín EÚ sa zostavuje pomocou sumárneho inovačného indexu (SII - Summary Innovation Index). SII sa využíva na hodnotenie celkovej národnej inovačnej výkonnosti a vypočíta sa na základe najnovších dostupných štatistík z Eurostatu a iných medzinárodne uznávaných zdrojov v čase analýzy.

Európsky inovačný rebríček poskytuje hodnotenie inovačnej výkonnosti členských štátov EÚ, ich relatívne silné a slabé stránky v oblasti výskumu a inovácií. Sleduje inovačné trendy členských štátov celej Európskej únie. Na meranie inovačnej výkonnosti členských štátov EÚ

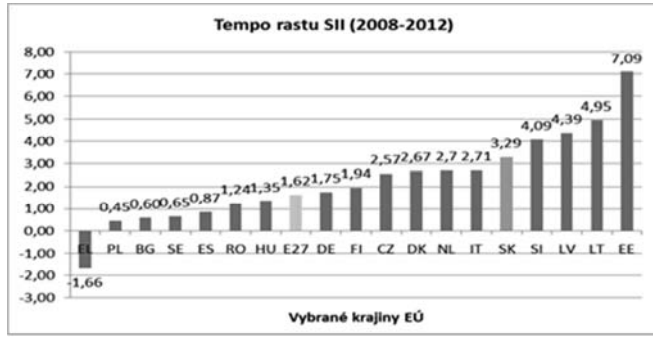
[1] Príspevok bol spracovaný v rámci riešenia úlohy VEGA č. 1/1164/12 „Možnosti uplatnenia informačných a komunikačných technológií na zvyšovanie efektívnosti medzinárodnej spolupráce malých a stredných podnikov SR v oblasti inovácií“.

sa používajú 3 základné typy ukazovateľov (vstupy – firemné aktivity – výstupy) s 8 inovačnými rozmermi zachytávajúce celkom 25 rôznych indikátorov. Vstupy zachytávajú hlavné faktory inovačnej výkonnosti, ktoré nepatria podniku (externé vplyvy) a pokrýva 3 inovačné dimenzie: ľudské zdroje (Human resources – 3 indikátory), oblasť výskumu (Research systems – 3 indikátory) a financovanie a podporu (Finance and support – 2 indikátory). Firemné aktivity monitorujú inovačné úsilie firiem, zoskupené do 3 inovačných rozmerov: firemné investície (Firm investments – 2 indikátory), väzby a podnikanie (Linkages & entrepreneurship – 3 indikátory) a duševný majetok (Intellectual assets – 4 indikátory). Výstupy pokrývajú vplyv inovačných aktivít firiem v oblasti inovácií s dvomi rozmermi: zlepšovateľov (Innovators – 3 indikátory) a ekonomické efekty (Economic effects – 5 indikátorov).

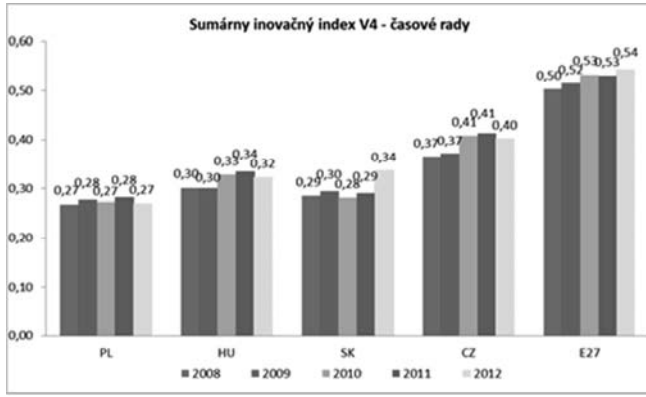
Krajiny zahrnuté do EIS 2013 sú na základe ich inovačnej výkonnosti (SII) rozdelené do štyroch skupín. Inovační lídri: Dánsko, Fínsko, Nemecko, Švédsko, ktorých výkonnosť je 20% alebo viac nad priemerom EÚ-27. Tempo rastu SII (2008-2012) inovačných lídrov je 1,8%. Dánsko je v tejto skupine líder tempa rastu SII s hodnotou 2,7%, potom nasleduje Fínsko a Nemecko s miernym tempom rastu SII a pomalé tempo rastu SII má Švédsko (0,6%). Švédsko je však stále najinovatívnejšou krajinou EÚ predovšetkým z dôvodu silných inovačných vstupov. Inovační nasledovníci sú: Rakúsko, Belgicko, Cyprus, Estónsko, Francúzsko, Írsko, Luxembursko, Holandsko, Slovinsko a Veľká Británia. Ich výkonnosť je v rozmedzí: 20% nad priemerom EÚ-27 -10% pod priemerom EÚ-27. Tempo rastu SII (2008-2012) inovačných nasledovníkov je 1,9%. Lídrami tempa rastu v tejto skupine sú Estónsko (7,1%) a Slovinsko (4,1%). Medzi miernych inovátorov patrí: Česká republika, Grécko, Maďarsko, Taliansko, Litva, Malta, Portugalsko, Slovensko a Španielsko. Ich výkonnosť je pod priemerom EÚ-27 (t.j. 10% až 50% pod priemerom EÚ-27). Tempo rastu SII (2008-2012) miernych inovátorov je 2,1%. Lídrom tempa rastu SII v tejto skupine je Litva s hodnotou 5,0%. Potom nasleduje Malta a Slovensko s miernym tempom SII rastu 3,3%, ale s druhým najvyšším po Litve. Medzi dobiehajúce krajiny patrí: Bulharsko, Lotyšsko, Poľsko a Rumunsko, ktoré sú významne pod priemerom EU27 (t.j. 50% a viac pod priemerom EÚ-27). Tempo rastu dobiehajúcich krajín SII (2008-2012) je 1,7%. Lídrom tempa rastu SII v tejto skupine je Lotyšsko (4,4%), (zdroj (European Commission, 2013).

V grafoch sú použité tieto skratky krajín EÚ: BE – Belgicko, BG – Bulharsko, CZ – Česká republika, DE – Nemecko, DK – Dánsko, EE – Estónsko, EL – Grécko, ES – Španielsko, FI – Fínsko, HU – Maďarsko, IT – Taliansko, LV – Lotyšsko, LT – Litva, NL – Holandsko, PL – Poľsko, RO – Rumunsko, SE – Švédsko, SK – Slovensko, SI –Slovinsko. MT – Malta, PT – Portugalsko.

Graf 1 znázorňuje vývoj SII v časových radoch 2008-2012 vo vybraných krajinách EÚ v porovnaní s inovačnými lídrami, ktorí sú zoskupený na konci grafu.



Graf 2 zobrazuje tempo rastu SII tých istých vybraných krajín EÚ uvedených v Grafe 1.

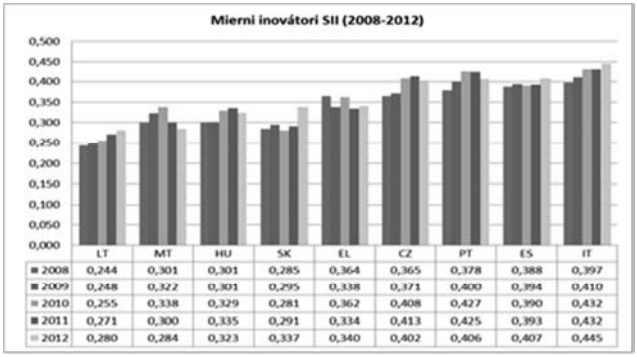


Graf 3 zobrazuje sumárny inovačný index (SII) v krajinách V4. Časové rady zobrazujú vývoj indexu SII medzi rokmi 2008 až 2012.

V rámci krajín V4 najvyššie tempo rastu indexu SII medzi rokmi 2008 až 2012 dosiahla Slovenská republika (3,29%), naopak najnižší rast indexu SII bol pozorovaný v Poľsku (0,45%). Tempo rastu indexu SII E-27 bol 1,62%. Pri pohľade na samotný index SII v roku 2012 najvyššiu hodnotu dosiahla Česká republika SIICZ2012 = 0,4 a najnižšiu Poľsko SIIPL2012 = 0,27. Priemer EU-27 bol 0,53. Z toho vyplýva, že všetky krajiny V4 sa radia do skupiny miernych inovátorov okrem PL. V tejto skupine je na čele Taliansko s hodnotou SIIIT2012 = 0,45. Z krajín východného bloku sú veľmi úspešné krajiny Estónsko s hodnotou indexu SIIEE2012 = 0,5 a Slovinsko s hodnotou indexu SIISI2012 = 0,51. Obe krajiny patria do skupiny inovačných nasledovníkov. Na čele tejto skupiny je Holandsko s hodnotou indexu SIINL2012 = 0,65. Lídrom tempa rastu SII v tejto skupine je však Estónsko = 7,1%. Jeho tempo rastu je dokonca najvyššie z celej EÚ. Za svoj úspech vďačí Estónsko hlavne strane vstupov, t.j. vo výdavkoch na vedu a výskum (VaV) vo verejnom sektore a vo výstupe, predovšetkým v oblasti dimenzie: zlepšovateľov. Švédsko bolo do roku 2006 na prvom mieste, v roku 2007 ho predbehlo Švajčiarsko, ale z krajín EÚ-27 je dlhodobo (2008 až 2012) prvom mieste. Švédsko má hodnotou indexu SIISE2012 = 0,75.

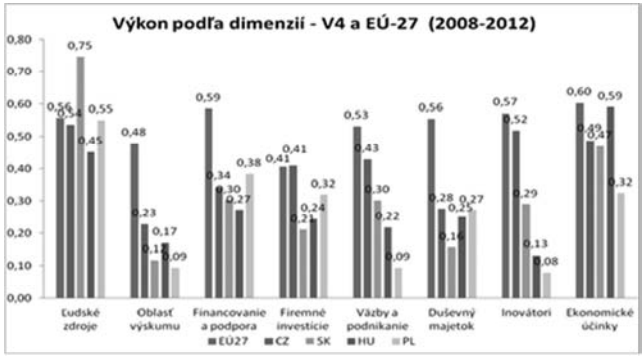
3. Hodnotenie inovačnej výkonnosti SR v rámci EÚ a V4

Pri hodnotení inovačnej výkonnosti Slovenska vychádza tento príspevok z informácií uverejnených v Innovation Union Scoreboard 2013 (European Commission, 2013).



	LT	MT	HU	SK	EL	CZ	PT	ES	IT
2008	0,244	0,301	0,301	0,285	0,364	0,365	0,378	0,388	0,397
2009	0,248	0,322	0,301	0,295	0,338	0,371	0,400	0,394	0,410
2010	0,255	0,338	0,329	0,281	0,362	0,408	0,427	0,390	0,432
2011	0,271	0,300	0,335	0,291	0,334	0,413	0,425	0,393	0,432
2012	0,280	0,284	0,323	0,337	0,340	0,402	0,406	0,407	0,445

Graf 4 zobrazuje sumárny inovačný index (SII) miernych inovátorov krajín EÚ. Časové rady zobrazujú vývoj indexu SII medzi rokmi 2008 až 2012.



Graf 5 zobrazuje inovačnú výkonnosť krajín V4 rozloženú do ôsmych inovačných dimenzií.

Vstupmi sú: *ľudské zdroje* (napr. dostupnosť vysokokvalifikovanej a vzdelanej pracovnej sily so stredoškolským a vysokoškolským stupňom vzdelania vrátane doktorandov), *oblasť výskumu* (napr. medzinárodná konkurencieschopnosť vedeckej základne a najviac citované publikácie) a *financovanie a podpora* (napr. výdavky na VaV a investície do rizikového kapitálu). Do firemných aktivít sú zahrnuté: *firemné investície* (napr. výdavky na VaV v podnikateľskom sektore), *väzby a podnikanie* (napr. MSP inovujúce doma ako aj ich vzájomná spolupráca) a duševný majetok (napr. aplikácie PCT patentov). Na strane výstupov sú firmy: *inovátori* (napr. inovácie produktu alebo procesu, marketingové inovácie) a *ekonomické efekty* (napr. zamestnanosť v činnostiach náročných na znalosti alebo prínos do obchodnej bilancie vývozom MHT produktov).

V skupine miernych inovátorov päťročný rast výrazne poklesol v prípade Grécka, Malty a Portugalska. Iba Českej republike, Litve a Slovensku sa podarilo zvýšiť svoje tempo rastu SII

za obdobie 2008-2012 v porovnaní s 2006-2010. Slovensko a Litva, oba mierni inovátori, sú svojím výkonom v oblasti *ľudských zdrojov* nad priemerom v tejto skupine. Slovensko v oblasti ľudských zdrojov je výrazne na čele ako v skupine miernych inovátorov, tak i v skupine krajín V4. Tento úspech Slovenska možno pripísať hlavne jeho rastom v oblasti nových absolventov doktorandského štúdia a mládeže so stredným vzdelaním. SR zaznamenala dokonca najvyšší nárast nových absolventov doktorandského štúdia zo všetkých členských štátov EÚ (rast tohto indikátora v SR bol zaznamenaný vo výške 22,0%, v ČR iba 2,0%). Rozmer ľudských zdrojov ako najdôležitejší stimulátor inovácie teda výrazne prispel k výške hodnoty sumárneho inovačného indexu Slovenska v roku 2012 (SIISK2012 = 0,34). Na dobrú výkonnosť SR v oblasti *financií a podpory* sa podpísal výrazne indikátor výdavkov na VaV vo verejnom sektore, ktorého priemerný rast bol 11,3%.

Relatívne nedostatky Slovenska sú v *oblasti výskumu* a *duševného vlastníctva*. V dimenzii duševného vlastníctva bol zaznamenaný vysoký rast žiadostí o patenty PCT (Patent Cooperation Treaty - zmluva o patentovej spolupráci) v oblasti spoločenských výziev a ochrannej známky spoločenstva. Pod výrazne nízku hodnotu Slovenska v inovačnej dimenzii duševného vlastníctva (najnižšia z krajín V4: 0,16) sa podpísal indikátor aplikácie PCT patentov, ktorého priemerný rast bol negatívny -4,7%.

Silný pokles bol zaznamenaný aj v oblasti *firemných investícií* z dôvodu výdavkov firiem na inovácie, kde priemerný rast tohto indikátora bol pre SR negatívny -19,2%. Takisto indikátor výdavkov na VaV v podnikateľskom sektore zaznamenal priemerný rast iba 8,6%. V oblasti firemných investícií má SR najnižšiu hodnotu inovačnej výkonnosti (0,21) zo všetkých krajín V4, pričom ČR má najvyššiu hodnotu inovačnej výkonnosti (0,41) v tejto oblasti z krajín V4 a rovnakú ako EÚ-27.

V oblasti *inovátorov* SR zaostáva za priemerom EÚ-27 aj ČR. V inováciách produktu alebo procesu MSP dosiahla SR priemerný rast 4,9% a v marketingových alebo organizačných inováciách MSP priemerný rast 6,1%. Pod nepriazeň v inovačnom výkone v oblasti *ekonomických efektoch* sa podpísal indikátor príjmov z licencií a patentov zo zahraničia (priemerný rast -38,4%). Prínos do obchodnej bilancie SR exportom MHT produktov (MHT: stredné a high-tech produkty) bol nízky – priemerný rast tohto indikátora bol iba 0,5%. Z celkového vývozu služieb priemerný rast indikátora vývozu služieb založeného na vedomostiach sa rovnal hodnote -0,2%. V inovačná výkonnosť ekonomických efektov SR zaostáva vo výkone vo V4 za ČR i Maďarskom.

Slovensko je jedným z miernych inovátorov s podpriemernou výkonnosťou. Priemerná výkonnosť indexu SII E27 v rozmedzí rokov 2008 až 2012 mala hodnotu 0,53, pričom Slovensko v rozmedzí tých istých rokov dosiahlo v tomto indexe priemernú hodnotu iba 0,30.

6. Záver

Aj keď Slovensko na základe uvedených výsledkov v oblasti hodnotenia inovácií nepatrí v rámci Európy medzi krajiny najhoršie, je žiaduce oveľa viac ako doposiaľ, venovať pozornosť kreativite a inováciám. Rozhodujúcimi aktérmi nových inovácií a kreativity sú vláda, univerzity, podniky, média a mimovládne organizácie. Po stagnácií, ktorá na Slovensku ale aj v Európe v posledných rokoch pretrváva, si budúcnosť vyžaduje oveľa väčšiu spoluprácu uvedených aktérov a uvoľniť ešte viac finančných prostriedkov od štátu do vzdelávania, vedy a výskumu a podpory malého a stredného podnikania. Takisto, ale podstatne viac finančných prostriedkov, by mal podnikateľský sektor nalievať do sféry firemných inovácií a firemných výdavkov na VaV malých a stredných podnikov.

Literatúra

EUROPEAN COMMISSION. 2013. Innovation Union Scoreboard 2013. Dostupné na: http://ec. europa.eu/ enterprise/ policies/ innovation/files/ ius-2013_en.pdf.

LUČKANIČOVÁ, M. – MALIKOVÁ, Z. 2011. Porovnanie krajín V4 podľa vybraných inovačných indexov. Dostupné na: *http: //www3.ekf.tuke.sk /mladivedci2011/ herlany_zbornik2011 /malikova_zuzana.pdf.*

EUROPEAN COMMISSION. 2009. Pro Inno Europe Paper N. 15. European Innovation Scoreboard (EIS) 2009. Dostupné na: *http://www.proinno-europe.eu/sites/default/files/page/10/07/I981-DG%20ENTR- Report%20EIS.pdf.*

SABADKA, D. 2011. Výskumno-vývojový potenciál a inovačná kapacita Európskej únie. Dostupné na: *http://www.sjf.tuke.sk/kpiam/TaIPvPP/2011/index.files/clanky/ Dusan%20Sabadka%20Vyskumno.pdf.*

SPIŠÁKOVÁ, E. – SUHÁNYI, L. 2009. Porovnanie krajín EU podľa sumárneho inovačného indexu. Dostupné na: *http://www.sjf.tuke.sk/transferinovacii/pages/archiv/transfer/14-2009/pdf/147-152.pdf.*

Adresa autora

Milan Potančok, Ing. Mgr. PhD.
ÚM STU – OEMP
Vazovova 5, 812 43 Bratislava
milan.potancok@stuba.sk

Teoretický, metodický a technický prístup k meraniu nerovnosti
Theoretical, methodical and technical approaches to inequality measurement

Ľubica Sipková, Juraj Sipko

Abstract: In this study we discuss theoretical, methodical and technical approaches to income inequality measurement, which are developed from the principles of probability modelling using inverse distribution functions – quantile functions. We compare technical advantages of measurement in the two program systems STATA and DAD.

Abstrakt: V príspevku diskutujeme teoretické, metodické a technické prístupy k meraniu príjmovej nerovnosti pri aplikácii parametrického prístupu merania príjmovej nerovnosti, ktorý vychádza z princípov pravdepodobnostného modelovania inverznými distribučnými funkciami – kvantilovými funkciami. Porovnávame technické možnosti, výhody a nevýhody aplikácií merania nerovnosti v dvoch programových systémoch s názvom STATA a DAD.

Key words: Income Inequality, Gini Indices, Quantile Function, Income Probability Models.

Kľúčové slová: nerovnosť_príjmov, Giniho_indexy, kvantilová_funkcia, modely_príjmov.

JEL classification: J31, C14, C46

1. Úvod

V súčasnosti, v podmienkach globalizujúceho sa sveta sa aj v slovenskej spoločnosti prejavujú podobné tendencie, napr. zväčšovanie rozdielov v príjmoch medzi bohatými a chudobnými, prehlbovanie regionálnych diferencií v príjmoch, alebo postupné posuny príjmových vrstiev v celkovom príjmovom rozdelení. Tieto tvrdenia sú platné podľa meraní s veľkou úrovňou zjednodušení a zovšeobecnení použitím konkrétneho matematicko-štatistického aparátu s využitím rôznych modulov programových systémov pri kvantifikácii štatistických mier na rôznych základoch.

2. Teoretické predpoklady, zjednodušenia a zovšeobecnenia

„Nerovnosť“, vychádzajúca z pojmov „bohatstvo“ a „chudoba“, sú všetko pojmy, ktoré nie sú jednoducho definovateľné, merateľné a porovnateľné. Dôvodom je veľké množstvo hľadísk, ktoré je pri ich vymedzení potrebné zohľadniť. Rôzne prístupy a spoločné črty koncepcií ich charakterizovania je zložité už len načrtnúť, nie ich jednoznačne vymedziť a nájsť ich prienik. Súčasné snahy o meranie úrovne sociálnej nerovnosti v krajinách, so zahrnutím viacerých rovín súčasne vedie k veľmi metodologicky komplikovaným, ťažko merateľným konceptom s primerane náročným metodologickým štatisticko-matematickým, programovým a technickým aparátom. Merateľnosť je len východiskom, ktoré slúži pre potreby porovnávania, ktoré prináša ďalšie abstrakcie a zjednodušenia konceptov.

Stanovenie indikátorov merania sociálnej nerovnosti je závislé od zvolených uhlov pohľadu, ktorých prienik je ešte zrozumiteľný a vyžaduje stanoviť úroveň zjednodušenia meraného konceptu. Terajšie prístupy k posúdeniu úrovne sociálnej nerovnosti sú determinované a obmedzené súčasným pochopením a sociálno-ekonomickým vymedzením pojmu „bohatstvo“. Najčastejšie sa v ekonomickej literatúre stretávame s vyjadrením bohatstva krajiny prostredníctvom ukazovateľa hrubý domáci produkt, alebo národný produkt, prepočítaný na obyvateľa. Problém je v tom, že odzrkadľuje len čiastočne ekonomickú a sociálnu situáciu obyvateľov krajiny. Dá sa povedať, že vyjadruje len istú mieru existujúceho potenciálu krajín, nie však mieru jeho využitia pre široký spoločenský

prospech a už vôbec nie veľkosť a perspektívy rozvojového potenciálu pre celé obyvateľstvo krajiny.

Meranie sociálno-ekonomickej situácie jednotlivcov v krajine vyžaduje veľký stupeň zjednodušenia a zovšeobecnenia. Sociálny blahobyt jednotlivca spoločnosti meraný a porovnávaný len podľa úrovne jeho príjmov je už samo o sebe takmer hraničné akceptovateľným zjednodušením. Posúdenie sociálnej situácie jednotlivca spoločnosti podľa jeho príjmu abstrahuje napr. od odlišností v spotrebe a v prístupe k tovarom[1]. Napr. súčasný systém merania chudoby a nerovnosti príjmov v rámci EÚ, ako aj národné analýzy príjmových rozdelení v rôznych sociálno-ekonomických štruktúrach, sú založené na predpoklade, že všetci jednotlivci v rámci EÚ využívajú príjem na zabezpečenie svojej životnej úrovne rovnakým spôsobom, s rovnakým efektom a za rovnakých podmienok. Určujúci je teda predpoklad, že rozdiely v životnej úrovni jednotlivcov v rámci EÚ môžu byť vyjadrené z ekonomického pohľadu, a to pomocou rozdielnych rozdelení príjmov (tzv. income poverty paradigm).

Napriek veľkému zjednodušeniu prístup v sebe zahŕňa veľké množstvo parciálnych problémov, ktoré je potrebné najskôr primerane vyriešiť a komplikácií, na ktorých spôsob odstránenia dosiaľ nie je primeraný jednotný názor. Sú nimi napr. vystihnutie „ekvivalentnosti“ pomocou vhodných stupníc[2] pri posúdení štruktúry domácností, stanovenie úrovne hraníc ešte „akceptovateľného“ príjmu[3] v spoločnosti, úrovne agregácie, napr. regionálnej[4], v jednotnom stanovení úrovne tejto hranice, ktorá by mala súvisieť so stupňom „relatívnosti“ konštruovaných mier nerovnosti, ako aj s úrovňou „averzie voči nerovnosti“ ktoré sú akceptované v danej spoločnosti. Hranica ohrozenia chudobou, ako 60 % mediánového príjmu v spoločnosti, metodologicky správne neumožňuje vhodne posúdiť vývoj chudoby a porovnať v čase zmeny rozsahu a hĺbky chudoby v krajine v absolútnom vyjadrení. Zmena asymetrie príjmového rozdelenia v čase s posunom mediánu napr. smerom k jej dolnému koncu môže viesť k zlepšujúcim sa hodnotám absolútnych mier chudoby napriek zhoršeniu situácie nielen „chudobných“, ale aj jednotlivcov v „strednej časti“ príjmového rozdelenia.

Taktiež výber jednotky merania nerovnosti v spoločnosti determinuje výsledné hodnotenie. Môže byť ňou rôzne volená jednotka spoločnosti, napr. súkromná domácnosť, jej určitým spôsobom vymedzený člen, napr. podľa jeho demografických atribútov, napr. vek,

[1] Jednotný prístup k posúdeniu sociálnej situácie obyvateľov v rámci Európskej únie zahŕňa ďalšie predpoklady, ktorých primeranosť je otvoreným problémom. Predpokladá sa, že jednotky analýz (jednotlivci alebo domácnosti) majú tiež zhodné preferencie, požiadavky na spoločnosť, očakávania, rovnaké predstavy o blahobyte, ako aj napr. rovnaké zdravie a zodpovedajúce výdavky na jeho udržiavanie.

[2] Ekvivalentná stupnica slúži na zohľadnenie odlišnej „potreby príjmu“ rôznych domácností podľa sociálno-demografickej štruktúry pri porovnateľnej životnej úrovni. Dopadom používanej jednotnej ES v analýzach EU SILC na indikátory chudoby a nerovnosti sa zaoberá aj viacero článkov autorov Bartošovej, Stankovičovej, Želinského, a ďalších (pozri napr. Bartošová, 2013; Želinský, 2012; Bartošová – Bína, 2010).

[3] Odlišný a kontrastný s vyššie uvedeným jednotným prístupom v EÚ, predstavuje metóda definovania relatívnych hraníc chudoby (hraníc rizika chudoby, alebo peňažno-príjmových hraníc chudoby), ako šesťdesiat percent *národného* mediánového ekvivalentného disponibilného príjmu domácnosti. Chudoba na Slovensku, ako členskej krajiny EÚ je meraná a porovnávaná použitím rovnakých mier ako v inom členskom štáte EÚ, ale vyjadrovaná z iného základu (úrovne relatívnej hranice chudoby v SR). Znamená to, že sú vzájomne porovnávané relatívne miery, ktoré sú vyjadrené z národnej úrovne, t. j. relatívne počítané z rozdielnych základov.

[4] Otázkou je, či stanovenie úrovní relatívnych hraníc chudoby podľa hraníc štátov EÚ je primerané a zodpovedá východiskovému chápaniu relatívnej chudoby podľa Amartya Sena. V súčasnosti, v čase neobmedzovanej migrácie v rámci EÚ, rýchleho šírenia informácií a jednoduchej komunikácie, keď hranice štátov v rámci EÚ nie sú takmer žiadnou prekážkou, obmedzením alebo vymedzením pre občanov EÚ, je neopodstanené aby relatívne hranice chudoby boli stanovené pre oblasti určené politicky, podľa hraníc štátov.

pohlavie, dosiahnuté vzdelanie, alebo iná sociálno-spoločenská determinácia jednotlivca a jeho skupinová príslušnosť k spoločnosti, napr. zamestnanecký pomer, príslušnosť k národnosti, náboženstvu, atď.

Predpokladom uplatňovaného štatistického prístupu je naviac zjednodušenie, že príjmy v krajine sú výsledkom stochastických procesov s bázou v primerane plynulom pravdepodobnostnom tvare. Ďalší predpoklad je, že toto rozdelenie možno úplne charakterizovať pomocou vhodného pravdepodobnostného modelu ich rozdelenia s odhadom jeho parametrov na základe výberových údajov. Takýto parametrický prístup k výpočtu štatistických mier nerovnosti predpokladá, že v malom počte parametrov analytického tvaru štatisticko-matematického modelu sú zosumarizované vlastnosti príjmového rozdelenia, ktorého absolútna a relatívna nerovnosť je predmetom štúdia.

Parametrický prístup k definovaniu mier nerovnosti je v doterajšej slovenskej literatúre uvádzaný prevažne s použitím distribučnej funkcie príjmového rozdelenia. Upozorňujeme na novú metodológiu definovania známych mier na báze kvantilovej štatistiky.

3. Kvantilový prístup k meraniu nerovnosti

Typická grafická prezentácia miery odchýlenia sa od „rovnosti“, reprezentovanej diagonálov, je Lorenzova krivka (Lorenz Curve, $L(p)$), z ktorej vychádza aj odvodenie známej miery nerovnosti, ktorou je Giniho koncentračný koeficient, známy aj ako Giniho index (angl. Ginni Coefficient, G). Klasický prístup k ich odvodeniu s východiskom v kumulatívnej distribučnej funkcii ponúka rozsiahla domáca aj zahraničná literatúra, napr. Bartošová (2013; 45). Definícia Lorenzovej krivky je možná aj na základe inverznej distribučnej funkcie príjmového rozdelenia, tzv. kvantilovej funkcie.

Prehľadnú systematizáciu deskriptívnych mier nerovnosti a jej zmien spolu s ich základnou metodikou výpočtu a ich vlastnosťami uvádza napr. Labudová (2010, 2012). Miery nerovnosti začleňuje do dvoch skupín. Do prvej dáva miery v tvare pomeru príjmov prepočítaných na obyvateľa alebo domácnosť dvoch skupín obyvateľov: najbohatšej a najchudobnejšej. Druhú skupinu tvoria miery nerovnosti definované na základe rozdelení príjmov.

V členení podľa Sala-i-Martina sú v prvej z troch skupín „ad-hoc“ idexy (napr. Giniho index, Bonferroniho index, DeVergottiniho index), stredná kvadratická odchýlka príjmu a stredná kvadratická odchýlka logaritmu príjmu. V druhej skupine je Atkinsonov index nerovnosti a iné indexy, ktoré sú založené na „sociálnej funkcii blahobytu“ (angl. Social Welfare Function Indexes). Miery zovšeobecnenej (generalizovanej) entrópie (angl. Generalized Entropy Indices of Inequality, *GE* alebo *GEI*), napr. Theilove indexy nesúladu v tvare *L* alebo *T*, ktoré tiež vyhovujú tzv. axiomatickému princípu nerovnosti a preto patria do tretej skupiny (pozri napr. Labudová, 2012; 108).

Aplikujeme nový prístup k definíciám známych mier nerovnosti pomocou kvantilovej funkcie. *Kvantilová (pravdepodobnostná) funkcia* vyjadruje hodnoty x_p, pre $0 \le p \le 1$ resp. p-kvantil príjmovej premennej X, ako funkciu pravdepodobnosti p, t. j. pravdepodobnosti, že náhodná premenná X nadobúda hodnoty menšie ako hodnota kvantilu x_p. Z matematického zápisu kvantilovej funkcie:

$$F^{-1}(p) = Q(p) = x_p \qquad (1)$$

pre každé reálne p, pre ktoré platí $0 \le p \le 1$ je zrejmé, že výstupom nie je pravdepodobnosť, ako v prípade kumulatívnej distribučnej funkcie $F(Q(p)) = p$, ale je to priamo hodnota príjmu.

Podiel úhrnu príjmov pod *p*-kvantilom z celkových príjmov v spoločnosti udáva Lorentzova krivka v proporcii *p*, teda $L(p)$. Čím menší je tento podiel, tým je rozdelenie príjmov v spoločnosti nerovnomernejšie. Viaceré miery nerovnosti sú odvodené od hodnôt *p*-proporcií, v ktorých $L(p)$ má určitú špecifikovanú vlastnosť. Platí to aj o Giniho indexe. Za predpokladu, že by v spoločnosti všetky jednotky rozdelenia mali rovnaký príjem, kumulácia podielu príjmovo najnižšej *p*-časti populácie na celkovom príjme by bola tiež *p* (kumulatívny percentuálny podiel príjmov časti populácie pod *p*-percentilom by bol 100*p* %). Lorenzova krivka by bola v tvare $L(p) = p$ a podiel populácie by bol rovnakým podielom *p* ako podiel úhrnu ich príjmov. V reálnom príjmovom rozdelení je teda najdôležitejšia informácia obsiahnutá v $L(p)$ o jej odchýlení od tejto rovnomernosti, alebo skutočnej rovnosti v rozdeľovaní, t. j. o vzdialenosti $p - L(p)$ Pri porovnaní s úplnou rovnosťou v rozdelení príjmov, nerovnosť reálneho rozdelenia odníme časť $p - L(p)$ celkových príjmov spoločnosti práve *p*-časti spoločnosti s najnižšími príjmami. Čím väčšia je odňatá časť $p - L(p)$, alebo chýbajúca časť $p - L(p)$príjmov v nízkopríjmovej časti populácie, tým je väčšia nerovnosť rozdelenia príjmov, ktorú definuje Giniho index cez celé rozdelenie takto:

$$\frac{G(p)}{2} = \int_0^1 [p - L(p)]\,\mathrm{d}\,p \qquad (2)$$

Giniho index podľa tejto definície implicitne predpokladá, že všetky rozdiely $p_i - L(p_i)$, pre všetky jednotky rozdelenia $i = 1, 2 \ldots N$ sú rovnako dôležité, t. j. majú rovnakú váhu v rozdelení.

Výpočet deskriptívnej miery $\tilde{L}(p)$ z empirického rozdelenia je jednoduché usporiadaním hodnôt príjmov od najmenšej hodnoty po najväčšiu hodnotu príjmu $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ a dopočítaním proporcií $p_{1:n}, p_{2:n}, \ldots, p_{n:n}$ napríklad podľa vzťahu $p_{i:n} = i/n$ tak, aby platilo $\tilde{Q}(p_i) = x_i$ pre $i = 1, 2 \ldots n$. Diskrétna Lorenzova krivka je potom definovaná takto:

$$L(p_i = \frac{i}{n}) = \frac{1}{n\mu} \sum_{j=1}^{i} Q(p_j) \qquad (3)$$

Ak je to potrebné, hodnoty diskrétnej Lorenzovej krivky medzi vypočítanými hodnotami podľa vzťahu (3) sa dajú získať interpoláciou. S-Giniho index je definovaný takto:

$$I(p;\rho) = \int_0^1 (p - L(p))\kappa(p;\rho)\,\mathrm{d}\,p \qquad (4)$$

kde $\rho \geq 1$ a $0 < p \leq 1$.

V prípade, keď $1 < \rho < 2$, väčšia váha je daná rozdielom $p_i - L(p_i)$ v hornej časti rozdelenia s väčším p a opačne, keď $\rho > 2$, väčšia váha je daná rozdielom v dolnom konci rozdelenia.

Indexy nerovnosti, ktoré majú vlastnosť rozložiteľnosti (dekompozície) na medzi-skupinovú a vnútri-skupinovú nerovnosť sa nazývajú indexami zovšeobecnenej (generalizovanej) entrópie (angl. Generalized entropy indices).

Označme ich všeobecne $I(\theta)$ a pomenujme konkrétne tvary v závislosti od hodnoty parametra θ. Všeobecný vzťah na ich výpočet je definovaný takto:

$$I(\theta)=\begin{cases}\dfrac{1}{\theta(\theta-1)}\displaystyle\int_0^1\left(\frac{Q(p)^{\theta}}{\mu}\right)\mathrm{d}\,p & \text{pre } \theta\neq 0;\ \theta\neq 1\\ \displaystyle\int_0^1\ln\left(\frac{\mu}{Q(p)}\right)\mathrm{d}\,p & \text{pre } \theta=0\\ \displaystyle\int_0^1\frac{Q(p)}{\mu}\ln\left(\frac{Q(p)}{\mu}\right)\mathrm{d}\,p & \text{pre } \theta=1\end{cases} \qquad (5)$$

Jednotlivé známe tvary podľa hodnôt θ sú takéto:

$\theta\leq 1$, pričom $\theta=1-\varepsilon$; $I(\theta)$ je v tomto prípade „ordinálne ekvivalentný“ Atkinsonovým indexom, čo znamená, že poradie rozdelení podľa veľkosti nerovnosti bude v prípade $I(\theta)$ rovnaké ako podľa Atkinsonových indexov,

$\theta=0$; $I(\theta=0)$ je priemernou logaritmickou odchýlkou, jej definícia je takáto:

$$\int_0^1\ln\left(\frac{\mu}{Q(p)}\right)\mathrm{d}\,p=\int_0^1\left(\ln\mu-\ln Q(p)\right)\mathrm{d}\,p \qquad (6)$$

$\theta=1$; $I(\theta=1)$ je Theilov index nerovnosti,

$\theta=2$; $I(\theta=2)$ je polovicou druhej mocniny variačného koeficienta.

Kvantilový spôsob definovania viacerých známych mier nerovnosti pozri napr. Araar – Duclos, (2009).

4. Aplikácia v programovom systéme STATA a DAD

V DAD programových aplikáciách je možné urobiť výpočet prípustných chýb odhadu 95 %-ných intervalov spoľahlivosti jednotlivých mier nerovnosti a ich komponentov v dekompozíciách, napr. Giniho koeficientu podľa zložiek. Systém DAD umožňuje aj váženie hodnôt kalibračnými váhami, ktoré sú v údajovej základni stratifikovaných výberov. Dostupné programové moduly, ktoré sú vhodné na výpočet mier nerovnosti sú pod systémom STATA (napr. ineqdeco, inequal7) často umožňujú do príkazov vložiť aj premennú obsahujúcu kalibračné váhy, ale procedúra descogini v systéme STATA túto možnosť neposkytuje. Výhodou DAD aplikácie je aj výstup s tabuľkou diferencií hodnôt dekompozície v prípade, keď sa porovnávajú súbory za dve obdobia, prípadne za dva rozdielne súbory.

Výsledné hodnoty absolútnych a relatívnych mier nerovnosti sa v obidvoch systémoch nelíšia s presnosťou na 4. desatinné miesta. Niekedy však výstupy aplikácie rovnakých metód analýzy v týchto dvoch systémoch obsahujú rôzne charakteristiky a pre porovnanie hodnôt výstupov je potrebné urobiť vzájomné prepočty. Nie je jednoduché zorientovať sa v rôznych označeniach charakteristík v programových produktoch systému STATA. Aj v tomto ohľade hodnotíme DAD ako vhodnejší produkt, v ktorom celý systém sociálnych analýz má jednotnú symboliku a metodické základy programových aplikácií sú rozpracované v knižnej publikácii uznávanými svetovými autoritami v skúmanej oblasti.

Systém DAD je veľmi komplexný, poskytuje aj rôznorodé názorné grafické výstupy pre mnohé aplikované kvantitatívne analýzy. Poskytuje možnosť porovnávania dvoch rozdelení a správnu aplikáciu váženia pri stratifikácii. Nevýhodou je však získavanie výsledkov len jedného „kroku aplikácie“. Rovnaké kroky je nutné opakovať pri zadávaní rôznych vstupných parametrov procedúr, čo je zdĺhavé. Výstupné zostavy zo systému STATA je možné odoslať do textového súboru. Práca s grafmi a ich úpravy sú však v systéme STATA v porovnaní so systémom DAD náročnejšie, ale poskytujú väčšie možnosti voľby rôznych foriem grafických výstupov.

5. Záver

Dôležitou informáciou je, že nové *definície prevažnej väčšiny štatistických mier chudoby, nerovnosti a blahobytu v teoretickom základe využívajú práve kvantilové funkcie*. Definovanie mier chudoby, nerovnosti a blahobytu pomocou jednotnej symboliky prostredníctvom kvantilových funkcií značne uľahčuje pochopenie ich podstaty, súvislostí ako aj ich grafickú prezentáciu.

Používanie uvedených softvérových produktov pri výpočte mier nerovnosti nie je triviálne ani jednoduché. Treba oceniť pomocné manuály a textové materiály, ktoré tvorcovia pre prácu s modulmi a so systémom DAD pripravili a sprístupnili na internete.

Literatúra

ARAAR, A. - DUCLOS, J. Y. 2009. DAD: A software for poverty and distributive analysis. In *Journal of Economic and Social Measurement*, IOS Press, 2009, Volume 34, Number 2-3 / 2009, 175 s. 189 s. ISSN 0747-9662 (Print), ISSN 1875-8932 (Online). DOI 10.3233/JEM-2009-0315.

BARTOŠOVÁ, J. 2013. *Finanční potenciál domácností: kvantitatívní metody a analýzy*. Kamil Marík Profesional publishing, powerpoint Praha, ČR, 2013, 264 s. ISBN 978-80-7431-107-9.

BARTOŠOVÁ, J. – BÍNA,V. 2010. Influence of the Calibration Weights on Results Obtained from Czech SILC Data. Paris 22.08.2010-27.08.2010. In COMPSTAT 2010. Paris : Physica-Verlag, 2010, s. 753-760. ISBN 978-3-7908-2603-6.

LABUDOVÁ, V. 2010. Miery príjmovej nerovnosti. In *Forum statisticum Slovacum* : vedecký časopis Slovenskej štatistickej a demografickej spoločnosti, Bratislava : Slovenská štatistická a demografická spoločnosť, 2010, Roč. 6, č. 5, s. 127-131. ISSN 1336-7420.

LABUDOVÁ, V. 2012. Miery príjmovej nerovnosti. In Nerovnosť a chudoba v Európskej únii a na Slovensku : zborník statí z vedeckej konferencie, Herľany, 26. septembra 2012. - Košice : Ekonomická fakulta, Technická univerzita v Košiciach, 2012. ISBN 978-80-553-1225-5, s. 107-112.

RAVALLION, M. 2003. The debate on globalization, poverty and inequality: Why measurement matters. In International Affairs, 2003, Vol. 79, no. 3, pp. 739-753.

SEN, A. K. 1973. *On economic inequality*. Clarendon Press, Oxford.

STANKOVIČOVÁ, I. 2010 Regionálne aspekty monetárnej chudoby na Slovensku, In: *Sociálny kapitál, ľudských kapitál a chudoba v regiónoch Slovenska*, Košice : Ekonomická fakulta TU, 67-75. ISBN 978-80-553-0573-8.

ŽELINSKÝ, T. 2012. Citlivosť vybraných mier príjmovej nerovnosti na voľbu ekvivalentnej stupnice. In *Forum statisticum Slovacum* : vedecký časopis Slovenskej štatistickej a demografickej spoločnosti, Bratislava : Slovenská štatistická a demografická spoločnosť, 2012, Roč. 7, č. 8, s. 203-208. ISSN 1336-7420.

Príspevok je riešením vedeckého projektu VEGA 1/0127/11.

Adresa autorov:

Ing. Ľubica Sipková, PhD.
Ekonomická univerzita v Bratislave
Dolnozemská 1, 852 35 Bratislava
lubica.sipkova@euba.sk

Doc. Juraj Sipko, M.B.A., PhD.
Paneurópska vysoká škola
Tematínska 10, 851 05 Bratislava 5
juraj.sipko@uninova.sk

Predikcia predčasného ukončenia poistnej zmluvy pomocou podmienených stromových štruktúr

Lapse prediction using conditional inference tree-based methods

Mária Stachová, Lukáš Sobíšek

Abstract: We estimate lapse prediction models built on real customer's data set that comes from a Czech insurance company. We focused on conditional inference tree-based models and compared these models with classification tree model (CART) and Breiman's random forest.

Abstrakt: Cieľom nášho príspevku je tvorba a porovnanie predikčných modelov. Modely sú budované na reálnych zákazníckych dátach pochádzajúcich z českej poisťovne. Zameriavame sme sa na podmienené stromové štruktúry a porovnávame ich s klasifikačnými stromami a Breimanovým náhodným lesom.

Key words: lapse prediction model, classification, decision tree based model, conditional inference forest

Kľúčové slová: model odhadu predčasného ukončenia zmluvy, klasifikačné modely rozhodovacích stromov, podmienené klasifikačné štruktúry

JEL classification: C19

1. Úvod

Životné poistenie je produkt, ktorý ma za cieľ finančne zabezpečiť poisteného a jeho rodinu v nepriaznivých životných situáciách ako je smrť, invalidita, alebo dlhodobé ochorenie poisteného. Pri tradičnom rizikovom poistení poisťovňa vyplatí poistnú čiastku v prípade, že nastane poistná udalosť (vyššie uvedené situácie). Toto poistenie je pre poisteného (ďalej klienta) relatívne lacné, pretože z poistného sa pokrývajú náklady poisťovne (administratívne, počiatočné, vrátane získateľskej provízie) a riziková zložka, z ktorej tvorí poisťovňa technickú rezervu na pokrytie nastania poistnej udalosti.

Poisťovne ponúkajúce životné poistenie však štandardne ešte aj ponúkajú kapitálové a investičné životné poistenie (ďalej KŽP a IŽP). Tieto produkty rovnako pokrývajú zmienené poistné riziká, ešte však klientovi umožňujú časť poistného alokovať do sporiacej zložky, s ktorou poisťovňa hospodári v prospech klienta. Tieto produkty sú výrazne drahšie ako rizikové poistenie, pretože získateľská provízia je niekoľkonásobne vyššia. Produkty sú konštruované tak, aby v prvých rokoch poistenia bola väčšina poistného spotrebovaná na náklady poisťovne a až v neskorších rokoch sa postupne stále viac z poistného alokuje do sporiacej zložky. Z tohto dôvodu sú tieto typy produktov kontraktované (uzatvárané) na dlhú dobu (niekoľko desiatok rokov), aby klient v preddôchodkovom veku mal na svojom sporiacom účte prostriedky, s ktorými si prilepší na dôchodku.

Pri KŽP aj IŽP dochádza vo významnej miere k predčasnému ukončeniu zmluvy zo strany klienta (ďalej len storno). Rozdelenie stornovania zmlúv má dva vrcholy, okolo prvého mesiaca po uzavretí, alebo po prvom roku. V prvých týždňoch klient často zmluvu stornuje z dôvodu, že pod vplyvom obchodných praktík finančného sprostredkovateľa uzavrel zmluvu, ktorú by sám od seba neuzavrel. Vysoká stornovosť po prvom roku (teda po tom, čo je klientovi zaslaný prvý výročný výpis s jeho zostatkom na sporiacom účte) reflektuje situáciu, kedy klient uzavrel zmluvu s inými očakávaním, ako produkt ponúka. Ďalej v období dvoch až troch rokov od uzavretia zmluvy je zvýšené riziko storna z dôvodu tzv. "kanibalizačného efektu". Poisťovne vyplácajú sprostredkovateľom províziu v prvých dvoch rokoch a vyžadujú

vrátenie provízie iba v prvých dvoch až troch rokoch, potom sprostredkovateľ opäť navštívi klienta s "výhodnou" ponukou uzavretia nového produktu a stornujú súčasnú poistku.

Pre poisťovne je storno nevýhodné, pretože znižuje zisk na produkte a zvyšuje jej reputačné riziko. Z tohto dôvodu je modelovanie stornovosti pre poisťovne zásadné. Miera stornovosti portfólia zmlúv vstupuje do modelov cash-flow, napr. MCEV (Market Consistent Embedded Value), MCVNB (Market Consistent Value of New Business), čo sú dôležité ukazovatele pre akcionárov, manažérov, zákazníkov a regulátora. Miera storna tak isto vstupuje ako z parametrov do cenotvorby produktu.

Pripravili sme predikčné modely pre dva typy neživotného poistenia (zvlášť pre IŽP a zvlášť pre KŽP) stornovosti do dvoch rokov od uzavretia zmluvy pre nemenovanú poisťovňu v Českej republike. Klasifikační model priraďuje každej zmluve číslo z množiny {0,1} v zmysle 0-storno do dvoch rokov, 1-zotrvá dlhšie ako dva roky, pričom istota správneho priradenia závisí od predikčnej schopnosti modelu. Poisťovňa môže znalosť storna využiť pre retenčné aktivity, kedy sa bude snažiť zmluvy s vysokou pravdepodobnosťou storna proaktívne zachrániť. Model ďalej odhaduje významné faktory-prediktory stornovosti. Znalosť týchto prediktorov môže poisťovňa využiť počas underwritingového procesu a rizikové návrhy zmlúv zamietnuť. V neposlednom rade pomáha znalosť faktorov pri tvorbe obchodnej stratégie.

V data miningovej praxi sa využívajú viaceré prístupy pri tvorbe predikčných modelov: lineárne regresné modely, zovšeobecnené lineárne regresné modely, neurónové siete, modely založené na stromových štruktúrach a iné. Každý prístup má svoje výhody a nevýhody.

Výhodou GLM je možnosť do modelu zahrnúť apriornú znalosť o dátach a umožňuje vyberať vhodné premenné, cez algoritmus variable selection, alebo shrinkage. Výhodou neurónových sietí je napr. možnosť ich dodatočného doučenia a naopak nevýhodou je ich náročná interpretovateľnosť (tzv. black box model). My sme sa rozhodli pre stromové štruktúry, pretože sa nám javia interpretačne jednoduchšie a teda pre poisťovňu použiteľnejšie. Predikčné modely sme odhadli metódou podmieneného náhodného stromu a podmieneného náhodného lesa. Porovnali sme ich efektívnosť a náročnosť s ostatnými často používanými stromovými štruktúrami - klasifikačný strom a náhodný les.

V sekcii 2 popisujeme spracovávané dáta. Metodológia, zameraná na podmienený strom a podmienený náhodný les je popísaná v sekcii 3. Výsledky prezentujeme v sekcii 4 a v záverečnej 5. sekcii uvádzame zhrnutie a naznačujeme náš budúci smer výskumu.

2. Dáta

V prvej fáze sme museli dáta pripraviť pre analýzu. Z dátového súboru sme vymazali riadky s chybnými a nezmyselnými hodnotami. Po vyčistení nám k dispozícii zostalo celkom 64 320 zmlúv IŽP, z toho: 23 155 stornovaných a 41 165 aktívnych, a 18 670 zmlúv KŽP, z toho: 5 440 stornovaných a 13 230 aktívnych. Celý súbor sa skladal z 82 990 riadkov (zmlúv) a 22 stĺpcov (premenných).

Dichotomická vysvetľovaná premenná Y klasifikuje zmluvy na aktívne a stornované. Storno sme odhadovali pomocou 21 vysvetľujúcich premenných. Tieto premenné vyjadrujú rôzne charakteristiky zmluvy a klienta. Informácie, ktoré obsahujú môžeme rozdeliť do týchto kategórií: demografické údaje, údaje o výške poistného, platobnej morálke, dĺžke trvania zmluvy, distribučného kanálu, spôsobe platby, príspevku zamestnávateľa, počtu poistných udalostí (akceptovaných a zamietnutých), počtu sprostredkovateľov, prepoistenosti (počet ďalších zmlúv v klientovom portfóliu) a frekvencii platenia. Premenné sú rôzneho typu - kategoriálne s vyšším počtom levelov, dichotomické, numerické.

Keďže sa naša analýza týka citlivých dát poistiteľa, nemôžeme publikovať väčší detail o dátach.

3. Metodológia

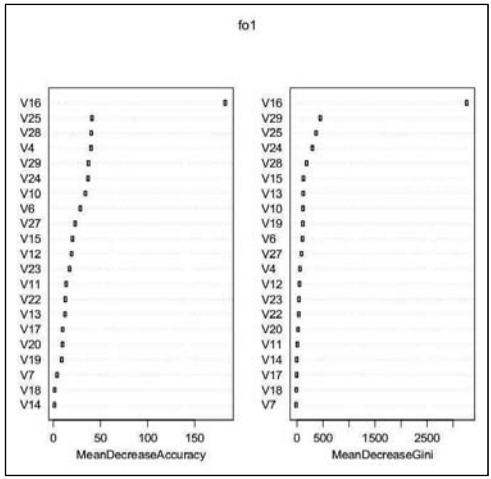
S použití Bernoulliho rozdelenia pravdepodobnosti sme dátovú množinu rozdelili na trénovaciu a testovaciu množinu. Trénovacia množina obsahovala 80 % všetkých prípadov a testovacia 20 %.

Kvôli jednoduchému interpretovaniu výsledkov a pomerne vysokej predikčnej schopnosti sme sa pri budovaní klasifikačných modelov zamerali na tzv. stromové štruktúry a to predovšetkým na podmienený náhodný strom a podmienený náhodný les (Strasser, Weber, 1999; Hothorn, Hornik, Zeilis, 2006). Podmienený náhodný les vychádza z myšlienky Breimanovho náhodného lesa (Breiman, 2001) avšak je tvorený z regresných, resp. klasifikačných stromov, ktoré odhadujú regresný vzťah pomocou rekurzívneho delenia dát v podmienenej inferenčnej štruktúre (Hothorn, Hornik, Zeilis, 2006). V tomto algoritme sa testuje celková nulová hypotéza nezávislosti medzi všetkými vstupnými premennými a závislou premennou. Ak sa hypotéza nezamieta algoritmus skončí, ak sa ale nulová hypotéza zamieta, algoritmus vyberie vstupnú premennú s najsilnejšou asociáciou na závislú premennú. V ďalšom kroku je binárne delenie realizované na vybranej vstupnej premennej. Tieto kroky sa rekurzívne opakujú. Kritériom pre ukončenie algoritmu je najmenšia hladina testu (p-hodnota), ktorej by sme ešte danú hypotézu zamietli. Podmienený náhodný strom sme vytvorili pomocou štatistického systému R (R Core Team, 2013) a jeho funkcie ctree(), ktorá je súčasťou balíčka `party` (Strobl et al., 2008; Strobl et al., 2007). Vylepšením cTree algoritmu je už spomínaný podmienený náhodný les, pričom tento je robustnejší. Podmienený náhodný les sme vybudovali pomocou funkcie cforest(), ktorá je tak isto súčasťou balíčka `party`.

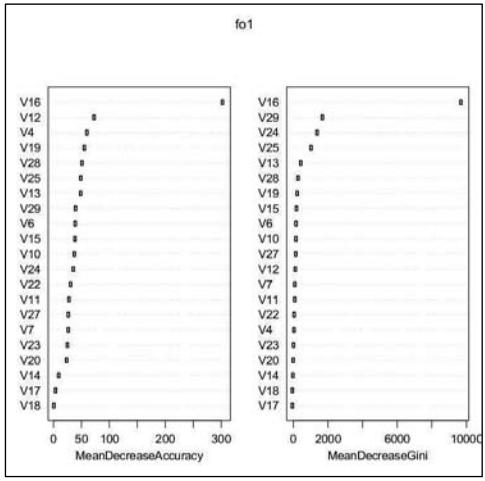
Schopnosť predikcie vybraných podmienených štruktúr sme porovnali s dobre známym modelom klasifikačného stromu a Breimanovym náhodným lesom. Klasifikačný strom bol vytvorený pomocou funkcie rpart() v rovnomennom balíčku `rpart` (Therneau, Atkinson, Ripley, 2013) a náhodný les pomocou funkcie randomForest() z balíčka `randomForest` (Liaw, Wiener, 2002).

4. Výsledky

Kvôli predstave o tom, ktoré premenné zohrávajú úlohu v predikčnom modeli, sme si nechali vykresliť graf dôležitosti premenných: Obr. 1 a Obr. 2. Náhodný les určil za najdôležitejšiu premennú „V16“ pre oba typy poistenia, čo je výška mesačného poistného. V prípade KŽP poistenia sú ďalšími dôležitými prediktormi: počet rokov, počas ktorých je zmluva zadlžená a maximálny dlh. V prípade IŽP je okrem vyššie spomenutých dôležitý aj kraj, z ktorého klient pochádza a príspevok zamestnávateľa.



Obr. 1: dôležité premenné určené náhodným lesom pre predikciu stornovania zmlúv v prípade KŽP



Obr. 2: dôležité premenné určené náhodným lesom pre predikciu stornovania zmlúv v prípade IŽP

Kvalitu jednotlivých modelov popisujú ich tabuľky (matice) úspešnosti, z ktorých môžeme vyčítať schopnosť modelu predpovedať. Riadky v maticiach predstavujú skutočné hodnoty zaradenia objektov a maticové stĺpce predstavujú predikované hodnoty. Diagonálne hodnoty patria správne zaradeným objektom a krížové diagonálne hodnoty patria nesprávne zaradeným objektom. V Tab. 1 a v Tab. 2 sú tabuľky úspešnosti jednotlivých modelov pre oba typy poistných zmlúv. Pre všetky modely bola vypočítaná aj chybovosť a taktiež je uvedená v Tab. 1 a v Tab. 2.

Tab. 1: klasifikačné tabuľky s úspešnosťou predikčných modelov pre KŽP

Klasifikačný strom

	FALSE	TRUE
FALSE	1045	151
TRUE	43	2495
error rate		**5.20%**

Podmienený klasifikačný strom

	FALSE	TRUE
FALSE	1033	163
TRUE	21	2517
error rate		**4.93%**

Náhdoný les

	FALSE	TRUE
FALSE	985	156
TRUE	16	2244
error rate		**5.06%**

Podmienený náhodný les

	FALSE	TRUE
FALSE	1011	185
TRUE	22	2516
error rate		**5.54%**

Tab. 2: klasifikačné tabuľky s úspešnosťou predikčných modelov pre IŽP

Klasifikačný strom

	FALSE	TRUE
FALSE	4596	277
TRUE	35	7956
error rate		**2.43%**

Podmienený klasifikačný strom

	FALSE	TRUE
FALSE	4562	311
TRUE	83	7908
error rate		**3.06%**

Náhdoný les

	FALSE	TRUE
FALSE	8923	503
TRUE	24	14587
error rate		**2.19%**

Podmienený náhodný les

	FALSE	TRUE
FALSE	4494	379
TRUE	26	7965
error rate		**3.15%**

5. Záver

Podmienené stromové štruktúry majú porovnateľnú predikčnú schopnosť ako klasické stromové štruktúry, k tomu podmienené stromy nie sú také citlivé na malé zmeny v dátach a tak odpadá potreba orezávania. Ďalšou nevýhodou klasických klasifikačných stromov je ich zašumenie (bias) (viac v Hothorn, Hornik, Zeilis, 2006), ktoré vzniká pri výbere deliaceho prediktora. Tento šum sa prenáša aj do modelu náhodného lesa, ktorý je kombináciou klasifikačných stromov. Nevýhodou náhodného lesa je aj jeho časová a výpočtová náročnosť. Pri našich výpočtoch sme zaznamenali až 12-krát vyššie nároky na internú pamäť počítača (RAM), čo je pri veľkých dátových štruktúrach zásadný a často pri použití bežných počítačoch ťažko riešiteľný problém.

V ďalšom výskume by sme radi odhadovali predikčné modely aj s využitím iných prístupov a porovnali výsledky s prezentovanými stromovými štruktúrami, z čoho by vzniklo odporučenia pre štatistikov, analytikov, alebo výskumníkov, aký prístup si pre tento typ úlohy vybrať.

Literatúra:

BREIMAN, L. 2001. Random forests. In: *Machine Learning*. Roč. 45, s. 5 – 32.

HOTHORN, T. – HORNIK, K. – ZEILIS, A. 2006. Unbiased recursive partitioning: a conditional inference framework. In: *Journal of Computational and Graphical Statistics*. Roč. 15, č. 3, s. 651 – 674.

LIAW, A. – WIENER, M. 2002. Classification and Regression by randomForest. In: *R News*, Roč. 2, č. 3, s. 18-22.

R CORE TEAM. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL : http://www.R-project.org/.

STRASSER, H. – WEBER, C. 1999. On the Asymptotic Theory of Permutation Statistics. In: *Mathematical Methods of Statistics*. Roč. 8, s. 220-250.

STROBL, C. – BOULESTEIX, A.-L. – KNEIB, T. – AUGUSTIN, T. –ZEILEIS, A. 2008. Conditional Variable Importance for Random Forests. In: *BMC Bioinformatics*. Roč. 9, č. 307, URL : http://www.biomedcentral.com/1471-2105/9/307.

STROBL, C. – BOULESTEIX, A.-L. – ZEILEIS, A. – HOTHORN, T. 2007. Bias in Random Forest Variable Importance Measures, Illustrations, Sources and a Solution. In: *BMC Bioinformatics*, Roč. 8, č. 25, URL: http://www.biomedcentral.com/1471-2105/8/25.

THERNEAU, T. – ATKINSON, B. – RIPLEY, B. 2013. rpart: Recursive Partitioning. R package version 4.1-1, 2013.

Adresa autorov:

Mária Stachová, Mgr., PhD.
EF UMB
Tajovského 10, 974 00 Banská Bystrica
maria.stachova@umb.sk

Lukáš Sobíšek, Ing.
VŠE Praha
nám. W. Churchilla 4, 130 67 Praha 3, ČR
lukas.sobisek@vse.cz

Poďakovanie: Tento príspevok bol vypracovaný za podpory projektu IGA VSE F4/17/2013 a vďaka podpore v rámci operačného programu Vzdelávanie pre projekt: Mobility - podpora vedy, výskumu a vzdelávania na UMB, kód ITMS: 26110230082, spolufinancovaný zo zdrojov Európskeho sociálneho fondu.

Modelovanie rizika v leasingu automobilov
Credit Scoring modelling in automobile leasing

Iveta Stankovičová, Martin Řezáč

Abstract: The article discusses the use of quantitative methods in the risk management and risk prediction. The aim of this work is the use of predictive modelling to determine the risk of clients in the field of automobile leasing. Subsequently, the paper outlines the data mining techniques and their use in the process of credit scoring. Finally, these findings are applied to a selected leasing company operating in the Slovak Republic and the results of predictive modelling are interpreted with regard to the risk management when underwriting new leasing contracts.

Abstrakt: Článok pojednáva o využití kvantitatívnych metód v manažmente rizík a predikovaní rizikovosti klientov. Cieľom práce je využitie prediktívneho modelovania na určenie rizikovosti klientov v oblasti leasingu automobilov. Približuje techniky data miningu a ich využitie v procese kredit skóringu. Poznatky sú aplikované na vybranú leasingovú spoločnosť pôsobiacu v Slovenskej republike a výsledky prediktívneho modelovania sú interpretované vzhľadom na riadenie rizika pri uzatváraní nových leasingových zmlúv.

Key words: automobile leasing, risk management, credit scoring models, scorecard

Kľúčové slová: leasing automobilov, manažment rizika, modely rizika, skórovacia karta

JEL classification: G2, C88

1. Úvod

Každé rozhodnutie vo firme nesie so sebou aj určitú mieru rizika. Riziko vyjadrujeme pomocou pravdepodobnosti. Za účelom identifikácie rizika a minimalizácie jeho dôsledkov bola vytvorená nová vetva manažmentu, manažment rizika. Dnes má každá firma k dispozícii veľké množstvo informácií operatívneho charakteru. Ich správne vyhodnotenie a využitie však prekračuje schopnosti jednotlivca. Boli vyvinuté viaceré metódy a programové nástroje, ktoré pomáhajú efektívne riadiť riziká. Cieľom tohto príspevku je ukázať ako prediktívne modelovanie pomáha pri určovaní rizikovosti klientov v oblasti leasingu automobilov.

2. Leasing automobilov na Slovensku

Leasing je osobitá forma financovania investícií a v súčasnej dobe má stále rastúci význam. Pojem leasing naše právne predpisy nepoznajú a nahradzujú ho spojením nájomný pomer alebo prenájom. S definíciou leasingu sa však môžeme oboznámiť v medzinárodnom účtovnom štandarde IAS 17 Leases (Prenájmy), ktorý nadobudol platnosť 1.1.1984 a ktorý hovorí, že „leasingom sa rozumie zmluva, ktorou prenajímateľ prenáša na nájomcu právo užívať po dohodnutú dobu určitý majetok za jednu alebo sériu platieb“.[1] Štandard rozlišuje dva druhy leasingu: operatívny a finančný leasing. *Operatívny leasing* je v slovenskej legislatíve nájomná zmluva, ktorú spomína zákon o daniach z príjmov a Občiansky zákonník. *Finančný leasing* je čiastočne upravený ako zmluva o nájme veci s právom kúpy prenajatej veci v zákone o daniach z príjmov, kde ho definuje ako nájom majetku s dojednaným právom kúpy prenajatej veci, pri ktorom bez zbytočného odkladu po ukončení doby nájmu prechádza vlastnícke právo k predmetu nájmu z prenajímateľa na nájomcu a náležitosti zmluvy definuje Obchodný zákonník.

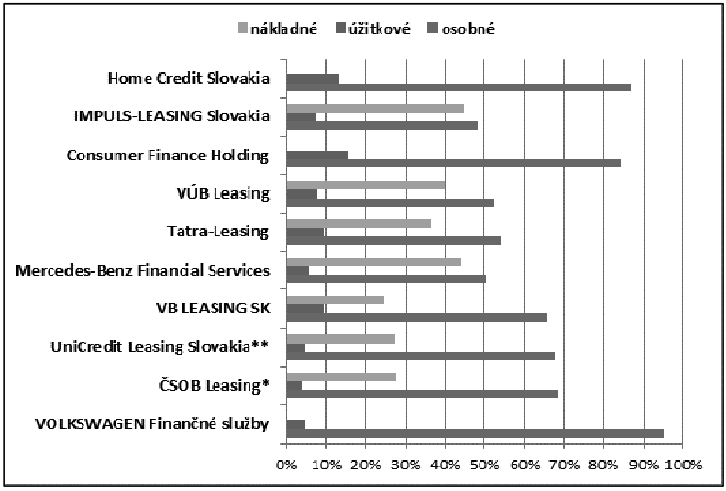
[1] International accounting standard 17 – *Leases*, 2010 [online]. Európska únia : Európska komisia, 24.3.2010 Dostupné na internete : ‹http://ec.europa.eu/internal_market/accounting/docs/consolidated/ ias17_en.pdf›

Rozhodujúcim subjektom umožňujúcim realizáciu leasingových transakcií sú leasingové spoločnosti, ktorých hlavným cieľom je dosiahnutie čo najväčšieho počtu a objemu leasingových prenájmov. Tento cieľ sa snažia dosiahnuť pri minimálnych nákladoch a s optimálnym hospodárskym výsledkom, pričom využívajú rôzne kvantitatívne a kvalitatívne metódy. Keďže ich hlavnou činnosťou je poskytovanie leasingových zmlúv, manažment rizika leasingových spoločností sa venuje hlavne predikcii rizikovosti potenciálnych klientov, kde je úverové riziko definované ako strata spôsobená platobnou neschopnosťou nájomcu.

Asociácia leasingových spoločností SR združuje na základe dobrovoľnosti od roku 1992 cez 40 leasingových spoločností. V Európe jej patrí 0,72% trhový podiel. V roku 2012 na Slovensku lízingový trh stagnoval na rovnakej úrovni ako v roku 2011, a to 1841 mil. €, ale počet nových zmlúv medziročne vzrástol o 10%, na 73369 zmlúv. Hlavným rozdielom oproti roku 2011 bola zmena štruktúry využívaných finančných produktov, a to nárast financovania spotrebiteľov o 23%, čiže úverov, splátkového predaja a operatívneho leasingu na úkor finančného leasingu.

Najvýznamnejšou komoditou leasingu na Slovensku sú osobné automobily. Objem leasingu áut dosiahol v roku 2012 až 833 mil. €, čo zodpovedá 9% rastu oproti roku 2011. Tento nárast bol spôsobený najmä segmentom ojazdených a nových áut pre spotrebiteľov (fyzické osoby), a to o viac ako 25%, kde z 10 áut 4 financujú leasingové spoločnosti. Financovanie áut pre podnikateľov (právnické osoby) vzrástlo iba o 4%. Podiel zapojenia leasingových spoločností pri financovaní nákupu osobných áut je už dlhodobo stabilný a to 50%. Pri úžitkových automobiloch je vyšší, leasing predstavuje až 2/3 ich financovania.

Podľa typu predmetu môžeme leasing automobilov rozdeliť na leasing osobných, úžitkových a nákladných vozidiel. Samotná štruktúra jednotlivých leasingových spoločností pri tomto delení sa líši, ale vo všeobecnosti sa dá povedať, že najväčšie percento predstavuje leasing osobných vozidiel pre fyzické a právnické osoby (viď Obr. 1).



* spoločné výsledky ČSOB Leasing + PSA Finance Slovakia: 87 573 tis. €
** vrátane výsledkov UniCredit Fleet Management
Zdroj údajov: Asociácia leasingových spoločností SR (http://www.lizing.sk)

Obr. 27: Štruktúra leasingu automobilov v najväčších leasingových spoločnostiach SR (za 1.-2. štvrťrok 2013)

3. Manažment rizika v leasingu

Pod pojmom riziko rozumieme pravdepodobnosť alebo hrozbu výskytu škody, poranenia, zodpovednosti, straty alebo nepriaznivého vývoja, spôsobenú vonkajšou alebo vnútornou zraniteľnosťou. Ide teda o možnosť nastania nepriaznivej situácie, ktorá má alebo môže mať negatívny vplyv na existenciu, fungovanie, konanie alebo výsledky subjektu. Dôsledkom je negatívny vývoj vnútropodnikových alebo vonkajších vzťahov a činností.

Existuje viacero druhov rizík, niektoré môžu byť zmiernené alebo dokonca eliminované pri správne zvolenom postupe manažmentu rizík. Manažment rizika (z ang. *Risk Management*) zahŕňa identifikáciu udalostí, ktoré by mohli mať nepriaznivé finančné dôsledky a následné uskutočnenie krokov na zabránenie a/alebo minimalizovanie škody spôsobenej týmito udalosťami. Jedná sa o kontinuálny proces vyhodnocovania vývoja prostredia, preferencií a dostupných informácií.

Podnikateľské subjekty čelia finančným a nefinančným rizikám. Finančné riziká sú také, ktoré zvyšujú pravdepodobnosť, že skutočná návratnosť určitej investície bude nižšia ako očakávaná. Zdrojmi finančného rizika môžu byť napríklad zmeny úrokových sadzieb, výmenných kurzov, rizikového kapitálu, refinancovania alebo nesplatenia (Saenko, 2011).

Vo všeobecnosti môžeme pri uzatváraní leasingových zmlúv identifikovať tieto skupiny rizík: 1. Vedomé poskytnutie nepravých, falošných alebo upravených informácií alebo dokumentov so zámerom získania zdrojov. 2. Zámerné nadhodnotenie prenajímaného majetku po dohode s dodávateľom. 3. Porušovanie podmienok prevádzky prenajatého majetku. 4. Riziko (default risk, riziko nesplatenia) straty v dôsledku neplnenia finančných záväzkov zo strany dlžníka v čase ich splatnosti.

4. Modelovanie leasingového rizika

V manažmente rizika sa dnes často využíva prediktívne modelovanie pomocou rôznych programových nástrojov (Berry a Linoff, 2011). Ide o proces tvorby kreditskóringových modelov rôznych typov, v ktorých sa na tvorbu výslednej skórovacej karty (resp. kariet) využívajú rôzne štatistické a neštatistické metódy.

V našej analýze bol využívaný softvér od spoločnosti SAS, konkrétne SAS Enterprise Miner 12.1. (SAS EM), ktorý obsahuje aj 4 uzly pre tvorbu kreditskóringových modelov (Interactive Grouping, Scorecard, Reject Inference, Credit Exchange). V uzle *Scorecard* je na tvorbu skórovacej karty implementovaná metóda logistickej regresie (viď Obr. 2).

Údaje pre analýzu boli poskytnuté leasingovou spoločnosťou, ktorá pôsobí v oblasti leasingu osobných a úžitkových áut v SR. Ide o retailové portfólio vyše 27 000 zmlúv. Súbor pozostáva z uzatvorených zmlúv, ktorých plynutie začalo v období medzi rokmi 2004 a 2007 a skončilo medzi rokmi 2007 až 2013. Štatistickou jednotkou je zmluva, ktorá bola uzatvorená s klientom a to buď súkromnou osobou (fyzická osoba, podiel 50,9% zo všetkých zmlúv v súbore) alebo podnikateľom (právnická osoba, podiel 49,1% zo všetkých zmlúv). Na základe tohto delenia zmlúv, môžeme získané premenné o štatistickej jednotke rozdeliť do nasledujúcich skupín:

- *Časové údaje:* dátum prevzatia, dátum ukončenia, dátum najstaršej dlžnej splátky.
- *Predmet leasingu:* druh vozidla (osobné/úžitkové), individuálna doprava, nové vozidlo, názov verzie vozidla, výkon motora v kW, objem motora v ccm.
- *Finančné údaje:* cena vozidla, percento akontácie, počet splátok, bonitná skupina.
- *Demografické údaje:* a) súkromná osoba: pohlavie, rodinný stav, mesto, PSČ; b) podnikateľ: typ firmy, typ účtovníctva, mesto, PSČ.

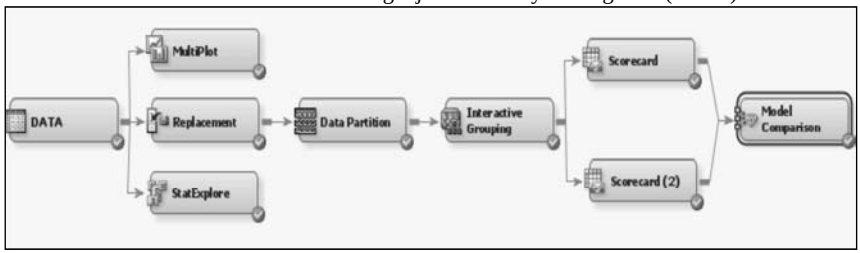
K poskytnutým premenným sme vytvorili niekoľko nových a odvodených premenných, napr. z PSČ sme odvodili kraj a okres. Premenná percento financovania bola vytvorená podielom financovanej čiastky a celkovej ceny v eurách, čo odzrkadľuje reálnu cenu leasingu.

Namiesto premennej verzia vozidla, ktorá bola veľmi neprehľadná, sme vytvorili premennú výrobca vozidla (5 skupín). Keďže sa jedná o retailové portfólio, vytvorili sme aj premennú počet zmlúv, ktorá kvantifikuje koľko má daný klient zmlúv u leasingovej spoločnosti. Poslednou novou premennou bolo vytvorenie kategórie auta (7 skupín), ktorá je vyjadrením sily vozidla, pomocou kombinácie objemu motora v ccm a výkonu motora v kW. V kategórii 1 sa nachádzajú vozidlá s najslabším výkonom a tento výkon rastie v závislosti od kategórie. Kategória 7 teda zodpovedá veľmi silnému vozidlu vzhľadom na výkon a objem motora.

Modelovanou cieľovou premennou bola binárna premenná default, ktorá vyjadruje, či daný klient platil (default = 0, t.j. dobrý klient) alebo neplatil (default = 1, t.j zlý klient) svoje leasingové splátky. Samozrejme v praxi sú aj klienti, ktorí môžu byť nedefinovaní, prípadne s príliš krátkou platobnou históriou, zamietnutí alebo vyradení. Takýto klienti boli z modelovanej vzorky údajov vyradení.

Výsledkom modelovania bude aplikačná skórovacia karta (*application scorecard*), ktorá hodnotí kredibilitu nových potenciálnych klientov na základe pravdepodobnosti, že klient bude stratový, prípadne bude dlžníkom. Aplikačný skóring kvantifikuje riziko, spojené so žiadosťami o leasing, na základe vyhodnotenia sociálnych, demografických, finančných a iných údajov v čase samotnej žiadosti. Zjednodušuje rozhodnutia súvisiace s akvizíciou daného klienta, tým, že napomáha k zautomatizovaniu celého procesu vzniku leasingu.

Boli vytvorené 2 výsledné modely, zvlášť pre fyzické a zvlášť pre právnické osoby, pretože ide o dva rôzne druhy správania v oblasti leasingu automobilov. Postup modelovania v SAS EM na základe SEMMA metodológie je znázornený na diagrame (Obr. 2).



Obr. 28: Diagram postupu modelovania v SAS EM

Uzol *Interactive Gruping* umožňuje automatickú ale aj ručnú kategorizáciu vstupných premenných tak, aby mali čo najvyššiu prediktívnu silu. Na tento účel sa využíva charakteristika WOE (*Weight of Evidence*), ktorá sa vypočíta v i-tej kategórii príslušnej s-tej premennej podľa vzťahu:

$$WOE_i^s = \ln\left(odds_{ratio_i}^{\ s}\right) = \ln\left(\frac{good_i^s}{bad_i^s} / \frac{good}{bad}\right) = \ln\left(\frac{good_i^s}{good} / \frac{bad_i^s}{bad}\right)$$

pričom počet dobrých klientov označíme ako ($good_i^s$), resp. zlých ako (bad_i^s) v i-tej kategórii príslušnej s-tej premennej. Priebeh WOE pre s-tú premennú má byť monotónny.

V uzle *Scorecard* sú do logistického modelu vybraté len významné premenné. Na tento účel sú využívané v SAS EM dve štatistiky, konkrétne Giniho index a informačná hodnota (IV=*Information Value*). Prednastavená (default) štatistika je IV s hodnotou 0,1.

Giniho index nadobúda hodnoty od 0 po 1, resp. od 0 do 100. Za významnú premennú v modeli aplikačného skóringu považujeme premennú vtedy, keď dosiahne hodnotu Giniho indexu vyššiu ako 0,25 (resp. 25). Giniho index sa vypočíta podľa vzťahu:

$$Gini = 1 - \sum_{k=2}^{n+m} \left(F_{m.Z_k} - F_{m.Z_{k-1}}\right).\left(F_{n.D_k} - F_{n.D_{k-1}}\right)$$

kde $F_{m.Z_k}$ ($F_{n.D_k}$) je k-tá hodnota empirickej distribučnej funkcie dobrých (zlých) klientov (Řezáč a Řezáč, 2011).

Informačná hodnota (IV) je index, ktorý vyjadruje prediktívnu silu premennej vo vzťahu k modelovanej premennej a vypočíta sa podľa vzťahu:

$$I_{Val} = \int_{-\infty}^{\infty} (f_D(x) - f_Z(x)) \ln\left(\frac{f_D(x)}{f_Z(x)}\right) dx$$

kde $f_D(x)$ (resp. $f_Z(x)$) označujú rozdelenie pravdepodobnosti (funkcie hustoty) dobrých, resp. zlých klientov (Řezáč, 2011). Za významnú premennú v modeli považujeme premennú vtedy, keď dosiahne informačnú hodnotu (IV) vyššiu ako 0,1.

Premenné s najväčšou silou predikcie sú podľa údajov v tabuľkách 1 a 2: percento financovania, počet splátok leasingu, kategória auta, financovaná hodnota, nové vozidlo, okres a cena (dosiahli vyššiu informačnú hodnotu ako 0,1). Silný predikčný vzťah k cieľovej premennej má financovaná hodnota vozidla a kategória auta. Najvyššiu prediktívnu silu majú podľa tohto hodnotenia počet splátok a percento financovania vozidla, obidve s informačnou hodnotou vyššou ako 0,5. Z výsledkov vyplýva, že rodinný stav, pohlavie klienta, počet leasingových zmlúv klienta, výrobca alebo druh vozidla výrazne neovplyvňujú to, či bude klient leasing splácať alebo nie.

Tab. 10: Kvalita premenných na základe Giniho indexu a IV (súkromné osoby)

Premenná	Gini	IV	Typ premennej
Percento financovania	**49.207**	**1.081**	**INTERVAL**
Počet splátok	**43.398**	**0.706**	**NOMINAL**
Kategória auta	**34.563**	**0.452**	**NOMINAL**
Financovaná hodnota	**27.212**	**0.278**	**INTERVAL**
Nové vozidlo	**19.332**	**0.261**	**BINARY**
Okres	**25.539**	**0.252**	**NOMINAL**
Cena	**18.939**	**0.128**	**INTERVAL**
Rodinný stav	9.268	0.055	NOMINAL
Počet zmlúv	1.280	0.009	NOMINAL
Druh vozidla	0.494	0.005	BINARY
Výrobca	2.808	0.004	NOMINAL
Pohlavie	0.669	0.000	BINARY
Individuálny dovoz	0.066	0.000	BINARY

Tab. 11: Kvalita premenných na základe Giniho indexu a IV (podnikatelia)

Premenná	Gini	IV	Typ premennej
Počet splátok	**48.059**	**1.031**	**NOMINAL**
Okres	**28.942**	**0.442**	**NOMINAL**
Percento financovania	**33.326**	**0.389**	**INTERVAL**
Nové vozidlo	**17.957**	**0.255**	**BINARY**
Cena	**19.894**	**0.173**	**INTERVAL**
Typ firmy	12.245	0.081	NOMINAL
Financovaná hodnota	7.833	0.048	INTERVAL
Kategória auta	10.528	0.039	NOMINAL
Typ účtovníctva	8.431	0.029	BINARY
Počet zmlúv	7.253	0.027	INTERVAL
Výrobca	4.102	0.021	NOMINAL
Druh vozidla	2.697	0.003	BINARY
Individuálny dovoz	0.638	0.003	BINARY

Kvalitu predikcie modelov môžeme vyhodnotiť na základe matice chybovosti (confusion matrix), ktorá vyjadruje chyby pri predikovaní cieľovej premennej, v našom prípade chyby pri predikovaní zlých klientov (1), teda toho, že klient nebude splácať. Z tabuľky 3 vyplýva, že skórovacia karta pre súkromné osoby identifikuje správne 96% dobrých klientov a 97% zlých klientov. Skórovacia karta pre podnikateľov identifikuje správne 98% dobrých klientov a 98,7% zlých klientov.

Prehľad ďalších štatistík kvality modelov uvádzame v tabuľke 4. Hodnota kumulatívneho liftu napríklad znamená, že výsledné modely sú lepšie viac ako 5-krát v porovnaní s náhodným výberov klientov pre schválenie leasingu. Na základe Kolmogorovej-Smirnovej štatistiky sme zistili aj hraničné počty bodov pre schválenie leasingu (tzv *cut-off value*). Pre

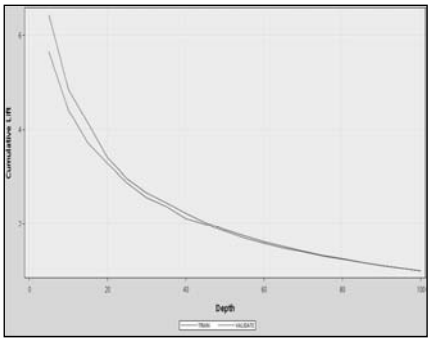
súkromné osoby je to 193 bodov a pre podnikateľov 212 bodov (hodnoty zistené na validačnej množine). Ide o hodnoty určené na základe štatistických kritérií a v praxi sa ešte musia modifikovať na základe finančných prepočtov. Celkovo môžeme konštatovať, že kvalita oboch modelov je vyhovujúca a môžu sa použiť v praxi leasingovej spoločnosti.

Tab. 12: Matica chybovosti výsledných modelov

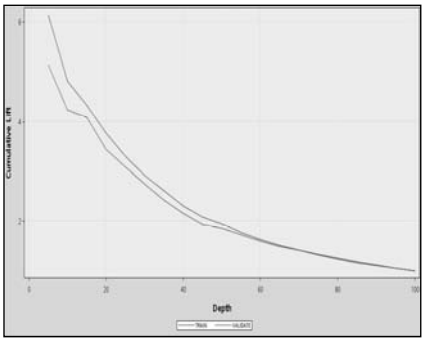
		Predikcia			
		Súkromné osoby		**Podnikatelia**	
		Dobrí (0)	**Zlí (1)**	**Dobrí (0)**	**Zlí (1)**
Skutočnosť	**Dobrí (0)**	96.37%	3.73%	98.08%	1.92%
	Zlí (1)	0.25%	97.49%	1.03%	98.66%

Tab. 13: Štatistiky kvality výsledných modelov

Štatistika	**Súkromné osoby**		**Podnikatelia**	
	Trénovacie údaje	**Validačné údaje**	**Trénovacie údaje**	**Validačné údaje**
Misclassification Rate	0.04	0.04	0.02	0.02
Kolmogorov-Smirnov Statistic	0.54	0.50	0.59	0.49
Area Under ROC	0.85	0.82	0.86	0.77
Gini Coefficient	0.70	0.65	0.72	0.53
Accuracy Ratio	0.70	0.65	0.72	0.53
Cumalative Lift (10. Percentil)	4.86	4.75	4.81	4.66



Obr. 29: Kumulatívny lift (súkromné osoby)



Obr. 30: Kumulatívny lift (podnikatelia)

Využitie výsledkov modelovania demonštrujeme na nasledovnom príklade. Z databázy vyberieme náhodného klienta - súkromnú osobu, ktorá žiada financovanie osobného ojazdeného vozidla. Cena tohto vozidla je 7525€ a požadované percento financovania je 50%, teda 3763€. Ide o vozidlo Škoda Octavia Tour s výkonom motora 75kW a objemom 1596ccm, čo vozidlo zaraďuje do našej 4. kategórie sily auta. Klientom je slobodný muž z Vranova nad Topľou, z Prešovského kraja a ide o jeho prvú leasingovú zmluvu.

Ak by sme na tohto klienta aplikovali náš predikčný model pre súkromné osoby, jeho bodové skóre by bolo 191. Hodnota cut-off pri súbore súkromných osôb bola stanovená na 193 bodov, preto by tento klient nebol akceptovaný. Hlavným dôvodom je, že sa jedná o leasing ojazdeného vozidla, ako aj o vyššie percento financovania. Postup skórovania môžeme vidieť v tabuľke 5. V tomto prípade by bola predikcia správna, keďže sa tento zákazník ukázal naozaj ako zlý klient a na jeho zmluve sa default vyskytol.

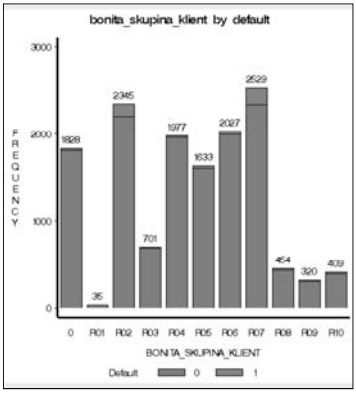
Tab. 14: Príklad skórovania

Charakteristika	Hodnota	Body skórovacej karty
Cena	7525	30
Financovaná hodnota	3763	34
Percento financovania	50%	35
Počet splátok	24	67
Nové vozidlo	N	-28
Kategória auta	4	27
Individuálny dovoz	N	Nemá vplyv
Druh vozidla	O	Nemá vplyv
Pohlavie	muž	Nemá vplyv
Rodinný stav	21 - slobodný	Nemá vplyv
Okres	Vranov nad Topľou	26
Kraj	Prešovský	Nemá vplyv
Výrobca	ŠKODA	Nemá vplyv
Počet zmlúv	1	Nemá vplyv
Spolu body:		**191**

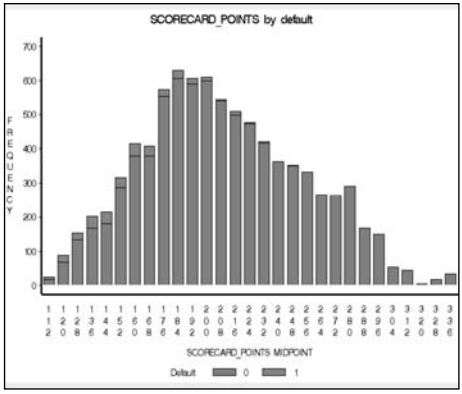
5. Porovnanie výsledkov modelovania s expertným hodnotením klientov

Poskytnutý súbor údajov obsahoval aj informáciu zaradení do *bonitnej skupiny* jednotlivých klientov v leasingovej spoločnosti, ktorá bola urobená na základe expertnej metódy. Ide o kategoriálnu premennú s 10-timi kategóriami, kde 1 predstavuje najmenej rizikovú skupinu a 10 najviac rizikovú. Hodnota 0 vyjadruje skutočnosť, že klientovi nebola priradená žiadna skupina, išlo o klientov, ktorým bol poskytnutý leasing v rokoch 2004 a 2005, keď tento systém ešte nebol zavedený. Distribúciu klientov podľa bonitnej skupiny uvádzame na obrázku 5 a je zrejmé, že takéto rozdelenie klientov bolo veľmi nerovnomerné. Môžeme si tiež všimnúť vysoké percento výskytu defaultu (nesplatenia) v druhej skupine, ktorá by mala predstavovať skupinu s relatívne nízkym rizikom.

Rozdelenie klientov podľa nášho modelu pre súkromné osoby uvádzame na ďalšom obrázku (Obr. 6). Toto rozdelenie je má vrchol v hodnote 193 bodov (cut-off hodnota) a podiel defaultu (nesplatenia) s rastúcim počtom bodov klesá. Podobný výsledok môžeme sledovať aj pri súbore podnikateľov, ale výsledky tu už neuvádzame z priestorových dôvodov.



Obr. 31: Bonitné skupiny klientov (expertná metóda)



Obr. 32: Rozdelenie bodov klientov podľa modelu skórovacej karty (súkromné osoby)

6. Záver

Skóringové technológie môžu byť použité ako objektívny nástroj manažmentu rizika, ktorý napomáha zabezpečiť centralizovaný, jednotný, viac konzistentný a spoľahlivý manažment rozhodovania v organizácii. Jeho hlavnými prínosmi sú: 1. zvyšovanie ziskovosti každodenných operatívnych rozhodnutí, napr. rozhodnutia súvisiace so získavaním a udržaním si klientov`2. zníženie času potrebného na rozhodovanie, zjednodušenie úverových operácií; 3. zabezpečenie individuálneho a zároveň automatizovaného prístupu ku každému zákazníkovi; 4. zautomatizovanie hromadných operačných rozhodnutí a zníženie nákladov na prácu; 5. zabezpečenie pripravenosti na meniace sa podmienky trhu.

Literatúra

BERRY, M. J. A. – LINOFF, G. S. 2011. *Data mining Techniques. For Marketing, Sales, and Customer Relationship management*. 3.vydanie. USA: Wiley Publishing, Inc., Indianapolis, s. 888.

Default Risk. [online]. USA: WebFinance Business Dictionary. Dostupné na internete: ‹http://www.businessdictionary.com/›

Definition of Risk. [online]. USA: WebFinance Business Dictionary, Dostupné na internete: http://www.businessdictionary.com/definition/ risk.html

ŘEZÁČ, M. – ŘEZÁČ, F. 2011. How to Measure the Quality of Credit Scoring Models. *Finance a úvěr - Czech Journal of Economics and Finance*, Praha: UK FSV Praha, roč. 61, č. 5/2011, s. 486-507.

ŘEZÁČ, M. 2011. Measuring Quality of Scoring Models Using Information Value. In: *Journal of Communication and Computer*, Libertyville, Illinois, USA: David Publishing Company, roč. 8, č. 3/2011, s. 234-239.

SAENKO, O. A. 2011. *Risk management of leasing company*. Ukrajina: Luhansk Taras Shevchenko National University, s. 5. Dostupné na internete: http://dspace. nbuv.gov.ua/bitstream/handle/123456789/24164/32-Saenko.pdf?sequence=1

Adresa autorov:

Iveta Stankovičová
Univerzita Komenského
Fakulta managementu
Katedra informačných systémov
Odbojárov 10, 820 05 Bratislava
iveta.stankovicova@fm.uniba.sk

Martin Řezáč
Masarykova univerzita
Přírodovědecká fakulta
Ústav matematiky a statistiky
Kotlářská 2, 611 37 Brno
mrezac@math.muni.cz

Poďakovanie:

Ďakujeme Nadácii Tatry banky za poskytnutie grantu Kvalita vzdelávania, z finančných prostriedkov ktorého mohla Univerzita Komenského zakúpiť univerzitnú licenciu softvéru SAS pre rok 2013. V module SAS Enterprise Miner boli uskutočnené všetky výpočty použité v tomto článku.

Vývoj mier monetárnej chudoby na Slovensku
Trend of monetary poverty measures in Slovakia

Iveta Stankovičová, Róbert Vlačuha

Abstract: The EU statistics on income and living conditions (EU SILC) is the reference source for comparative statistics on income distribution and social inclusion in the European Union (EU). We used Slovak EU SILC data for empirical analysis of monetary poverty measures. We computed monetary poverty measures, namely 3 FGT indexes and Watts index. The aim of this paper is to analyse trends for these indicators in Slovakia in the period 2009-2012 and compare results by socio-economic factors: NUTS2 regions, type of households, age groups and economic activity.

Abstrakt: Výberové zisťovanie o príjmoch a životných podmienkach domácností EU SILC je zdrojovou základňou pre porovnávanie distribúcie príjmov a sociálnej inklúzie na úrovni EÚ. V príspevku sme použili slovenské EU SILC údaje na empirické analýzy mier monetárnej chudoby. Vypočítali sme miery monetárnej chudoby (3 FGT indexy a Watts index). Cieľom príspevku je analýza vývoja týchto indikátorov na Slovensku v rokoch 2009 až 2012 a porovnanie výsledkov podľa socio-ekonomických faktorov: regióny NUTS2, typ domácnosti, vekové skupiny a ekonomická aktivita.

Key words: Monetary Poverty, FGT Indexes, Watts Index, EU SILC Database.

Kľúčové slová: monetárna chudoba, FGT indexy, Wattsov index, EU SILC databáza.

JEL classification: O15, C46, I32

1. Úvod

Chudoba patrí aj v dvadsiatom prvom storočí medzi problémy, ktorými je potrebné sa zaoberať. Skúmanie chudoby a sociálneho vylúčenia je potrebné zasadiť do kontextu špecifického hospodárskeho a sociálneho vývoja na konkrétnom území. Zmenený medzinárodný a politický kontext v Európe priniesol pre Slovensko nové výzvy aj v skúmaní chudoby. Stali sme sa súčasťou spoločenstva, v ktorom sú otázky nepriaznivej životnej situácie a nerovnakých životných šancí v centre pozornosti výskumu i politického rozhodovania.

V súčasnosti sa vyskytuje v oblasti skúmania chudoby pojem „nová chudoba“. Ide o označenie paradoxného javu, kedy aj napriek relatívnemu blahobytu a prosperite vo vyspelých krajinách, žije súčasne 10 až 20% ich populácie stále v chudobe (Godschalk, 1991). Tento pojem sa spája tiež so skutočnosťou, že chudoba je vo väčšej časti európskeho priestoru predovšetkým relatívnou a nie absolútnou. Chudoba je v rozvinutých krajinách Európy primárne spájaná s nedostatkom pracovných príležitostí a s dlhodobou nezamestnanosťou. „Nová“ chudoba je častejšie zastúpená v určitých sociálnych kategóriách, koncentruje sa na určitom mieste (resp. priestore), má trvalý charakter a často sa reprodukuje z generácie na generáciu. Súvisí s trhom práce a dotýka sa nezanedbateľnej časti populácie. Najviac ohrozenými „novou“ chudobou sa stávajú nezamestnaní (najmä dlhodobo nezamestnaní), ľudia dlhodobo znevýhodnení resp. vylúčení na trhu práce (v dôsledku napríklad zníženej fyzickej či psychickej schopnosti – ťažko zdravotne postihnutí; nízkej kvalifikácie; v dôsledku diskriminácie – ženy, etnické minority, starí ľudia). Taktiež ľudia, ktorým sa nepodarilo adaptovať sa na nové podmienky, ale aj osoby s nízkymi príjmami, tzv. pracujúci chudobní („working poor“), ktorí sa v dôsledku nízkej miery dosiahnutej kvalifikácie uplatňujú na sekundárnom trhu práce ako nekvalifikovaná pracovná sila. Medzi „ohrozených“ patria aj obyvatelia žijúci v menej rozvinutých (marginalizovaných) regiónoch.

2. Miery monetárnej chudoby

Pre potreby merania monetárnej chudoby bolo vyvinutých veľké množstvo rôznych ukazovateľov (mier), ktoré sú schopné zachytiť rôzne stránky chudoby. V tomto článku sme sa sústredili na výpočet mier chudoby, ktoré sú známe pod názvom FGT indexy (Foster, Greer a Thorbecke, 1984) a aj na Wattsov index (1968).

Nech $\mathbf{y} = (y_1, y_2, ..., y_n)$ je vektor príjmov populácie usporiadaných vzostupne a n je celkový počet populácie v súbore. Predpokladajme, že $z > 0$ je určená hranica chudoby a q je počet chudobnej populácie, pre ktoré platí $y_1 \leq y_2 \leq ... \leq y_q \leq z$. FGT indexy sa dajú vyjadriť nasledovným všeobecným vzorcom:

$$P_{\propto}(y, z) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^{\propto} \quad (1)$$

Vzorec (1) umožňuje zapísať rad mier (indexov), ktoré závisia od stupňa α.

Ak α = 0, dostávame mieru, ktorá sa nazýva aj *„jednoduchá miera rizika chudoby"* a vyjadruje podiel populácie, ktorá sa nachádza pod hranicou chudoby. Na jednej strane je to jednoduchá a najviac používaná miera chudoby, na druhej strane má však určité nedostatky. Po prvé, neberie do úvahy hĺbku, resp. intenzitu chudoby. Po druhé nič nehovorí o distribúcii chudoby. Jednoduchá miera chudoby sa napr. nemení, ak ľudia pod hranicou chudoby sa stávajú chudobnejší.

Pre α = 1 sa tento index volá *„hĺbka (intenzita) chudoby"* a vyjadruje, ako nízko pod hranicou chudoby je priemerný príjem ľudí vystavených chudobe. Čím je táto miera bližšie k nule, tým je priemerná hĺbka chudoby nižšia a teoreticky na vymanenie populácie z chudoby by bolo potrebných menej finančných prostriedkov. Naopak, čím je táto miera bližšia k hodnote 1, tým je situácia medzi chudobnou populáciou horšia a ich prepad pod hranicou chudoby je väčší.

Pre α = 2 dostávame tzv. *„závažnosť chudoby (vážená priemerná hĺbka chudoby)"*. V tejto podobe index kombinuje informácie o chudobe s príjmovou nerovnosťou chudobnej populácie. Hovorí nám teda o distribúcii populácie pod hranicou chudoby.

Vyššie tri spomínané miery (indexy) sa označujú ako P0, P1 a P2 a ich kombinácia nám dáva komplexný pohľad na výskyt, hĺbku a závažnosť chudoby.

Ďalšou mierou (indexom), ktorá je citlivá na výskyt a zároveň distribúciu chudoby bola navrhnutá v roku 1968 a nazýva sa podľa autora Watts index. Dá sa vyjadriť nasledujúcim všeobecným vzorcom:

$$W = \frac{1}{n}\sum_{i=1}^{q}[ln(z) - ln(y_i)] \quad (2)$$

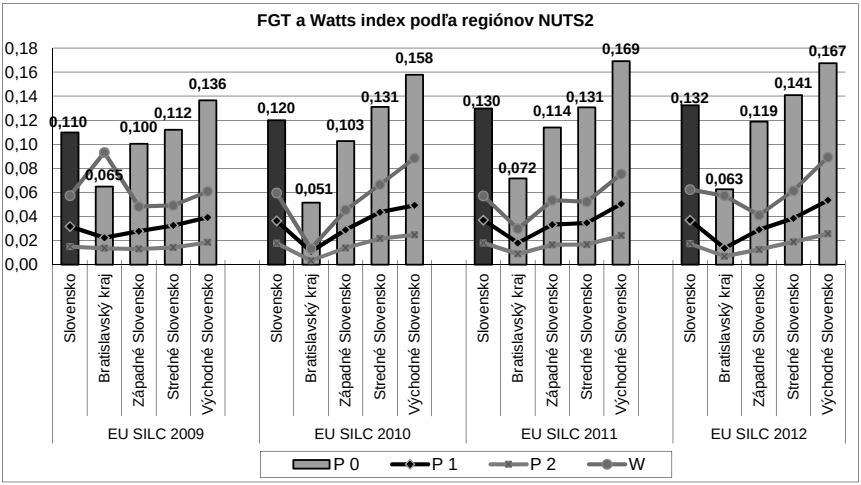
kde n je celkový počet populácie v súbore zoradenej vzostupne podľa príjmu (resp. výdavkov), $\mathbf{y} = (y_1, y_2, ..., y_n)$ je vektor príjmov populácie, $z > 0$ je určená hranica chudoby a q je počet chudobnej populácie. Vo všeobecnosti platí, že čím je táto miera bližšie k nule, tým je výskyt a distribúcia chudoby nižšia a naopak, čím je táto miera bližšia k hodnote 1, tým je situácia medzi chudobnou populáciou horšia.

3. Vývoj mier monetárnej chudoby na Slovensku v rokoch 2009 až 2012

Chudobu ovplyvňuje množstvo faktorov. Za najvýznamnejšie faktory môžeme vo všeobecnosti považovať regionálne hľadisko, vek, typ domácnosti či status ekonomickej aktivity. V tejto časti príspevku sa zameriame na analýzu mier monetárnej chudoby na Slovensku podľa všetkých vyššie spomínaných faktorov.

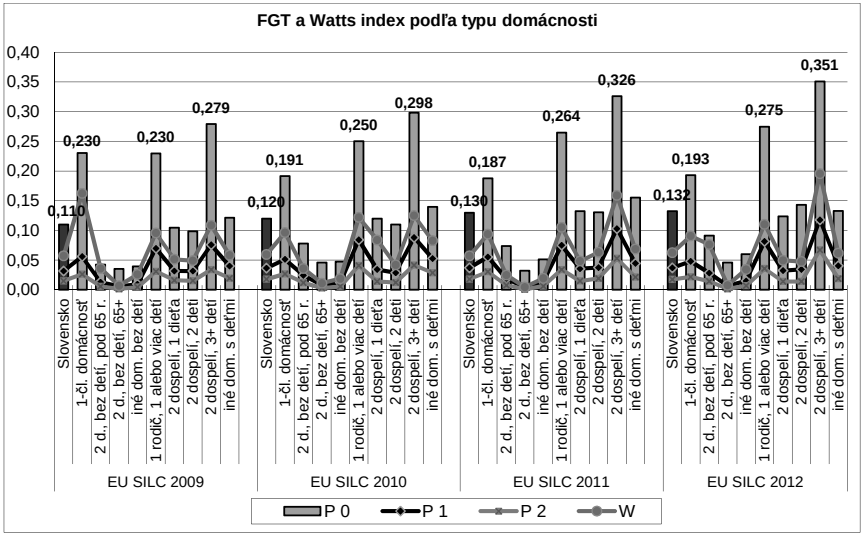
Uvádzame niekoľko dokumentov zaoberajúcich sa chudobou a jej vývojom na Slovensku. Ide napríklad o práce autorov Bartošová a Želinský (2013), Želinský a Stankovičová (2012), Ivančíková a Vlačuha (2010).

Pri výpočte FGT indexov a Watts indexu sme v príspevku vychádzali z výberového štatistického zisťovania EU SILC. Výberové zisťovanie o príjmoch a životných podmienkach domácností EU SILC realizuje od roku 2005 na Slovensku Štatistický úrad Slovenskej republiky. Ide o harmonizované zisťovanie členských štátov EU, ktorého úlohou je zabezpečiť produkciu pravidelných, včasných a kvalitných údajov o príjmoch, chudobe a sociálnom vylúčení.



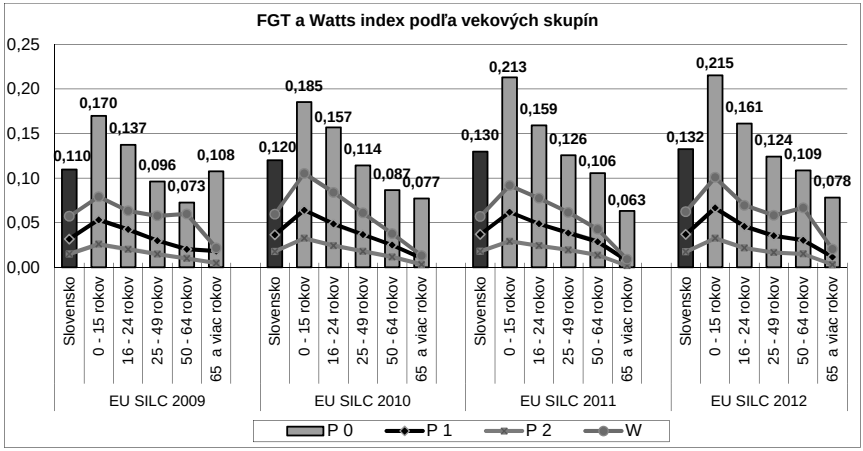
Obr. 33: FGT a Watts index podľa regiónov NUTS2

Z hľadiska výskytu chudoby (index P0) môžeme na Slovensku v rokoch 2009 až 2012 pozorovať negatívny trend. V roku 2009 bolo mierou monetárnej chudoby na Slovensku ohrozených 11,0% obyvateľstva a postupne táto miera narástla na 13,2% obyvateľstva v roku 2012. Ako je zrejmé z Obr. 1 a Tab. 1, na výskyte monetárnej chudoby sa výrazne prejavili regionálne disparity. Počas celého sledovaného obdobia boli najmenej ohrození obyvatelia Bratislavského kraja a najviac boli ohrození obyvatelia Východného Slovenska. Najväčšia disparita bola pozorovaná v roku 2010, kde rozdiel medzi týmito dvoma regiónmi predstavoval 10,7 percentuálnych bodov (p. b.). Čo sa týka hĺbky (P1) a závažnosti chudoby (P2), tieto kopírovali trend výskytu chudoby a nezaznamenali sme výraznejšie odchýlky. Zaujímavý bol však vývoj Watts indexu, ktorý v roku 2009 vyšiel paradoxne najvyšší v Bratislavskom kraji. V rokoch 2010 a 2011 síce Watts index odzrkadľoval podobný trend ako FGT indexy, v roku 2012 bol Watts index pre Západné Slovensko nižší ako pre Bratislavský kraj.



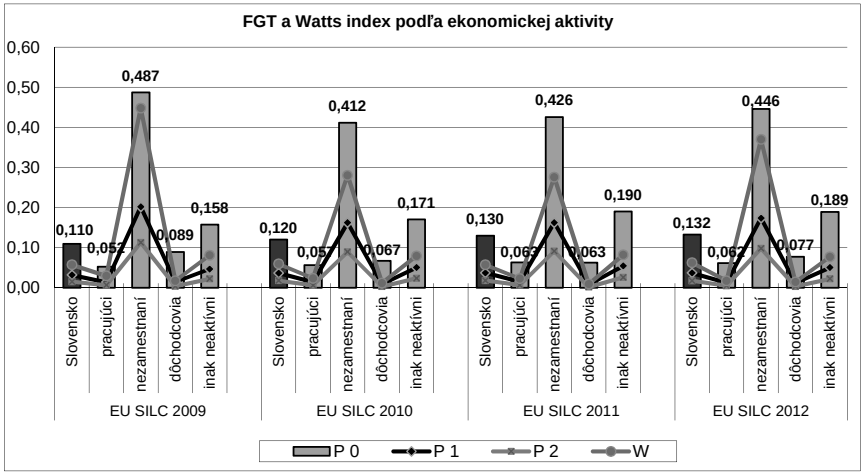
Obr. 2: FGT a Watts index podľa typu domácnosti

Ďalším z významných faktorov vplývajúcim na mieru chudoby je typ domácnosti (Obr. 2, Tab. 2). Z analyzovaných rokov sme na Slovensku identifikovali 3 typy najviac ohrozených domácností. Pri jednočlenných domácnostiach sme v prvých troch rokoch zaznamenali mierne pozitívny trend vo vývoji výskytu chudoby (P0) na Slovensku, keď miera ohrozenia príjmovou chudobou klesla z 23,0% na 18,7%. V roku 2012 stúpol počet ohrozených jednočlenných domácností na 13,3%. Ďalšou ohrozenou skupinou boli neúplné domácnosti, teda domácnosti s jedným rodičom a jedným alebo viacerými deťmi. Pri tomto type domácnosti sme zaznamenali negatívny trend počas všetkých sledovaných rokoch a výskyt monetárnej chudoby sa u nich zvýšil z 23,0% v roku 2009 až na 27,5% v roku 2012. Najviac ohrozenou skupinou sa ukázali viacdetné domácnosti, teda domácnosti s 2 rodičmi a 3 alebo viac deťmi. Ich miera ohrozenia chudobou narástla z 27,9% v roku 2009 až na 35,1% v roku 2012. Pri viacdetných domácnostiach sa výraznejšie prejavil aj nárast Watts indexu, ktorého hodnota bola v roku 2012 na úrovni 19,5%. Čo sa týka hĺbky (P1) a závažnosti chudoby (P2), tieto opäť kopírovali trend výskytu chudoby a nezaznamenali sme výraznejšie odchýlky. Môžeme vo všeobecnosti povedať, že hĺbka a závažnosť chudoby neboli pri jednotlivých typoch domácností až také výrazné.



Obr. 3: FGT a Watts index podľa vekových skupín

Pri analýze výskytu chudoby podľa vekových skupín (Obr. 3, Tab. 3) sa potvrdil trend vysokého výskytu chudoby pri viacdetných domácnostiach a najviac ohrozenými boli deti vo veku do 15 rokov. Ich miera ohrozenia monetárnou chudobou (P0) narástla zo 17,0% v roku 2009 až na 21,5% v roku 2012. Najmenej ohrozenými boli obyvatelia vo veku 65 rokov a viac. Hĺbka a závažnosť chudoby vo všetkých sledovaných rokoch úplne presne kopírovali trend vývoja výskytu chudoby. Je zaujímavé, že síce pri deťoch sme pozorovali nárast výskytu chudoby, neprejavilo sa to však negatívne na hĺbke a závažnosti chudoby tejto časti populácie. Pri Watts indexe pozorujeme, až na malé výnimky, vo všeobecnosti podobný trend ako pri FGT indexoch. Aj podľa tohto indexu boli najviac ohrozené deti vo veku do 15 rokov a najmenej ohrození obyvatelia vo veku 65 rokov a viac.



Obr. 4: FGT a Watts index podľa ekonomickej aktivity

Posledným faktorom, podľa ktorého sme hodnotili vývoj monetárnej chudoby na Slovensku, bola ekonomická aktivita obyvateľstva (vo veku 16 rokov a viac, Obr. 4, Tab. 4). Z výsledkov vyplýva, že najviac ohrozenou skupinou sú nezamestnaní, keď takmer každý druhý nezamestnaný bol ohrozený rizikom monetárnej chudoby. Najmenej ohrozenou skupinou boli pracujúci a dôchodcovia. Negatívny trend v prvých troch rokoch vo výskyte chudoby u pracujúcich a pozitívny trend u dôchodcov v tomto období spôsobil, že ich miera chudoby sa v roku 2009 vyrovnala a dostala na úroveň 6,3%. V poslednom roku sa opäť medzi nimi prejavil rozdiel a dôchodcovia boli viac ohrození výskytom chudoby oproti pracujúcim o 1,5 p.b.. U nezamestnaných sa oproti ostatným skupinám obyvateľstva výraznejšie prejavila aj hĺbka (P1) a závažnosť (P2) chudoby. Taktiež veľmi negatívne hodnoty boli u nezamestnaných zaznamenané aj pri Watts indexe. Podľa výsledkov sa jednoznačne potvrdilo, že aktívna účasť na trhu práce a sociálna ochrana formou poskytovania starobných dávok sú faktory, ktoré zohrávajú dôležitú úlohu v boji proti chudobe, pretože pomáhajú konkrétnym skupinám obyvateľstva neprepadnúť pod hranicu rizika chudoby.

4. Záver

Predložený príspevok mal za cieľ poskytnúť prehľad o vývoji vybraných monetárnych mier chudoby na Slovensku. Pri analýze chudoby sme vychádzali z dát výberového štatistického zisťovania EU SILC za roky 2009 až 2012. Vybrané monetárne miery chudoby sme v jednotlivých rokoch analyzovali podľa faktorov, ktoré majú na Slovensku vo všeobecnosti najväčší dopad na výskyt, hĺbku a závažnosť chudoby. Tieto faktory boli regionálne hľadisko, vek, typ domácnosti a status ekonomickej aktivity.

Pri analýze mier chudoby podľa regionálneho hľadiska sa vo všetkých mierach výrazne prejavili regionálne disparity na Slovensku. Rozdiel vo výskyte chudoby bol v niektorých rokoch medzi Bratislavským krajom a Východným Slovenskom až 3-násobný. Pri analýze podľa typu domácnosti sme identifikovali 3 typy najviac ohrozených domácností: jednočlenné, neúplné a domácnosti 2 rodičov s troma a viac deťmi. Trend vysokého výskytu chudoby pri viacdetných domácnostiach sa prejavil aj pri analýze podľa vekových skupín, keď najviac ohrozenými boli deti vo veku do 15 rokov. Najväčší vplyv na výskyt, hĺbku a závažnosť chudoby má na Slovensku jednoznačne ekonomická aktivita. Takmer každý druhý nezamestnaný je na Slovensku ohrozený chudobou a oproti ostatným skupinám obyvateľstva sa u nich výraznejšie prejavila aj hĺbka a závažnosť chudoby.

Z analýzy údajov zo štatistické zisťovania EU SILC sa jednoznačne potvrdila relevantnosť FGT indexov na komplexné hodnotenie výskytu, hĺbky aj závažnosti chudoby na Slovensku, ktoré sme naviac analyzovali podľa vybraných socio-ekonomických faktorov. Watts index sa nie vždy ukázal ako vhodné meradlo miery monetárnej chudoby. Vo všeobecnosti môže byť použitý ako doplnková informácia ku FGT indexom, ale neodporúčame ho použiť na hodnotenie monetárnej chudoby na Slovensku samostatne.

Poďakovanie

Príprava príspevku bola podporená Slovenskou Vedeckou grantovou agentúrou ako súčasť výskumného projektu *VEGA 1/0127/11 Priestorová distribúcia chudoby v EÚ.*

Ďakujeme Štatistickému úradu SR, ktorý v súlade s článkom 30 Zákona č.540/2001 Z.z. o štátnej štatistike, poskytol na vedecké a výskumné účely anonymizované údaje z výberových zisťovaní EU SILC 2009 až 2012.

Ďakujeme Nadácii Tatry banky za poskytnutie grantu Kvalita vzdelávania, z finančných prostriedkov ktorého mohla Univerzita Komenského zakúpiť univerzitnú licenciu softvéru SAS pre rok 2013. V systéme SAS boli uskutočnené všetky výpočty použité v tomto článku.

Literatúra

BARTOŠOVÁ, J. – ŽELINSKÝ, T. 2013. The extent of poverty in the Czech and Slovak Republics 15 years after the split. In: *Post-Communist Economies*, *25*(1), s. 119-131. Dostupné: http://dx.doi.org/10.1080/14631377.2013.756704

EURÓPSKA KOMISIA. 2003. *'Laeken Indicators' - Detailed calculation methodology*. Luxembourg: European Commission - Eurostat.

EUROSTAT. 2009. *Algorithms to compute indicators in the streamlined Social Inclusion Portfolio based on EU-SILC and adopted under the Open Method of Coordination (OMC).* Luxembourg: Eurostat.

FOSTER, J., GREER, J., THORBECKE, E. 1984. A Class of Decomposable Poverty Measures. In*: Econometrica.* Vol. 52, No. 3, s. 761-766.

FOSTER, J., GREER, J., THORBECKE, E. 2010. The Foster–Greer–Thorbecke (FGT) poverty measures: 25 years later. In: *The Journal of Economic Inequality*, *8*(4), s. 491-524.

GODSCHALK, J. 1991. Moderní sociální stát a chudoba. In: *Všeobecné otázky sociální politiky.* Bratislava: VÚPSV, s. 6 - 23.

HAGENAARS, A. J. M. 1986. *The perception of poverty.* Amsterdam, North Holland.

IVANČÍKOVÁ, Ľ – VLAČUHA, R. 2010, Chudoba a sociálne vylúčenie v regiónoch Slovenska, Herľany

KAKWANI, N. - SILBER, J. 2008. *Quantitative approaches to multidimensional poverty measurement*. (2nd ed., p. 265). New York: Palgrave Macmillan.

World Bank Institute (2005). *Poverty Manual, All, JH revision of August 8, 2005*. Retrieved from website: http://siteresources.worldbank.org/PGLP/Resources/PovertyManual.pdf

ZHENG, B. 1997. Aggregate Poverty Measures. In: *Journal of Economic Survey, 11(2), s.* 123-62.

ŽELINSKÝ, T. – STANKOVIČOVÁ, I. 2012. Spatial aspect of poverty in Slovakia. In: *The 6th International Days of Statistics and Economics. Conference Proceedings. September 13–15, 2012. Prague, Czech Republic.* Dostupné: http://msed.vse.cz/msed_2012/en/

Adresa autorov:

Iveta Stankovičová
Univerzita Komenského v Bratislave
Fakulta managementu
Odbojárov 10, 820 05 Bratislava
iveta.stankovicova@fm.uniba.sk

Róbert Vlačuha
Štatistický úrad SR
Miletičova 3, 824 76 Bratislava 26
robert.vlacuha@statistics.sk

Tab. 15: FGT a Watts index podľa regiónov NUTS2

	2009				2010				2011				2012			
	P0	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**
SK	**11,0**	**3,2**	**1,5**	**5,7**	**12,0**	**3,6**	**1,8**	**5,9**	**13,0**	**3,7**	**1,8**	**5,7**	**13,2**	**3,7**	**1,7**	**6,2**
SK01	6,5	2,2	1,4	9,3	5,1	1,1	0,3	1,3	7,2	1,8	0,9	3,0	6,3	1,4	0,7	5,7
SK02	10,0	2,8	1,3	4,8	10,3	2,9	1,4	4,5	11,4	3,3	1,6	5,4	11,9	2,9	1,2	4,1
SK03	11,2	3,3	1,4	4,9	13,1	4,4	2,2	6,6	13,1	3,5	1,7	5,2	14,1	3,8	1,9	6,1
SK04	13,6	3,9	1,8	6,1	15,8	4,9	2,5	8,8	16,9	5,0	2,4	7,5	16,7	5,3	2,6	8,9

SK01 - Bratislavský kraj, **SK02** - Západné Slovensko, **SK03** - Stredné Slovensko, **SK04** - Východné Slovensko

Tab. 2: FGT a Watts index podľa typu domácnosti

	2009				2010				2011				2012			
	P0	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**
SK	**11,0**	**3,2**	**1,5**	**5,7**	**12,0**	**3,6**	**1,8**	**5,9**	**13,0**	**3,7**	**1,8**	**5,7**	**13,2**	**3,7**	**1,7**	**6,2**
1	23,0	5,6	2,6	16,2	19,1	5,1	2,7	9,6	18,7	5,5	3,0	9,4	19,3	4,8	2,2	9,1
2	4,2	1,3	0,6	3,6	7,8	2,5	1,2	3,6	7,4	1,9	0,8	2,5	9,1	2,8	1,4	7,6
3	3,5	0,6	0,2	0,7	4,6	0,8	0,3	1,1	3,2	0,4	0,1	0,5	4,6	0,7	0,2	0,8
4	3,9	0,9	0,4	2,6	4,8	1,2	0,6	1,9	5,1	1,3	0,6	1,9	6,0	1,5	0,7	3,4
5	23,0	7,0	3,1	9,6	25,0	8,4	4,1	12,2	26,4	7,5	3,4	10,5	27,5	8,1	3,6	11,0
6	10,5	3,2	1,6	5,1	12,0	3,4	1,4	8,4	13,2	3,5	1,5	4,8	12,4	3,2	1,4	5,0
7	9,9	3,1	1,5	4,9	11,0	2,9	1,2	4,2	13,1	3,7	1,9	6,3	14,3	3,4	1,4	4,8
8	27,9	7,6	3,4	10,9	29,8	8,8	4,2	12,5	32,6	10,3	5,3	15,9	35,1	11,8	6,7	19,5
9	12,2	4,0	1,9	5,8	14,0	5,3	2,9	8,3	15,5	4,5	2,1	6,9	13,3	4,0	1,9	6,2

1 - jednočlenná domácnosť, **2** - dvaja dospelí, bez detí, obaja menej ako 65r., **3** - dvaja dospelí, bez detí, aspoň jeden vo veku 65r. a viac, **4** - iné domácnosti bez detí, **5** - jeden rodič, jedno alebo viac detí, **6** - dvaja dospelí, jedno dieťa, **7** - dvaja dospelí, dve deti, **8** - dvaja dospelí, tri alebo viac detí, **9** - iné domácnosti s deťmi

Tab. 3: FGT a Watts index podľa vekových skupín

	2009				2010				2011				2012			
	P0	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**
SK	**11,0**	**3,2**	**1,5**	**5,7**	**12,0**	**3,6**	**1,8**	**5,9**	**13,0**	**3,7**	**1,8**	**5,7**	**13,2**	**3,7**	**1,7**	**6,2**
1	17,0	5,3	2,6	7,9	18,5	6,4	3,3	10,5	21,3	6,2	2,9	9,2	21,5	6,7	3,3	10,1
2	13,7	4,2	2,0	6,3	15,7	4,9	2,4	8,4	15,9	4,9	2,4	7,8	16,1	4,6	2,2	7,0
3	9,6	3,0	1,5	5,8	11,4	3,7	1,8	6,1	12,6	3,8	2,0	6,2	12,4	3,5	1,7	5,8
4	7,3	2,0	1,0	6,0	8,7	2,5	1,2	3,8	10,6	2,9	1,4	4,3	10,9	3,0	1,5	6,7
5	10,8	1,8	0,5	2,2	7,7	1,0	0,3	1,3	6,3	0,8	0,2	0,9	7,8	1,2	0,3	2,0

1 – (0 - 15 r.), **2** – (16 - 24 r.), **3** – (25 - 49 r.), **4** – (50 - 64 r.), **5** – (65 a viac rokov)

Tab. 4: FGT a Watts index podľa ekonomickej aktivity

	2009				2010				2011				2012			
	P0	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**	**P0**	**P1**	**P2**	**W**
SK	**11,0**	**3,2**	**1,5**	**5,7**	**12,0**	**3,6**	**1,8**	**5,9**	**13,0**	**3,7**	**1,8**	**5,7**	**13,2**	**3,7**	**1,7**	**6,2**
1	5,2	1,4	0,6	3,0	5,7	1,4	0,6	2,4	6,3	1,5	0,7	2,3	6,2	1,3	0,5	1,7
2	48,7	20,2	11,3	44,8	41,2	16,2	8,9	28,0	42,6	16,3	9,1	27,6	44,6	17,4	9,8	37,0
3	8,9	1,5	0,4	1,7	6,7	0,9	0,3	1,2	6,3	0,8	0,2	0,9	7,7	1,2	0,3	1,4
4	15,8	4,7	2,2	8,1	17,1	5,0	2,4	8,0	19,0	5,4	2,6	8,2	18,9	5,0	2,2	7,7

1 - pracujúci**, 2** - nezamestnaní, **3** - dôchodcovia, **4** - inak neaktívni

Kvantilová regresia pre biologické data pomocou SAS-u
Quantile regression for biological data using SAS

Beáta Stehlíková, Ján Brindza

Abstract: Ordinary least squares regression models the relationship between one or more covariates X and the conditional mean of a response variable Y. Quantile regression provides more complete picture . The aim of this paper is to describe the QUANTREG procedure in SAS, which computes estimates and related quantiles for quantile regression. The calculation is demonstrated on biological data. Paper contains a very detailed interpretation of the results.

Abstrakt: Obvykle používané regresné modely využívajúce metódu najmenších štvorcov popisujú vzťah medzi jedným alebo viacerými premennými X a podmienenej strednej hodnoty závisle premennej Y. Quantilová regresia poskytuje úplnejší obraz. Cieľom tejto práce je popísať procedúru QUANTREG v SAS-e, ktorá vypočítava odhady pre kvantilovú regresiu. Výpočet je demonštrovaný na biologických dátach. Príspevok obsahuje veľmi podrobnú interpretáciu výsledkov.

Key words: quantile regression, SAS, cornelian cherry

Kľúčové slová: kvantilová regresia, SAS, drieň obyčajný

JEL classification: C21 C31

1. Úvod

Hľadanie a skúmanie závislosti premenných patrí medzi dôležité úlohy štatistiky. Regresná analýza je jednou z často využívaných štatistických metód, ktorá rieši túto úlohu – na základe nameraných hodnôt $\boldsymbol{X}$ predikujeme závisle premennú Y pomocou vhodnej funkcie h, ktorá závisle premennú Y dobre aproximuje v určitom zmysle. Kvalita predikcie sa posudzuje pomocou vhodnej tzv. stratovej funkcie L. Takýto prístup však poskytuje iba čiastočný pohľad na vzťah medzi premennými, pretože často nás môže zaujímať o popis vzťahov v rôznych bodoch podmienenej distribúcii premennej Y. Kvantilová regresia dokáže odpovedať na takto formulovanú otázku.

2. Materiál a metódy

Drieň obyčajný (*Cornus mas* L.) pochádza z juhovýchodnej Európy a malej Ázie. Na Slovensku sa vyskytuje v jeho južných oblastiach. Nemá vysoké nároky na pestovanie – je odolný voči suchu, chorobám ako aj škodcom. Plody drieňa sa konzumujú v čerstvom stave, slúžia k výrobe štiav, vyrábajú sa z neho kompóty a džemu, tiež víno a drienkovica.

Z morfometrickej analýzy jednotlivých častí rastlín sa získali rozsiahle experimentálne údaje pri každom genotype ako aj v celej kolekcii hodnotených genotypov. V experimentoch sa testovalo 238 ekotypov drieňa obyčajného (*Cornus mas* L.) za účelom určenia výťažnosti dužiny pri technologickom spracovaní plodov. V analýze sú použité zbrané znaky: PHMOT - hmotnosť plodu (g), KHMOT – hmotnosť kôstky (g).

V klasickej regresnej analýze sa často za funkciu h sa volí lineárna funkcia, t.j.

$$h(\boldsymbol{X}) = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad , \tag{1}$$

a parametre β_i (i = 1, 2, ..., k) sa nazývajú regresné koeficienty. Odhadujú sa pomocou metódy najmenších štvorcov alebo metódou maximálnej vierohodnosti. Stratová funkcia sa volí $L(u) = u^2$a odhaduje sa podmienená stredná hodnota $E(Y|\boldsymbol{X})$.

Nech $F(y) = P(Y \leq y)$ je distribučná funkcia pravdepodobnostného rozdelenia náhodnej premennej Y a α je ľubovoľné číslo z intervalu (0, 1). α-kvantil y_α rozdeľuje definičný obor náhodnej premennej Y na dve časti tak, že platí $P(Y \leq y_\alpha) = \alpha$ a $P(Y \geq y_\alpha) = 1 - \alpha$.

Kvantilová funkcia je funkcia daná predpisom $Q(\alpha) = F^{-1}(\alpha) = inf\{y \in R: F(y) \geq \alpha\}$.

V prípade, že stratová funkcia je tvaru $L(u) = \frac{1}{2}|u|$, jedná sa o mediánovú regresiu a hľadá sa odhad podmieneného mediánu ako minimalizovaním výrazu $E[L(Y-\theta)|\,\mathbf{X}] = \frac{1}{2}E[|Y-\theta|\,|\mathbf{X}]$ vzhľadom na parameter θ. Odhad regresných koeficientov sa získa minimalizovaním hodnoty $\sum L(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}) = \frac{1}{2}\sum|y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|$.

Keď sa medián nahradí α - kvantilom a stratová funkcia je tvaru

$$L(u) = \begin{cases} \alpha u & \text{pre } u \geq 0, \\ (\alpha - 1)u & \text{pre } u < 0, \end{cases} \tag{2}$$

jedná sa o kvantilovú regresiu. Stratová funkcia sa dá vyjadriť pomocou funkcie

$$\chi_A(u) = \begin{cases} 1 \text{ pre } u \in A, \\ 0 \quad \text{inak}, \end{cases} \tag{3}$$

vzťahom

$$\rho_\alpha(u) = (\alpha - 1)u\chi_{(-\infty,0)}(u) + \alpha\, u\, \chi_{[0,\infty)}(u). \tag{4}$$

Odhady podmienených α – kvantilov sa získajú minimalizáciou výrazu $E[L(Y-\theta)|\,\mathbf{X}] = E[\rho_\alpha(Y-\theta)\,|\mathbf{X}]$ vzhľadom na parameter θ. Odhad parametrov $\boldsymbol{\beta}$ vedie k minimalizačnej úlohe

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^k} \sum_{i=1}^n \rho_\alpha\,(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}), \tag{5}$$

ktorá sa rieši napríklad pomocou metód lineárneho programovania (simplexova metóda, metóda vnútorného bodu a iné).

K výpočtom bol použitý procedúra QUANTREG programu SAS. Z optimalizačných metód bola použitá simplexova metóda. Príkazy v SASe použité pri výpočtoch kvantilovej regresie sú nasledovné:

```
ods html;
   ods graphics on;
      proc quantreg data=drien alpha=0.05 ci=resampling;
          model PHMOT = KHMOT / quantile= 0.05 to 0.95 by 0.05
          plot=quantplot;
          run;
    ods graphics off;
ods html close;
```

Príkazy `ods html` na začiatku a `ods html close` na konci zabezpečujú výstup vo forme webovej stránky. Nie je problém napísať príkaz aj pre viacnásobnú kvantilovú regresiu. V prípade, že by sme chceli skúmať závislosť hmotnosti plodu – PHMOT od šírky kôstky - KSIR a hrúbky kôstky – KHR stačí zameniť príkaz

```
model PHMOT = KHMOT /
```

príkazom

```
model PHMOT = KSIR KHR /
```

Je potrebné pripomenúť, že nakoľko sa pri odhade regresných koeficientov v kvantilovej regresii jedná o optimalizačný problém, pridanie ďalších premenných nemá za následok vylepšenie modelu.

3. Výsledky a diskusia

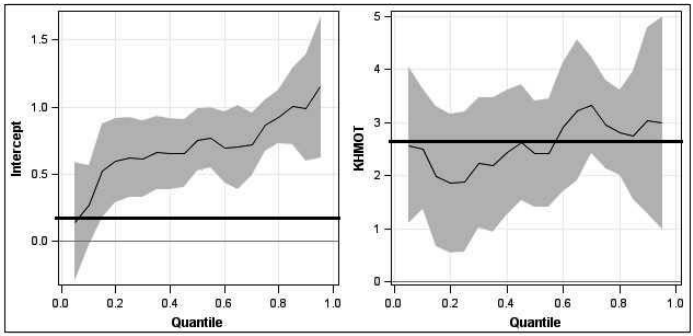
Základné štatistické ukazovatele variability vybratých hodnotených znakov z kolekcie genotypov drieňa obyčajného (*Cornus mas* L.) sú uvedené v tabuľke 1.

***Tab. 2: Popisné štatistiky pre vybrané znaky drieňa obyčajného* (*Cornus mas* L.)**

Hodnotené znaky	n	min	max	Priemer	Medián	Štandardná odchýlka
Hmotnosť plodu (g)	238	0,110	2,717	13,041	1,300	0,3909
Hmotnosť kôstky (g)	238	0,100	0,507	0,2331	0,227	0,0653

Zdroj: Vlastné výpočty

Klasický regresný model odhaduje, ako v priemere jednotlivé charakteristiky vplývajú na hmotnosť plodu. Zodpovedá na otázku, či je hmotnosť kôstky, šírka kôstky, hrúbka kôstky štatisticky signifikantne ovplyvňujú hmotnosť plodu. Klasický lineárny model však nedokáže zodpovedať na otázku, či hmotnosť kôstky ovplyvňuje rozdielne hmotnosť plodu keď je hmotnosť plodu veľká alebo malá. Regresné koeficienty kvantilovej regresie odhadujú zmenu v danom kvantile vysvetľovanej závisle premennej vyvolané jednotkovou zmenou vysvetľujúcej premennej. Týmto spôsobom je možné zistiť ako jednotlivé percentily hmotnosti plodu môžu byť viac ovplyvnené charakteristikami kôstky ako iné percentily veľkosti plodu. Toto sa odráža v zmene regresných koeficientov.



Zdroj: Vlastné zobrazenie

Obr. 1: Odhady regresných koeficientov pre rôzne kvantily hmotnosti plodu

Krivky na obrázku 1 znázorňujú zmeny hodnôt regresných koeficientov odhadnutých pomocou kvantilovej regresie v závislosti na jednotlivých kvantiloch hmotnosti plodu *drieňa obyčajného* (*Cornus mas* L.). Šedá oblasť predstavuje ich 95 percentný interval spoľahlivosti. Vodorovná čierna čiara predstavuje odhad koeficientov pomocou metódy najmenších štvorcov. Z obrázku vidíme, že hodnota absolútneho člena má rastúci trend. Hodnota regresného koeficienta pre závisle premennú hmotnosť kôstky je v prípade klasickej regresie približne do 0,60 kvantilu hmotnosti plodu nadhodnotená a pre kvantily vyššie ako hodnota 0,60 je hodnota regresného koeficientu pre závisle premennú hmotnosť kôstky

podhodnotená. Šedá oblasť 95 percentného intervalu spoľahlivosti pre absolútny člen pretína priamku $y = 0$ iba pre nízke kvantily (približne do 0,10 kvantilu) hmotnosti plodu. Všade inde je prienik priamky $y = 0$ so šedou oblasťou nulový, t.j. koeficienty sú štatisticky signifikantné na hladine významnosti $\alpha = 0{,}05$.

Tab. 2: Výsledky kvantilovej regresie závislosti hmotnosti plodu od hmotnosti kôstky

Parameter	Ukazovateľ	Klasická lineárna regresia	Kvantilová regresia				
			0,1 kvantil	0,2 kvantil	0,5 kvantil	0,8 kvantil	0,9 kvantil
Intercept	Odhad	0,1456	0,2711	0,5999	0,755	0,9273	0,9897
	Štandardná chyba	0,2248	0,1508	0,1584	0,1166	0,1019	0,2002
KHMOT (hmotnosť kôstky)	P-hodnota	0,5178	0,0735	0,0002	< 0,0001	< 0,0001	< 0,0001
	Odhad	25,685	24,950	18,529	24,107	28,095	30,339
	Štandardná chyba	0,7472	0,5763	0,6691	0,5054	0,4111	0,8882
	P-hodnota	0,0007	< 0,0001	0,0061	< 0,0001	< 0,0001	0,0007

Zdroj: Vlastné výpočty

4. Záver

Cieľom príspevku je dať podrobný návod na využitie procedúry QUANTREG v SAS-e a tým prispieť k jej širšiemu využitiu. K naplneniu tohto cieľa napomáha aj detailná interpretácia číselných výstupov ako aj korešpondujúcej obrazovej časti výstupu. Postup výpočtu je demonštrovaný na experimentálnych dátach získaných na Slovenskej poľnohospodárskej univerzite v Nitre. Biodiverzita patrí medzi najunikátnejšie javy prírody. Nie je jednoduché ju podchytiť. Kvantilová regresia je jedným z krôčikov, ako sa priblížiť k reálnemu popisu závislostí znakov biologického materiálu. Kvantilová regresia je mimoriadne vhodná aj všade tam, kde sa prirodzene vyskytujú aj extrémne údaje a z povahy problému nie je vhodné ich vylúčiť.

Literatúra

SAS INSTITUTE INC. SAS 9.1. 2000. Help and Documentation.

BRINDZA, J. et al. 2005. Informačný systém pre evidenciu a hodnotenie genetických zdrojov rastlín. [cit. 2013-11-10]. Dostupné na internete:

http://www.fem.uniag.sk/uveu2005/zbornik/zbornik/sekcia_3/brindza.pdf

KOENKER, R. – HALLOCK, K. 2001. Quantile Regrression: An Introduction. In: *Journal of Economic Perspectives*, roč. 15, č. 4, s. 43-56

SAS 9.1 Proc Quantreg Documentation

Adresa autorov:

Beáta Stehlíková, prof. RNDr. CSc.
Paneurópska vysoká škola
Fakulta ekonomiky a podnikania
Tematínska 10, 851 03 Bratislava
stehlikovab@gmail.sk

Ján Brindza, doc. Ing. PhD.
Slovenská poľnohospodárska univerzita
Katedra genetiky a šľachtenia rastlín FAPZ
Tr. A. Hlinku 2, 949 76 Nitra
brindza.jan@gmail.sk

Schröterova trieda rozdelení
Schröter's Class of Distributions

Gábor Szűcs

Abstract: This article deals with modelling of claim number distribution in the collective risk model and focuses particularly on Schröter's class of discrete distributions. Paper contains an overview of basic notations and definitions, various calibration methodologies and solution of a model example. The article includes links to external resources which contain implementation of aforementioned calibration methods in the statistical software R (R Core Team, 2013).

Abstrakt: Článok sa zaoberá modelovaním rozdelenia počtu poistných plnení v modeli kolektívneho rizika, pričom sa zameriava predovšetkým na Schröterovu triedu diskrétnych rozdelení. Príspevok obsahuje prehľad základných označení a definícií, rôzne metodiky na kalibráciu parametrov Schröterovej triedy a riešenie motivačného príkladu. V článku sú uvedené odkazy na externé zdroje, ktoré obsahujú programovú implementáciu spomínaných kalibračných metód v rámci štatistického softvéru R (R Core Team, 2013).

Key words: claim number distributions, Schröter's class of distributions, calibration methods

Kľúčové slová: rozdelenia počtu poistných plnení, Schröterova trieda rozdelení, kalibračné metódy

JEL classification: C16, C88

1. Úvod

Model kolektívneho rizika patrí medzi najpopulárnejšie modely neživotného poistenia. Aplikuje sa predovšetkým na modelovanie celkovej výšky poistných plnení a definuje sa vzťahom

$$S = \sum_{i=1}^{N} X_i, \tag{1}$$

kde $X_i; i = 1,2, \ldots$ sú nezávislé a rovnako rozdelené náhodné premenné popisujúce výšky individuálnych poistných plnení, N predstavuje počet poistných plnení za určité obdobie (N je diskrétna náhodná premenná nezávislá od $X_i; i = 1,2, \ldots$), kým S je celková výška poistných plnení za zvolenú časovú jednotku. Ak platí vzťah (1), tak hovoríme, že náhodná premenná S má tzv. *zložené rozdelenie*.

S modelom kolektívneho rizika sa zaoberajú mnohé odborné publikácie. Jedným zo základných dielov v tejto oblasti je kniha Mikosch (2006), ktorý obsahuje všeobecnejšiu definíciu modelu (celková výška a počet plnení sa modelujú pomocou náhodných procesov). Ďalšou kvalitnou publikáciou je Dickson (2005), ktorá sa podrobne zaoberá modelovaním počtu poistných plnení. Práve to bude cieľom aj nášho príspevku: uviesť najdôležitejšie typy a triedy diskrétnych rozdelení, ktoré sa môžu používať na popísanie náhodnej veličiny N a detailne predstaviť tzv. Schröterovu triedu rozdelení. Poznamenáme, že táto publikácia je súčasťou väčšieho výskumného projektu zaoberajúceho sa s tzv. Schröterovou rekurziou, ktorá sa používa pri hľadaní pravdepodobnostného rozdelenia zloženej náhodnej premennej S.

Tento príspevok obsahuje tri hlavné kapitoly. V druhej časti sa definuje tzv. $\mathcal{R}_k$-trieda diskrétnych rozdelení a jej špeciálne prípady: Panjerova a Schröterova trieda. Tretia kapitola obsahuje rôzne metódy, ktoré sa môžu používať pri kalibrácii parametrov Schröterovho rozdelenia. V záverečnej časti je uvedené riešenie ilustračného príkladu a porovnanie výsledkov.

2. Rozdelenia počtu poistných plnení

Ako sme už spomínali v úvode, pri modelovaní počtu poistných plnení sa obvykle používajú diskrétne pravdepodobnostné distribúcie, ako napr. Poissonovo, binomické, negatívne binomické alebo logaritmické rozdelenie. V praxi sa však ukázalo, že tieto rozdelenia majú síce dobrú interpretáciu a vhodné štatistické vlastnosti, ale na druhej strane nie sú dostatočne „bohaté" nato, aby prijateľne popísali reálny vývoj počtu plnení v poistných portfóliách. Práve preto sa začali skúmať iné, všeobecnejšie typy a triedy rozdelení, ktoré zaručia kvalitnejší fit modelu. Jedným z možných prístupov, ktorý preferujeme aj v tomto článku, uvažuje nasledovnú rekurentnú definíciu diskrétnych rozdelení.

Definícia 1. (Dickson, 2005) Uvažujme diskrétnu náhodnú premennú N s nekumulatívnym pravdepodobnostným rozdelením $\{p_n\}_{n=0}^{\infty}$, kde $p_n = \Pr(N = n)\,; n = 0,1,2, \ldots$ Hovoríme, že rozdelenie $\{p_n\}_{n=0}^{\infty}$ patrí do rekurzívnej triedy $\mathcal{R}_k$, ak je splnený vzťah

$$p_n = \sum_{i=1}^{k} \left(a_i + \frac{b_i}{n}\right) p_{n-i} \quad \text{pre } n = 1,2, \ldots, \tag{2}$$

kde k je prirodzené číslo, a_i, b_i sú reálne parametre pre $i = 1,2, \ldots, k$ a $p_n = 0$ pre všetky $n < 0$.

V nasledujúcej časti uvedieme niektoré špeciálne prípady $\mathcal{R}_k$-tried, ktoré sa skúmali osobitne vo viacerých odborných publikáciách.

Trieda $\mathcal{R}_1$ sa nazýva **Panjerova trieda** rozdelení a definuje sa rovnicou

$$p_n = \left(a_1 + \frac{b_1}{n}\right) p_{n-1} \quad \text{pre } n = 1,2, \ldots \tag{3}$$

Ak rozdelenie náhodnej premennej N pochádza z triedy $\mathcal{R}_1$, tak používame označenie $N \sim Pan(a_1, b_1)$. V nasledujúcej tabuľke ponúkame prehľad najznámejších distribúcií patriacich do Panjerovej triedy.

Tab. 16: Rozdelenia patriace do Panjerovej triedy

rozdelenie	*označenie*	*vzťah s rozdelením* $Pan(a_1, b_1)$
Poissonovo	$Po(\lambda), \lambda > 0$	$Po(\lambda) \Leftrightarrow Pan(0, \lambda)$
Binomické	$Bin(m,p)$, $m \in \mathbb{N}, p \in (0;1)$	$Bin(m,p) \Leftrightarrow Pan\left(-\frac{p}{1-p}, \frac{(m+1)p}{1-p}\right)$
Negatívne binomické	$NegBin(r,p)$, $r > 0, p \in (0;1)$	$NegBin(r,p) \Leftrightarrow Pan(p, (r-1)p)$
Geometrické	$Geom(p), p \in (0;1)$	$Geom(p) \Leftrightarrow Pan(p, 0)$

Ďalšie podrobnosti o Panjerovej triede sa dajú nájsť napríklad v Dickson (2005) a Szűcs (2011).

Vráťme sa k Definícii 1. a uvažujme prípad $k = 2$, teda triedu $\mathcal{R}_2$. Podľa definície platí

$$p_n = \left(a_1 + \frac{b_1}{n}\right) p_{n-1} + \left(a_2 + \frac{b_2}{n}\right) p_{n-2} \quad \text{pre } n = 1,2, \ldots \tag{4}$$

Ak položíme $a_1 = a, a_2 = 0, b_1 = b, b_2 = c$, tak dostaneme podtriedu triedy $\mathcal{R}_2$ známu pod názvom **Schröterova trieda**. Ak rozdelenie veličiny N patrí do Schröterovej triedy, tak používame zápis $N \sim Schr(a,b,c)$ a platí

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \quad \text{pre } n = 1,2, \ldots, \tag{5}$$

kde a, b, c sú reálne parametre rozdelenia a $p_{-1} = 0$. Ako vidíme, Schröterova trieda je už trochu abstraktnejšia ako $\mathcal{R}_1$. V nasledujúcej vete je uvedený príklad, ako skonštruovať netriviálne rozdelenie patriace do Schröterovej triedy.

Veta 1. Uvažujme nezávislé náhodné premenné N_1 a N_2, pričom $N_1 \sim NegBin(r,p)$ a $N_2 \sim Po(\lambda)$. Definujme náhodnú veličinu $N_3 = N_1 + N_2$. Potom rozdelenie premennej N_3 patrí do Schröterovej triedy, t. j. $N_3 \sim Schr(a,b,c)$, s parametrami

$$a = 1-p, \quad b = (1-p)(r-1) + \lambda, \quad c = -\lambda(1-p). \tag{6}$$

Dôkaz. Viď v knihe Dickson (2005) na stranách 76-77 (treba položiť $\alpha = (1-p)$ a $\beta = (1-p)(r-1)$).

Vidíme, že Schröterovo rozdelenie sa dá „vyrobiť" pomerne jednoduchým spôsobom. Stretávame sa však aj s takou situáciou, keď $Schr(a,b,c)$-rozdelenie nevznikne ako súčet náhodných premenných pochádzajúcich zo známych teoretických distribúcií (stáva sa to najmä v rôznych praktických aplikáciách). Dá sa ukázať, že ku každej trojici (a,b,c) sa dá nájsť prislúchajúce diskrétne rozdelenie $\{p_n\}_{n=0}^{\infty}$ tak, aby platil vzťah (5).

3. Kalibrácia parametrov Schröterovej triedy

Uvažujme teraz, že máme k dispozícií historické dáta $\boldsymbol{n_1, n_2, \dots, n_d}$: počty poistných plnení v uplynulom období. Na základe týchto údajov dokážeme zostrojiť empirické pravdepodobnostné rozdelenie náhodnej premennej $\boldsymbol{N}$, označme ho ako $\boldsymbol{\Pi = (\pi_0, \pi_1, \pi_2, \dots, \pi_m)}$, kde $\boldsymbol{m = \max\{n_1, n_2, \dots, n_d\}}$. Našou úlohou je nájsť také rozdelenie patriace do triedy $\boldsymbol{Schr(a,b,c)}$, ktoré „je najbližšie" k nášmu empirickému rozdeleniu, t. j. potrebujeme nakalibrovať parametre $\boldsymbol{a,b,c}$. Najprv uvedieme počiatočnú kalibračnú metódu (tzv. *metódu kalibračných pätíc*), potom ďalšie postupy, ktoré sa líšia predovšetkým voľbou funkcie, ktorá slúži na určenie vzdialenosti medzi empirickým a fitovaným rozdelením.

Uvažujme vyššie zavedené označenia a definujme kalibračné pätice $\kappa_n = (\pi_{n-2}, \pi_{n-1}, \pi_n, \pi_{n+1}, \pi_{n+2})$ pre $n \in (2,3,\dots,m-2)$. Parametre Schröterovej triedy a,b,c hľadáme ako riešenie sústavy lineárnych rovníc typu

$$\pi_n = \left(a + \frac{b}{n}\right)\pi_{n-1} + \frac{c}{n}\pi_{n-2}, \qquad \pi_{n+1} = \left(a + \frac{b}{n+1}\right)\pi_n + \frac{c}{n+1}\pi_{n-1},$$

$$\pi_{n+2} = \left(a + \frac{b}{n+2}\right)\pi_{n+1} + \frac{c}{n+2}\pi_n.$$

Za predpokladu, že π_{n-2} a π_{n-1} sa nerovnajú nule (v danej kalibračnej pätici κ_n), neznáme a,b,c sa dajú jednoznačne vyjadriť a vypočítať z vyššie uvedeného systému rovníc. Postupným použitím každej kalibračnej pätice κ_n pre $n = 2,3,\dots,(m-2)$ dostaneme $(m-3)$ trojíc parametrov a_n, b_n, c_n. Vzniká otázka: ktorú trojicu a_n, b_n, c_n, $n = 2,3,\dots,(m-2)$ by sme mali zvoliť za finálny odhad koeficientov Schröterovej triedy? Odpoveď na túto otázku nám dá nasledovná pomocná funkcia:

$$\Sigma(a_n, b_n, c_n) = \sum_{j=1}^{m} \left| \pi_j - \left(a_n + \frac{b_n}{j}\right)\pi_{j-1} - \frac{c_n}{j}\pi_{j-2} \right| \quad \text{pre } n = 2,3,\dots,m-2. \tag{7}$$

Za definitívny odhad koeficientov zoberieme tú trojicu a_n, b_n, c_n, ktorá minimalizuje funkciu $\Sigma(a_n, b_n, c_n)$, teda môžeme písať

$$\left(\hat{a}, \hat{b}, \hat{c}\right) = \arg\min_{n \in 2,3,\dots,m-2} \Sigma(a_n, b_n, c_n). \tag{8}$$

Ako sme už naznačili, tento odhad bude slúžiť ako prvotný odhad parametrov Schröterovej triedy. Pri hľadaní trojice $\left(\hat{a}, \hat{b}, \hat{c}\right)$ používame program `schr.dist.calib`, ktorý sme vytvorili v prostredí štatistického softvéru R (R Core Team, 2013). V nasledovnej tabuľke ponúkame prehľad ďalších kalibračných metód, ktoré sú založené na optimalizácii (minimalizácii) uvedených pomocných funkcií. Prvý stĺpec obsahuje označenia metodík korešpondujúcich s názvami metód v spomínanom programe `schr.dist.calib`.

Tab. 2: Kalibračné metódy (na odhad parametrov Schröterovej triedy)[1]

názov metódy	*pomocná funkcia*	*minimalizačná úloha*
`abs1`	$\Sigma_1(a,b,c) = \sum_{j=1}^{m} \left\lvert \pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2} \right\rvert$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_1$
`abs2`	$\Sigma_2(a,b,c) = \sum_{j=1}^{m} \frac{\left\lvert \pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2} \right\rvert}{\left\lvert \left(a + \frac{b}{j}\right)\pi_{j-1} + \frac{c}{j}\pi_{j-2} \right\rvert}$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_2$
`abs3`	$\Sigma_3(a,b,c) = \sum_{j=1}^{m} \frac{\left\lvert \pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2} \right\rvert}{\pi_j}$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_3$
`quad1`	$\Sigma_4(a,b,c) = \sum_{j=1}^{m} \left(\pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2}\right)^2$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_4$
`quad2`	$\Sigma_5(a,b,c) = \sum_{j=1}^{m} \frac{\left(\pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2}\right)^2}{\left\lvert \left(a + \frac{b}{j}\right)\pi_{j-1} + \frac{c}{j}\pi_{j-2} \right\rvert}$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_5$
`quad3`	$\Sigma_6(a,b,c) = \sum_{j=1}^{m} \frac{\left(\pi_j - \left(a + \frac{b}{j}\right)\pi_{j-1} - \frac{c}{j}\pi_{j-2}\right)^2}{\pi_j}$	$\min_{(a,b,c)\in\mathbb{R}^3} \Sigma_6$

4. Riešenie ilustračného príkladu

Uvažujme určitý typ neživotného poistenia, pri ktorom sledujeme počet poistných plnení za zvolenú časovú jednotku: jeden týždeň. Máme k dispozícii historické dáta predstavujúce týždenné počty poistných plnení pripadajúcich na tisíc platných poistných zmlúv. Poznamenáme, že historické dáta sú v tomto prípade generované zo zmesi troch Poissonových rozdelení (nie sú to reálne dáta).[2]

Tab. 3: Výberové charakteristiky polohy dátového súboru

Minimum	Prvý kvartil	Medián	Priemer	Tretí kvartil	Maximum
0,00	6,00	11,0	11,5	16,0	33,0

Pre úplnosť by sme dodali aj ďalšie parametre súboru: dĺžka dátového vektora $d = 1000$, maximum $m = 33$. Počet týždenných poistných plnení, ozn. N, by sme chceli modelovať vhodným rozdelením z triedy $Schr(a,b,c)$, potrebujeme preto odhadnúť parametre triedy. Používame pritom vytvorenú funkciu `schr.dist.calib`[3], ktorá na základe zadaných údajov zostrojí empirickú nekumulatívnu distribúciu, vypočíta prvotné odhady $\hat{a}, \hat{b}, \hat{c}$ (metóda

[1] *Poznámka.* Ak v metódach `abs2` resp. `quad2` nastane, že $\left(a + \frac{b}{j}\right)\pi_{j-1} + \frac{c}{j}\pi_{j-2} = 0$ pre nejaké $j \in \{1,2,\dots,m\}$, tak tieto pravdepodobnosti vynecháme z funkcie Σ_2 resp. Σ_5. Rovnako postupujeme aj v prípade `abs3` resp. `quad3`, ak $\pi_j = 0$ pre nejaké $j \in \{1,2,\dots,m\}$.

[2] Dátový súbor sa dá stiahnuť z adresy: http://www.iam.fmph.uniba.sk/ospm/Szucs/data/schroter-data02.txt

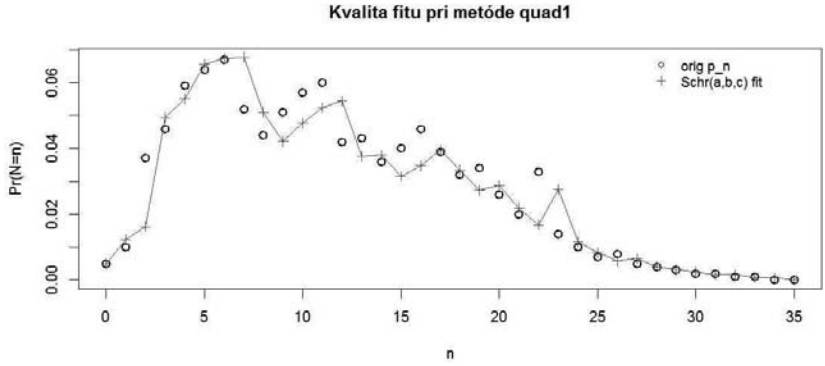
[3] Zdrojový kód funkcie `schr.dist.calib` je dostupný na internetovej adrese: http://www.iam.fmph.uniba.sk/ospm/Szucs/Schroter/schr.dist.calib.txt

`pentad`, t. j. metóda kalibračných pätíc), prípadne vylepší kalibráciu pomocou zvolenej kalibračnej metódy[4]. Súhrn výsledných odhadov uvádzame v nasledujúcej tabuľke.

Tab. 4: Nakalibrované parametre Schröterovho rozdelenia

par./metóda	`pentad`	`abs1`	`abs2`	`abs3`	`quad1`	`quad2`	`quad3`
$\hat{a}$	0,3767506	0,6916422	0,5564786	0,6217798	0,7584524	0,7321354	0,6498033
$\hat{b}$	0,9042428	1,4571205	1,4435217	1,3782206	1,7062588	1,9899676	1,5682988
$\hat{c}$	3,3806140	0,7314266	3,3594223	1,5728211	0,0632068	0,6206355	1,0774116
W^2 *(C-vM)*	4,9923	1,1097	4,5113	0,6921	0,4318	0,8676	0,7059
D (K-S)	0,1756	0,0726	0,0939	0,065	0,047	0,046	0,065

V predposlednom riadku *Tab. 4* sme uviedli hodnoty testovacej štatistiky Cramérovho-von Misesovho testu dobrej zhody za predpokladu (za platnosti nulovej hypotézy), že dáta pochádzajú z fitovaného rozdelenia. V poslednom riadku *Tab. 4* sú uvedené hodnoty testovacej štatistiky Kolmogorovovho-Smirnovovho testu (podrobnejšie viď v článku Arnold-Emerson, 2011). Pri výbere najvhodnejšej trojice parametrov $\hat{a}, \hat{b}, \hat{c}$ používame spomínané „meradlá vzdialenosti" medzi empirickým rozdelením a fitovanou distribúciou. Najlepší výsledok v zmysle Cramér-von Misesovej vzdialenosti W^2 nám dá kalibračná metóda `quad1`, kým Kolmogorovova-Smirnovova vzdialenosť D je najmenšia v prípade `quad2` resp. `quad1`. Za finálny odhad parametrov Schröterovho rozdelenia (v tomto konkrétnom ilustračnom príklade) by sme teda mohli zvoliť trojicu zo stĺpca `quad1`: $\hat{a} = 0{,}7584524\,; \hat{b} = 1{,}7062588; \hat{c} = 0{,}0632068$.



Obr. 34: Porovnanie empirického a fitovaného rozdelenia pri metóde quad1

5. Záver

Môžeme skonštatovať, že Schröterove rozdelenia by mohli byť vhodné na modelovanie počtu poistných plnení v modeli kolektívneho rizika. Ak nahradíme empirické rozdelenie nakalibrovaným Schröterovým rozdelením, tak samozrejme stratíme nejakú časť informácií a dopustíme sa nepresností. Na druhej strane však získame účinný nástroj na modelovanie

[4] Podrobná dokumentácia k funkcii `schr.dist.calib` je dostupná na stránke: http://www.iam.fmph.uniba.sk/ospm/Szucs/Schroter/schr.dist.calib-manual.pdf

budúcich počtov poistných plnení a hlavne zrýchlime výpočty (napr. kalkuláciu pravdepodobnostného rozdelenia celkovej výšky plnení). Ako sme už spomínali v úvode, výsledky uvedené v tomto článku sa používajú v rámci širšieho výskumného projektu. Ukázalo sa, že práve kvalitné modelovanie počtu plnení a čo najpresnejšia kalibrácia parametrov Schröterovej triedy sú kľúčovou úlohou pri používaní Schröterovej rekurzie a hľadaní rozdelenia celkovej výšky poistných plnení.

Poďakovanie

Tento článok vznikol s podporou grantu VEGA č. 2/0038/12.

Literatúra

ARNOLD, T. B. - EMERSON, J. W. 2011. Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions. *The R Journal*, č. 3/2, s. 34 – 39. ISSN 2073-4859.

DICKSON, D. 2005. *Insurance. Risk and Ruin*. Cambridge: Cambridge Univesity Press. ISBN 0-521-84640-4.

MIKOSCH, T. 2006. *Non-Life Insurance Mathematics*. Corrected Second Printing. Copenhagen: Springer, University of Copenhagen. ISBN-10 3-540-40650-6.

R CORE TEAM. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SZŰCS, G. 2011. *Panjerove rekurzie v neživotnom poistení*. Diplomová práca, Bratislava: Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave.

Adresa autora:

Gábor Szűcs, Mgr.
Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
Mlynská dolina, 842 48 Bratislava
szucs@fmph.uniba.sk

Demografické změny krajů České republiky mezi lety 2006–2011 z pohledu shlukové analýzy[1]
Demographic Changes in Regions of the Czech Republic between 2006–2011 as seen by Cluster Analysis

Ondřej Šimpach, Jitka Langhamrová

Abstract: The aim of the paper is to examine the similarity of regions in the Czech Republic according to various indicators from the area of demographic indicators using hierarchical cluster analysis method. Czech Republic has 14 regions in total. Regions are merged into the clusters according to the selected indicators using Euclidean distances. Selected attributes include the number of inhabitants, percentage of population aged 65+ in the total population, live births, deaths total, number of immigrants and number of emigrants (all in 31 Dec). The hierarchical clusterization of regions is calculated for each year based on data of 2006–2011 and next comparison is performed. Ascertained outputs can be used to plan community development and for urban planning such as transport and communications constructions, building of nurseries and basic schools and for decisions about placement of cultural facilities.

Abstrakt: Cílem předkládaného článku je prozkoumat podobnosti krajů České republiky podle různých ukazatelů z oblasti demografie s pomocí metody hierarchického shlukování. Česká republika má celkem 14 krajů. Tyto kraje budou spojeny do několika shluků v závislosti na vybraných indikátorech s využitím Euklidovské vzdálenostní metriky. Zvolené atributy zahrnují počty obyvatel v kraji, procentní zastoupení osob 65+ v populaci, živě narození celkem, zemřelí celkem, počet přistěhovalých a počet vystěhovalých (vše k okamžiku 31. prosince). Hierarchické shlukování krajů je vypočteno pro každý rok z období 2006–2011 a odlišné výsledky jsou spolu vzájemně porovnány.

Key words: Demographic indicators, Ward's method, Euclidean distances, Hierarchical Cluster analysis.

Klíčová slova: Demografické ukazatele, Wardova metoda, Euklidovské vzdálenosti, Shluková analýza.

1. Úvod

Nejenom pro účely územního plánování a rozhodování o investicích ve veřejném sektoru, ale i pro zjednodušování administrativních a ekonomických procesů je výhodné, známe-li podobnost vybraných územních celků navzájem na základě znalostí určitých socio-ekonomických faktorů. O investicích ve veřejném sektoru pojednává např. Nutt, (2006), který byl inspirací pro analýzu politiky soudržnosti určitých územních celků (viz např. Pechrová, Kolářová, 2012). Předkládaná studie čerpá inspiraci zejména od Lv et al., (2011), kteří ve své analýze využili obdobné socio-ekonomické ukazatele pro vytvoření shluků daných územních celků, nicméně jejich analýza byla zaměřena na populace pouze městského typu. Autoři Ozus et al., (2012) využili hierarchického shlukování pro hodnocení efektivnosti výstaveb multifunkčních obchodních center na území města. K jejich analýze bylo zapotřebí statistik o vývoji počtů zaměstnaných a nezaměstnaných osob v letech 1970–2000 a dále statistik z cestovního ruchu. Cílem této studie je prozkoumání podobnosti krajů České republiky, podle vybraných demografických ukazatelů (viz Lv et al., 2011) na základě hierarchické shlukové analýzy (Ward, 1963). Zjištěná podobnost může být využita k vysvětlení některých souvislostí (či naopak protikladů), se kterými se můžeme setkat v regionální socio-

[1] Článek byl podpořen z projektu Vysoké školy ekonomické v Praze IGA 6/2013 „Hodnocení výsledků metod shlukové analýzy v ekonomických úlohách".

hospodářské statistice, či administrativních a rozhodovacích procesech veřejného sektoru, (kterými se mj. zabýval např. Feldstein, 1964).

Česká republika má celkem 14 krajů (jednotek NUTS 3 příslušné klasifikace), jejichž výčet je uveden v Tabulce 1.

Tab. 17: Kraje České republiky (s definovanými zkratkami)

Hl. m. Praha	Hlavní město Praha	KHR	Královéhradecký
STČ	Středočeský	PAR	Pardubický
JIČ	Jihočeský	VYS	Vysočina
PLZ	Plzeňský	JIM	Jihomoravský
KVA	Karlovarský	OLM	Olomoucký
ÚST	Ústecký	ZLN	Zlínský
LIB	Liberecký	MSL	Moravskoslezský

Kraj Hlavní město Praha vychází ve většině publikovaných prací jako odlehlé pozorování, v případě shlukování dojde nejspíše k vytvoření jediného a vzdáleného samostatného shluku (viz např. Řezanková et al., 2011 nebo Löster, 2012). Hierarchické shlukování krajů bude vypočteno na základě vybraných údajů (Lv et al., 2011 nebo Arnio, Baumer, 2012) z let 2006, 2007, … a 2011, přičemž data byla pořízena z databáze Českého statistického úřadu (ČSÚ) a databáze Ministerstva zemědělství (MZe). Databáze MZe s podrobností na obce byla v minulosti využita např. i k analýze vybraných okresů (Šimpach, 2013). Vývoj těchto shluků bude tedy možné porovnat v šestiletém časovém horizontu.

2. Metodika a data

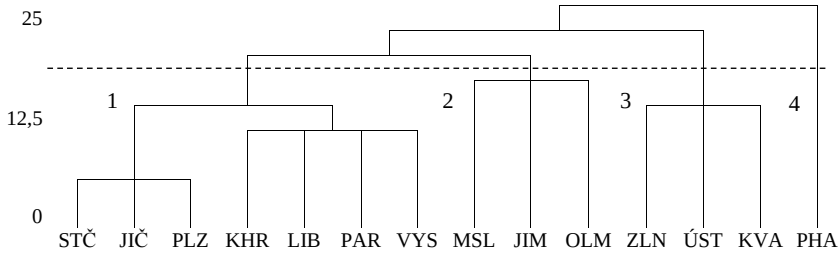
Vzdálenosti mezi jednotlivými kraji České republiky jsou vypočteny s využitím Euklidovské metriky (čtvercových vzdáleností). Poté jsou na základě známých matic vzdáleností rozděleny do 4–5 shluků (s ohledem na vhodnost zařazení do příslušného shluku (viz např. Löster, 2012)) a pochopitelně v závislosti na vybraných demografických ukazatelích, s využitím hierarchického shlukování a Wardovy metody (viz např. Danielson, 1980 nebo Bavaud, 2010). Počty shluků vychází ve většině případů 4 a vyplývají z dendrogramů, jejichž řez byl proveden vždy na stejné vzdálenosti, aby byly jednotlivé výsledky mezi sebou srovnatelné a dále pak z doporučení udané CHF indexem, zvaným též pseudo *F* index (viz Calinski, Habarasz, 1974 a dále aplikace Löstera, 2011), založeném na podílu průměrné mezishlukové a průměrné vnitroshlukové variability. Ze zmíněných datových matic, pořízených z databází ČSÚ a MZe, byly vybrány na základě zmiňovaných literárních zdrojů statistiky o

- počtu obyvatelích v daném kraji,
- podílu osob 65+ v populaci,
- počtu živě narozených celkem,
- počtu zemřelých celkem,
- počtu přistěhovalých a
- počtu vystěhovalých,

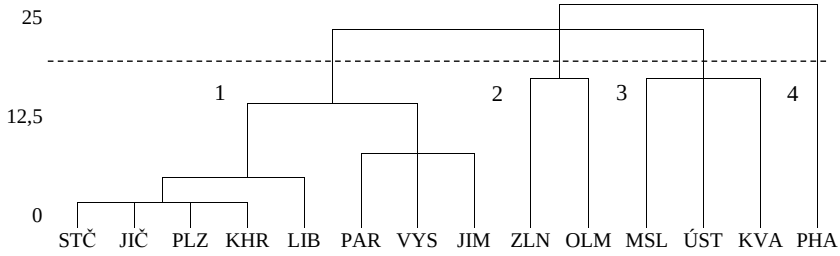
vše aktuální k 31. prosinci roků 2006, 2007, … a 2011. Výpočty vzdálenostních matic byly prováděny v systému IBM SPSS Statistics, na základě nichž byly konstruovány dále prezentované dendrogramy.

3. Výsledky

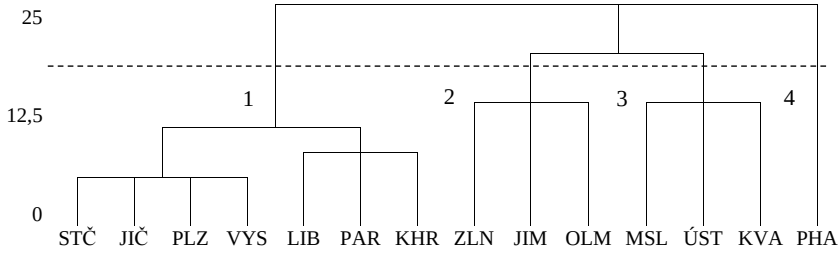
S využitím z-transformace, hierarchického shlukování založeném na Wardově metodě (Ward, 1963) a čtvercových Euklidovských vzdálenostních metrikách byly vypočteny shluky pro kraje České republiky v letech 2006, 2007, … a 2011. Řezy dendrogramy byly provedeny na vzdálenosti 19 jednotek, čímž došlo ve většině případů k vytvoření 4 shluků, v jednom 5. Dendrogramy jsou postupně zobrazovány v obrázcích 1–6.



Obr. 1: Dendrogram pro kraje ČR v roce 2006. (zdroj: vlastní výpočet a konstrukce)



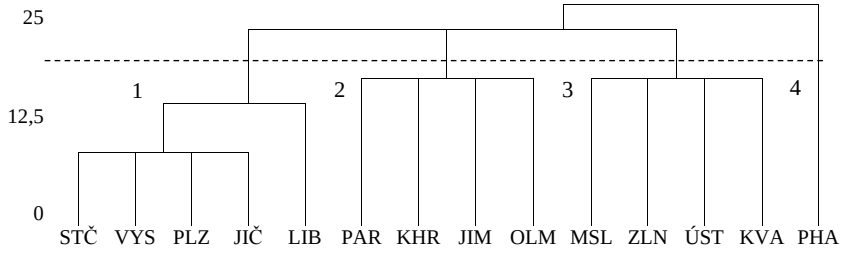
Obr. 2: Dendrogram pro kraje ČR v roce 2007. (zdroj: vlastní výpočet a konstrukce)



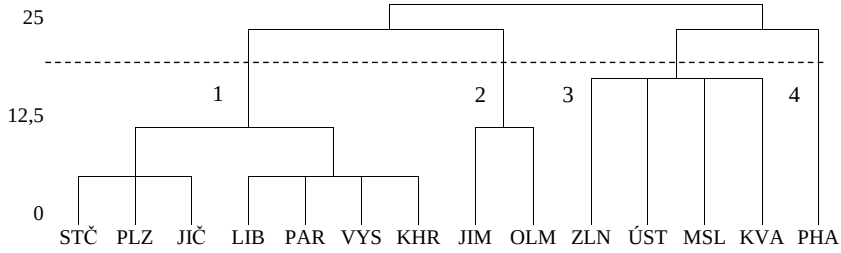
Obr. 3: Dendrogram pro kraje ČR v roce 2008. (zdroj: vlastní výpočet a konstrukce)

Při pohledu na situaci v roce 2006 vidíme pod označením „1“ velký shluk sedmi pouze českých krajů. Ty jsou podobné zejména vyšším podílem seniorů nad 65 let v populaci a nižšími počty živě narozených dětí. Také se jedná o kraje migračně atraktivní a převažuje u

nich kladné migrační saldo. Kraj Karlovarský a Ústecký je zahrnut ve shluku číslo „3“ spolu s jedním moravským krajem – Zlínským. Tyto kraje spolu souvisí zejména vyšší mírou emigrace než ostatní kraje. Kraj Hlavní město Praha tvoří jeden samostatný shluk proto, že je ve většině srovnávaných statistik výrazně odlišný od ostatních krajů. Ve shluku číslo „2“ se nachází kraje pouze moravské – Moravskoslezský, Jihomoravský a Olomoucký. V roce 2007 je zajímavé pozorovat, že do velkého shluku s označením „1“ vstupuje Jihomoravský kraj. Statistiky, které vstoupily do analýzy, byly v roce 2007 v Jihomoravském kraji obdobné, jako u zmiňovaných českých krajů. Shluk číslo „2“ je tedy zmenšen o jednoho člena, shluk číslo „3“ zůstal nezměněn, pouze se mírně změnily hodnoty v matici vzdáleností. Kraje Moravskoslezský, Ústecký a Karlovarský mají každoročně mnohem vyšší míry emigrace než všechny ostatní kraje, je to způsobeno zejména horšími pracovními podmínkami v regionech a méně rozvinutou infrastrukturou. V roce 2008 se některé kraje přeuspořádaly v rámci velkého shluku „1“, Jihomoravský kraj se přesunul od českých krajů k moravským do shluku „2“ (ke Zlínskému a Olomouckému). Shluk číslo „3“ zůstal v roce 2008 od předchozího roku nezměněn. Kraj Hlavní město Praha je nejvíce odlišný od ostatních krajů zejména vyšším podílem seniorů nad 65 let v populaci, (jde o regresivní typ populace), a dále vyšší mírou imigrace.

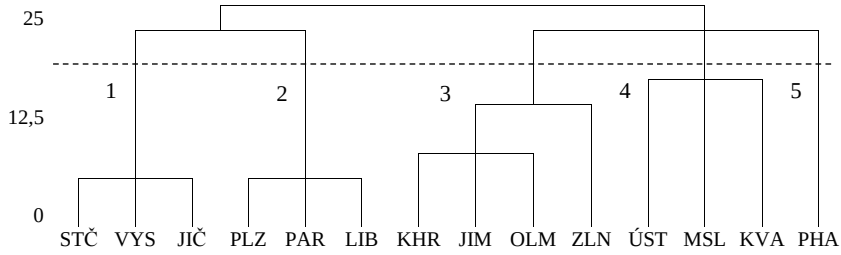


Obr. 4: Dendrogram pro kraje ČR v roce 2009. (zdroj: vlastní výpočet a konstrukce)



Obr. 5: Dendrogram pro kraje ČR v roce 2010. (zdroj: vlastní výpočet a konstrukce)

Rok 2009 byl ve znamení přeuspořádání členů mezi jednotlivými shluky. Uspořádání na obrázku 4 je mnohem rovnoměrnější a shluky jsou podobně velké. Byl to rok silného ekonomického poklesu, který rozhodně měl vliv i na zmíněné ukazatele z oblasti demografie, které rozhodovaly o takovémto výsledku. Nejvíce rovnoměrné byly v tomto roce statistiky imigrace a emigrace jednotlivých krajů. Počty zemřelých se dlouhodobě mění jen nepatrně, ale počty živě narozených na zmíněné události reagovaly.



Obr. 6: Dendrogram pro kraje ČR v roce 2011. (zdroj: vlastní výpočet a konstrukce)

Srovnáme-li rok 2009 s rokem 2010, zjistíme, že rok 2010 byl z pohledu výsledků mnohem více variabilní. Do soustavy se navrátil velký shluk číslo „1“, opět tvořený pouze českými kraji. Jihomoravský a Olomoucký kraj (oba z Moravy) tvoří shluk číslo „2“, ve shluku číslo „3“ zůstaly stejné kraje jako v roce 2009. U Zlínského, Ústeckého, Moravskoslezského a Karlovarského kraje pravděpodobně nedošlo k žádným významnějším změnám z pohledu sledovaných ukazatelů. Hlavní město Praha si drží svou suverenitu vždy ve čtvrtém samostatném shluku, jak v roce 2009, tak i ve zbývajících letech. Poslední dendrogram na obrázku 6 je jediným, kde došlo k rozdělení krajů České republiky do pěti shluků. Původní velký shluk se rozpadl na dva menší. K Jihomoravskému a Olomouckému kraji, které byly v roce 2010 ve shluku číslo „2“ (v roce 2011 už číslo „3“), přibyl Královéhradecký a Zlínský kraj. Ve shluku dříve označeném „3“ (nyní „4“), bývaly a i se nyní nachází kraje Ústecký, Moravskoslezský a Karlovarský. Tyto kraje v roce 2011 opět zaznamenávaly nejvyšší záporná migrační salda – mechanický úbytek obyvatelstva.

4. Diskuse a závěr

Výsledky této studie jsou do velké míry ovlivněny použitou metodikou. Při využití jiné než Wardovy metody bychom dostali shluky jiné (viz např. Löster, 2011). Obdobně pak nenormalizovaná vstupní data mění výsledky zásadní měrou. V našem případě však normalizace byla zapotřebí, neboť vstupní data nebyla ve srovnatelných jednotkách.

Demografické ukazatele, které spolu nejvíce souvisí a spojují jednotlivé kraje, jsou především podíl obyvatel 65letých a starších v celkové populaci a dále pak statistiky migrace. Kraje, které mají již dlouhodobě regresivnější populační strukturu, bývají spojovány do stejných shluků a kraje, které dlouhodobě působí jako migračně neatraktivní pak také. Počty obyvatel a počty zemřelých celkem se v dlouhém období mění jen velmi nepatrně, proto přesouvání krajů mezi jednotlivými shluky nehrozí ze strany těchto statistik. Studie by v budoucnu mohla být rozšířena o další socio-ekonomické ukazatele či národohospodářské indikátory. Např. studie Löstera a Langhamrové (2011) poskytuje mnohé informace o vývoji nezaměstnanosti, a právě míra nezaměstnanosti a další statistiky z trhu práce by byly adekvátním doplňkem pro další kalkulace.

Literatura

ARNIO, Ashley N., BAUMER, Eric P. (2012). Demography, foreclosure, and crime: Assessing spatial heterogeneity in contemporary models of neighborhood crime rates, *Demographic Research*, Vol. 26, (May 2012), p. 449-486.

BAVAUD, F. (2010). Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs, *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, Volume 6321, 2010, pp 103-118.

CALINSKI, T., HARABASZ, J. (1974). A Dendrite Method for Cluster Analysis, *Comunications in Statistics*, No. 3, 1974, pp. 1-27.

DANIELSON, Per-Erik (1980). Euclidean distance mapping, *Computer Graphics and Image Processing*, Volume 14, Issue 3, November 1980, Pages 227–248.

FELDSTEIN, Martin S. (1964). Net Social Benefit Calculation and the Public Investment Decision, *Oxford Economic Papers New Series*, Vol. 16, No. 1 (Mar., 1964), pp. 114-131.

LÖSTER, T. (2011). *Hodnocení výsledků metod shlukové analýzy*. (Doktorská disertační práce). Praha : FIS VŠE v Praze, 2011, 137 s.

LÖSTER, T. (2012). Kritéria pro hodnocení výsledků shlukování se známým zařazením do skupin založená na konfuzní matici. *Forum Statisticum Slovacum [online]*, 2012, roč. 8, č. 7, s. 85–89. ISSN 1336-7420. URL: http://ssds.sk/casopis/archiv/2012/fss0712.pdf.

LÖSTER, T., LANGHAMROVÁ, J. (2011). Analysis of Long-Term Unemployment in the Czech Republic. Praha 22.12.2011–23.12.2011. *In: LÖSTER, Tomáš, PAVELKA, Tomáš (ed.). International Days of Statistics and Economics*. Slaný : Melandrium, 2011, s. 228–234. ISBN 978-80-86175-77-5.

LV, J., LIU, QM., REN, YJ., GONG, T., WANG, SF. and LI, LM. (2011). Socio-demographic association of multiple modifiable lifestyle risk factors and their clustering in a representative urban population of adults: a cross-sectional study in Hangzhou, China, *International Journal of Behavioral Nutrition and Physical Activity*, 2011, Vol. 8 : 40.

NUTT, Paul, C. (2006). Comparing Public and Private Sector Decision-Making Practices, *Journal of Public Administration Research and Theory*, (April 2006) 16 (2): 289-318.

OZUS, E., AKIN, D., and ÇIFTÇI, M. (2012). Hierarchical Cluster Analysis of Multicenter Development and Travel Patterns in Istanbul, *Journal of Urban Planning and Development*, Volume 138, Issue 4 (December 2012), 303–318.

PECHROVÁ, M., KOLÁŘOVÁ, A. (2012). Does the cohesion policy mitigate the disparities among the regions in the Czech Republic?. Karviná 09.11.2012. *In: Mezinárodní vědecká konference doktorandů a mladých vědeckých pracovníků [CD-ROM]*. Opava : Slezská univerzita, 2012, p. 224–234. ISBN 978-80-7248-800-1.

ŘEZANKOVÁ, H., LÖSTER, T., HÚSEK, D. (2011). Evaluation of Categorical Data Clustering. Fribourg 26.01.2011–28.01.2011. *In: Advances in Intelligent Web Mastering – 3*. Berlin : Springer Verlag, 2011, s. 173–182. ISBN 978-3-642-18028-6. ISSN 1867-5662.

ŠIMPACH, O. (2013). Application of Cluster Analysis on the Demographic Development of Municipalities in the Districts of Liberecky Region. Prague 19.09.2013 – 21.09.2013. *In: International Days of Statistics and Economics at VŠE, Prague*. Prague : VŠE, 2013, s. 1390–1399. ISBN 978-80-86175-87-4.

WARD, J. H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236–244.

Adresa autorů:

Ondřej Šimpach, Ing.
KDEM FIS VŠE v Praze
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
ondrej.simpach@vse.cz

Jitka Langhamrová, doc., Ing., CSc.
KDEM FIS VŠE v Praze
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
langhamj@vse.cz

Vliv poslední dekády na průměrnou délku vzdělání v České republice[1]
Last decade impact on Average Length of Education in the Czech Republic

Petra Švarcová, Pavla Tůmová

Abstract: It is generally known, that the number of university graduates has increased dramatically in recent years in the Czech Republic. In our paper we analyse how this increase is reflected in the structure of education and related characteristics - the average length of education. This paper follows previous studies focused on changes in the nineties, when also some shift in education system was done. Because only census is relevant and reliable source of data about education structure, new results from the last census in 2011 were used. The paper develops previous studies by analysis of education structure by size of municipality, what could describes concentration of high educated people in bigger cities and could causes problems to smaller regions.

Abstrakt: Je obecně známo, že počet absolventů vysokých škol v České republice v posledních letech se dramaticky zvýšil. V naší studii analyzujeme, jak se tento nárůst promítl do struktury vzdělání a s tím související charakteristiky -průměrné délky vzdělání. Článek navazuje na předchozí studie, které se zabývaly změnami v devadesátých letech, kdy také došlo k určitému posunu ve vzdělávacím systému. Protože jediným spolehlivým a dostupným zdrojem pro zjištění vzdělanostní struktury obyvatelstva je census, bylo využito dat z posledního Sčítání lidu, domů a bytů v roce 2011. Článek rozšiřuje původní studie o analýzu vzdělanostní struktury podle velikosti obce, což je charakteristika, která ukazuje na koncentraci více vzdělaných osob do větších měst. To může způsobovat problémy zejména menším regionům.

Key words: Average Length of Education, Census 2011, HumanCapital

Klíčová slova: Průměrná délka vzdělání, SLDB 2011, Lidský kapitál

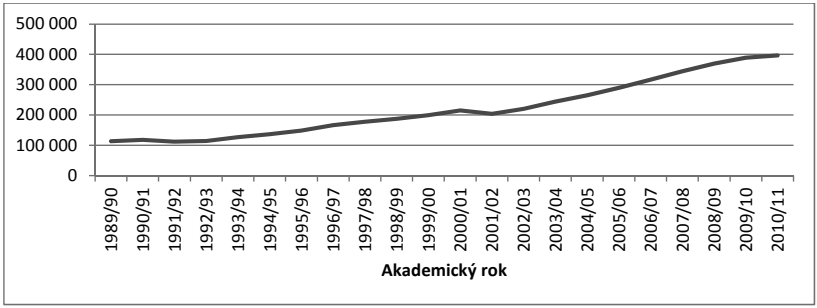
JEL classification: I21, I24, J24

1. Úvod

Lidský kapitál a vzdělání byly spojovány již od vzniku termínu lidský kapitál (A.Smith, 1776). V souvislosti s cíli Bílé knihy (2001) rozšířit počet studentů asi o 12%[2] a dalšími doporučeními v následujícím období došlo v posledních letech v České republice k dramatickému nárůstu absolventů vysokých škol, jak veřejných, tak i soukromých. Ve snaze vyrovnat se podílem vysokoškolsky vzdělaných západním zemím se počet studentů téměř zčtyřnásobil, jak lze vidět na obr. 1.

[1] Příspěvek vznikl za podpory Vysoké školy ekonomické v Praze - Interní grantové agentury; projekt č. IGA 9/2013 "Kvantifikace dopadů vzdělávací politiky poslední dekády ve světle výsledků SLDB 2011."

[2]Bílá kniha (2001): "Celkový počet studentů v terciérní sféře vzělávání bude činit 250 000 (asi 195 000 v bakalářských a magisterských studijních programech, 15 000 doktorandů, 30 000 studentů vyších odborných škol a 10 000 studentů na soukromých vysokých škoách) oproti současnému stavu 223 000, což představuje navýšení celkových počtů oproti současnoti asi o 12 %.", online na http://aplikace.msmt.cz/pdf/bilakniha.pdf



Obr. 17Počet studentů veřejných a soukromých VŠ celkem, vč. doktorského studia

Fenomén růstu počtu studentů a absolventů se netýká pouze České republiky, podle článku Barra a Lee (2001) došlo k zlepšení přístupu k vysokoškolskému studiu (ale nejen k němu, lze pozorovat zlepšení přístupu ke vzdělání jako celku) téměř v celém světě. Při rozdělení zemí do 9 skupin byly vyčleněny země, u kterých došlo k hospodářskému zvratu. U těchto zemí byl mezi roky 1990 – 2000 zaznamenán, jako u jediných, mírný pokles v ukazateli průměrná délka vzdělání, u ostatních tento ukazatel rostl.

Růst vysokoškolsky vzdělaného obyvatelstva (obecně se nemusí jednat pouze o nejvyšší vzdělání, ale může jít i posun mezi nižšími stupni) vede k růstu hodnot u charakteristik, které měří úroveň lidského kapitálu. V této práci bude využit ukazatel „průměrná délka vzdělání" (Average Length of Education). Práce navazuje na výpočty z publikace Mazoucha a Fischera (2011), kde byl sledován vývoj průměrné délky vzdělání pro jednotlivé kraje mezi lety 1991 a 2001.

Pro analýzy týkající se vzdělanostní struktury obyvatelstva existuje velmi omezené množství zdrojů. Nejlépe využitelným zdrojem pro tento typ analýzy jsou data ze Sčítání domů, lidu a bytů (SLDB) a právě nová data ze SLDB z roku 2011 umožňují aktualizovat ukazatele jako průměrnou délku vzdělávání. Z již existujících studií např. Mazoucha a Fischera (2011) víme, že struktura a úroveň lidského kapitálu není ve všech krajích ČR na stejné úrovni. Protože v jednotlivých regionech je také různé zastoupení různých velikostních skupin obcí, dá se předpokládat, že na rozdělení lidského kapitálu má také vliv velikost obce. Vysokoškolsky vzdělané obyvatelstvo bude pravděpodobně alokováno ve větších městech, která jim mohou nabídnout vyšší možnosti uplatnění.

Předmětem tohoto příspěvku není pouze analýza vzdělanostní struktury obyvatelstva a jejího vývoje, resp. změny v čase, ale chceme také zohlednit délku jednotlivých stupňů studia a proto volíme jako nejvhodnější charakteristiku průměrnou délku vzdělání (ALE).

2. Data

Jak je uvedeno výše, data pro analýzy týkající se struktury vzdělání existují pouze v omezené míře. Jedním z mála šetření, které zjišťuje taková data, je Sčítání lidu, domů a bytů. Nevýhodou šetření je skutečnost, že probíhá pouze jednou za deset let, na druhou stranu však poskytuje vyčerpávající údaje o obyvatelstvu, které by jindy nebylo možné zjistit. Poslední sčítání na území ČR proběhlo v březnu roku 2011. Rozhodným okamžikem byla půlnoc z 25. na 26. března.

Povinnost občanů zúčastnit se šetření je dána zákonem č. 296/2009 Sb. o sčítání lidu, domů a bytů v roce 2011, přesněji podle § 7 - Povinnost poskytnout údaje.V rámci sčítání občané ČR vyplňují až tři sčítací formuláře a to Sčítací list osoby, Domovní list a Bytový list.

Data pro naši analýzu jsou získána přímo z otázky týkající se nejvyššího ukončeného vzdělání (otázka 12, Sčítací list osoby)[3]. Zjišťování a následné zpracování dat zajišťuje Český statistický úřad.

3. Metodika výpočtu

Metodika výpočtu průměrné délky vzdělání je převzata z publikace Mazoucha a Fischera (2011). Při výpočtu průměrné délky vzdělání je v první řadě nutné definovat délky jednotlivých stupňů studia. Každý stupeň formálního studia je ohodnocen příslušným počtem let jeho standardní délky.

Tento ukazatel je vhodným a snadným ukazatelem pro hodnocení lidského kapitálu, který má kvantitativní povahu. Nevýhodou tohoto ukazatele však může být volba koeficientů l_k (viz vzorec 1), představující ohodnocení jednotlivých stupňů studia. Jedná se o skutečnost, že v průběhu let se měnil počet let strávených na jednotlivých stupních studia. Je nutno respektovat, že zde existují ročníky absolventů, které např. na základní škole strávily na místo dnešních 9 let pouze 8 let. I když zde tento problém existuje, ve výpočtu budeme uvažovat délky studia, které odpovídají současnému stavu.

V naší práci uvažujeme pět stupňů studia s následujícím počtem let:

Tab. 10 Jednotlivé stupně vzdělání a jejich počet let studia

Úroveň vzdělání	Skupina k	Počet let délky studia l_k
Bez vzdělání	1	0
Základní vzdělání včetně neukončeného	2	9
Střední bez maturity	3	12
Střední s maturitou *	4	13,7**
Vysokoškolské	5	18

*Včetně nástavbového a vyššího odborného

**Počet let pro 4. skupinu byl vypočten na základě expertního odhadu podle podílů absolventů různých typů středního vzdělání s maturitou ve společnosti

Průměrná délka vzdělání vychází ze statistických metod pro výpočet střední hodnoty pomocí relativní četnosti. Úpravou vzorce pro střední hodnotu získáváme vzorec pro výpočet průměrné délky vzdělání v následujícím tvaru:

$$ALE = \sum_{k=1}^{5} f_k * l_k, \quad (1)$$

kde l_k je celkový počet let studia nutný k dosažení příslušného vzdělanostního stupně,

f_k je relativní četnost příslušné vzdělanostní skupiny.

Průměrná délka vzdělání byla následně vypočtena pro všech 14 krajů České republiky, dále jsme porovnávaly průměrnou délku vzdělání podle velikosti obce. Pro naše účely byly obce rozděleny do 5 kategorií podle následujícího klíče:

1 – 999 obyvatel

1 000 – 4 999 obyvatel

[3]Sčítání lidu, domů a bytů - http://www.czso.cz/sldb2011/redakce.nsf/i/o_scitani

5 000 – 19 999 obyvatel
20 000 – 99 999 obyvatel
Nad 100 000 obyvatel

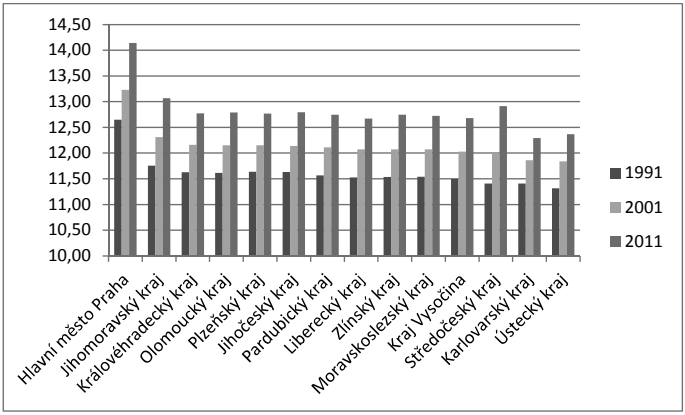
Do výpočtu byli zahrnuti pouze občané starší 25 let.

4. Výsledky

Mezi roky 1991 a 2001 došlo k nárůstu průměrné délky vzdělávání (ALE) v jednotlivých krajích, a to v průměru 0,54 roku (Mazouch, Fischer, 2011). Hlavním cílem bylo zjistit a porovnat hodnotu průměrné délky vzdělávání na základě dat ze SLDB 2011. V první části naší analýzy jsme se zaměřily na porovnání vývoje ALE v čase. Vývoj pro jednotlivé kraje nám ukazuje následující graf. Na něm můžeme vidět, že ve všech krajích došlo k nárůstu průměrné délky studia. V průměru se jedná o růst 0,54 let mezi lety 1991 a 2001 a 0,66 rokumezi lety 2001 a 2011.

Vidíme, že nejvyšší ALE má ve všech sledovaných letech Hlavní město Praha, kde hodnota vzrostla od roku 1991 z 12,65 let na 13,23 let v roce 2001 a 14,14 let v roce 2011.U tohoto kraje můžeme sledovat i druhý největší růst ALE mezi sledovanými roky. Tato změna představuje nárůst ALE mezi roky 1991 a 2011 o 1,49 roku. Největší nárůst ALE měl kraj Středočeský, kde průměrná délka vzdělání vzrostla od roku 1991 z původních 11,4 přes 12,01 let v roce 2001 až na současných 12,9 v roce 2011. Tato skutečnost může být dána stěhováním obyvatel co nejblíže k Praze kvůli větším možnostem nalézt kvalifikovanou práci.

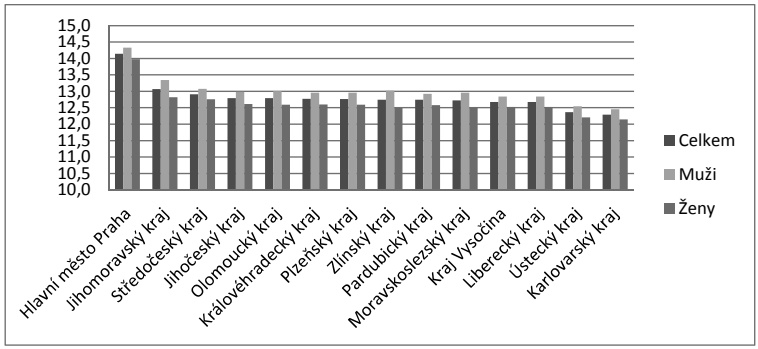
K růstu průměrné délky vzdělání došlo ve všech krajích ČR. Všechny kraje kromě Jihomoravského, Středočeského, Karlovarského a Ústeckého jsou nyní s průměrnou délkou vzdělání na stejné hodnotě, přibližně 12,7 let. U Jihomoravského a Středočeského kraje je nárůst ALE až na 13 let. Nejnižší hodnoty ve všech sledovaných obdobích mají kraje Karlovarský a Ústecký, kde se hodnoty ALE v roce 1991 pohybují kolem 11,4 let, v roce 2001 dochází k růstu na 11,9 roku a v roce 2011 na současných 12,3 let.



Obr. 18 Průměrná délka vzdělání v krajích ČR v letech 1991, 2001 a 2011

V případě, že sledujeme rozdělení průměrné délky vzdělání v roce 2011, můžeme ji sledovat nejen podle krajů, ale také podle pohlaví. Následující obrázek 3 nám ukazuje rozdělení ALE podle krajů a pohlaví.

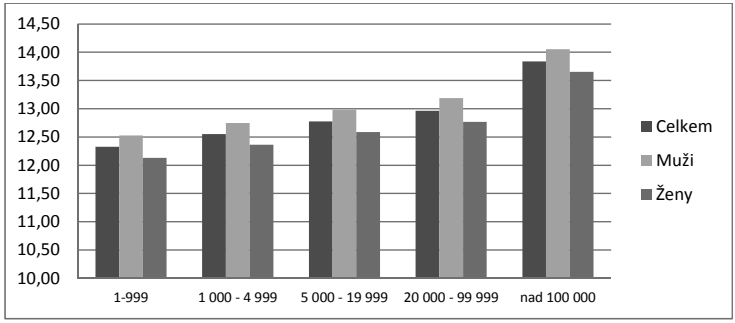
Můžeme vidět, že ve všech krajích včetně Hl. města Prahy mají vyšší délku vzdělání muži než ženy, a to v průměru o 0,38 roku. Absolutně nejnižší průměrnou délku vzdělávání můžeme nalézt u žen v Karlovarském kraji, kde tato hodnota nabývá pouze 12,14 let. Můžeme říci, že vzdělanostní struktura žen v Karlovarském kraji odpovídá struktuře vzdělání jako v ostatních krajích v roce 2001, tedy o deset let dříve.



Obr. 19 Průměrná délka vzdělávání podle pohlaví a kraje v roce 2011

Na základě výsledků z krajů a silného postavení Prahy můžeme usoudit, že průměrná délka vzdělávání se liší nejen podle kraje, ale také podle velikosti obce. Vysoká hodnota průměrné délky vzdělání v Hl. m. Praze nastiňuje problém urbanizace lidského kapitálu, kdy vzdělanější lidé odcházejí do větších měst, které jim může nabídnout lepší uplatnění svých znalostí nežli menší obce. Zároveň se však může jednat o studenty, kteří ve větších městech studovali a po studiu již v daném městě zůstali.

Tuto teorii potvrzuje následující graf, který ukazuje rozdělení průměrné délky vzdělání opět podle pohlaví a podle velikosti obce, přesněji podle počtu obyvatel obce. Vidíme, že v nejmenších obcích do 1000 obyvatel je průměrná délka vzdělání pouze 12,33, u mužů pak 12,53 a u žen pouze 12,13 let. Naopak s rostoucím počtem obyvatel v obci roste i průměrná délka vzdělání jak celková, tak i podle jednotlivých pohlaví. Nejvýraznější úroveň je pak u skupiny obcí nad 100 000 obyvatel, kde se průměrná délka vzdělání celkově pohybuje kolem 13,8 let, u mužů pak dokonce více než 14 let. U žen je to pak opět o něco méně, a to 13,65 let.



Obr. 20 Průměrná délka vzdělání podle vzdělání a velikosti obce

5. Závěr

Tento příspěvek měl za cíl sledovat úroveň vzdělání v krajích ČR a úroveň vzdělání podle velikost obce. Jako vhodný ukazatel jsme volily průměrnou délku vzdělání (ALE), která hodnotí nejen strukturu vzdělání ale současně i délku studia jednotlivých stupňů vzdělání. Jako vstupní data byla využívána data ze Sčítání, lidu a domů a bytů roku 2011.

Výsledky analýzy ukazují, že průměrná délka vzdělání v posledních deseti letech vzrostla, v průměru o 0,66 let. Je možná překvapivé, že tato hodnota není větší, díky velkému nárůstu studentů a absolventů vysokých škol, viz Obr. 1. Růst ALE výrazně nepřekročil hodnotu růstu mezi lety 1991 a 2001 (0,54 roku), což naznačuje na vysokou robustnost tohoto ukazatele.

Opět se ukázalo, že v roce 2011 i nadále dominuje Hl. město Praha, které rostlo od roku 1991 o 1,49 roku. Také je zřejmé, že pomyslné nůžky mezi kraji se nesvírají, ale naopak spíše rozevírají a kraje tak spíše divergují než konvergují, což může způsobit v budoucnu nemalé problémy.

Literatura

BARRO, R. J. – LEE J.-W. 2001. International Data on Educational Attainment: Updates and Implications, Oxford economic papers, ISSN 0030-7653, 2001, Vol. 53, No. 3, Special Issue on Skills Measurement and Economic Analysis, pp. 541-563,

MAZOUCH, P. – FISCHER, J. 2011. Lidský kapitál: měření, souvislosti, prognózy. Vyd. 1. Praha: C. H. Beck, 2011. 116 s. Beckova edice ekonomie. ISBN 978-80-7400-380-6

SMITH, A.: AnInquiryintotheNature And CausesoftheWealthofNationsBook 2 – OftheNature, Accumulation, and EmploymentofStock; Published 1776.05/2007, ISBN 0857710028, p. 199

Sčítání lidu, domů a bytů -http://www.czso.cz/sldb2011/redakce.nsf/i/o_scitani [15. 11. 2013]

Národní program rozvoje vzdělávání v České republice - Bílá kniha, Praha 2001, ÚIV, nakladatelstvíTaurus, ISBN 80-211-0372-8 –onlinehttp://aplikace.msmt.cz/pdf/bilakniha.pdf [16. 11. 2013]

Adresa autorov:

Ing. Petra Švarcová
Fakulta informatiky a statistiky
Vysoká škola ekonomická v Praze
Náměstí W. Churchilla 4
Praha 130 67 Praha 3
Petra.svarcova@vse.cz

Bc. Pavla Tůmová,
Fakulta informatiky a statistiky
Vysoká škola ekonomická v Praze
Náměstí W. Churchilla 4
Praha 130 67 Praha 3
tumova.pavla@gmail.com

Vývoj úverového trhu a ekonomický rast na Slovensku
Development of the credit market and economc growth in Slovakia

Michaela Urbanovičová, Beáta Gavurová

Abstract: Existence of efficient and well-functioning financial sector is an assumption of the growth of economic performance. On the other hand, development of an economy can encourage, but also hobbles the financial sector. The aim of this paper is to examine the relation between credit market development and economic growth for Slovakia using the linear regression and Granger causality. We have used the date on monthly basis and the analysed period lasts from January 2009 to November 2012. Credit market is represented by the volume of granted new-credits and economic growth by the index of industrial production. Granger causality test have shown that causality runs from economic growth to credit market, but not in the opposite direction. The empirical results have indicated that development of the economy in Slovakia positively encourages the credit market.

Abstrakt: Existencia efektívneho a dobre fungujúceho finančného sektora je predpokladom rastu výkonnosti ekonomiky. Naopak, rozvinutosť ekonomiky môže stimulovať, ale aj brzdiť rozvoj finančného sektora. Cieľom príspevku je skúmať vzťah medzi vývojom úverového trhu a ekonomickým rastom v podmienkach Slovenska využitím lineárnej regresie a Grangerovej kauzality. Pri analýze sú použité údaje na mesačnej báze v období od januára 2009 do novembra 2012. Premennou reprezentujúcou úverový trh je objem poskytnutých nových úverov a ekonomický rast reprezentuje index priemyselnej produkcie. Test Grangerovej kauzality preukázal, že ekonomický rast ovplyvňuje úverový trh, no neplatí to v opačnom smere. Skúmaním sme dospeli k záveru, že vývoj ekonomiky pozitívne podporuje úverový trh.

Key words: Credit market, economic growth, Granger causality.

Kľúčové slová: Úverový trh, ekonomický rast, Grangerova kauzalita.

JEL classification: O16

1. Úvod

Dôležitým odvetvím hospodárstva na Slovensku je bankový sektor. Počiatky jeho transformácie sú spojené s vytvorením dvojstupňovej bankovej sústavy pozostávajúcej z centrálnej a komerčných bánk. Tento krok znamenal začiatok procesu postupnej demonopolizácie bankovníctva, teda zmenšovania zásahov štátu do činnosti komerčných bánk. Vzhľadom na vzájomnú interakciu bankového sektora a ekonomiky je pre rast výkonnosti ekonomiky nevyhnutná efektívnosť a hospodárnosť bankového sektora, ktoré boli neskôr zabezpečené reštrukturalizáciou bankovníctva.

Interakciou medzi finančným systémom a ekonomickým rastom sa už v roku 1911 zaoberal Joseph Schumpeter, ktorý tvrdil, že služby poskytované finančnými sprostredkovateľmi sú nevyhnutné pre technologické inovácie a ekonomický vývoj. Levine (2002) zdôrazňuje významnosť bankového sektora pre ekonomický rast a poukazuje na situácie, kedy sa môžu komerčné banky aktívne podieľať na podpore inovácií a budúceho ekonomického rastu identifikovaním a financovaním produktívnych investícií. Niektorí ekonómovia zastávajú názor, že financie sú relatívne nevýznamným faktorom ekonomického vývoja. Robinson (1952) tvrdí, že finančný vývoj jednoducho nasleduje ekonomický rast. Lucas (1988) označuje vzťah medzi finančným a ekonomickým vývojom za príliš zdôrazňovaný.

2. Vzťah úverového trhu a ekonomického rastu

Vzájomný vzťah medzi vývojom úverového trhu a ekonomickým rastom je predmetom skúmania mnohých autorov, ktorí poukazujú na existenciu kauzality medzi nimi. Otázkou zostáva, či úverový trh ovplyvňuje ekonomický rast alebo naopak, ekonomický rast ovplyvňuje vývoj úverového trhu, príp. existuje komplementárny vzťah medzi nimi.

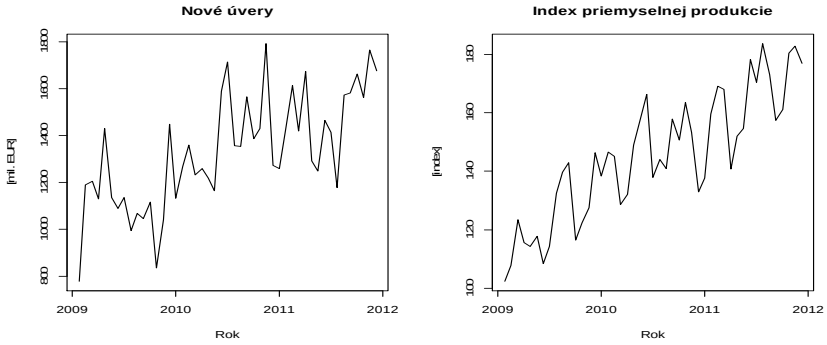
Koivu (2002) sa zaoberal vzťahom medzi finančným sektorom a ekonomickým rastom v transformujúcich sa krajinách. Za premenné reprezentujúce finančný sektor zvolil rozpätie medzi úrokovými sadzbami na úvery a vklady a objem úverov poskytnutých súkromnému sektoru ako podiel na HDP. Dospel k záveru, že prvá menovaná premenná negatívne ovplyvňuje ekonomický rast, čo je v súlade s teoretickými modelmi. No na druhej strane preukázal, že rast poskytnutých úverov neakceleruje ekonomiku. K podobnému záveru dospeli aj Cristea a Dracea (2010). V podmienkach Rumunska preukázali, že rast úverov poskytnutých súkromnému sektoru nepodporuje ekonomický rast, a teda vedie k relatívnemu poklesu miery ekonomického rastu. King a Levine (1993) taktiež skúmali predmetný vzťah a zistili, že vývoj bankového sektora môže poháňať ekonomický rast v dlhom období. Vazakidis a Adamopoulos (2011) preverovali interakciu medzi vývojom úverového trhu a ekonomického rastu v podmienkach Grécka. Ich výsledky potvrdili, že krátkodobý rast ekonomiky vyvoláva rast bankových úverov. Smer vplyvu od ekonomického rastu k úverového trhu potvrdil aj Adamapoulos (2010) v podmienkach Španielska.

3. Údaje a metodológia

Predmetom tohto príspevku je skúmať vzťah medzi vývojom úverového trhu v podmienkach Slovenska využitím lineárnej regresie. Ako premennú reprezentujúcu vývoj úverového trhu sme zvolili objem poskytnutých nových úverov (NU), nakoľko odzrkadľujú skutočný objem novoposkytnutých úverov v aktuálnom mesiaci. Pre ekonomický rast bol vybraný index priemyselnej produkcie (IPP) vzhľadom na to, že priemyselná produkcia tvorí veľkú časť produkcie krajiny a je vhodnou proxy premennou hrubého domáceho produktu (Eller, Frömmel a Srzentic, 2010).

Dáta sú čerpané z webových stránok Národnej banky Slovenska a Štatistického úradu Slovenskej republiky. Pri analýze sú použité údaje na mesačnej báze v období od januára 2009 do novembra 2012.

Grafické znázornenie vývoja a základná deskriptívna štatistika skúmaných premenných sú prezentované na obrázku (Obr. 1) a v tabuľke (Tab. 1).



Obr. 35: Grafická analýza premenných v modeli

Zdroj: Vlastné spracovanie podľa údajov NBS a ŠÚSR

Tab. 18: Deskriptívna štatistika premenných v modeli

	Nové úvery	**Index priemyselnej produkcie**
Počet pozorovaní	47	47
Stredná hodnota	1330,51	145,16
Smerodajná odchýlka	239,70	21,88
Rozptyl	57454,97	478,81
Medián	1292,88	145,00
Minimum	779,15	102,40
Maximum	1791,24	183,70

4. Výsledky testovania

Vzťah ekonomického rastu a úverového trhu analyzujeme pomocou modelu jednoduchej lineárnej regresie využitím Grangerovej kauzality. Za účelom odhadnúť a následne skúmaný vzťah kvantifikovať je potrebné otestovať prítomnosť jednotkového koreňa v časových radoch, ktoré do modelu vstupujú. Na testovanie jeho prítomnosti sme zvolili rozšírený Dickey Fuller test (ADF). Výsledky testovania sú poskytnuté v tabuľke (Tab. 2).

Tab. 19: ADF test

	NU	IPP
ADF – level	T	T
(testovacia štatistika)	-4.4839 ***	-5.8349 ***

T – trend a konštanta, C – konštanta, N – bez trendu a konštanty

Hladina významnosti: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zdroj: Vlastné spracovanie v programe R

Použitím ADF testu sme preukázali stacionaritu časových radov s konštantou a trendom na hladine významnosti 1 %, a teda sme prijali alternatívnu hypotézu o neprítomnosti jednotkového koreňa.

Korelačná matica, ktorú uvádzame v tabuľke 3, potvrdzuje silnú pozitívnu koreláciu medzi novými úvermi poskytnutými v bankovom sektore a vývojom ekonomického rastu.

Tab. 20: Korelácia

	Nové úvery	*Index priemyselnej produkcie*
Nové úvery	1	0,65
Index priemyselnej produkcie	0,65	1

Vzájomný vzťah medzi úverovým trhom a ekonomických vzťahom ďalej skúmame pomocou testu Grangerovej kauzality, ktorého výsledky sú uvedené v tabuľke (Tab. 4).

Tab. 21: Grangerova kauzalita

Kauzalita	F - štatistika
NU --> IPP	0,05
IPP --> NU	18,90 ***

Zdroj: Vlastné spracovanie v programe R

Na základe výsledkov testovania zamietame nulovú hypotézu a môžeme konštatovať, že vývoj ekonomického rastu grangerovsky vysvetľuje objem novoposkytnutých úverov, no

tento záver neplatí v opačnom smere. Teda vývoj úverov grangerovsky nevysvetľuje vývoj ekonomického rastu.

Z dosiahnutých výsledkov stanovíme rovnicu lineárnej regresie v tvare:

$$NU_t = \beta_0 + \beta_1 IPP_t \quad (1)$$

Stanovený model následne otestujeme na prítomnosť heteroskedasticity pomocou Breusch-Pagan testu a autokorelácie pomocou Durbin-Watson testu.

Tab. 22: Testovanie modelu

Normatita rozdelenia	Jarque – Bera test – p hodnota	0,78
Heteroskedasticita	Breusch-Pagan test – p hodnota	0,91
Autokorelácia	Durbin-Watson test – DW štatistika	1,63
Špecifikácia modelu	Reset test – p hodnota	0,48

Zdroj: Vlastné spracovanie v programe R

Výsledky testov poskytnuté v tabuľke (Tab. 5) preukázali, že model nie je zaťažený heteroskedasticitou, ani autokoreláciiou. Na základe Jarque – Bera testu môžeme konštatovať, že reziduá pochádza z normálneho rozdelenia a model je špecifikovaný správne.

Kvantifikovanie a testovanie významnosti indexu priemyselnej produkcie, ako aj upravený index determinácie sú prezentované v tabuľke (Tab. 6).

Tab. 23: Testovanie významnosti nezávislej premennej

	Odhad	Smerodajná odchýlka	t hodnota	p hodnota
Intercept	296,231	182,035	1,627	0,111
IPP_t	7,125	1,240	5,745	7,46e-07 ***
Upravený R^2	0,4103			
*Hl. významnosti: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1*				

Zdroj: Vlastné spracovanie v programe R

Z uvedeného vyplýva, že ak index priemyselnej produkcie vzrastie o jednu jednotku objem novoposkytnutých úverov porastie o 7,125 mil. EUR. Teda stimulácia ekonomického rastu podporuje úverový trh a s rastom produkcie rastie aj potreba zdrojov na financovanie nových aktivít v ekonomike.

5. Záver

Na zabezpečenie výkonnosti ekonomiky je nevyhnutná existencia efektívneho a dobre fungujúceho finančného sektora. Zdravý bankový sektor podporuje rast ekonomickej výkonnosti krajiny, no nefunkčný bankový systém môže negatívne ovplyvniť ekonomický rast. Naopak, rozvinutosť ekonomiky môže stimulovať, ale aj brzdiť rozvoj finančného sektora. Ich vzájomná interakcia je teda zrejmá.

Vzťah medzi ekonomickým rastom a vývojom úverového trhu je rozsiahlym predmetom empirických štúdií. Skúmaním uvedeného vzťahu v podmienkach Slovenskej republiky sme dospeli k záveru, že vývoj ekonomiky grangerovsky vysvetľuje objem novoposkytnutých úverov, no tento záver neplatí v opačnom smere. Rast ekonomiky, ktorý je reprezentovaný indexom priemyselnej produkcie, pozitívne podporuje objem poskytnutých nových úverov.

Príspevok prezentuje predbežné výsledky výskumu v súlade s podporeným projektom VEGA č. 1/1050/12 „Návrh systému merania výkonnosti v zdravotníckych zariadeniach na Slovensku a implementácia metrík výkonnosti.“

Literatúra

ADAMOPOULOS, A. 2010. The Relationship between Credit Market Development and Economic Growth. In: *American Journal of Applied Sciences*, roč. 7, č. 4, s. 518 – 526. ISSN 1546-9239.

CRISTEA, M. – DRACEA, R. 2010. Does credit market accelerate economic growth in Romania? Statistical approches. RePEc: EconPaper, roč. 1, č. 11. s. 184 – 190. Dostupné na internete: <http://www.oenb.at/de/img/feei_2011_q4_studies_2_tcm14-241681.pdf>

ELLER, M. – FRÖMMEL, M. – SRZENTIC, N. 2010. Private Sector Credit in CESEE: Long-Run Relationships and Short-Run Dynamics. In: Österreichische Nationalbank: *Focus on European Economic Integration*, č. Q2/10, s. 50 – 78.

KING, R.G. – LEVINE, R. 1993. Finance and Growth: Schumpeter might be right. In: *The Quarterly Journal of economics*, roč. 108, č. 3, s. 717 – 737. ISSN 1531-4650.

KOIVU, T. 2002. Do efficient banking sectors accelerate economic growth in transition countries? Bank of Finland: Discussion paper No. 14. [online]. Dostupné na internete: <http://www.suomenpankki.fi/bofit/tutkimus/tutkimusjulkaisut/dp/Documents/dp1402.pdf>.

LEVINE, R. 2002. Bank Based or Market-based financial system: Which is better? NBER: Working paper No. 9138. [online]. Dostupné na internete: <http://www.nber.org/papers/w9138.pdf>.

LUCAS, R.E. 1988. On the Mechanics of Economic Development. In: *Journal of Monetary economics*, roč. 22, č. 1, s. 3 – 42. ISSN 0304-3932.

NÁRODNÁ BANKA SLOVENSKA: Poskytnuté úvery v aktuálnom mesiaci a ich úrokové miery - nové obchody. Bratislava: NBS. [online]. Dostupné na internete: <http://www.nbs.sk/sk/statisticke-udaje/menova-a-bankova-statistika/zdrojove-statisticke-udaje-penaznych-financnych-institucii/uvery>.

ROBINSON, J. 1952. The Rate of Interest or other Essays. In *The American Economic Review*, roč. 43, č. 4, s. 636 – 641. ISSN 0002-8282.

SCHUMPETER, J.A. 1911. The Theory of Economic Development. Cambidge, MA: Harvard University Press.

ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY: Index priemyselnej produkcie oproti priemernému mesiacu roku 2005, očistený o vplyv počtu pracovných dní. Bratislava: ŠÚSR. [online]. Dostupné na internete: <http://portal.statistics.sk/showdoc.do?docid=37591>.

VAZAKIDIS, A. – ADAMOPOULOS, A. 2011. Credit market development and credit growth an empirical analysis for Greece. In. *American Journal of Applied Sciences*, roč. 8, č. 6, s. 584 – 593. ISSN 1546-9239.

Adresa autorov:

Michaela Urbanovičová, Ing.
Ekonomická fakulta
Technická univerzita v Košiciach
Němcovej 32, 041 01 Košice
michaela.urbanovicova@tuke.sk

Beáta Gavurová, doc., Ing., PhD., MBA
Ekonomická fakulta
Technická univerzita v Košiciach
Němcovej 32, 040 01 Košice
beata.gavurova@tuke.sk

Identifikácia súčiniteľa kapilárnej vodivosti z experimentálnych údajov
Identification of the capillary conduction coefficient from experimental data

Jiří Vala, Petra Jarošová

Abstract: The paper comes from the critical evaluation of contemporary computational approaches to the identification of the capillary transfer coefficient, namely of porous building materials, based on the approximate solution of an inverse problem for one nonlinear partial differential equation, related to the physical principle of mass conservation and to the Fick constitutive law. It demonstrates how the generalization and combination of such approaches leads to the design to an effective computational algorithm in the MATLAB environment.

Abstrakt: Príspevok vychádza z kritického zhodnotenia súčasných výpočtových prístupov k identifikácii súčiniteľa kapilárnej vodivosti, najmä pórovitých stavebných materiálov, z experimentálnych údajov, založených na približnom riešení inverzného problému pre jednu nelineárnu parciálnu diferenciálnu rovnicu, odkazujúceho na fyzikálny princíp zachovania hmotnosti a na Fickov konštitutívny vzťah. Ukazuje, ako zobecnenie a kombinácia týchto prístupov vedie k návrhu efektívneho výpočtového algoritmu v prostredí MATLABu.

Key words: building materials, capillary conductivity, inverse problems, numerical modelling.

Kľúčové slová: stavebné materiály, kapilárna vodivosť, inverzné problémy, numerické modelovanie.

JEL classification: C63

1. Introduction

Identification of capillary transfer properties, namely in the case of porous building materials, is a serious nontrivial problem, both from the experimental and the computational point of view. An increasing moisture content in materials and structures can make their thermal insulation and accumulation properties much worse, which forces unexpected high energy consumption, and contribute to the deterioration of their mechanical properties substantially, too. Moreover, some technologies, as in cooling and freezing plants, should not admit presence of humidity from external environment at all. Reliable computational modelling and simulation, supported by well-considered organization of experiments, is thus very desirable.

Unlike the linearized theory of heat conduction, elastic deformation, etc., even in the case of a (seemingly easy) direct problem of capillary conduction (with a priori known material properties), the nonlinear and nonstationary character of the analyzed physical process brings non-negligible difficulties to all computational tools. Most identification approaches try to overcome such obstacles, as well as still other difficulties, typical for inverse problems, namely ill-possedness, numerical instability, etc. – cf. [Isakov], p. 20, using very simple model configurations of experiments. However, simplified considerations related to one-dimensional specimens of infinite lengths and other phenomena, not observable in the real world, but forcing numerical integration over infinite sets from small data sets with hidden (both system and random) errors from various sources, do often not lead to satisfactory results. This is the principal motivation for our checkup of existing approaches and for some recommendations to their improvement with the aim of derivation of effective computational algorithms producing credible results for building design.

2. Physical and mathematical model

In general, let us consider an open set W in the 3-dimensional real Euclidean space R^3, supplied with the Cartesian coordinate system $x := (x_1,x_2,x_3)$, and some open set $I = (0,a)$ where a is allowed to be a finite positive real number or ∞, containing the time t. Let $\partial\Omega$ be the boundary of Ω in R^3 where the unit outward normal vector $\boldsymbol{\nu}(x) := (\nu_1(x),\nu_2(x),\nu_3(x))$ can be introduced (almost) everywhere; this enables us to define, following [Roubíček], p. 14, standard Lebesque, Sobolev, Bochner, etc. spaces, as $L^2(\Omega)$, $W^{1,2}(\Omega)$, $L^2(\partial\Omega)$, $L^2(I,W^{1;2}(\Omega))$ and similar ones. We shall consider that Ω consists of two disjoint parts, Γ and Θ. For arbitrary functions $\varphi(x,t)$ and $\psi(x,t)$, or $\varphi(t)$ and $\psi(t)$ (alternatively), with $x \in \Omega$ or (in the sense of traces) with $x \in \partial\Omega$, whose products are integrable in the required sense, we shall introduce the notations

$$(\varphi, \psi) := = \int_\Omega \varphi(x;t)\psi(x;t)dx \tag{1}$$

$$\langle\varphi, \psi\rangle := = \int_{\partial\Omega} \varphi(x;t)\psi(x;t)ds(x) \tag{2}$$

$$[\varphi, \psi] := = \int_I \varphi(x;t)\psi(x;t)dt \tag{3}$$

all integrals occurring in (1) and (2) still depend on t. In particular: i) for $\varphi(t),\psi(t) \in L^2(\Omega)$ and t fixed (1) refers to the standard scalar product in $L^2(\Omega)$, ii) for $\varphi(t),\psi(t) \in L^2(\partial\Omega)$ and t fixed (2) refers to such product in $L^2(\partial\Omega)$, iii) for $\varphi, \psi \in L^2(I)$ (3) refers to such product in $L^2(I)$. However, we can have e.g. $\varphi \in L^p(\Omega)$ and $\psi \in L^{p/(p-1)}$ in (1) with a real $p \geq 1$, including the case $\varphi \in L(\Omega)$ and $\psi \in L^\infty(\Omega)$. The notation (1) can be applied also to vector variables $\boldsymbol{\varphi}(x,t) := (\varphi_1(x,t),\varphi_2(x,t), \varphi_3(x,t))$ instead of scalar ones, working with the scalar product $\varphi(x,t)$. $\psi(x,t)$ in R^3 for all needed x and t.

The most important physical quantity of all our considerations will be the mass fraction $u(x,t)$ of liquid water (or other contaminant of non-variable density in similar applications) with no source in Ω, only with surface source on $\partial\Omega$, introduced, using boundary mass fractions, as $u(x,t) = \bar{u}(x,t)$ on Θ for prescribed values $\bar{u}(x,t)$, or, alternatively, using relative boundary mass fluxes, some prescribed values $\bar{\jmath}(x,t)$ on Γ, related to corresponding relative mass fluxes on Ω thanks to the relation $j.\nu = \bar{\jmath}$ (the dependence of such quantities on x or t is not highlighted for simplicity).

For a positive bounded real differentiable function b there is natural to consider $V := \{v \in W^{1,2}(\Omega)$: $v = 0$ on $\Theta\}$, $H := \{ v \in W^{1,2}(\Omega)$: $v = \bar{u}$ on $\Theta\}$, $\mathbf{j} \in L^2(I,W^{1;2}(\Omega)^3)$, $\bar{\jmath} \in L^2(I,L^2(\Gamma))$, $u - \bar{u} \in V$ for (almost) all $t \in I$ (some extension of $\bar{u}(.,t)$, from Θ to Ω must be available) and $u \in L^2(I,H)$. Applying one additional notation $U := \{v \in L^2(I,H)$: $\dot{v} \in L^2(I,L^2(\Omega))\}$, using the dot symbols for time derivatives everywhere, assuming some a priori known function β, the direct problem is now to find $u \in U$ satisfying, following [Bermúdez], p. 137, the physical mass balance condition

$$(v, \dot{u}) = \langle v, \bar{\jmath}\rangle + (\nabla v, \mathbf{j}) \tag{4}$$

for any $v \in V$, together with the empirical Fick constitutive relation

$$\mathbf{j} = -\nabla\beta(u) \tag{5}$$

interpretable as a very special case of the Kirchhoff transformation in the sense of [Roubíček], p. 253. Some initial condition $u(.,0) = u_0$ with $u_0 \in L^2(I,L^2(\Omega))$ is considered to be prescribed, too.

The equations (4) and (5) together give

$$(v, \dot{u}) + (\nabla v, \nabla\beta(u)) = \langle v, \bar{\jmath}\rangle. \tag{6}$$

Since $\nabla\beta(u) = \beta'(u)\nabla(u)$ (the prime symbol is used for the ordinary derivatives), we can define $\kappa(u) := \beta'(u)$, which is the most frequently introduced capillary transfer coefficient; thus an alternative form of (6) is

$$(v, \dot{u}) + (\nabla v, \kappa(u)\,\nabla u) = \langle v, \bar{\jmath}\rangle. \tag{7}$$

The Green - Ostrogradskiĭ theorem, applied to (7), yields

$$(v, \dot{u}) - \nabla.(\kappa(u)\,\nabla u) = \langle v, \bar{\jmath} + \kappa(u)\nabla u.\, v\rangle. \tag{8}$$

at least in the sense of distributions. Inserting the Dirac distribution (generalized function) $\delta(x - x_0)$ as v locally, for x_0 from Ω or $\partial\Omega$, we can see the well-known consequence of (8)

$$\dot{u} = \nabla(\kappa(u)\nabla u) \text{ on } \Omega, \quad \kappa(u)\nabla u.\, v + \bar{\jmath} = 0 \text{ on } \Gamma$$

(with x instead of x_0 again). For sufficiently smooth functions β we can also write $\nabla(\kappa(u)\nabla u) = \kappa(u)\Delta u + \kappa'(u)\,\nabla u.\nabla u$, which, applied to (8), gives

$$(v, \dot{u}) - (v, \kappa(u)\,\Delta u) - (v\nabla u, \kappa'(u)\nabla u) = \langle v, \bar{\jmath} + \beta(u).\, v\rangle. \tag{9}$$

Finally, another application of the Green - Ostrogradskiĭ theorem to (6) gives

$$(v, \dot{u}) - \big(\Delta v, \beta(u)\big) = \langle v, \bar{\jmath}\rangle + \langle \nabla v.\, v, \beta(u)\rangle. \tag{10}$$

(which is rarely used, but useful in this paper).

The analysis of existence and uniqueness of problems (6) – (10), including the convergence of sequences of approximate solutions from finite-dimensional spaces, is not easy, unlike the case with κ independent of u, which is the (not very hard) exercise for the application of the Lax - Milgram theorem and the Euler implicit method. Nevertheless, the detailed analysis for both classical and variational solutions, starting from (6) can be found in [Roubíček], p. 239, in regard of inverse problems then in [Vala]. Much more limited results for (9) can be found in [Rincon etal.].

For the identification of the dependence of β or κ on u we need some reasonable inverse formulation. Let us assume that some decomposition $\kappa(u)=c_i\,\kappa_i(u)$ for the Einstein summation index $i \in \{1,...,n\}$ and an integer n (finite in practical calculations, admitting $n \to \infty$); here $\kappa_1,...,\kappa_n$ should be prescribed (linearly independent) functions and $c_1,...,c_n$ a priori unknown real coefficients. To determine such coefficients, usually some values of u, in addition to boundary and initial conditions, are known from indirect non-destructive or low-invasive (e.g. microwave) measurements, whereas u_0, $\bar{u}$ and $\bar{\jmath}$ come from direct ones; for more details to experimental settings see [Škramlík etal.]. Then e.g. the solution of (7) can be found correctly (at least theoretically because the discretization leads to nonlinear algebraic systems), up to n unknown parameters $c_1,...,c_n$, i.e.

$$(v, \dot{u}) + (\nabla v, \kappa_i(u)\nabla u)c_i = \langle v, \bar{\jmath}\rangle. \tag{11}$$

The partial or complete knowledge of $u(x,t)$ could be asserted now to determine $c_1,...,c_n$.

3. Variational and general integration approaches

For simplicity, let us suppose that we are able to reconstruct the complete distribution of u on $\Omega \times$ I satisfying (6) (at least with sufficient accuracy for practical evaluations) from the above mentioned data. Inserting $v = \beta_j(u)$, with respect to the relation $\beta'_j(u) = \kappa_j(u)$, for any $j \in \{1,...,n\}$ into (11), integrating (11) over I, we obtain

$$[(\kappa_j(u)\nabla u, \kappa_i(u)\nabla u)]c_i = [\langle v, \bar{\jmath}\rangle - (\beta_j(u), \dot{u})]. \tag{12}$$

This enables us to identify $\kappa(u)$ without any additional optimization, from a positive definite system of n linear algebraic equations. However, the numerical integration, in general in R^4, is rather difficult and requires a sufficiently rich data set.

The basic ideas of two other general integration approaches can be taken from [Stenlund]. The first approach starts (in our notation) from (10) with $v = |x - x_0|^{-1}$. Taking into account the well-known relation from the theory of distributions

$$\Delta(|x - x_0|^{-1}) + 4\pi\delta(x - x_0) = 0,$$

the seemingly simple resulting formula (valid in Ω only, not on Γ, all arguments are highlighted here because of the risk of misunderstanding) is

$$\beta(u(x_0, t)) = -\frac{1}{4\pi}\int_\Omega \frac{\dot{u}(x,t)}{|x - x_0|}\,\mathrm{d}x. \tag{13}$$

Nevertheless, the visible disadvantage of the implementation of (13) for numerical evaluations is its singularity at all integration points $x \in \Omega$, in analytical calculations typically avoided by the Cauchy principal value of singular integrals; for a method of (rather hard) negotiation of such obstacle see [Ghanbari etal.].

The second announced approach starts from (9) with $v = \delta(x - x_0)$, without any reference to the theory of distributions. Utilizing the notation $f(u) := \Delta u/(\nabla u.\nabla u)$ and $g(u) := \dot{u}(\nabla u.\nabla u)$, from the classical differential calculus (we have only one linear differential equation with constant coefficients now) we are able to conclude

$$\kappa(u) = K \exp(-\hat{f}(u)) + \int_{u_0}^{u} g(\tilde{u})\,\mathrm{d}\tilde{u}; \tag{14}$$

here $\hat{f}(u)$ is an arbitrary primitive function to $f(u)$ and, thanks to the initial condition, $K := \kappa(u_0)\exp(-\hat{f}(u_0))$; the knowledge of the value $\kappa(u_0)$ is needed. Moreover, the numerical evaluation of (14), namely of its right-hand-side integral, is very complicated.

Both (13) and (14) return discrete values of $\kappa(u(x,t))$ only, not the required explicit description of the real function $\kappa(.) = c_i\kappa_i(.)$. This can be done by minimization of some quadratic cost function $[(c_i\kappa_i(u) - \kappa(u), w(c_i\kappa_i(u) - \kappa(u)))]$ with an appropriate weight w in $L^2(\Omega \times \boldsymbol{I})$ theoretically (in certain finite-dimensional space in practice) with the obvious result

$$[(\kappa_j(u), w\kappa_i(u))]c_i = [(\kappa_j(u), w\kappa(u))] \tag{15}$$

for all $\boldsymbol{j} \in \{1,...,n\}$. In particular, the choice $w = \nabla u.\nabla u$ causes the minimization of a sum of all available mass fluxes; for other appropriate choices cf. [Bochev etal.], p. 89. Then $c_1,...,c_n$ can be obtained from a positive definite system of n linear algebraic equations (15) easily.

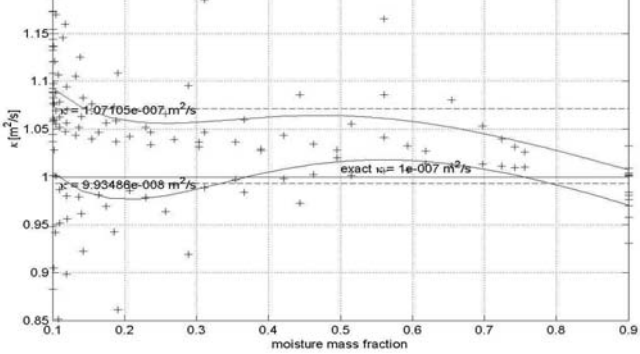


Fig. 36: Computational identification of a function κ from experimental data

4. One-dimensional approaches for special experimental configurations

The third integration method, mentioned even in the title of [10], relies, in general, on the choice $v = \chi(M)$ in (10) where $\chi(M)$ is the characteristic function of some carefully selected subset M of Ω. Nevertheless, its effective implementation is related to the following (very special) geometrical configuration: $\Omega = \{x \in R^3: x_3 > 0\}$, $\Theta = \{x \in R^3: x_3 = 0\}$ and $M = \{x \in R^3: x_3 > x_0\}$ for some prescribed positive x_0. Moreover, $\bar{u}$ on Θ is supposed to be independent of x_1 and x_2. Then, working with x_3 instead of x only (omitting the index 3), we have

$$\kappa(u(x_0,t)) = u(x_0,t)^{-1}\int_{x_0}^{\infty}\dot{u}(x,t)\ \mathrm{d}x; \tag{16}$$

the zero partial derivative of u with respect to x for $x\to\infty$ is needed here, which is not trivial – cf. [Micu etal.]. All values from (16) can be inserted into (15) easily. Especially, under the assumption of constant $\bar{u}$ in time, the similar result

$$\kappa(u(x_0,t)) = (2tu'(x_0,t))^{-1}\int_{x_0}^{\infty} xu'(x,t)\,\mathrm{d}x; \tag{17}$$

referenced in the technical literature as the Boltzmann - Matano method, has been derived by Matano (1933) in another way, explained in [Černý etal.], p. 63, coupling the classical differential calculus with the Boltzmann transform $X = x/(2\sqrt{t})$ (1894); for more historical remarks see [Stenlund].

An alternative choice, avoiding the numerical integration over an infinite set, can be $M = \{x\in R^3: 0< x_3 < x_0\}$ with Θ replaced by Γ. The resulting relation, analogous to (16), is

$$\kappa(u(x_0,t)) = u(x_0,t)^{-1}\Big(\int_0^{x_0}\dot{u}(x,t)\,\mathrm{d}x - \bar{\jmath}(t)\Big). \tag{18}$$

Another interesting approach, developed in [Černý etal.], p. 62, coming from (8), relies on the special choice $\kappa_i(u) = \kappa(\overline{w}_i)$ for $\boldsymbol{i}\in\{1,...,n\}$ with a priori prescribed set of positive constants w_i such that $w_0<w_1<...<w_n$ (which are estimates of finite mass fraction levels) where $\overline{w}_i = (w_{i-1}+w_i)/2$, accompanied by the setting $\nu = \chi(M_i)$ for $M_i = \{(x,t)\in\Omega\times I:\ w_{i-1} < u(x,t) \le w_i\}$; consequently $c_i = \mathrm{meas}(M_i)$, avoiding all additional cost functions like (15). This can be formulated generally, but no practical numerical application except simple one-dimensional configurations is known. The reason is clear: the crucial point of this method is the sufficiently precise analysis of isolines of $u(x,t)$, characterized by parametric equations, utilized in complicated integrals (at least double ones, on $\Omega\times I$). More technical details are discussed in [Černý etal.], p. 63; consequently [Černý etal.], p. 66, presents the successful usage of some of genetic algorithms, too. For other alternative optimization approaches, as differential evolution, particle swarm or simulated annealing methods, see [Colaço etal.].

The illustrative Figure 1 shows the results of identification of $\kappa(u)$ for the case of its nearly constant expected values; such case can be used to verify the correctness of the original software code, written in MATLAB, with no references to additional packages. Under the assumption of constant k there are several possibilities how to derive analytical solutions for simple experimental configurations, as the standard Laplace or Fourier transform or the (above mentioned) Boltzmann transform together with the means of classical differential calculus (substitutions and separation of variables), e. g. for constant $u(0,.) = u_1$ and $u(.,0) = u_0$ with the result

$$u = u_0 + (u_1 - u_0)\mathrm{erfc}(x/(2\sqrt{\kappa t})). \tag{19}$$

More generally, the basis $\kappa_i(r) = \exp(-(i-1)r)$ has been considered for some positive r, $n = 5$, $\boldsymbol{i}\in\{1,...,5\}$, $u_0 = 0.1$, $u_1 = 0.9$ and experimental data u and $\bar{\jmath}$, obtained at the Faculty of Civil Engineering of the Brno University of Technology (FCE BUT). The upper curve shows the identification result by (17) and (15) with $w = 1$, the lower one the same one by (18) replacing (17), both in comparison with that predicted by (19).

5. Conclusions and generalizations

We have demonstrated how most contemporary methods of identification of capillary conduction coefficient can be derived from general formulations (6) – (10), including certain suggestions of their improvement, namely those related to applied numerical evaluations. Some generalizations are needed evidently: a posteriori error estimates, separating (as much as possible) i) influences of other physical processes, ii) numerical discretization and integration errors, iii) uncertainties of measurements, etc., as well as the proper existence and convergence analysis for particular algorithms.

The above sketched motivations for further both theoretical and laboratory study will be incorporated into the research project at BUT, referenced below, as its significant part. The analysis of much more extensive experimental data, oriented to the effective design of composite materials for civil engineering, is being prepared to be published in another paper soon.

6. Acknowledgements

This work on this paper has been supported by the project of specific university research at BUT, No. FAST-S-13-2088.

References

BOCHEV, P. B. – GUNZBURGER, M. D. 2009. Least-Squares Finite Element Methods. *Springer, Berlin*.

BERMÚDEZ DE CASTRO, A. 2005. Continuum Thermomechanics. *Birkhäuser, Basel*.

COLAÇO, M. J. – ORLANDE, H. R. B. – DULIKRAVICH, D. S. 2006. Inverse and optimization problems in heat transfer. In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 28, p. 1 – 24.

ČERNÝ, R., et al. 2010. Complex System of Methods for Directed Design and Assessment of Functional Properties of Building Materials: Assessment and Synthesis of Analytical Data and Construction of the System. *Czech Technical University, Prague*.

GHANBARI, M. – ASKARIPOUR, M. – KHEZRIMOTLAGH, D. 2010. Numerical solution of singular integral equations using Haar wavelet. In: *Australian Journal of Basic and Applied Sciences* 4, p. 5852 – 5855.

ISAKOV, V. 2006. Inverse Problems for Partial Differential Equations. *Springer, Berlin*.

Micu, S. – Zuazua, E. 2000. On the lack of null-controllability of the heat equation on the half-line with a Dirichlet boundary control. In: *Transactions of the American Mathematical Society* 353, p. 1635 – 1659.

RINCON, M. A. – LÍMACO, J. – LIU, I.-S. 2005. Existence and uniqueness of solutions of a nonlinear heat equation. In: *Tendências em Matemática Aplicada e Computacional* 6 , p. 273 – 284.

ROUBÍČEK,T. 2005. Nonlinear Partial Differential Equations with Applications. *Birkhäuser, Basel*.

STENLUND, H. 2004. Three methods for solution of concentration dependent diffusion coefficient. *Visilab Signal Technologies, technical report*, available at *www.visilab.fi/nonlinear_diffusion.pdf*.

ŠKRAMLIK, J. – NOVOTNÝ, M. – ŠUHAJDA K. 2012. The moisture in capillaries of building materials. In: *DPC Journal of Civil Engineering and Architecture* 2, p. 1536 – 1543.

VALA, J. 2013. On the computational identification of temperature-variable characteristics of heat transfer. In: *International Conference Applications of Mathematics (in honor of the 70th birthday of Karel Segeth) in Prague*, p. 215 – 224.

Authors' addresses:

Jiří Vala, prof. Ing., CSc.
Fakulta stavební VUT v Brně
Ústav matematiky a deskriptivní geometrie
Veveří 95, 602 00 Brno
vala.j@fce.vutbr.cz

Petra Jarošová, Ing. arch.
Fakulta stavební VUT v Brně
Ústav pozemního stavitelství
Veveří 95, 602 00 Brno
jarosova.p@fce.vutbr.cz

Formování vhodného způsobu zajištění vstupních informací pro modifikovanou metodu CPM
Forming a Suitable Way for Ensure the Input of Information to Modified CPM Method

Radek Vostál, Petr Vondráček, Pavel Foltin

Abstract: The Article discusses the possible approaches to the analysis of security factors influencing the environment of the logistics chains. A key prerequisite for analysis of the level and extent ware the identification of resource availability of relevant data and their ability for evaluation. The relevance ratio of each factor was examined by a modified method of critical paths and then to extend the portfolio security criteria.

Abstrakt: Příspěvek pojednává o možném přístupu k analýze bezpečnostních faktorů ovlivňujících prostředí realizace logistických řetězců. Klíčovým předpokladem analýzy úrovně a rozsahu byla identifikace dostupnosti zdrojů relevantních dat a jejich možnosti ohodnocení. Míra relevantnosti jednotlivých faktorů byla zkoumána prostřednictvím modifikované metody kritických cest a následně jejího rozšíření o portfolio bezpečnostních kritérií.

Key words: logistics chains, supply chain security, Critical Path Method, portfolio of security criterion.

Kľúčové slová: Logistický řetězec, bezpečnost dodavatelského řetězce, metoda kritických cest, portfolio bezpečnostních kritérií.

JEL classification: C44

1. Úvodní vymezení zkoumaného problému

Dynamické prostředí logistických řetězců lze považovat za jeden z velmi významných zdrojů jejich potenciálních narušení a následných nedostatků v jejich efektivním řízení. Přestože pravděpodobně existují rozdílné dopady neefektivního řízení logistického řetězce v podmínkách civilní organizace a ozbrojených sil[1], lze s jistotou konstatovat, že v obou případech se jedná o negativní fakt. Snahou odpovědných manažerů by mělo být včas rozpoznávat dynamičnost daného logistického řetězce, identifikovat míru rozdílnosti stavu plánovaného a reálného. Následně činit taková rozhodnutí, která budou naplňovat požadavky na efektivní řízení, při současném zohlednění identifikovaných změn v daném bezpečnostním prostředí.

V rámci realizovaného výzkumu byla pozornost zaměřena na bezpečnostní prostředí logistických řetězců, které je považováno za jeden z významných aspektů efektivního řízení logistických řetězců. V průběhu výzkumu byla provedena analýza bezpečnostního prostředí logistických řetězců, na jejímž základě bylo identifikováno portfolio bezpečnostních kritérií. O takto vytvořené portfolio byla dále rozšířena standardní metoda kritických cest (CPM), která vede k možnosti algoritmizace zkoumaných souvislostí do funkčního modelu (Foltin et al., 2013). Na základě vytvořeného portfolia bezpečnostních kritérií bylo hodnoceno působení identifikovaných faktorů bezpečnosti logistických řetězců na vytvořeném algoritmizovaném modelu. Jako rozhodující předpoklad úspěšného provedení objektivní simulace bezpečnosti určitého logistického řetězce byl identifikován požadavek na dostatečné množství relevantních informačních zdrojů, které budou využitelné v souladu se základním přístupem ke zkoumání problematiky.

[1] Na tomto místě není podstatná diskuze nad mírou závažnosti těchto dopadů.

Příspěvek sumarizuje vybranou část doposud dosažených výsledků specifického výzkumu „Aplikace modifikované metody CPM na portfolio bezpečnostních kritérií v logistických řetězcích“ (SV13-FEM-K101-07-SED), realizovaného pod záštitou Fakulty ekonomiky a managementu Univerzity obrany v Brně.

2. Cíl a metoda řešení

Cílem dílčí fáze realizovaného projektu specifického výzkumu bylo zajistit dostatečné množností relevantních informačních vstupů pro verifikaci realizovaného přístupu a dále také pro samotnou simulaci vývoje bezpečnostního prostředí daného logistického řetězce pomocí algoritmizovaného modelu. Pro počáteční zjednodušení, avšak také především pro detailnější zaměření pozornosti na optimální způsob zajištění požadovaných informací bylo zkoumání zaměřeno na jediný logistický uzel, a to námořní přístav Koper (Slovinsko). Očekávaným přínosem tohoto omezení je již zmíněné nalezení vhodného způsobu zajišťování informačních vstupů, s čímž je velmi úzce spojena identifikace těch informačních zdrojů, jejichž využití bude objektivně možné pro více různých uzlů nebo hran logistického řetězce. Současně je očekáváno, že po vymezení optimálního přístupu k zajištění informačních vstupů, bude z velké části možné postupovat analogicky pro ostatní uzly, potažmo hrany logistického řetězce.

Za základní metodu řešení vymezeného problému lze považovat internetovou rešerši potenciálně vhodných zdrojů. S ohledem na potřebu zachycení dynamiky logistického řetězce, ve zvoleném případě vybraného logistického uzlu, je pozornost věnována dvěma aspektům hledaných informací, a to jejich věrohodnosti a frekvenci aktualizace poskytovaných dat.

3. Diskuse získaných poznatků

V úvodní fázi realizace popisovaného řešení byl systém získávání informací pro naplnění jeho cíle rozdělen do dvou relativně oddělených částí. Z hlediska frekvence aktualizace poskytovaných dat lze totiž identifikovat dva základní typy informací. Za prvé se jedná se o statické informace, jejichž platnost je v čase stabilní a poskytuje dlouhodobě konstantní parametry logistického uzlu (např. geografická poloha, provozní kapacita, klimatické omezení plného provozu, apod.). Tento typ dat byl označen za výchozí (nulové) hodnoty daného logistického uzlu.

Druhým typem informací o témže uzlu jsou informace v čase se vyvíjející, které je možné označit za dynamické informace. Za takové informace je možné označit ostatní informace, které nevykazují statickou povahu informací a v čase se mění. Tento typ dat byl pro naplnění základního výzkumného konceptu považován za stěžejní.

a. Statické informace

Z provedené internetové rešerše vyplynulo, že zajištění vhodného informačního vstupu statických informací je relativně snadné, neboť výskyt těchto informací je v síti internet častý. V rámci realizovaného výzkumu byly identifikovány dva zdroje statických dat. Prvním dostatečně věrohodným zdrojem byly internetové stránky daného logistického uzlu např. z Port Handbook jednotlivých přístavů (Luka Koper, 2005). Jako druhý zdroj statických dat byly identifikovány stránky organizací či agentur sdružující informace o daných typech uzlů. Rešeršní metodou byl potvrzen předpoklad dostupnosti obou typů statických informací i u dalších logistických uzlů, obdobného zaměření.

Zvolený postup potvrdil, že získávání reliabilních statických dat o logistických uzlech přímo na internetových portálech daných uzlů je možné považovat za optimální a použitelné při dalších fázích realizovaného výzkumu.

b. Dynamické informace

V případě dynamických informací byly jako výchozí požadavky realizované internetové rešerše definovány požadavky na dostupnost informací, jejich současná validita a reliabilita a frekvence jejich průběžné aktualizace. Ve srovnání se statickým typem dat vyplynulo, že dostupnost dynamických informací vyhovujících definovaným požadavkům je značně omezená.

V rámci internetové rešerše bylo identifikováno několik dostupných zdrojů, jejichž informace poskytovaly primární informace o možném narušení bezpečnostního prostředí v daném logistickém uzlu či jinak informovaly o reálném dění ve zvoleném uzlu. Rešerší však bylo zjištěno, že ve většině případů dostupné zdroje nebyly vyhovující jak z hlediska validity, reliability a nebyly ani v dostatečné frekvenci průběžně aktualizovány. V oblasti zajišťování dynamických informacích identifikován pouze jeden zdroj informací naplňující definované požadavky. Jedná se o informační výstup projektu MarineTraffic, který se zabývá sběrem a prezentací reálných dat v námořní dopravě. Věrohodnost tohoto projektu (prezentovaných dat) a aktualizace disponibilních informací je podložena napojením na Automatic Identification System (AIS). AIS je systém provozovaný pod záštitou Mezinárodní námořní organizace (IMO) a slouží k monitorování pohybu (trasy, aktuální rychlosti, aktuálního kurzu a dalších údajů) podstatné většiny námořních plavidel. (MarineTraffic.com, 2013).

Na základě omezeného přístupu k identifikaci dynamických informací byl pro testování přístupu zvolen jeden dynamický faktor bezpečnosti logistického řetězce, resp. zvoleného konkrétního logistického uzlu. Vybraným faktorem byl faktor počasí, neboť v jeho případě je možné identifikovat dostatečné množství dostupných zdrojů informací, zajistit jejich validitu a reliabilitu v téměř reálném čase. Jako vhodné zdroje dynamických informací o počasí v oblasti daného logistického uzlu byla identifikována řada zdrojů. Z hlediska věrohodnosti dostupných dat a jejich aktualizace 8x denně, včetně budoucí predikce, byl zvolen portál „My weather 2“ (Weather2, 2013). Portál poskytuje detailní informace o logistickém uzlu (tj. přístav Koper), a to přímo z daného přístavu, včetně klimatických parametrů spojených pouze s námořní dopravou.

4. Závěr

V rámci realizovaného projektu specifického výzkumu byly identifikovány požadavky na zdroje informací modifikované metody kritických cest (CPM) rozšířené o portfolio bezpečnostních kritérií. Požadované informační vstupy byly rozděleny do dvou kategorií, a to statické a dynamické informace. Zvolený přístup byl testován na konkrétním logistickém uzlu, tj. přístav Koper. Byly zjištěny významné rozdíly mezi statickými a dynamickými informacemi logistického uzlu, a to z hlediska dostupnosti zdrojů informací, validity a reliability zveřejňovaných informací a současně frekvence jejich aktualizace. Dílčím poznatkem realizovaného výzkumu u dynamických informací bylo zjištěno, že objektivnost požadovaných dynamických informací o logistických uzlech i hranách je oproti statickým informacím značně omezená.

V dalším fázi realizovaného výzkumu bude na základě identifikovaných zdrojů dynamických informací testován algoritmus modifikované CPM a verifikována jeho platnost. Zvolený postup bude v obecné podobě následně aplikovatelný na další typy logistických uzlů, včetně jejich vazeb.

Literatura

FOLTIN, P., SEDLAČÍK M. – ONDRYHAL, V. 2013. Bezpečnostní aspekty logistických řetězců. In: *Manažment, teória, výučba a prax 2013: Zborník z príspevkov z medzinárodnej vedecko-odbornej konferencie*. Liptovský Mikuláš: Akadémia ozbrojených síl, 2013, s. 88-96. ISBN 978-80-8040-477-2.

Port Handbook: Ship Guide. PORT OF KOPER. 2005. *Luka Koper*. [cit. 2013-11-10]. Dostupné z: http://www.luka-kp.si/eng/port-handbook/ship-and-truck-guide/ship-guide/104#2.

Přístavy. MARINETRAFFIC.COM. 2013. *MarineTraffic.com* [cit. 2013-11-10]. Dostupné z: http://www.marinetraffic.com/ais/cz/default.aspx.

Slovenia-Koper: Marine Weather. WEATHER2. 2013. *Weather2* Glam Entertainment, [cit. 2013-11-10]. Dostupné z: http://www.myweather2.com/Marine/Global-Ports/Slovenia/Koper.aspx.

Adresy autorů:

Bc. Radek Vostál (2. ročník-MN)
Fakulta ekonomiky a managementu
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
radek.vostal@unob.cz

Bc. Petr Vondráček (2. ročník-MN)
Fakulta ekonomiky a managementu
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
petr.vondracek@unob.cz

Dr. habil. Ing. Pavel Foltin, Ph.D.
Fakulta ekonomiky a managementu
Univerzita obrany
Kounicova 65, 662 10 Brno, CZ
pavel.foltin@unob.cz

Citlivosť miery majetkovej chudoby na úrokovú mieru
Sensitivity of asset poverty rate to interest rate

Tomáš Želinský

Abstract: The aim of paper is to analyse sensitivity of asset poverty to interest rate. Two at risk of asset poverty rates are estimated: extension of income by the annual lifetime annuity value of its current net worth and estimation of share of persons with limited access to asset sources not allowing them to make their ends meet during the given period. The results indicate that poverty rates estimated using the first approach are strongly sensitive to the values of interest rates, and the first approach is robust to the interest rate.

Abstrakt: Cieľom príspevku je analyzovať citlivosť miery majetkovej chudoby na úrokovú mieru. V práci sú použité dve miery rizika majetkovej chudoby: rozšírenie príjmu o teoretický tok z čistej hodnoty majetku a odhad podielu osôb, ktorých prístup k majetkovým zdrojom je nedostatočný na to, aby im umožnil uspokojiť potreby počas zvoleného obdobia. Výsledky naznačujú, že kým použitie prvého prístupu môže byť do istej miery výrazne citlivé od hodnoty použitej úrokovej miery, druhý prístup je voči použitej úrokovej miere v zásade robustný.

Key words: Asset poverty, sensitivity analysis, HFCS, Slovakia.

Kľúčové slová: Majetková chudoba, citlivostná analýza, HFCS, Slovensko.

JEL classification: I32

1. Úvod

Analýza chudoby býva najčastejšie uskutočňovaná na základe údajov o príjmoch, resp. spotrebe. Posudzovanie životných podmienok a finančnej situácie len na základe príjmov nie je úplné, nakoľko domácnosti s rovnakým príjmom môžu disponovať rôznym majetkom. Aj z toho dôvodu sa odporúča dopĺňať údaje o príjmoch údajmi o majetku domácností. Avšak kým získať prehľad (vo forme individuálnych údajov) o príjmovej situácii domácností je relatívne jednoduché, a to či už na základe výberových zisťovaní, alebo administratívnych zdrojov údajov, v prípade úspor a iných foriem majetku je to zložitejšie.

Príjem je nestála miera, minulé príjmy nemusia nevyhnutne indikovať, aké zdroje má jednotlivec k dispozícii v súčasnosti, keďže príjem mohol byť utratený relatívne rýchlo a nakúpené statky mohli byť rýchlo spotrebované. Na druhej strane, majetok je stabilnejším indikátorom statusu alebo pozície v spoločnosti a reprezentuje nahromadenú kúpnu silu. Navyše, majetok na rozdiel od príjmu je kumulovaný dlhodobo a k výrazným zmenám dochádza zriedka (spravidla vo výnimočných situáciách). Majetok predstavuje úspory a investície, ktoré môžu byť čerpané v budúcnosti v prípade potreby. (Oliver a Shapiro, 1990) Potreba zohľadniť majetok ako indikátor blahobytu teda vychádza z poznania, že majetok prináša jeho vlastníkovi výhodu v živote a je zároveň zdrojom spotreby, pretože je možné vymeniť ho za hotovosť v časoch ekonomickej záťaže spôsobenej napríklad nezamestnanosťou, chorobou a pod. (Caner a Wolff, 2004).

Uvedené predpoklady možnosti výmeny majetku za hotovosť možno považovať za platné v prípade individuálnych nepriaznivých udalostí, no nie v prípade systematických udalostí ako napr. hospodárska kríza. Ak by sa v jednom čase pokúsilo riešiť svoju nepriaznivú situáciu predajom majetku veľké množstvo ľudí, predajná cena by mohla byť výrazne podhodnotená, prípadne u veľkej časti z nich by vôbec nemuselo dôjsť k predaju.

Weisbrod a Hansen (1968) sú považovaní za prvých, ktorí sa pokúsili analyzovať chudobu tak, že údaje o príjmoch doplnili údajmi o čistom majetku spotrebnej jednotky. Ako ale

samotní autori uvádzajú, nimi navrhnutá miera je založená na predpoklade, že aktuálny príjem a aktuálny čistý majetok sú dôležité, avšak nie jediné determinanty ekonomickej pozície spotrebnej jednotky. Keďže príjem je toková veličina, čistý majetok stavová veličina, autori navrhli jednoduchú mieru, ktorá transformuje čistý majetok na teoretický tok príjmu z tohto majetku:

$$Y^* \equiv Y + NW \frac{r}{1-(1+r)^{-n}}, \tag{1}$$

kde

Y je aktuálny ročný príjem,

NW je hodnota čistého majetku,

r je úroková miera,

n je očakávaná dĺžka života osoby.

Druhý štandardný prístup k odhadu miery majetkovej chudoby je založený na myšlienke, či osoba po strate zdroja príjmu má dostatok majetku na zabezpečenie minimálnej životnej úrovne v určitom (zvyčajne krátkom) časovom období (Ivančíková a Vlačuha, 2012). Odhad doby, počas ktorej má osoba žijúca v domácnosti s príslušnou hodnotou čistého majetku zabezpečenú minimálne požadovanú životnú úroveň (vyjadrenú v peniazoch), je založený na vzťahu (1), pričom uvažujeme predlehotnú mesačnú rentu:

$$NW = z\left(1+\frac{r}{12}\right)\frac{1-\left(1+\frac{r}{12}\right)^{-m}}{\frac{r}{12}}, \tag{2}$$

kde

NW je ekvivalentná hodnota čistého majetku,

z je hranica chudoby (tzn. mesačná suma, o ktorej sa predpokladá, že osoba bude mať zabezpečenú minimálnu životnú úroveň),

r je ročná úroková miera,

m je počet mesiacov.

Vyjadrením m z rovnice (2) dostávame počet mesiacov, počas ktorých čistý majetok zabezpečí vybranej osobe minimálnu životnú úroveň.

Z oboch vzťahov je zrejmé, že odhad mier majetkovej chudoby je okrem iných premenných závislý aj na úrokovej miere. Cieľom príspevku je analyzovať citlivosť oboch mier majetkovej chudoby na použitú úrokovú mieru.

2. Metodika

2.1 Zdroj údajov

Ako hlavný zdroj údajov použitých na odhad miery majetkovej chudoby boli údaje zisťovania HFCS (Prieskum finančnej situácie a spotreby domácností (z angl. *Household Finance and Consumption Survey*). HFCS je spoločným projektom centrálnych bánk Eurosystému. Ide o harmonizované zisťovanie v krajinách eurozóny a poskytuje reprezentatívne údaje na národnej úrovni. Cieľom zisťovania je poskytnúť detailné údaje o finančnej situácii na úrovni domácností (vlastníctvo reálnych a finančných aktív, zadlženosť, príjmy, výdavky a ďalšie). (ECB, 2013)

Na Slovensku realizovala Národná banka Slovenska prieskum prvýkrát v roku 2010 (čo je zároveň referenčným obdobím použitých údajov). Vzorku tvorí 2 057 domácností vybratých pomocou kvótového výberu tak, aby bola vzorka reprezentatívna na úrovni krajiny a zároveň na úrovni krajov. Chýbajúce odpovede na dôležité otázky boli doplnené (imputované) použitím vhodných metód tak, aby sa zachovala pôvodná štruktúra dát a vzájomné distribučné vlastnosti všetkých premenných. Chýbajúce údaje boli imputované piatimi rôznymi metódami v prostredí programovacieho jazyka STATA. (Senaj a Zavadil, 2012)

Podrobnejšie informácie týkajúce sa metodiky výberu, váženia a imputácie, ako aj organizácie a samotného priebehu zisťovania je možné nájsť v (Senaj a Zavadil, 2012; ECB, 2013).

Všetky odhady a výpočty v práci uskutočnené na základe údajov HFCS sú založené na údajovej databáze NBS (2012).

2.2 Odhad mier majetkovej chudoby

V práci sú uskutočnené dva odhady miery majetkovej chudoby:

- *Príjmovo-majetková chudoba:* Odhad miery rizika chudoby založený na rozšírení ekvivalentného disponibilného príjmu o ročnú anuitu z ekvivalentnej hodnoty čistého majetku. Vo výpočtoch podľa vzťahu (1) je pre očakávanú dĺžku života[1] *n* použitá hodnota z úmrtnostných tabuliek pre Slovenskú republiku za rok 2010 pre príslušný vek a pohlavie osoby. Miera rizika chudoby je odhadnutá v súlade s metodikou Eurostatu
- *Majetková chudoba:* Odhad podielu osôb, ktorých prístup k majetkovým zdrojom je nedostatočný na to, aby im umožnil uspokojiť potreby počas zvoleného obdobia podľa vzťahu (2).

V oboch prípadoch sú vo výpočtoch zohľadnené štyri formy majetku:

1. vklady na bežných a sporiacich účtoch,
2. bod 1. *plus* hodnota podielových fondov, dlhopisov, akcií, ďalších investícií držaných domácnosťami a pohľadávky u ostatných domácností,
3. bod 2. *plus* reálne aktíva okrem nehnuteľností *mínus* dlžná suma prečerpania na účtoch, dlžná suma na kreditných kartách, dlžná suma ostatných nezabezpečených úverov,
4. bod 3. *plus* hodnota nehnuteľností (na bývanie aj ostatných) *mínus* dlžná suma pri úveroch zabezpečených nehnuteľnosťami.

2.3 Softvérové spracovanie

Všetky analýzy uskutočnené v práci boli spracované v prostrední softvéru `R` (R Core Team, 2012). Na prácu so štandardnými ukazovateľmi odhadovanými v súlade s metodikou Eurostatu bola použitá predovšetkým knižnica `laeken`. V práci boli ďalej použité knižnice `reshape`, `shape`, `lattice`, `ineq`, `gmodels`.

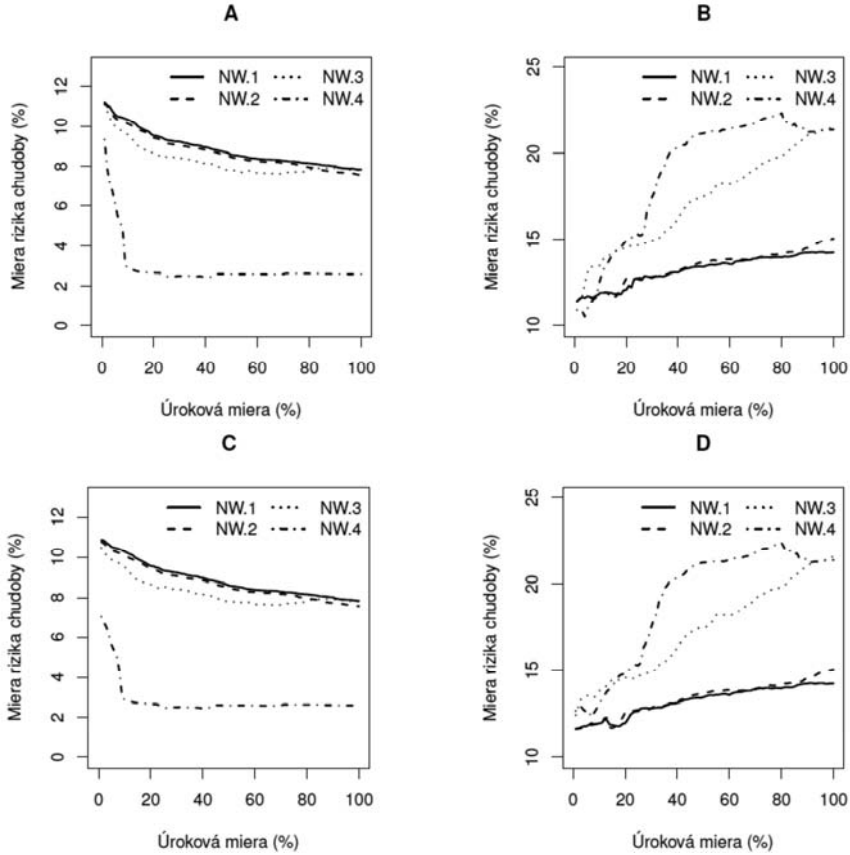
3. Výsledky a diskusia

3.3 Príjmovo-majetková chudoba

Postupom popísaným vyššie sú odhadnuté dve miery rizika chudoby (obr. 1), pričom sú zohľadnené dva prístupy k stanoveniu hranice chudoby: 1. hranica chudoby je definovaná ako 60 % mediánu ekvivalentného disponibilného príjmu odhadnutého na základe údajov HFCS (zjednodušene: príjmová hranica); 2. hranica chudoby je definovaná ako 60 % mediánu

[1] K diskusii o strednej dĺžke života a pri narodení v porovnaní so strednou dĺžkou života v zdraví pozri napr. Megyesiová, Lieskovská a Gazda (2012) a Megyesiová a Lieskovská (2013).

ekvivalentného disponibilného príjmu rozšíreného o anuitu z čistého príjmu (zjednodušene ju budeme označovať ako *príjmovo-majetková hranica*). Okrem príjmu rozšíreného o ročnú doživotnú anuitu definovaného vzťahom (1) je zohľadnený aj prepočet s predpokladom $n \to \infty$, s ktorým pracovali aj Brandolini, Magri a Smeeding (2010). Výsledkom je tak komparácia štyroch rôznych prístupov k vyjadreniu príjmovo-majetkovej chudoby.



Vysvetlivky: A: večná renta, príjmová hranica; B: večná renta, príjmovo-majetková hranica; C: doživotná renta, príjmová hranica; D: doživotná renta, príjmovo-majetková hranica.

Obr. 1: Citlivosť miery majetkovo-príjmovej chudoby na úrokovú mieru (vlastné spracovanie na základov údajov HFCS)

Na prvý pohľad je zrejmé, že NW.1 a NW.2 poskytujú veľmi podobné výsledky, čo možno vysvetliť relatívne nízkym podielom osôb žijúcich v domácnostiach s inými finančnými aktívami, akými sú bežné a sporiace účty. Použitím výlučne príjmovej hranice chudoby zároveň platí, že relatívne podobné výsledky získame aj zahrnutím hodnoty reálnych aktív bez nehnuteľností (NW.3). Vo všetkých troch prípadoch (NW.1, NW.2, NW.3) sa miera rizika chudoby približuje k úrovni 8 %. Zohľadnením celkovej hodnoty čistej majetku (NW.4)

na intervale úrokovej miery 1 -- 10 % p. a. rýchlo klesá z 9 % na cca 3 % a od tejto úrovne je necitlivá na výšku úrokovej miery.

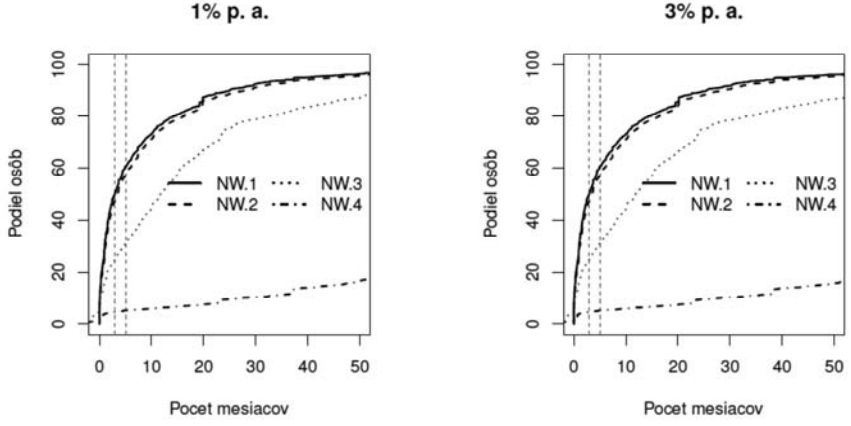
Iným prípadom je použitie hranice príjmovo-majetkovej chudoby, kedy pre každú úroveň príjmu zvýšenú o doživotnú/večnú ročnú anuitu čistej hodnoty príslušnej skupiny aktív je odhadnutá nová hranica chudoby (ako 60 % mediánu). Rastom úrokovej miery dochádza k rastu čistej súčasnej hodnoty majetku, čomu zodpovedá nový medián a nová hranica chudoby, a z toho dôvodu podiel osôb ohrozených rizikom chudoby narastá. Použitie NW.1 a NW.2 opäť poskytuje takmer rovnaké výsledky, avšak NW.3 a NW.4 vykazujú výrazne vyššiu citlivosť na úrokovú mieru ako NW.1 a NW.2. To je zapríčinené rádovo vyššími čistými hodnotami reálnych aktív ako finančných aktív.

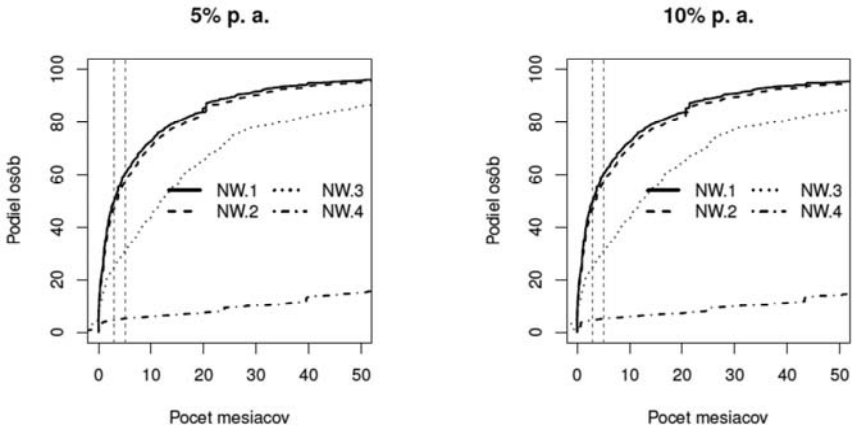
Použitie NW.1 a NW.2 poskytuje takmer rovnaké hodnoty, ktoré sa pohybujú v intervale cca 10 – 12 % bez ohľadu na typ renty, hranice chudoby a úrokovej miery (z intervalu 1 – 15 % p. a.). Zároveň je zrejmé, že na zvolené hodnoty úrokovej miery je najcitlivejšia čistá hodnota celkového majetku (NW.4).

3.4 Majetková chudoba

Nie je prekvapujúce, že medzi mierami rizika majetkovej chudoby odhadnuté použitím NW.1 a NW.2 existujú zanedbateľné rozdiely (obr. 2). Miera chudoby na základe NW.1 a NW.2 rastie veľmi rýchlym tempom -- napríklad ak by bolo minimálne požadované obdobie stanovené na 5 mesiacov, miera rizika chudoby by sa pohybovala na úrovni okolo 60 %. To sa dá vysvetliť samozrejme tým, že miera úspor na Slovensku je veľmi nízka (čo je determinované nízkou úrovňou príjmov). Z obr. 2 je zrejmé, že voľba úrokovej miery významným spôsobom neovplyvní odhad miery rizika chudoby (čo dokonca platí aj pre výrazne vyššie hodnoty úrokovej miery).

Použitie čistej hodnoty majetku domácnosti (NW.4) ako vstupu na určenie miery rizika chudoby je potrebné brať s istou dávkou abstrakcie. Ide totiž o prípad, kedy by domácnosť musela predať nehnuteľnosť, v ktorej býva, čo je skôr výnimočné.





Vertikálne prerušované čiary naznačujú obdobie 3 a 5 mesiacov

Obr. 2: Citlivosť miery majetkovej chudoby na úrokovú mieru a dĺžku obdobia (vlastné spracovanie na základov údajov HFCS)

4. Záver

Cieľom príspevku bolo analyzovať vplyv použitej úrokovej miery na odhad mier majetkovej chudoby. V práci boli použité dve miery majetkovej chudoby: rozšírenie príjmu o teoretický tok z čistej hodnoty majetku a odhad podielu osôb, ktorých prístup k majetkovým zdrojom je nedostatočný na to, aby im umožnil uspokojiť potreby počas zvoleného obdobia.

Výsledky naznačujú, že kým použitie prvého prístupu môže byť do istej miery výrazne závislé od hodnoty použitej úrokovej miery, druhý prístup je voči použitej úrokovej miere v zásade robustný.

5. Poďakovanie

Príspevok bol napísaný s podporou Vedeckej grantovej agentúry MŠ SR a SAV v rámci riešenia vedecko-výskumného projektu *VEGA 1/0127/11 Priestorová distribúcia chudoby v EÚ*.

Literatúra

BRANDOLINI, A. – MAGRI, S. – SMEEDING, T. M. (2010). Asset-Based Measurement of Poverty. *Journal of Policy Analysis and Management*, roč. 29, č. 2, s. 267–284.

CANER, A. – WOLFF, E. N. (2004). Asset Poverty in the United States, 1984-99: Evidence from the Panel Study of Income Dynamics. *Review of Income and Wealth*, roč. 50, č.4, s. 493–518.

ECB (2013). *The Eurosystem Household finance and Consumption Survey: Methodological Report for the First Wave*. Frankfurt am Main: European Central Bank.

IVANČÍKOVÁ, Ľ. – VLAČUHA, R. (2012). Možnosti merania majetkovej chudoby na Slovensku. In: Pauhofová, I. a Želinský, T. (eds.): *Nerovnosť a chudoba v Európskej únii a na Slovensku: Zborník statí*, s. 39–48. Košice: Ekonomická fakulta TUKE.

MEGYESIOVÁ, S. – LIESKOVSKÁ, V. (2013). Gesunde Lebensjahre bei der Geburt in den EU-Mitgliedstaaten. In: *Mitteilungen der Deutschen Gesellschaft für Demographie*, roč. 12, č. 15, s. 13.

MEGYESIOVÁ, S. – LIESKOVSKÁ, V. – GAZDA, V. (2012). Stredná dĺžka života pri narodení v porovnaní so strednou dĺžkou života v zdraví. In *Forum Statisticum Slovacum*, roč. 8, č. 2, s. 123-129.

NBS (2012). HFCS 2010. Bratislava: Národná banka Slovenska.

OLIVER, M. L. – SHAPIRO, T. M. (1990). Wealth of a Nation: A Reassessment of Asset Inequality in America Shows At Least One Third of Housebolds Are Asset-Ppoor. *The American Journal of Economics and Sociology*, roč. 49, č. 2, s. 129–151.

R CORE TEAM (2012). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

SENAJ, M. – ZAVADIL, T. (2012). *Výsledky prieskumu finančnej situácie slovenských domácností*. Bratislava: Národná banka Slovenska, príležitostná štúdia 1/2012.

WEISBROD, B. A. – HANSEN, W. L. (1968). An Income-Net Worth Approach to Measuring Economic Welfare. *The American Economic Review*, roč. 58, č. 5, s. 1315–1329.

Adresa autora:

Tomáš Želinský, doc. Ing. PhD.
Ekonomická fakulta, TU Košice
Němcovej 32, 040 01 Košice
tomas.zelinsky@tuke.sk

Zo života SŠDS

From live of SSDS

FERNSTAT 2013

V dňoch 12. a 13. septembra 2013 sa v horskom hoteli Šachtička pri Banskej Bystrici uskutočnil už IX. ročník medzinárodnej vedeckej konferencie FernStat (**F**inancie, **E**konómia/**E**konomika, **R**iadenie/**R**egióny, **N**ázory). Konferenciu zorganizovala banskobystrická pobočka Slovenskej štatistickej a demografickej spoločnosti (SŠDS) v spolupráci s Ekonomickou fakultou Univerzity Mateja Bela v Banskej Bystrici. Predsedom programového výboru bol Vladimír Úradníček a predsedníčkou organizačného výboru bola Mária Kanderová.

Na konferencii sa zúčastnili odborníci zo Slovenska, Poľska a Českej republiky, ktorí pre konferenciu pripravili viac ako 20 príspevkov tematicky zameraných na aplikovanú matematickú, ekonomickú, demografickú a výpočtovú štatistiku.

Počas rokovania konferencie a pri, do noci trvajúcej, panelovej diskusii si vymenili svoje odborné názory, skúsenosti, poznatky a zručnosti zástupcovia z Univerzity v Olsztyne (Poľsko), Univerzity Pardubice (Česká republika), Bankovního institutu vysoká škola Praha, zahraničná vysoká škola Banská Bystrica, Žilinskej univerzity, Univerzity Komenského v Bratislave, Akadémie ozbrojených síl generála Milana Rastislava Štefánika v Liptovskom Mikuláši a z Ekonomických fakúlt Technickej univerzity Košice a UMB v Banskej Bystrici.

Obr. 1: Benedykt Puczkowski z Univerzity v Olsztyne pri svojom vystúpení na konferencii

Osobitne zaujala sekcia veľmi kvalitných vystúpení mladých štatistikov (doktorandov) z Ekonomickej fakulty Technickej univerzity v Košiciach, Fakulty managementu Univerzity Komenského Bratislava, Slovenskej poľnohospodárskej univerzity v Nitre a z domácej Ekonomickej fakulty UMB v Banskej Bystrici. Je určite potešiteľné, že finančno-ekonomická štatistická veda má šikovnú mladú generáciu, ktorá vytvára reálne predpoklady pre možnosť ďalšieho dynamického napredovania tejto vednej oblasti v blízkej budúcnosti na Slovensku. Treba len dúfať, že spoločensko-ekonomické podmienky a nielen deklaratívne akcentovanie znalostnej ekonomiky neodradia týchto šikovných mladých ľudí od ich ďalšieho pôsobenia aj na poli štatistickej, príp. finančno-ekonomickej vedy. Našou spoločnou snahou by mohlo byť aj užšie prepojenie vedeckej spolupráce mladých vedeckých pracovníkov domácich a zahraničných vysokých škôl a pravidelné vytváranie priestoru na ich vzájomné stretnutia

a publikovanie výsledkov ich výskumu. Pekným príkladom môže byť v tejto súvislosti tradičná Prehliadka prác mladých štatistikov a demografov, ktorá sa v decembri 2013 uskutočňuje už po šestnásty krát.

Obr. 2: Pohľad na časť auditória pri rokovaní konferencie

Zhoršené povetrnostné podmienky neumožnili tento rok absolvovanie tradičnej dlhšie trvajúcej vychádzky cez Španiu Dolinu na Staré Hory, resp. na Donovaly, ale napriek tomu mohli účastníci konferencie diskutovať prerokovávanú problematiku aj aspoň počas kratšej prechádzky v bezprostrednom okolí Šachtičiek.

Sínusoidový entuziazmus organizátorov konferencie FernStat má ambíciu nájsť širšie zázemie pri organizovaní tejto konferencie. Sme si vedomí toho, že kvantitatívne narastajúci počet akcií Slovenskej štatistickej a demografickej spoločnosti nesmie byť na úkor kvality akcií, realizovaných regionálnymi pobočkami SŠDS. Za prínosné by sme preto považovali rozšírenie organizačných zložiek aj o inú regionálnu pobočku SŠDS, resp. o nového zahraničného partnera. Veríme, že tradícia konferencie FernStat bude pokračovať aj jej jubilejným, desiatym ročníkom s viacerými (kvalitatívne plusovými) novinkami.

Vladimír Úradníček
Ekonomická fakulta UMB

9. stretnutie štatistických spoločností v Ľubľani
9th meeting of the Statistical Societies in Ljubljana

Účastníci stretnutia – zľava Constantin Anchelage (Rumunsko), Mocja Noč Razinger (Slovinsko), Peter Mach (Slovensko), Constantin Mirut (Rumunsko), Hana Řezanková (Česko), Andrej Blejec (Slovinsko) Éva Laczka a Lorinc Soos (Maďarsko).

Z iniciatívy Maďarskej štatistickej spoločnosti sa pred deviatimi rokmi (2005) stretli zástupcovia šiestich štatistických spoločností zo strednej Európy – českej, maďarskej, rakúskej, rumunskej, slovinskej a slovenskej v Budapešti na prvom regionálnom stretnutí. Stretnutie popoludní pokračovalo vo Višegráde a preto sa táto skupina začala neformálne označovať V6. Vo Višegráde podpísali zástupcovia spoločností Dohodu o spolupráci, ktorá okrem iného deklarovala, že sa zástupcovia spoločností budú každoročne stretať, aby sa informovali o svojej činnosti a vymenili si názory na aktuálne otázky.

Deviate stretnutie predstaviteľov štatistických spoločností stredoeurópskeho regiónu V6 sa uskutočnilo 25. 10. 2013 v Ľubľane. Na stretnutí sa zúčastnili zástupcovia štatistických spoločností z Česka, Maďarska, Slovenska, Slovinska a Rumunska. Zástupcovia Rakúskej štatistickej spoločnosti sa ospravedlnili, lebo súčasne prebiehali Rakúske štatistické dni, hlavná akcia ich spoločnosti.

Stretnutie, ktoré sa konalo na pôde Slovinského štatistického úradu SR, otvoril a viedol predseda Slovinskej štatistickej spoločnosti Andrej Blejec.

Účastníkov stretnutia privítala zástupkyňa generálnej riaditeľky Slovinského štatistického úradu Karmen Hren. Informovala prítomných o práci úradu, ktorý má okolo 350 zamestnancov. Všetci zamestnanci pracujú priamo v novom sídle úradu (úrad nemá žiadne terénne pracoviská). Práca úradu je založená predovšetkým na využívaní administratívnych

zdrojov a registrov, čo predstavuje významné úspory pri zbere údajov. Ako príklad takýchto úspor Karmen Hren uviedla sčítanie obyvateľov v roku 2011, ktoré bolo kompletne realizované na základe údajov z registrov. Podľa jej vyjadrenia úrad takto ušetril 20 mil. eur. Spolupráca úradu so Slovinskou štatistickou spoločnosťou je veľmi dobrá, podpredsedníčkou spoločnosti je zamestnankyňa úradu Mocja Noč Razinger, ktorá sa tiež na stretnutí zúčastnila. Najvýznamnejším spoločným podujatím úradu a spoločnosti sú Štatistické dni, ktoré sa tradične konali v kúpeľnom mestečku Radenci, tohto roku sa však budú konať 19. 11. 2013 v mestečku Brdo.

Ďalším pravidelným bodom stretnutí je informácia o činnosti spoločností za uplynulý rok. Generálna tajomníčka Maďarskej štatistickej spoločnosti Éva Laczka v rámci svojej prezentácie informovala aj o niektorých akciách Rakúskej štatistickej spoločnosti. Prítomní si tiež pripomenuli bývalého predsedu Rakúskej štatistickej spoločnosti Joachima Lamela (jedného zo signatárov dohody o spolupráci V6), ktorý v tomto roku zomrel. Prácu Českej štatistickej spoločnosti stručne prezentovala jej predsedníčka Hana Řezánková, činnosť Rumunskej štatistickej spoločnosti jej zástupcovia Constantin Anchelage a Constantin Mirut, o činnosti Slovinskej štatistickej spoločnosti informoval jej predseda Andrej Blejec. Z činnosti Slovenskej štatistickej a demografickej spoločnosti som podrobnejšie informoval o oslavách 45. výročia vzniku spoločnosti, ktoré sa konali na slávnostnej konferencii pod záštitou predsedníčky ŠÚ SR Ľudmily Benkovičovej 20. 3. 2013 v Sládkovičove.

Ďalším bodom programu bola informácia o zasadaní Federácie európskych národných štatistických spoločností (FENStatS), ktoré sa konalo v auguste 2013 počas Svetového štatistického kongresu v Honk Kongu. O zasadaní informoval predseda Slovinskej štatistickej spoločnosti Andrej Blejec. V súčasnosti je už 5 spoločností z V6 členmi FENStatS. Prítomní podporili aj záujem Rumunska pripojiť sa k FENStatS, čím sa V6 stane špecifickou podmnožinou FENStatSu. V diskusii o otázkach spoločného záujmu sa hovorilo ďalej najmä o potrebe prehĺbiť informovanosť a koordináciu pri organizovaní medzinárodných akcií, aby nedochádzalo ku kolízii termínov, ktorá znemožňuje účasť. Pre zlepšenie výmeny informácií a skúseností medzi spoločnosťami by sa mohla uskutočniť napr. virtuálna konferencia, ktorá by pomohla pri výmene informácií. Spoločným problémom všetkých spoločností, o ktorom sa tiež hovorilo v diskusii, je zapájanie mladších štatistikov do činnosti spoločností.

Začiatkom roku 2012 sa Slovinský štatistický úrad presťahoval do novej modernej budovy. Účastníci stretnutia mali možnosť si prezrieť priestory, kde zamestnanci pracujú, školiace pracoviská, call centrum, výpočtové stredisko a tlačové centrum. Všetci pracovníci úradu majú dnes spoločné moderné pracovisko, technologicky vybavené na vysokej úrovni.

Na záver stretnutia podpísali účastníci spoločné vyhlásenie, ktoré potvrdzuje význam spolupráce skupiny V6. Spoločné vyhlásenie zdôrazňuje význam národných štatistických konferencií, ako aj vzájomnú výmenu informácií a význam zapájania mladých štatistikov do aktívnej práce v štatistických spoločnostiach.

Jubilejné desiate stretnutie spoločností sa uskutoční budúci rok v Prahe.

Peter Mach
podpredseda SŠDS pre medzinárodné styky

Z HISTÓRIE SEMINÁROV VÝPOČTOVÁ ŠTATISTIKA 2013
FROM THE HISTORY OF SEMINARS COMPUTATIONAL STATISTICS 2013

Pri príležitosti 22. ročníka seminára Výpočtová štatistika uvádzame stručnú chronológiu predošlých ročníkov.

Prvý seminár sa uskutočnil 9. - 10. 12. 1986 z iniciatívy zamestnancov Katedry štatistiky VŠE v Bratislave a Katedry statistiky VŠE v Prahe zaoberajúcimi sa problematikou využitia výpočtovej techniky v riešení štatistických úloh. Príspevky účastníkov boli uverejnené v Informáciách SDŠS č. 3 a č. 4 v roku 1986.

Miestom konania Seminárov bola do roku 2011, budova Infostat-u, v roku 2012 kongresová sála ŠÚ SR na Hanulovej 5/c v Bratislave. Väčšina seminárov sa organizovala v spolupráci so Štatistickým úradom SR (resp. SŠU v Bratislave) a Infostat-om Bratislava (resp. VUSEIaR Bratislava), v roku 2012 je spoluorganizátorom aj Prírodovedecká fakulta UK Bratislava. V aktuálnom 22. ročníku seminára bolo miesto konanie Aula Prírodovedeckej fakulty UK v Bratislave a druhá časť akcie pre mladých: Pohľady do analytiky - Analytika očami profesionálov - pásmo prednášok v zasadacej miestnosti Fakulty managementu UK, Odbojárov 10, Bratislava.

Druhý seminár prebehol 8. 12. 1987, tretí seminár 11. - 12. 12.1990. Potom nastala prestávka v organizácii seminárov Výpočtovej štatistiky a 4. seminár sa uskutočnil 7. - 8. 12. 1994.

Od 5. seminára uskutočneného 5. - 6. 12. 1996 sa už realizuje každoročne ako medzinárodný seminár.

6. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4.- 5. 12. 1997,

7. medzinárodný seminár Výpočtová štatistika sa uskutočnil 3. - 4. 12. 1998,

8. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 1999,

9. medzinárodný seminár Výpočtová štatistika sa uskutočnil 7. – 8. 12. 2000,

10. medzinárodný seminár Výpočtová štatistika sa uskutočnil 6. – 7. 12. 2001,

11. medzinárodný seminár Výpočtová štatistika sa uskutočnil 5. - 6. 12. 2002,

12. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4. - 5. 12. 2003,

13. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 2004,

14. medzinárodný seminár Výpočtová štatistika sa uskutočnil 1. - 2. 12. 2005,

15. medzinárodný seminár Výpočtová štatistika sa uskutočnil 7. - 8. 12. 2006,

16. medzinárodný seminár Výpočtová štatistika sa uskutočnil 6. - 7. 12. 2007,

17. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4. - 5. 12. 2008,

18. medzinárodný seminár Výpočtová štatistika sa uskutočnil 3. - 4. 12. 2009,

19. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 2010,

20. medzinárodný seminár Výpočtová štatistika sa uskutočnil 1. - 2. 12. 2011,

21. medzinárodný seminár Výpočtová štatistika sa uskutočnil 6. - 7. 12. 2012 a

22. medzinárodný seminár Výpočtová štatistika sa uskutočnil 5. - 6. 12. 2013.

Zameraním seminára je problematika na rozhraní počítačových vied a štatistiky.

Tematické okruhy posledných seminárov sa nemenia:

- praktické využitie paketov štatistických programov,
- práca s rozsiahlymi súbormi údajov,
- vyučovanie výpočtovej štatistiky a príbuzných predmetov,

- praktické aplikácie výpočtovej štatistiky,
- iné.

V čase konania seminára Výpočtová štatistika sa uskutočňuje aj ***prehliadka prác mladých štatistikov a demografov***. Táto akcia prebieha od 7. seminára. Na 8. medzinárodnom seminári prezentovalo svoje práce 5 mladých štatistikov a demografov, na 9. medzinárodnom seminári už bolo 20 prác mladých štatistikov a demografov, na 10. bolo prihlásených 26 prác a na 11. bolo prihlásených 18 prác, ale vzhľadom na niekoľko prác vypracovaných skupinou autorov bol počet účastníkov vyšší než predošlý rok. Na 12. seminári bolo prihlásených 19 prác, pričom niektoré sú prácou viacerých autorov. Na ďalšom 13. seminári bolo prihlásených 9 prác od 12 autorov. V rámci 14. seminára bolo prihlásených 15 sólových prác mladých autorov. Na 15. seminári bolo prihlásených 20 prác mladých autorov. V rámci 16. seminára bolo prihlásených 17 sólových prác mladých autorov. V rámci 17. seminára bolo prihlásených 15 sólových prác mladých autorov. V 18. ročníku bolo prihlásených 12 sólových prác mladých autorov. V 19. ročníku bolo prihlásených 15 prác autorov. V 20. ročníku seminára bolo prihlásených 15 prác mladých autorov. V 21. ročníku seminára bolo prihlásených 19 prác mladých autorov, z toho jedna práca sú dvaja autori. V aktuálnom ročníku 22. seminára bolo prihlásených 12 prác mladých autorov, z toho dve práce sú napísané dvoma autormi.

Prípadní záujemcovia z radov mladých štatistikov a demografov (za mladých považujeme štatistikov a demografov pred ukončením vysokej školy) môžu získať informácie na www.ssds.sk , blok akcie a na e-mailových adresách:

chajdiak@statis.biz ; jan.luha@fmed.uniba.sk ; iveta.stankovicova@fm.uniba.sk.

Informácie o najbližšom seminári získate na webovskej stránke SŠDS www.ssds.sk.

doc. Ing. Jozef Chajdiak, CSc.
STU Bratislava
predseda SŠDS

RNDr. Ján Luha, CSc.
LFUK Bratislava
vedecký tajomník SŠDS

doc. Ing. Iveta Stankovičová, PhD.
FM UK Bratislava
predsedníčka Programového a organizačného výboru seminára Výpočtová štatistika

OBSAH
CONTENTS

Máme zákazníkov v 135 krajinách.

Spolu ich je viac ako 60 000.

90 z top 100 firiem rebríčka Fortune Global 500 používa SAS.

Od nášho vzniku v roku 1976 sme najväčšia súkromne vlastnená softvérova spoločnosť na svete.

Až 25% z obratu investujeme do výskumu a vývoja.

SAS má vyše 13 500 zamestnancov vo viac ako 400 pobočkách po celom svete.

SAS je popredným svetovým poskytovateľom riešení v oblasti

Rýchle spracovanie obrovských objemov údajov vedie k rýchlejším a presnejším strategickým podnikateľským rozhodnutiam, ziskovejším vzťahom so zákazníkmi a dodávateľ

Na Slovensku má SAS svoje zastúpenie od roku 1995 a špecializuje sa hlavne na finančný sektor, te

ani na Slovensku, kde od roku 1999 spolupracujeme s najvýznamnejšími vysokými školami s cieľom podporiť vzdelávanie študentov nielen teoreticky,

Viac informácií nájdete na:
www

The Power to KNOW®

„Vďaka tomu, že študenti využívajú vysoko profesionálny analytický softvér, získavajú predstavu o tom, v akých situáciách sa dajú analýzy dát využiť. Učia sa, ako v praxi používať rôzne analytické metódy a interpretovať výsledky, ktoré zo softvéru vzídu. Rozhodovať iba na základe intuície sa dnes nedá. Analýzy dát a získavanie užitočných informácií zohrávajú v biznisových rozhodnutiach čoraz väčšiu úlohu."

doc. Ing. Iveta Stankovičová, PhD.
vysokoškolský pedagóg
Fakulta managementu
Univerzita Komenského v Bratislave

„Po zoznámení sa so softvérom SAS® sa pre mňa stala štatistika po množstve teoretických poznatkov oveľa atraktívnejšou. Zaujímavé grafické spracovanie i široká škála možností využitia boli hlavné dôvody prečo som sa rozhodol venovať väčšiu pozornosť práve tomuto softvéru."

Ondrej Dúžik
doktorand
Fakulta hospodárskej informatiky
Ekonomická univerzita v Bratislave

„Študenti sú neraz milo prekvapení, že môžu pracovať s aktuálnym softvérom, ktorý využíva toľko firiem po celom svete. Veľmi pozitívne vnímali aj prednášky SAS konzultantov u nás na fakulte, ktorí im priblížili využitie SAS-u v praxi. Som presvedčená, že vedieť aspoň niečo zo SAS-u, prináša študentom na pohovoroch značnú konkurenčnú výhodu oproti ostatným a nadobudnuté vedomosti im budú nápomocné v ich budúcom zamestnaní."

Ing. Renáta Prokeinová PhD.
vysokoškolský pedagóg
Fakulta ekonomiky a manažmentu
Slovenská poľnohospodárska univerzita v Nitre

sas® | THE POWER TO KNOW.®

SAS je popredným svetovým poskytovateľom riešení v oblasti biznis analytiky a odborných služieb. Celosvetovo softvér SAS® využíva viac ako 65 000 organizácií pre zlepšenie svojej výkonnosti pomocou spracovania obrovských objemov údajov, čo vedie napr. k rýchlejším a presnejším strategickým podnikateľským rozhodnutiam, ziskovejším vzťahom so zákazníkmi a dodávateľmi a dodržiavaniu regulačných požiadaviek. Spoločnosť SAS má viac ako 13 000 zamestnancov vo viac ako 400 pobočkách v 55 štátoch sveta. Na Slovensku má svoje zastúpenie od roku 1995 a špecializuje sa hlavne na finančný sektor, telekomunikácie a energetiku. Viac informácií nájdete na: **www.sas.com/slovakia.**

SAS Academic Program

Spoznajte softvér, ktorý je už 37 rokov celosvetovým lídrom v štatistike a analytike. Pridajte sa k vyše 3000 univerzitám z celého sveta, ktoré pri výučbe využívajú SAS.

SAS na vyučovaní

Softvér SAS® je vysokým školám poskytovaný pre účely výučby a výskumných projektov na základe platby ročných licenčných poplatkov. Pedagógovia v rámci našej spolupráce získavajú prístup k skriptám, cvičným dátam a rôznym učebným materiálom. Tak pedagógom pomáhame zaujať a učiť študentov praktické veci na najnovších technológiách.

SAS pre tvorbu záverečných prác

Podporujeme študentov k tvorbe záverečných prác v SAS-e. Študentom poskytujeme softvér počas doby realizácie výskumného projektu, ŠVOČ, bakalárskej, diplomovej alebo dizertačnej práce úplne zadarmo. Každý študent navyše získa možnosť zapojiť sa do súťaže o najlepšiu študentskú prácu.

Motivácia certifikátmi

Certifikát od SAS-u môže získať každý študent, ktorému pedagóg na záver semestra udelí z predmetu vyučovaného v softvéri SAS® hodnotenie "A". Aj takouto cestou chceme podporiť tých najšikovnejších študentov a pomôcť im pri uplatnení v praxi.

Prednášky pre študentov

Pravidelne sa stretávame so študentami na fakultách po celom Slovensku. Na prednáškach im poukazujeme na praktickú využiteľnosť biznis analytiky v rôznych priemyselných odvetviach. Naším cieľom je pomôcť im spoznať nielen teóriu, ale aj aplikáciu so softvérom (najmä SAS® Enterprise Miner™ a SAS® Enterprise Guide®), s ktorým sa stretnú v praxi.

Súťaže

Každoročne pre študentov a pedagógov pripravujeme rôzne súťaže s ich prácami, pri ktorých použili softvér SAS®. Hodnotíme možnosť praktického využitia výsledkov práce, správnu aplikáciu metód SAS pre riešenia daného problému a inovatívne využitie nástrojov SAS pre uchopenie danej problematiky.

Školenia

Vychádzame v ústrety študentom, ktorí chcú poznať SAS ešte viac a ponúkame im špeciálne školenia za študentské ceny. Predpokladáme, že absolvovanie takéhoto školenia bude pozitívne vnímané budúcimi zamestnávateľmi v životopisoch študentov a absolventov. V ponuke máme extra školenia aj pre pedagógov.

Pracovné ponuky

Vďaka týmto aktivitám pre študentov sme často v kontakte s tými najšikovnejšími. Máme s nimi konkrétne skúsenosti i referencie od ich pedagógov. Študenti sa potom na nás zvyknu obracať s prosbou o pomoc pri hľadaní práce/praxe/stáže u našich zákazníkov. Je našou prioritou vedieť tých najšikovnejších z nich uplatniť v praxi.

Kontakt:

Barbora Okruhľanská
Academic Program Coordinator
Barbora.Okruhlanska@sas.com
+421 2 5778 0949
www.sas.com/slovakia/academic

www.facebook.com/SASslovakia

MICROCOMP

MICROCOMP-Computersystém s.r.o.
je úspešným dodávateľom
informačných technológií a riešiteľom
projektov informačnej bezpečnosti.

systémová integrácia

dodávky hardvéru

dodávky dátových sietí

vývoj, úpravy a customizácia
informačných systémov

analytické práce

vytváranie a realizácia bezpečnostných
projektov informačných systémov

vzdelávanie, školenia

konzultácie pre zákazníkov

servisná podpora, záručný
a pozáručný servis

Sídlo
Kupecká 9
94901 Nitra
tel.: +421 37 6511306
fax: +421 37 6516166
obchod@microcomp.sk

Pobočka
Odborárska 5
83102 Bratislava
tel.: +421 2 53631221
fax: +421 2 53419854

Pobočka
Na troskách 16
97401 Banská Bystrica
tel.: +421 48 4143052
fax: +421 48 4143053

www.microcomp.sk

Pokyny pre autorov

Jednotlivé čísla vedeckého recenzovaného časopisu FORUM STATISTICUM SLOVACUM sú prevažne tematicky zamerané zhodne s tematickým zameraním akcií SŠDS. Príspevky v elektronickej podobe prijíma zástupca redakčnej rady na elektronickej adrese uvedenej v pozvánke na konkrétne odborné podujatie Slovenskej štatistickej a demografickej spoločnosti. Akceptujeme príspevky v slovenčine, češtine, angličtine, nemčine, ruštine a výnimočne po schválení redakčnou radou aj inom jazyku. Názov word-súboru uvádzajte a posielajte v tvare: ***priezvisko_nazovakcie.doc resp. docx***

Forma: Príspevky písané výlučne len v textovom editore MS WORD, verzia 6 a vyššia, písmo Times New Roman CE 12, riadkovanie jednoduché (1), formát strany A4, všetky okraje 2,5 cm, strany nečíslovať. Tabuľky a grafy v čierno-bielom prevedení zaradiť priamo do textu článku a označiť podľa šablóny. Bibliografické odkazy uvádzať v súlade s normou STN ISO 690 a v súlade s medzinárodnými štandardami. Citácie s poradovým číslom z bibliografického zoznamu uvádzať priamo v texte.

Rozsah: Maximálny rozsah príspevku je 6 strán.

Príspevky sú recenzované. Redakčná rada zabezpečí posúdenie príspevku oponentom.

Príspevky nie sú honorované, poplatok za uverejnenie akceptovaného príspevku je minimálne 30 €. Za každú stranu naviac je poplatok 5 €.

Štruktúra príspevku: ***(Pri písaní príspevku využite elektronickú šablónu: http://www.ssds.sk/ v časti Vedecký časopis, Pokyny pre autorov.). Časti v angličtine sú povinné!***

Názov príspevku v slovenskom jazyku (štýl Názov: Time New Roman 14, Bold, centrovať)

Názov príspevku v anglickom jazyku (štýl Názov: Time New Roman 14, Bold, centrovať)

Vynechať riadok

Meno1 Priezvisko1, Meno2 Priezvisko2 (štýl normálny: Time New Roman 12, centrovať)

Vynechať riadok

Abstrakt: Text abstraktu v slovenskom jazyku, max. 10 riadkov (štýl normálny: Time New Roman 12).

Abstract: Text abstraktu v anglickom jazyku, max. 10 riadkov (štýl normálny: Time New Roman 12).

Kľúčové slová: Kľúčové slová v slovenskom jazyku, max. 2 riadky (štýl normálny: Time New Roman 12).

Key words: Kľúčové slová v anglickom jazyku, max. 2 riadky (štýl normálny: Time New Roman 12).

JEL classification: Uviesť kódy klasifikácie podľa pokynov v:
<http://www.aeaweb.org/journal/jel_class_system.php>

Vynechať riadok a nastaviť si medzery odseku pre nadpisy takto: medzera pred 12 pt a po 3 pt. Nasleduje vlastný text príspevku v členení:

1. **Úvod** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať,)*
2. **Názov časti 1** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*
3. **Názov časti 1. . .**
4. **Záver** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*

Vlastný text jednotlivých častí je písaný štýlom Normal: písmo Time New Roman 12, prvý riadok odseku je odsadený vždy na 1 cm, odsek je zarovnaný s pevným okrajom. Riadky medzi časťami a odsekmi nevynechávajte. Nastavte si medzi odsekmi medzeru pred 0 pt a po 3 pt.

5. **Literatúra** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, číslovať)*

[1] Písať podľa normy STN ISO 690

[2] GRANGER, C.W. – NEWBOLD, P. 1974. Spurious Regression in Econometrics. In: Journal of Econometrics, č. 2, 1974, s. 111 – 120.

Adresa autora (-ov): ***Uveďte svoju pracovnú adresu!!!*** *(štýl Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, adresy vpísať do tabuľky bez orámovania s potrebným počtom stĺpcov a s 1 riadkom):*

Meno1 Priezvisko1, tituly1 (študenti ročník)
Pracovisko1 (študenti škola1)
Ulica1, 970 00 Mesto1
meno1.priezvisko1@mail.sk

Meno2 Priezvisko2 , tituly2 (študenti ročník)
Pracovisko2 (študenti škola2)
Ulica2, 970 00 Mesto2
meno2.priezvisko2@mail.sk

FORUM STATISTICUM SLOVACUM

vedecký recenzovaný časopis Slovenskej štatistickej a demografickej spoločnosti

Vydavateľ:
Slovenská štatistická a demografická spoločnosť
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Redakcia:
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Fax: 02/39004009

e-mail:
chajdiak@statis.biz
jan.luha@fmed.uniba.sk

Dátum vydania: november 2013

Registráciu vykonalo:
Ministerstvo kultúry Slovenskej republiky

Dátum registrácie: 22. 7. 2005

Evidenčné číslo: EV 3287/09

Tematická skupina: B1

Periodicita vydávania:
minimálne 2 krát ročne

Objednávky:
Slovenská štatistická a demografická spoločnosť
Miletičova 3, 824 67 Bratislava 24
Slovenská republika

IČO: 178764
DIČ: 2021504276
Číslo účtu: 0011469672/0900

ISSN 1336-7420

Ročník: IX.

Číslo: 7/2013

Cena výtlačku: 30 EUR
Ročné predplatné: 120 EUR