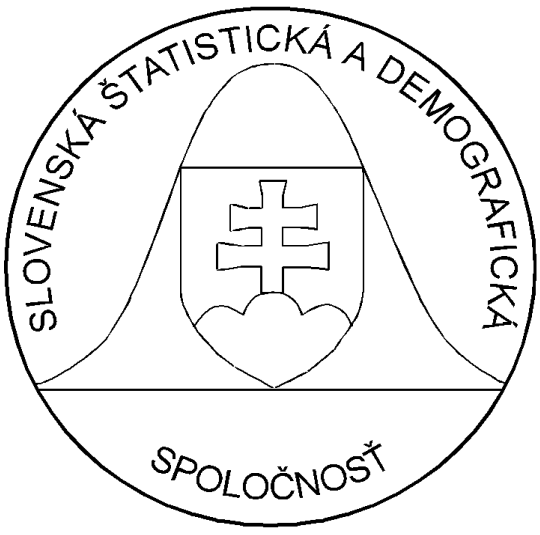
6/2007

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 Poriadané akcie)

SLÁVNOSTNÁ KONFERENCIA 40 ROKOV SŠDS,
27. 3. 2008, Bratislava

KONFERENCIA POHĽADY NA EKONOMIKU SLOVENSKA 2008,
tematické zameranie: *Vývoj HDP a vývoj dlhodobej nezamestnanosti*
15. 4. 2008, Bratislava, hotel Bôrik

EKOMSTAT 2008, 22. škola štatistiky
Tematické zameranie: *Štatistické metódy v praxi*
1. – 6. 6. 2008, Trenčianske Teplice

FernStat 2008
V. medzinárodná konferencia aplikovanej štatistiky
(**F**inancie, **E**konomika, **R**iadenie, **N**ázory)
tematické zameranie: *Aplikovaná, demografická, matematická štatistika, štatistické riadenie kvality*.
rok 2008, hotel Lesák, Tajov pri Banskej Bystrici

14. SLOVENSKÁ ŠTATISTICKÁ KONFERENCIA,
tematické zameranie: *Regionálna štatistika*
rok 2008, Žilinský kraj

12. SLOVENSKÁ DEMOGRAFICKÁ KONFERENCIA,
tematické zameranie: *Využitie GIS v demografii*
rok 2009, Trenčiansky kraj

ÚVOD

Vážené kolegyne, vážení kolegovia,
šieste číslo tretieho ročníka vedeckého časopisu Slovenskej štatistickej a demografickej spoločnosti (SŠDS) je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou 16. medzinárodného seminára Výpočtová štatistika a Prehliadkou prác mladých štatistikov a demografov. Tieto akcie sa uskutočili v dňoch 6. a 7. decembra 2007 na Infostate v Bratislave.

Akcie, z poverenia Výboru SŠDS, zorganizoval Organizačný a programový výbor: doc. Ing. Jozef Chajdiak, CSc. – predseda, RNDr. Ján Luha, CSc. – tajomník, Doc. RNDr. Beáta Stehlíková, CSc., Doc. RNDr. Bohdan Linda, CSc., Doc. Dr. Jana Kubanová, CSc., RNDr. Jitka Bartošová PhD., Ing. Vladimír Úradníček PhD., Ing. Iveta Stankovičová, PhD., RNDr. Samuel Koróny.

Na príprave a zostavení tohto čísla participovali: doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc. , Ing. Iveta Stankovičová, PhD.

Recenziu príspevkov zabezpečili: doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., Ing. Iveta Stankovičová, PhD., RNDr. Samuel Koróny.

Veľmi nás teší neustály záujem o seminár Výpočtová štatistika. Výbor SŠDS oceňuje aktivitu mladých v rámci Prehliadky prác mladých štatistikov a demografov, čo svedčí tiež o dobrej práci pedagógov a ich študentov. Dúfame, že možnosť prezentácie zvyšuje aj odbornú úroveň mladých štatistikov a demografov.

Organizátori seminára si považujú za milú povinnosť poďakovať za podporu a pomoc predsedníčke Štatistického úradu SR PhDr. Ľudmile Benkovičovej, CSc. a Infostatu.

Výbor SŠDS

Z HISTÓRIE SEMINÁROV VÝPOČTOVÁ ŠTATISTIKA

Pri príležitosti 16. ročníka semináru 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 vždy budova Infostat-u a väčšina seminárov sa organizovala v spolupráci so Štatistickým úradom SR (resp. SŠU v Bratislave) a Infostat-om Bratislava (resp. VUSEIaR Bratislava).

Druhý seminár prebehol 8. 12. 1987, tretí seminár 11. - 12. 12.1990. Pád socializmu a spoločenské zmeny spôsobili určitú prestávku v organizácii seminárov Výpočtovej štatistiky.
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 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 a
16. medzinárodný seminár Výpočtová štatistika sa uskutočnil 6. - 7. 12. 2007.

Príspevky 2. seminára boli opublikované v Informáciách SDŠS č. 1/1999 a od 3. seminára sa publikujú v samostatnom Zborníku príspevkov príslušného seminára. Od 14. seminára sú príspevky publikované vo vedeckom časopise SŠDS FORUM STATISTICUM SLOVACUM.

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 aktuálnom ročníku, v rámci 16. seminára bolo prihlásených 17 sólových prác mladých autorov.

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 resp. Jan.Luha@statistics.sk

Informácie o najbližšom seminári získate na webovskej stránke SŠDS http://www.ssds.sk/ resp. http://www.statistics.sk v bloku Slovenská štatistická a demografická spoločnosť.

Doc. Ing. Jozef Chajdiak, CSc.
vedecký tajomník SŠDS
RNDr. Ján Luha, CSc.
člen sekretariátu Výboru SŠDS

Modely mzdových rozdělení

Diana Bílková

Abstract: Lognormal distribution is considered as the one of fundamental distributions in statistics. Different methods can be used for the estimation of its parameters. As for examples, moment method, maximum likelihood method, quantile method, Kemsley method, Cohen method and graphical method can be mentioned. Pearson´s system of curves enables, due to its flexibility, to reach almost perfect agreement with empirical distribution. However, it is a case of purely formal method, where some difficulties with interpretation occur. Johnson´s system of curves is based on transforms into normal distribution and three-parametric lognormal curve, i.e. Johnson´s curve of S_L type, is its part as well.

Key words: Moment method, maximum likelihood method, quantile method, Pearson´s system, Johnson´s system

1. Úvod

Zájem ekonomů o mzdy obyvatelstva ve všech vyspělých zemích pramení ze snahy o objektivní řešení otázek vztahujících se k životní úrovni obyvatelstva. Předpovědi mzdových rozdělení zkonstruované na základě mzdových modelů umožňují úspěšně řešit otázky tohoto typu. Mzdových modelů lze využít při hodnocení životní úrovně nebo při mezioblastním či mezinárodním srovnávání životní úrovně. Pro správné hodnocení mzdové stránky životní úrovně obyvatelstva i pro správné rozhodování ohledně opatření v této oblasti je nezbytné znát úplné rozdělení mezd daného období, tj. znát obsazení ve všech mzdových skupinách. K rozdělení mezd obyvatelstva by měli přihlížet i podnikatelé při zvažování odbytových možností, tato rozdělení mohou být rovněž využita i při různých úvahách o daňovém zatížení apod. Znalostí rozdělení mezd lze nepřímo využít i k jiným účelům. Jedná se o zdokonalení postupů při výběrovém zjišťování jiných statistických znaků, které jsou lineárně závislé na mzdách. Jako příklad můžeme uvést výdaje domácností, vybavenost domácností, využití času, nákupní úmysly, ale rovněž postoje při různých sociologických průzkumech.

2. Konstrukce mzdového modelu

Nejdůležitějším úkolem je konstrukce statického mzdového modelu, který představuje křivku, jejíž tvar dobře zobrazuje průběh polygonu rozdělení četností. Při konstrukci takovéto křivky by mělo dojít k vyvážení dvou protichůdných požadavků spočívajících v tom, že uvedená křivka by měla mít z hlediska pružnosti a přizpůsobivosti se skutečnému tvaru rozdělení dostatečný počet parametrů, na druhé straně však, má-li křivka příliš velký počet parametrů, znamená to újmu na jejich ekonomické interpretovatelnosti a činí model méně stabilním v prostoru a čase. Jedním z nejpoužívanějších rozdělení v rámci modelování mzdových rozdělení je lognormální rozdělení. O lognormálním rozdělení lze uvažovat jako o vhodném modelu kdykoliv v situacích, ve kterých je potřeba model s kladnou šikmostí. Z hlediska počtu parametrů rozlišujeme dvouparametrické, tříparametrické a čtyřparametrické lognormální rozdělení. V případě mzdových rozdělení nachází uplatnění především dvouparametrické a tříparametrické lognormální rozdělení. Jako doplněk lognormálních modelů je možno využít křivek z Pearsonova nebo Johnsonova systému.

2.1 Tříparametrické a dvouparametrické lognormální rozdělení

Náhodná veličina X má tříparametrické lognormální rozdělení LN(μ,σ^2,θ) s parametry μ, σ^2 a θ, kde $-\infty < \mu < \infty$, $\sigma^2 > 0$ a $-\infty < \theta < \infty$, jestliže její hustota pravděpodobnosti $f(x; \mu,\sigma^2,\theta)$ má tvar

$$f(x;\mu,\sigma^2,\theta) = \frac{1}{\sigma (x-\theta)\sqrt{2\pi}} e^{-\frac{[\ln(x-\theta)-\mu]^2}{2\sigma^2}}, \quad x > \theta, \qquad (1)$$
$$= 0, \quad \text{jinak}.$$

náhodná veličina

$$Y = \ln(X - \theta) \qquad (2)$$

má normální rozdělení N(μ,σ^2) a náhodná veličina

$$U = \frac{\ln(X-\theta)-\mu}{\sigma} \qquad (3)$$

má normované normální rozdělení N(0;1). Parametr μ je střední hodnota náhodné veličiny (2) a parametr σ^2 je rozptyl této náhodné veličiny. Parametr θ představuje teoretické minimum náhodné veličiny X. Z výsledků našich výpočtů je zřejmé, že u mzdových rozdělení je v řadě případů hodnota parametru θ záporná, tedy tříparametrická lognormální křivka se často v začátku svého průběhu dostává pod nulu. To však, vzhledem ke skutečnosti, že zprvu má křivka velice těsný kontakt s osou x, nemusí vadit dobré shodě modelu se skutečným rozdělením. Jestliže do vztahů (1), (2) a (3) dosadíme $\theta = 0$, získáváme dvouparametrické lognormální rozdělení. Charakteristiky lognormálního rozdělení jsou popsány např. v [2].

Je nutné si uvědomit, že mzdy zpravidla vyčerpávajícím zjišťováním nelze zkoumat, a proto je třeba parametry modelu odhadnout na základě náhodného výběru, kdy využíváme některé z metod bodového odhadu, např. momentovou metodu, kvantilovou metodu, metodu maximální věrohodnosti apod. Je známo, že při použití momentové metody odhadu parametrů není zaručena maximální vydatnost odhadu. V případě mzdových rozdělení však problémy tohoto typu zpravidla nenastávají a momentová metoda je použitelná, neboť se zde zpravidla jedná o výběry značných rozsahů. V takovém případě každá konzistentní metoda odhadu parametrů, a tedy rovněž momentová metoda, přinese dobré výsledky.

Dále je třeba posoudit vhodnost zkonstruovaného modelu nebo vybrat jeden model z několika různých alternativ, což se provádí pomocí nějakého kritéria, jakým může být např. součet absolutních odchylek pozorovaných a teoretických četností za všechny intervaly

$$S = \sum_{j=1}^{k} \left| n_j - n \cdot \pi_j \right| \qquad (4)$$

nebo známé kritérium χ^2

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - n\cdot\pi_j)^2}{n\cdot\pi_j}, \qquad (5)$$

kde n_j jsou pozorované četnosti v jednotlivých intervalech, π_j jsou teoretické pravděpodobnosti příslušnosti statistické jednotky do j-tého intervalu, n je celkový rozsah statistického souboru, $n.\pi_j$ jsou teoretické četnosti v jednotlivých intervalech, $j = 1, 2, \ldots, k$, a k je počet intervalů.

Otázka vhodnosti dané křivky pro model rozdělení mezd není však zcela běžným matematicko-statistickým problémem, při kterém testujeme nulovou hypotézu

„H_0: Výběr pochází z předpokládaného teoretického rozdělení.“

proti alternativní hypotéze

„H_1: Není tomu tak.“,

neboť při testech dobré shody se v případě mzdových rozdělení zpravidla setkáváme s tím, že pracujeme s rozsáhlými soubory, a proto by test téměř vždy vedl k zamítnutí nulové hypotézy. Vyplývá to nejen ze samotného principu konstrukce testu, ale rovněž z toho, že při takto velkých rozsazích výběrů je při zvolené hladině významnosti tak velká síla testu, že test odkryje všechny sebenepatrnější odchylky skutečného mzdového rozdělení a modelu. Malé odchylky nás ale prakticky nezajímají, stačí pouze přibližná shoda modelu se skutečností a model, tj. křivku“ si v takovém případě „vypůjčíme“. Testové kritérium (5) lze v tomto směru použít pouze orientačně, při vyhodnocování vhodnosti modelu je třeba postupovat do značné míry subjektivně a opírat se o logický rozbor a zkušenost.

3. Datová základna

Datovou základnu pro naše výpočty představují osoby rozdělené podle pohlaví na muže a ženy a osoby celkem za obě pohlaví dohromady, a to za roky 2004 a 2005. Sledovanou proměnnou byly hrubé měsíční mzdy v Kč. Údaje potřebné pro výpočty pocházejí z internetových stránek Českého statistického úřadu.

Na základě získaných dat byly vypočteny výběrové charakteristiky, a to aritmetický průměr, rozptyl, směrodatná odchylka a variační koeficient, které se nacházejí v tabulce 1, a momentová míra šikmosti a špičatosti, které udává tabulka 2. Tabulka 3 obsahuje odhadnuté hodnoty kvartilů z intervalového rozdělení četností.

Tabulka 1: Aritmetický průměr $\bar{x}$ (v Kč), rozptyl s_x^2(v Kč^2) , směrodatná odchylka s_x (v Kč) a variační koeficient v_x (v %) hrubých měsíčních mezd

	Výběrové charakteristiky statistického souboru			
	$\bar{x}$	s_x^2	s_x	v_x
Rok 2005	20 300	75 784 932	8 705	42,88
Rok 2004	19 350	70 767 278	8 412	43,48
Rok 2005 – muži	22 196	80 191 300	8 955	40,35
Rok 2005 – ženy	17 779	58 793 830	7 668	43,13
Rok 2004 – muži	21 207	75 584 900	8 694	41,00
Rok 2004 – ženy	16 905	53 908 828	7 342	43,43

Tabulka 2: Momentová míra šikmosti b_1 a momentová míra špičatosti b_2 hrubých měsíčních mezd

	Výběrové charakteristiky statistického souboru	
	b_1	b_2
Rok 2005	0,908 501	0,339 528
Rok 2004	1,034 369	0,771 707
Rok 2005 – muži	0,763 941	-2,058 431
Rok 2005 – ženy	1,135 804	1,383 773
Rok 2004 – muži	0,898 242	0,246 010
Rok 2004 – ženy	1,254 582	1,948 976

Z tabulky 1 jsou patrné značné rozdíly v průměru hrubých měsíčních mezd mezi muži a ženami jak v roce 2004, tak v roce 2005. V roce 2004 je průměr hrubých měsíčních mezd žen o 4 302 Kč nižší než průměr hrubých měsíčních mezd mužů a v roce 2005 tento rozdíl činí 4 417 Kč. Z této tabulky je také patrný růst průměru hrubých měsíčních mezd mezi roky 2004 a 2005 o 950 Kč.

Tabulka 1 skýtá dále určitý přehled o rozdílech v příjmové diferenciaci. Je zde vidět, že hodnoty rozptylu mezi roky 2004 a 2005 rostou, a to ať už se jedná o celkový soubor mužů a žen dohromady nebo soubory zvlášť mužů a zvlášť žen. Data tedy zřejmě nelze považovat za homoskedastická ve smyslu stejné variability stejných rozdělení, neboť absolutní charakteristiky variability, v našem případě rozptyl, a tedy i směrodatná odchylka, v čase zřejmě porostou. Variační koeficient je charakteristikou relativní variability. Z tabulky 1 je zřejmé, že hodnoty variačního koeficientu se příliš neliší a mezi roky 2004 a 2005 spíše nepatrně klesají jak u celkového souboru mužů a žen dohromady, tak zvlášť u souboru mužů a zvlášť u souboru žen.

Tabulka 2 uvádí hodnoty charakteristik tvaru rozdělení. Z tabulky 2 je zřejmé, že rozdělení hrubých měsíčních mezd u všech souborů, které jsme sledovali, se, jak lze v případě mzdových rozdělení očekávat, vyznačují kladnou šikmostí. Hodnoty obou těchto charakteristik jsou u všech sledovaných souborů téměř shodné, výjimku tvoří pouze soubor mužů v roce 2005, kde vyšla momentová charakteristika špičatosti záporná, ve všech ostatních případech vychází tato charakteristika kladně. Dále pro lepší přehled o poloze rozdělení hrubých měsíčních mezd byly odhadnuty hodnoty kvartilů, které udává tabulka 3.

Tabulka 3: Odhadnuté hodnoty výběrových kvartilů (v Kč) z intervalového rozdělení četností

	Dolní kvartil	Medián	Horní kvartil
	$\tilde{x}_{25}$	$\tilde{x}_{50}$	$\tilde{x}_{75}$
Rok 2005	13 924	18 667	24 542
Rok 2004	13 318	17 803	23 258
Rok 2005 – muži	15 677	20 319	26 974
Rok 2005 – ženy	12 156	16 496	21 669
Rok 2004 – muži	14 897	19 388	25 507
Rok 2004 – ženy	11 563	15 643	20 373

4. Metody odhadu parametrů tříparametrického lognormálního rozdělení

Podstata momentové metody odhadu parametrů spočívá v tom, že se položí do rovnosti výběrové a příslušné teoretické momenty daného rozdělení. Obecné a centrální momenty je možné kombinovat. Nyní odhadujeme tři parametry, potřebujeme tedy soustavu tří momentových rovnic. První rovnici získáme tak, že položíme výběrový aritmetický průměr $\bar{x}$ do rovnosti střední hodnotě tříparametrického lognormálního rozdělení, v případě druhé rovnice výběrový druhý centrální moment m_2 položíme do rovnosti rozptylu tříparametrického lognormálního rozdělení. Třetí rovnici získáme tak, že položíme do rovnosti výběrový třetí centrální moment m_3 s teoretickým třetím centrálním momentem tříparametrického lognormálního rozdělení. Řešením soustavy tří momentových rovnic získáváme momentové odhady parametrů tříparametrického lognormálního rozdělení

$$\tilde{\sigma}^2 = \ln\left[\sqrt[3]{1+\frac{1}{2}b_1^2+\sqrt{\left(1+\frac{1}{2}b_1^2\right)^2-1}}+\sqrt[3]{1+\frac{1}{2}b_1^2-\sqrt{\left(1+\frac{1}{2}b_1^2\right)^2-1}}-1\right], \tag{6.1}$$

$$\tilde{\mu} = \frac{1}{2}\ln\frac{m_2}{e^{\tilde{\sigma}^2}(e^{\tilde{\sigma}^2}-1)}, \tag{6.2}$$

$$\tilde{\theta} = \bar{x} - e^{\tilde{\mu}+\frac{\tilde{\sigma}^2}{2}}. \tag{6.3}$$

V tabulce 4 se nacházejí odhadnuté hodnoty parametrů tříparametrického lognormálního rozdělení získané momentovou metodou včetně hodnoty testového kritéria (5).

Výhodou lognormálních modelů je, že parametry těchto modelů mají jednoduchou interpretaci. Jednoduchou interpretaci mají rovněž některé funkce parametrů lognormálních křivek. Z tabulky 4 je zřejmé, že ve všech případech je momentovou metodou odhadnutá hodnota parametru θ záporná, což znamená, že zpočátku průběhu se tato křivka dostává do záporných hodnot. Z hlediska záporných hodnot odhadů parametru θ nelze tomuto parametru přikládat žádnou interpretaci.

Tabulka 4: Odhadnuté hodnoty parametrů tříparametrického lognormálního rozdělení momentovou metodou a hodnota testového kritéria χ^2

	$\tilde{\mu}$	$\tilde{\sigma}^2$	$\tilde{\theta}$	χ^2
Rok 2005	10,253 203	0,083 084	-9 276,840	114 241
Rok 2004	10,086 031	0,104 881	-5 948,033	113 094
Rok 2005 – muži	10,458 184	0,060 376	-13 699,789	99 767
Rok 2005 – ženy	9,897 167	0,123 588	-3 361,532	23 713
Rok 2004 – muži	10,263 491	0,081 385	-8 650,678	191 785
Rok 2004 – ženy	9,751 210	0,146 556	-1 575,730	55 955

5. Závěr

Mnohé statistické analýzy konkrétních napozorovaných údajů naráží na smutnou všeobecně známou skutečnost, že data nejsou taková, jaká bychom pro analýzu potřebovali. Údaje jsou pro účely prezentace uspořádány do intervalů a tak také publikovány. V době počítačového zpracování je však škoda, že nejsou k dispozici rovněž relativně velké soubory dat v takové podobě, v jaké byly napozorovány. Odpadly by tím úvahy o tom, jak jsou mzdy v rámci intervalového rozdělení četností skutečně rozděleny a rovněž problémy s nevhodností krajních otevřených intervalů.

6. Literatura

[1] Aitchison, J. – Brown, J. A. C. 1957. The Lognormal Distribution with Special Reference to Its Uses in Economics. Cambridge: Cambridge University Press, 1957.

[2] Bílková, D. 1996. Příjmová rozdělení: Modelování v letech 1956 – 1992 a předpovědi pro roky 1995 a 1997 (doktorská disertační práce). Praha: VŠE, 1996.

[3] Hátle, J. – Hustopecký, J. – Novák, I. 1975. Modelování a krátkodobá předpověď příjmových rozdělení. Praha: VÚSEI/VŠE, 1975.

[4] Český statistický úřad: http//www.czso.cz/

Adresa autora:
Diana Bílková, Ing., Dr.
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4Ulica1
130 67 Praha 3
bilkova@vse.cz

Estimating the parameter θ of the Poisson distribution based on the divergence criteria

Martin Boďa[†]

Abstract: The focus of the article rests upon measures of discrepancy applicable in the estimation of the parameter θ of the Poisson distribution. The article summarizes the properties of the selected six measures of discrepancy and by way of simulation demonstrates the optimization approach and the iteration approach to constructing estimates for θ.

Keywords: Poisson distribution, chi-square, modified chi-square, Hellinger distance, Kullback-Leibler separator, Haldane's discrepancy, f-divergence, measure of discrepancy.

(To M. S.)

Introduction

It is indisputable that – from amongst methods of estimation – Ronald A. Fisher's maximum likelihood method and Karl Pearson's method of moments are accorded the greatest popularity. Information on these overriding approaches towards the parameter estimation is to be found in every reputable, if basic and general, book of statistics, such as in that of Rao (1973) or that of Anděl (1985). In addition to the said broad concepts, there is still a well-established method that lies in deriving parameter estimates from discrepancy criteria which measure the divergence of theoretical probabilities from empirical probabilities contained in the observations at hand. A brief account of this method is given in Rao (1973, pp. 352-3) where it is summarized under the term *minimum chi-square and associated methods.* In line with this ansatz, the article focuses upon some/possible ways of estimating the parameter θ of the Poisson distribution based on discrepancy criteria.

Motivation

Let us assume that $(\xi_1, \xi_2, ..., \xi_n)$ is a random sample from a Poissonian population, which implies that the underlying probability distribution is given by the formula

$$P_\theta\{\xi_\bullet = x\} = q_x = e^{-\theta}\frac{\theta^x}{x!} \quad \text{for} \quad x = 0, 1, \dots ,$$

where θ is an unknown parameter for which only positive values are admitted. The fact that the parameter is not known outlines the goal to construct an estimator founded on a suitable discrepancy criterion. In estimating θ from the observations the procedure delineated in Anděl (1985, p. 201) is followed. The sample $(\xi_1, \xi_2, ..., \xi_n)$ of n observations is sorted out in suitable classes in this fashion: [1.] the first class is represented by the variables which take values less or equal to r, [2.] the next $k-1$ classes are made up by individual variables the values of which are subsequently equal to $r+1, r+2, ..., r+k-2$, and [3.] the final class includes the variables with values greater or equal to $r+k-1$. The grouping into classes requires $r \geq 0$ and the number of classes $k \geq 3$. For each class the frequency of observations is determined, and denoted respectively $X_r, X_{r+1}, ..., X_{r+k-1}$ to be correspondent with the indexes of classes $r, r+1, ..., r+k-1$; so that it holds that $\sum X_\bullet = n$. Whilst allotting the sample $(\xi_1, \xi_2, ..., \xi_n)$ into k classes, it must be satisfied that each class is sufficiently saturated and each class must count at least 5 observations, that is $X_\bullet \geq 5$. For $P_\theta\{\xi_\bullet = i\}$ is denoted as q_i, the outcome of the grouping may be lucidly illustrated by Table 1.

[†] Ing. Martin Boďa. Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Tajovského 10, 975 90 Banská Bystrica. E-mail: martin.boda@umb.sk.

Class	r	$r+1$	…	$r+k-2$	$r+k-1$
Variables	$\xi_\bullet \le r$	$\xi_\bullet = r+1$	…	$\xi_\bullet = r+k-2$	$\xi_\bullet \ge r+k-1$
Frequency	X_r	X_{r+1}	…	X_{r+k-2}	X_{r+k-1}
Probability	$\sum_{i=0}^{i=r} q_i$	q_{r+1}	…	q_{r+k-2}	$\sum_{i=r+k-1}^{i=+\infty} q_i$

Table 1

For convenience, the notation

$$p_r \equiv \sum_{i=0}^{i=r} q_i\,, \qquad p_j = q_j \quad \text{for} \quad \forall j = r+1, \ldots, r+k-2\,, \qquad p_{r+k-1} \equiv \sum_{i=r+k-1}^{i=+\infty} q_i$$

is further employed, whilst it is borne in mind that probabilities $p_\bullet$ (as well as $q_\bullet$) remain a function of the unknown parameter θ to be estimated.

The empirical frequencies $X_\bullet$ are complemented by their theoretical counterparts $np_\bullet$, and it is but necessary to introduce a measure of discrepancy between the observed frequencies $X_\bullet$ and the theoretical ones $np_\bullet$, or, rather, between the empirical probabilities $X_\bullet/n$ and the theoretical probabilities $p_\bullet$, which is not an easy task as there are, in fact, infinitely many (reasonable) possibilities. Rao (1973, p. 352) states five measures of discrepancy (M. D.):

(a.) *Chi-square*

$$\chi^2 = \sum_{i=r}^{i=r+k-1} \frac{\left[X_i - np_i\right]^2}{np_i}\,,$$

(b.) *Modified chi-square*

$$\operatorname{mod}\chi^2 = \sum_{i=r}^{i=r+k-1} \frac{\left[X_i - np_i\right]^2}{X_i}\,,$$

(c.) *Hellinger distance*

$$H.D. = \arccos \sum_{i=r}^{i=r+k-1} \sqrt{\frac{X_i}{n} p_i}\,,$$

(d.) *Kullback-Leibler separator*

$$K.L.S. = \sum_{i=r}^{i=r+k-1} p_i \log \frac{p_i}{X_i/n}\,,$$

(e.) *Haldane's discrepancy*

$$D_k = \frac{(n+k)!}{n!} \sum_{i=r}^{i=r+k-1} \frac{X_i!}{(X_i+k)!} p_i^{k+1}\,.$$

The measures of discrepancy (a.) to (e.) constitute a utile vehicle to obtaining an estimate of θ produced by that value of θ° in the parametric space Θ (in this specific case $\Theta = (0, +\infty)$) in respect to which a measure of discrepancy is minimum, that is $\theta^\circ = \arg\min_{\theta\in\Theta} M.D.$

In Anděl (1985, p. 196-197) and in Mori and Koike (1978) a different approach to the modification of the chi-square is presented. The authors calculate the derivative of the chi-square measure (in the form of (a.) as defined above) and equate the derivative zero. At this point they observe that $(X_i - np_i)/np_i \xrightarrow[n]{P} 0$ for all classes $i = 1, 2, \ldots, k$ and obtain the system

$$\sum_{i=r}^{i=r+k-1} \frac{X_i}{p_i} \frac{dp_i}{d\theta}\bigg|_{\theta=\hat{\theta}} = 0$$

for the estimation of the unknown parameter θ, whereby they implicitly define the (modified) minimum chi-square statistic. It is straightforward that their modified chi-square explicitly reads

(f.) *Modified chi-square* *

$$\operatorname{mod}\chi_*^2 = -\sum_{i=r}^{i=r+k-1} \frac{X_i}{n} \log p_i\,.$$

Though diverse the measures of discrepancy (a.), (b.), (c.) and (d.) may seem, luckily, they may be clustered under a general family of discrepancy measures, which is presented e. g. in Vajda (1982) and in Basu and Sarkar (1994). The former terms them *f*-divergence measures, and the latter authors refer to them as disparity test statistics (since they treat them for the purpose of general goodness-of-fit testing). In this article the notion of *f*-divergence as established by Vajda (1982) is adopted. Inasmuch as *f*-divergence is originally defined as a divergence of two probability densities, it is suitable to customize it to fit the purpose of measuring the divergence of theoretical probabilities from empirical probabilities. In this spirit, f-divergence of the theoretical distribution $p_\bullet$ from the empirical values $X_\bullet/n$ is defined for every convex function $f\colon (0, +\infty) \to (-\infty, +\infty\rangle$, strictly convex in 1, by the formula

$$D_f =_{def} \sum_{\xi\in\,\Omega} p_\xi\, f\left(\frac{X_\xi/n}{p_\xi}\right) \equiv \sum_{i=r}^{i=r+k-1} p_i\, f\left(\frac{X_i/n}{p_i}\right) = \sum_{i=r}^{i=r+k-1} p_i(\theta)\, f\left(\frac{X_i/n}{p_i(\theta)}\right),$$

in which $f(1) = 0$. (All though Vajda's definition stipulates additional properties, it is not needful in this case as the aforesaid grouping has been accomplished.) Having the definition of f-divergence available, it only suffices to choose a suitable function form for f, and the discrepancy measures (a.), (b.), (c.) and (d.) readily obtained. The transformations are summarized in Table 2.

Measure of discrepancy	$f(\bullet)$	$D_f = \sum_{i=r}^{i=r+k-1} p_i\, f\left((X_i/n)/p_i\right)$
Chi-square	$f(u) = n\,(u-1)^2$	$\chi^2 = \sum_{i=r}^{i=r+k-1} \frac{\left[X_i - np_i\right]^2}{np_i}$
Modified chi-square	$f(u) = \frac{n}{u}\,(u-1)^2$	$\mathrm{mod}\,\chi^2 = \sum_{i=r}^{i=r+k-1} \frac{\left[X_i - np_i\right]^2}{X_i}$
Hellinger distance	$f(u) = 1-\sqrt{u}$	$H.D.^* = 1 - \sum_{i=r}^{i=r+k-1} \sqrt{\frac{X_i}{n}\,p_i}$
Kullback-Leibler separator	$f(u) = -\log(u)$	$K.L.S. = \sum_{i=r}^{i=r+k-1} p_i \log\frac{p_i}{X_i/n}$

Table 2

Upon studying Table 2 it strikes instantly that there is a slight, formal departure in case of Hellinger distance. However, the difference between Hellinger distance ($H.D.$) as defined in Rao (1973) and Hellinger distance ($H.D.^*$) in the generalizing framework of f-divergence is a mere formal issue for both $g(y) = \arccos(y)$ (in $H.D.$) and $g^*(y) = 1 - y$ (as for $H.D.^*$) are a decreasing function. When minimizing both the measures of discrepancy, the procedure yields the same result: the same estimate of θ.

Estimators for θ

Given a measure of discrepancy M. D. from amongst (a.) to (f.), the estimate θ° of θ is implicitly defined by the equation

$$\left.\frac{dM.D.}{d\theta}\right|_{\theta=\theta^\circ} = 0\,,$$

which leads to estimates presented in Table 3. In deriving estimates allowance is made for the facts that $dq_i/d\theta = q_i(i/\theta - 1)$, and consequently

$$\frac{dp_r}{d\theta} = \sum_{i=0}^{i=r} q_i\left(\frac{i}{\theta}-1\right), \qquad \frac{dp_j}{d\theta} = q_j\left(\frac{j}{\theta}-1\right) \quad \text{for} \quad \forall j = r+1, \ldots, r+k-2\,, \qquad \frac{dp_{r+k-1}}{d\theta} = \sum_{i=r+k-1}^{i=+\infty} q_i\left(\frac{i}{\theta}-1\right).$$

Measure of discrepancy	Estimator for θ
CHI-SQUARE	$\chi^2\,\theta^\circ = \dfrac{X_r^2 \dfrac{\sum_{i=0}^{i=r} iq_i}{\left(\sum_{i=0}^{i=r} q_i\right)^2} + \sum_{i=r+1}^{i=r+k-2} \dfrac{iX_i^2}{q_i} + X_{r+k-1}^2 \dfrac{\sum_{i=r+k-1}^{i=+\infty} iq_i}{\left(\sum_{i=r+k-1}^{i=+\infty} q_i\right)^2}}{\dfrac{X_r^2}{\sum_{i=0}^{i=r} q_i} + \sum_{i=r+1}^{i=r+k-2} \dfrac{X_i^2}{q_i} + \dfrac{X_{r+k-1}^2}{\sum_{i=r+k-1}^{i=+\infty} q_i}}$
MODIFIED CHI-SQUARE IN THE SENSE OF RAO	$\mathrm{mod}\,\chi^2\,\theta^\circ = \dfrac{\dfrac{\sum_{i=0}^{i=r} q_i \sum_{i=0}^{i=r} iq_i}{X_r} + \sum_{i=r+1}^{i=r+k-2} \dfrac{iq_i^2}{X_i} + \dfrac{\sum_{i=r+k-1}^{i=+\infty} q_i \sum_{i=r+k-1}^{i=+\infty} iq_i}{X_{r+k-1}}}{\dfrac{\left(\sum_{i=0}^{i=r} q_i\right)^2}{X_r} + \sum_{i=r+1}^{i=r+k-2} \dfrac{q_i^2}{X_i} + \dfrac{\left(\sum_{i=r+k-1}^{i=+\infty} q_i\right)^2}{X_{r+k-1}}}$
MODIFIED CHI-SQUARE IN THE SENSE OF ANDĚL	$\mathrm{mod}\,\chi^2\,\theta^\circ = \dfrac{1}{n}\left[X_r \dfrac{\sum_{i=0}^{i=r} iq_i}{\sum_{i=0}^{i=r} q_i} + \sum_{i=r+1}^{i=r+k-2} iX_i + X_{r+k-1} \dfrac{\sum_{i=r+k-1}^{i=+\infty} iq_i}{\sum_{i=r+k-1}^{i=+\infty} q_i} \right]$
HELLINGER DISTANCE	$H.D.\,\theta^\circ = \dfrac{\sqrt{\dfrac{X_r}{n}} \dfrac{\sum_{i=0}^{i=r} iq_i}{\sqrt{\sum_{i=0}^{i=r} q_i}} + \sum_{i=r+1}^{i=r+k-2} i\sqrt{\dfrac{X_i}{n}} q_i + \sqrt{\dfrac{X_{r+k-1}}{n}} \dfrac{\sum_{i=r+k-1}^{i=+\infty} iq_i}{\sqrt{\sum_{i=r+k-1}^{i=+\infty} q_i}}}{\sqrt{\dfrac{X_r}{n} \sum_{i=0}^{i=r} q_i} + \sum_{i=r+1}^{i=r+k-2} \sqrt{\dfrac{X_i}{n}} q_i + \sqrt{\dfrac{X_{r+k-1}}{n} \sum_{i=r+k-1}^{i=+\infty} q_i}}$
KULLBACK-LEIBLER SEPARATOR	$K.L.S.\,\theta^\circ = \dfrac{\left(\sum_{i=0}^{i=r} iq_i\right)\left(\log \dfrac{\sum_{i=0}^{i=r} q_i}{\frac{X_r}{n}} + 1\right) + \sum_{i=r+1}^{i=r+k-2} iq_i \left(\log \dfrac{q_i}{\frac{X_i}{n}} + 1\right) + \left(\sum_{i=0}^{i=+\infty} iq_i\right)\left(\log \dfrac{\sum_{i=r+k-1}^{i=+\infty} q_i}{\frac{X_{r+k-1}}{n}} + 1\right)}{\left(\sum_{i=0}^{i=r} q_i\right)\left(\log \dfrac{\sum_{i=0}^{i=r} q_i}{\frac{X_r}{n}} + 1\right) + \sum_{i=r+1}^{i=r+k-2} q_i \left(\log \dfrac{q_i}{\frac{X_i}{n}} + 1\right) + \left(\sum_{i=0}^{i=+\infty} q_i\right)\left(\log \dfrac{\sum_{i=r+k-1}^{i=+\infty} q_i}{\frac{X_{r+k-1}}{n}} + 1\right)}$
HALDANE'S DISCREPANCY	$D_k\,\theta^\circ = \dfrac{\dfrac{\left(\sum_{i=0}^{i=r} iq_i\right)\left(\sum_{i=0}^{i=r} q_i\right)^k}{\binom{X_r + k}{k}} + \sum_{i=r+1}^{i=r+k-2} \dfrac{i(q_i)^{k+1}}{\binom{X_i + k}{k}} + \dfrac{\left(\sum_{i=0}^{i=+\infty} iq_i\right)\left(\sum_{i=r+k-1}^{i=+\infty} q_i\right)^k}{\binom{X_{r+k-1} + k}{k}}}{\dfrac{\left(\sum_{i=0}^{i=r} q_i\right)^{k+1}}{\binom{X_r + k}{k}} + \sum_{i=r+1}^{i=r+k-2} \dfrac{(q_i)^{k+1}}{\binom{X_i + k}{k}} + \dfrac{\left(\sum_{i=r+k-1}^{i=+\infty} q_i\right)^{k+1}}{\binom{X_{r+k-1} + k}{k}}}$

Table 3

The equation for the estimators θ° is solved iteratively. The zeroth estimate

$$\theta_0^\circ = \frac{1}{n} \sum_{i=r}^{i=r+k-1} iX_i \ ,$$

is a medium to computing the probabilities q_i and obtaining the first estimate θ_1° and the procedure is repeated till the estimate $\theta_\bullet^\circ$ has converged satisfactorily.

Besides the procedure described, it is, of course, possible to minimize the criteria (a.) to (f.) numerically as a task of optimization.

Statistical properties of the measures of discrepancy

It is noted in Rao (1973) that the measures (a.) to (e.) provide reasonably good estimators and that, under certain regularity conditions, they are consistent and first-order efficient (after the

definition of Rao). Following the regularity conditions of Birch given in Basu and Sarkar (1994), it may be demonstrated that the measures (a.) to (d.) are best asymptotically normal (B. A. N.). If θ^* is the estimated parameter θ , the regularity conditions in the Poissonian case under investigation are trivial and read

(1.) The neighbourhood of θ^* is completely contained in $(0, +\infty)$.
(2.) $p_j > 0$, for $\forall\, j = r, ..., r + k - 1$.
(3.) p_j is differentiable at θ^*.
(4.) The Jacobian $(\partial p_j(\theta)/\partial\theta)$ is of full rank 1.
(5.) The inverse mapping of p_j^{-1} is continuous at $p_j(\theta)$.
(6.) The mapping p_j is continuous at every point $\theta \in (0, +\infty)$.

The conditions warrant that nonuniqueness and unboundedness of the measures (a.) to (d.) occurs with probability zero as n goes to infinity, and that these estimators are B. A. N. It is needless to say that these conditions are met.

Comparison of the measures by a simulation

So as to evaluate the estimation capacity of the measures under consideration, 50 samples of 250 observations of the Poisson distribution with the parameter $\theta^* = 10$ were generated by use of MAPLE. The subsequent procedures were effected in Microsoft ® Excel. The samples were sorted out in $k = 12$ classes with values from $r = 5$ to $r + k - 1 = 16$, and the competing estimates were constructed on the basis of the sorted-out samples. The results attached to the article contain: (1.) the average computed out of the original sample (referred to as original average), (2.) the average computed out of the sorted-out sample (referred to as sorted-out average) utilized as the zeroth step in the iteration process, the optimized estimate for each criterion, and the 16 initial iterations. The results are differentiated after their proximity to the true value of θ and after their capability to stand the iteration process. It transpires clear – without further evidence of words and numbers – that for the estimation it is suitable to employ either the straightforward optimization (minimization) of the selected criterion, or the iteration procedure with Hellinger distance, Kullback-Leibler separator, and the modified chi-square in the sense of Anděl. Obviously, the other criterions do not provide convergent results.

Conclusion

The presented measures of discrepancy possess under regularity conditions exquisite statistical properties for their best asymptotical normality. Even so, it is not possible to employ the iteration approach in the estimation of θ arbitrarily since it has been demonstrated that – save Hellinger distance, Kullback-Leibler separator, and the modified chi-square in the sense of Anděl – the iterated estimates do not necessarily converge.

References

ANDĚL, J. 1985. *Matematická statistika.* 2nd ed. Praha: SNTL/Alfa 1985. No ISBN.

BASU, A. – SARKAR, S. 1994. On disparity based goodness-of-fit tests for multinomial models. In: *Statistics & Probability Letters.* Vol. 19.No. 4. 1994. Pp. 307-312.

MORI, K. – KOIKE, T. 1978. *Parameters Estimation Based on Minimum Chi-square.* In: *Japan Industrial Management Association Series*. No 4. No. 4. 1978. Pp. 400-405.

RAO, C. R. 1973. *Linear Statistical Inference and Its Applications.* 2nd ed. New York. Wiley 1973. ISBN 0-471-2187-8.

VAJDA, I. 1982. *Teória informácie a štatistického rozhodovania.* Bratislava: Alfa 1982. No ISBN.

SAMPLE	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25
the true θ	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
original average	10.064	10.46	10.24	9.748	10.056	10.352	10.008	10.032	9.868	10.148	10.012	10.244	10.064	9.864	9.928	10.224	9.696	9.992	9.936	10.34	10.176	10.084	10.308	10.032	10.096
sorted-out average	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

CHI-SQUARE CRITERION

the optimalized θ	10.107	10.415	10.255	9.5337	10.048	10.32	9.9843	9.9313	9.7603	10.201	9.9162	10.13	9.8977	9.9342	9.8854	10.145	9.6443	9.9924	9.9085	10.343	10.426	10.051	10.371	10.028	10.17
the iteration #1	10.091	0.0909	10.264	9.8682	10.018	10.354	10.089	10.097	9.9968	10.129	10.053	10.37	10.16	9.7738	9.9493	10.264	9.727	10.065	9.9421	10.365	9.9585	10.104	10.254	10.078	10.038
the iteration #2	10.063	0.0633	10.227	9.7413	10.036	10.306	10.024	10.045	9.833	10.135	9.9973	10.208	10.062	9.8198	9.9348	10.18	9.691	9.9865	9.9146	10.269	10.266	10.06	10.301	10.037	10.095
the iteration #3	10.092	0.0917	10.265	9.8595	10.018	10.349	10.081	10.081	10.036	10.129	10.049	10.383	10.159	9.7857	9.9506	10.245	9.7243	10.075	9.9437	10.372	9.8248	10.104	10.242	10.081	10.04
the iteration #4	10.062	0.0625	10.226	9.7494	10.036	10.311	10.03	10.056	9.786	10.135	10.002	10.194	10.063	9.8109	9.9335	10.194	9.6935	9.9743	9.9129	10.261	10.461	10.059	10.315	10.034	10.093
the iteration #5	10.093	0.0925	10.266	9.8518	10.018	10.345	10.076	10.073	10.095	10.13	10.044	10.398	10.157	9.7922	9.952	10.234	9.7219	10.089	9.9454	10.381	9.5525	10.105	10.225	10.084	10.041
the iteration #6	10.062	0.0616	10.224	9.7564	10.036	10.314	10.035	10.061	9.7147	10.135	10.005	10.178	10.065	9.8061	9.9319	10.203	9.6957	9.9584	9.9111	10.252	10.862	10.059	10.336	10.03	10.092
the iteration #7	10.093	0.0934	10.267	9.8452	10.018	10.342	10.071	10.069	10.187	10.13	10.041	10.416	10.155	9.7958	9.9538	10.227	9.7199	10.108	9.9473	10.391	9.0185	10.105	10.199	10.088	10.042
the iteration #8	10.061	0.0607	10.223	9.7624	10.036	10.317	10.039	10.064	9.6075	10.134	10.008	10.159	10.067	9.8034	9.93	10.208	9.6976	9.9379	9.9092	10.241	11.643	10.058	10.369	10.026	10.091
the iteration #9	10.094	0.0943	10.269	9.8395	10.018	10.339	10.068	10.067	10.327	10.131	10.038	10.436	10.154	9.7978	9.9559	10.223	9.7182	10.132	9.9494	10.402	8.098	10.106	10.159	10.093	10.044
the iteration #10	10.06	0.0598	10.222	9.7677	10.035	10.32	10.043	10.065	9.448	10.134	10.011	10.137	10.068	9.802	9.9277	10.211	9.6991	9.9114	9.9069	10.229	12.893	10.058	10.419	10.021	10.09
the iteration #11	10.095	0.0953	10.27	9.8346	10.018	10.337	10.065	10.067	10.541	10.131	10.036	10.46	10.152	9.7989	9.9583	10.22	9.7167	10.162	9.9518	10.416	6.9958	10.106	10.097	10.098	10.045
the iteration #12	10.059	0.0588	10.22	9.7723	10.035	10.322	10.045	10.065	9.2157	10.134	10.013	10.112	10.069	9.8012	9.9251	10.213	9.7005	9.8771	9.9044	10.215	14.09	10.058	10.496	10.015	10.088
the iteration #13	10.096	0.0963	10.272	9.8303	10.018	10.335	10.063	10.066	10.862	10.131	10.034	10.489	10.151	9.7995	9.9612	10.219	9.7155	10.203	9.9544	10.432	6.3069	10.107	10.004	10.105	10.046
the iteration #14	10.058	0.0578	10.219	9.7763	10.035	10.323	10.047	10.066	8.8913	10.134	10.015	10.082	10.071	9.8007	9.9219	10.214	9.7016	9.833	9.9017	10.198	14.651	10.057	10.614	10.008	10.087
the iteration #15	10.097	0.0973	10.273	9.8265	10.018	10.334	10.061	10.066	11.328	10.131	10.032	10.521	10.15	9.7998	9.9646	10.218	9.7144	10.255	9.9573	10.45	6.076	10.107	9.8626	10.112	10.047
the iteration #16	10.057	0.0567	10.217	9.7797	10.035	10.324	10.049	10.066	8.4729	10.133	10.016	10.048	10.072	9.8005	9.9182	10.215	9.7026	9.7762	9.8986	10.179	14.809	10.057	10.793	10.001	10.086

MODIFIED CHI-SQUARE CRITERION IN THE SENSE OF RAO

the optimalized θ	10.107	10.415	10.255	9.5337	10.048	10.32	9.9843	9.9313	9.7603	10.201	9.9162	10.13	9.8977	9.9342	9.8854	10.145	9.6443	9.9924	9.9085	10.343	10.426	10.051	10.371	10.028	10.17
the iteration #1	10.025	10.427	10.204	9.9433	10.025	10.279	10.05	10.14	9.9457	10.07	10.074	10.298	10.242	9.7091	9.9815	10.169	9.7358	9.999	9.9227	10.216	9.98	10.068	10.229	10.05	10.021
the iteration #2	9.9526	10.442	10.159	10.376	10.002	10.236	10.119	10.424	10.092	9.9385	10.245	10.445	10.643	9.4209	10.069	10.197	9.8361	10.005	9.9352	10.103	9.6592	10.084	10.116	10.069	9.87
the iteration #3	9.8156	10.477	10.073	11.241	9.9576	10.148	10.261	11.109	10.352	9.6774	10.609	10.721	11.563	8.7841	10.238	10.259	10.047	10.015	9.959	9.8896	9.1569	10.114	9.9171	10.105	9.5755
the iteration #4	9.559	10.559	9.9123	12.817	9.871	9.9694	10.557	12.719	10.815	9.1708	11.391	11.247	13.715	7.5061	10.575	10.395	10.491	10.034	10.004	9.4873	8.4496	10.173	9.5773	10.171	9.0253
the iteration #5	9.083	10.748	9.6174	15.19	9.7038	9.6153	11.176	15.579	11.633	8.2408	13.018	12.241	17.09	5.6133	11.269	10.691	11.413	10.069	10.089	8.734	7.5834	10.289	9.0218	10.294	8.0964
the iteration #6	8.2175	11.189	9.0935	17.726	9.3888	8.9438	12.471	18.197	13.055	6.7729	15.62	13.997	19.37	4.0796	12.764	11.333	13.167	10.133	10.25	7.3697	6.6298	10.516	8.179	10.525	6.8004
the iteration #7	6.7499	12.22	8.2178	19.591	8.8218	7.7923	14.955	19.92	15.324	5.1365	18.061	16.389	20.775	3.2936	15.597	12.649	15.681	10.252	10.56	5.3835	5.608	10.973	7.0322	10.963	5.4049
the iteration #8	4.927	14.392	6.9023	20.87	7.888	6.1784	17.859	21.099	17.871	3.9607	19.77	18.527	21.755	2.8614	18.376	14.889	18.03	10.472	11.161	3.9141	4.5741	11.905	5.6917	11.804	4.2472
the iteration #9	3.7356	17.253	5.3111	21.8	6.5667	4.6113	19.776	21.97	19.714	3.2798	20.966	20.062	22.497	2.5858	20.109	17.424	19.726	10.877	12.342	3.2041	3.7259	13.774	4.4515	13.385	3.4809
the iteration #10	3.1214	19.33	4.0328	22.519	5.1077	3.6256	21.032	22.651	20.967	2.8708	21.856	21.167	23.089	2.3915	21.267	19.347	20.926	11.622	14.532	2.8065	3.1637	16.613	3.5936	15.89	3.0074
the iteration #11	2.7591	20.7	3.3007	23.099	3.9815	3.0747	21.938	23.205	21.878	2.6	22.551	22.004	23.579	2.2453	22.116	20.676	21.821	12.953	17.342	2.5482	2.8015	18.935	3.072	18.332	2.6977
the iteration #12	2.5172	21.677	2.8757	23.582	3.3015	2.7347	22.636	23.67	22.584	2.4061	23.118	22.667	23.995	2.13	22.779	21.643	22.522	15.09	19.4	2.3638	2.5533	20.443	2.7388	20.007	2.4796
the iteration #13	2.3417	22.423	2.5996	23.994	2.8879	2.5026	23.2	24.069	23.154	2.259	23.592	23.211	24.357	2.0362	23.318	22.388	23.092	17.564	20.756	2.2238	2.3719	21.494	2.5081	21.165	2.3168
the iteration #14	2.2072	23.021	2.4038	24.353	2.6134	2.3323	23.671	24.417	23.63	2.1426	23.999	23.67	24.675	1.9577	23.771	22.986	23.57	19.458	21.723	2.1128	2.2326	22.284	2.3378	22.022	2.1895
the iteration #15	2.0997	23.517	2.256	24.67	2.4168	2.2006	24.074	24.725	24.036	2.0476	24.354	24.065	24.959	1.8908	24.161	23.484	23.979	20.765	22.462	2.0219	2.1214	22.91	2.2058	22.695	2.0867
the iteration #16	2.0113	23.938	2.1394	24.953	2.2678	2.0949	24.424	25.002	24.39	1.9681	24.668	24.411	25.215	1.8328	24.501	23.907	24.336	21.715	23.055	1.9457	2.0301	23.425	2.0997	23.243	2.0013

HELLINGER DISTANCE CRITERION

the optimalized θ	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9085	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #1	10.078	10.432	10.238	9.7104	10.035	10.312	10.016	10.011	9.8574	10.152	9.9895	10.227	10.031	9.8551	9.9228	10.169	9.6861	10.003	9.9176	10.301	10.198	10.063	10.308	10.044	10.105
the iteration #2	10.085	10.437	10.243	9.6996	10.034	10.318	10.016	10.007	9.8539	10.159	9.9883	10.23	10.017	9.8642	9.916	10.176	9.6851	10.007	9.9185	10.315	10.213	10.065	10.316	10.046	10.11
the iteration #3	10.089	10.44	10.245	9.6942	10.034	10.32	10.016	10.005	9.8519	10.163	9.9876	10.232	10.01	9.8684	9.9124	10.179	9.6846	10.009	9.9189	10.321	10.221	10.065	10.321	10.047	10.113
the iteration #4	10.091	10.441	10.247	9.6915	10.033	10.322	10.016	10.004	9.8509	10.165	9.9873	10.233	10.007	9.8704	9.9105	10.18	9.6844	10.01	9.9192	10.325	10.225	10.066	10.323	10.048	10.114
the iteration #5	10.092	10.441	10.247	9.6901	10.033	10.323	10.016	10.003	9.8504	10.166	9.9872	10.233	10.005	9.8713	9.9096	10.181	9.6843	10.01	9.9193	10.327	10.228	10.066	10.324	10.048	10.114
the iteration #6	10.092	10.442	10.248	9.6894	10.033	10.323	10.016	10.003	9.8501	10.167	9.9871	10.234	10.004	9.8718	9.9091	10.181	9.6842	10.011	9.9194	10.328	10.229	10.066	10.325	10.048	10.115
the iteration #7	10.093	10.442	10.248	9.6891	10.033	10.323	10.016	10.003	9.8499	10.167	9.9871	10.234	10.004	9.872	9.9088	10.182	9.6842	10.011	9.9194	10.328	10.23	10.066	10.325	10.048	10.115
the iteration #8	10.093	10.442	10.248	9.6889	10.033	10.323	10.016	10.003	9.8498	10.167	9.987	10.234	10.003	9.8721	9.9087	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.048	10.115
the iteration #9	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8498	10.167	9.987	10.234	10.003	9.8721	9.9086	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.048	10.115
the iteration #10	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8721	9.9086	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #11	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9086	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #12	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9086	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #13	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9086	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #14	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9085	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #15	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9085	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115
the iteration #16	10.093	10.442	10.248	9.6888	10.033	10.323	10.016	10.003	9.8497	10.167	9.987	10.234	10.003	9.8722	9.9085	10.182	9.6842	10.011	9.9194	10.329	10.231	10.066	10.326	10.049	10.115

KULLBACK-LEIBLER SEPARATOR CRITERION

the optimalized θ	10.098	10.431	10.25	9.6408	10.037	10.322	10.005	9.9799	9.8213	10.179	9.9678	10.205	9.966	9.8947	9.8993	10.17	9.6726	10.005	9.916	10.333	10.298	10.061	10.342	10.043	10.133
the iteration #1	10.067	10.421	10.23	9.7228	10.036	10.302	10.015	10.015	9.8603	10.14	9.9895	10.219	10.05	9.8427	9.9326	10.157	9.6864	9.9968	9.916	10.281	10.182	10.06	10.296	10.04	10.1
the iteration #2	10.07	10.422	10.232	9.7145	10.036	10.304	10.014	10.011	9.8569	10.144	9.9873	10.217	10.042	9.8486	9.9294	10.159	9.685	9.9975	9.916	10.286	10.192	10.06	10.3	10.041	10.103
the iteration #3	10.073	10.423	10.234	9.707	10.036	10.306	10.013	10.008	9.8538	10.148	9.9853	10.216	10.034	9.8538	9.9266	10.16	9.6837	9.9982	9.916	10.291	10.201	10.06	10.304	10.041	10.106
the iteration #4	10.075	10.424	10.235	9.7003	10.036	10.308	10.012	10.004	9.851	10.151	9.9835	10.215	10.027	9.8584	9.9241	10.161	9.6826	9.9988	9.916	10.295	10.209	10.06	10.307	10.041	10.109
the iteration #5	10.077	10.425	10.236	9.6943	10.037	10.309	10.011	10.002	9.8484	10.154	9.9819	10.214	10.021	9.8625	9.9217	10.162	9.6816	9.9993	9.916	10.298	10.216	10.06	10.31	10.041	10.111
the iteration #6	10.079	10.426	10.238	9.6889	10.037	10.31	10.011	9.9991	9.8461	10.156	9.9805	10.213	10.015	9.8662	9.9196	10.163	9.6807	9.9998	9.916	10.301	10.223	10.061	10.313	10.041	10.113
the iteration #7	10.081	10.426	10.239	9.684	10.037	10.312	10.01	9.9969	9.8439	10.158	9.9792	10.213	10.01	9.8694	9.9177	10.164	9.6798	10	9.916	10.304	10.229	10.061	10.316	10.042	10.115
the iteration #8	10.083	10.427	10.24	9.6797	10.037	10.313	10.009	9.9949	9.842	10.16	9.978	10.212	10.005	9.8723	9.916	10.164	9.6791	10.001	9.916	10.307	10.235	10.061	10.318	10.042	10.117
the iteration #9	10.084	10.427	10.241	9.6757	10.037	10.314	10.009	9.9932	9.8402	10.162	9.977	10.211	10.001	9.8748	9.9144	10.165	9.6784	10.001	9.916	10.31	10.24	10.061	10.32	10.042	10.119
the iteration #10	10.085	10.428	10.242	9.6722	10.037	10.314	10.008	9.9917	9.8385	10.164	9.976	10.211	9.9977	9.8771	9.913	10.165	9.6778	10.001	9.916	10.312	10.245	10.061	10.322	10.042	10.12
the iteration #11	10.087	10.428	10.242	9.669	10.037	10.315	10.008	9.9903	9.8371	10.165	9.9752	10.21	9.9945	9.8791	9.9117	10.166	9.6773	10.002	9.916	10.314	10.249	10.061	10.324	10.042	10.121
the iteration #12	10.088	10.429	10.243	9.6662	10.037	10.316	10.008	9.9891	9.8357	10.167	9.9744	10.21	9.9916	9.8808	9.9106	10.166	9.6768	10.002	9.916	10.316	10.253	10.061	10.325	10.042	10.123
the iteration #13	10.089	10.429	10.244	9.6636	10.037	10.316	10.007	9.988	9.8344	10.168	9.9737	10.209	9.9889	9.8824	9.9095	10.167	9.6764	10.002	9.916	10.317	10.257	10.061	10.327	10.042	10.124
the iteration #14	10.089	10.429	10.244	9.6613	10.037	10.317	10.007	9.9871	9.8333	10.169	9.9731	10.209	9.9866	9.8838	9.9086	10.167	9.676	10.003	9.916	10.319	10.26	10.061	10.328	10.042	10.125
the iteration #15	10.09	10.429	10.245	9.6593	10.037	10.318	10.007	9.9863	9.8323	10.17	9.9726	10.208	9.9845	9.885	9.9077	10.167	9.6756	10.003	9.916	10.32	10.263	10.061	10.329	10.042	10.125
the iteration #16	10.091	10.43	10.245	9.6574	10.037	10.318	10.007	9.9855	9.8313	10.171	9.9721	10.208	9.9826	9.8861	9.9069	10.168	9.6753	10.003	9.916	10.321	10.266	10.061	10.331	10.043	10.126

SAMPLE	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25
the true θ	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
original average	10.064	10.46	10.24	9.748	10.056	10.352	10.008	10.032	9.868	10.148	10.012	10.244	10.064	9.864	9.928	10.224	9.696	9.992	9.936	10.34	10.176	10.084	10.308	10.032	10.096
sorted-out average	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

HALDANE'S DISTANCE CRITERION

the optimalized θ	10.012	10.483	10.39	9.4498	10.32	10.388	9.7253	10.139	9.6405	10.118	10.218	10.092	9.9419	10.247	10.156	10.138	9.9451	9.8681	10.11	10.318	10.678	10.246	10.503	9.8083	10.401
the iteration #1	10.091	9.7813	9.9213	11.893	9.0751	9.8084	10.83	8.8335	11.085	10.237	8.8747	10.636	10.891	6.9157	9.6057	10.389	8.2682	10.451	9.4965	10.177	7.5773	9.6304	9.7088	11.159	7.9208
the iteration #2	10.126	6.6701	8.8804	16.74	6.8724	7.8103	14.53	4.5861	13.682	10.821	5.3073	12.077	17.211	4.0132	8.7453	12.66	4.2201	11.545	8.4254	9.7797	6.7229	8.4625	7.9255	13.922	5.1301
the iteration #3	10.171	3.9708	5.8455	19.322	4.0281	4.1836	18.378	3.4377	18.081	13.743	3.6658	16.923	19.558	3.2096	6.6432	17.742	3.2975	16.483	4.2988	6.461	4.2401	4.3107	4.7664	18.162	3.615
the iteration #4	10.225	3.1907	3.8026	20.777	3.2161	3.2825	20.199	2.9271	20.028	18.101	3.0462	19.412	20.93	2.7979	3.9667	19.84	2.849	19.199	3.3292	3.9322	3.3057	3.334	3.5003	20.074	3.0205
the iteration #5	10.287	2.7868	3.113	21.77	2.8018	2.8404	21.361	2.6231	21.244	20.04	2.6989	20.835	21.882	2.5374	3.1889	21.116	2.5717	20.699	2.867	3.1733	2.8536	2.8697	2.9608	21.275	2.6828
the iteration #6	10.364	2.5298	2.7401	22.516	2.54	2.5659	22.203	2.4156	22.114	21.252	2.4692	21.812	22.603	2.3533	2.7857	22.019	2.3784	21.714	2.5837	2.7764	2.5748	2.5854	2.6448	22.138	2.4579
the iteration #7	10.481	2.3478	2.4978	23.109	2.3553	2.3742	22.857	2.2621	22.786	22.12	2.3026	22.549	23.18	2.2142	2.5291	22.711	2.2336	22.472	2.3871	2.5228	2.3807	2.3884	2.4311	22.805	2.2942
the iteration #8	10.756	2.2099	2.324	23.599	2.2157	2.2304	23.388	2.1425	23.33	22.791	2.1746	23.136	23.658	2.1041	2.3472	23.268	2.1197	23.073	2.2403	2.3426	2.2353	2.2413	2.2739	23.346	2.1679
the iteration #9	12.675	2.1006	2.1914	24.015	2.1053	2.1171	23.835	2.0457	23.786	23.334	2.072	23.621	24.066	2.0141	2.2094	23.733	2.027	23.569	2.1251	2.2058	2.1211	2.1259	2.1518	23.799	2.0665
the iteration #10	17.769	2.0112	2.0856	24.375	2.0151	2.0248	24.219	1.9653	24.176	23.789	1.9873	24.034	24.42	1.9385	2.1003	24.13	1.9495	23.989	2.0314	2.0973	2.0281	2.0321	2.0534	24.187	1.9828
the iteration #11	19.855	1.9361	1.9987	24.693	1.9394	1.9477	24.554	1.897	24.516	24.179	1.9158	24.392	24.732	1.874	2.0109	24.476	1.8834	24.353	1.9532	2.0084	1.9504	1.9538	1.9717	24.526	1.9119
the iteration #12	21.127	1.8718	1.9255	24.976	1.8747	1.8818	24.852	1.8379	24.818	24.519	1.8543	24.707	25.011	1.8179	1.9358	24.783	1.8261	24.673	1.8866	1.9338	1.8842	1.8871	1.9025	24.827	1.8509
the iteration #13	22.027	1.816	1.8627	25.231	1.8185	1.8247	25.119	1.7862	25.089	24.82	1.8006	24.989	25.263	1.7685	1.8716	25.057	1.7758	24.958	1.8289	1.8699	1.8268	1.8293	1.8427	25.097	1.7977
the iteration #14	22.717	1.7669	1.808	25.463	1.7691	1.7746	25.361	1.7404	25.333	25.091	1.7532	25.243	25.492	1.7246	1.8158	25.304	1.7311	25.215	1.7783	1.8143	1.7764	1.7786	1.7905	25.341	1.7506
the iteration #15	23.273	1.7231	1.7598	25.676	1.7251	1.73	25.582	1.6994	25.557	25.335	1.7109	25.474	25.703	1.6852	1.7667	25.53	1.691	25.448	1.7333	1.7653	1.7317	1.7336	1.7442	25.563	1.7085
the iteration #16	23.737	1.6838	1.7168	25.872	1.6856	1.6901	25.785	1.6624	25.762	25.558	1.6728	25.686	25.897	1.6495	1.723	25.737	1.6548	25.662	1.693	1.7217	1.6915	1.6933	1.7028	25.768	1.6707

MODIFIED CHI-SQUARE CRITERION IN THE SENSE OF ANDĚL

the optimalized θ	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #1	10.088	10.454	10.246	9.7316	10.03	10.325	10.028	10.025	9.876	10.156	10.002	10.257	10.041	9.8483	9.9193	10.193	9.6938	10.017	9.9226	10.324	10.171	10.071	10.309	10.052	10.098
the iteration #2	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #3	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #4	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #5	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #6	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #7	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #8	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #9	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #10	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #11	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #12	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #13	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #14	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #15	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098
the iteration #16	10.088	10.454	10.247	9.7315	10.03	10.325	10.028	10.025	9.8762	10.156	10.002	10.258	10.041	9.8485	9.919	10.194	9.6939	10.017	9.9227	10.325	10.171	10.071	10.309	10.053	10.098

SAMPLE	#26	#27	#28	#29	#30	#31	#32	#33	#34	#35	#36	#37	#38	#39	#40	#41	#42	#43	#44	#45	#46	#47	#48	#49	#50
the true θ	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
original average	9.892	10.136	10.188	10.084	10.356	10.212	9.784	10.252	9.872	10.372	10.04	9.784	9.768	10.284	9.868	9.824	9.8	10.06	9.956	9.68	9.94	9.636	9.896	9.772	9.488
sorted-out average	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

CHI-SQUARE CRITERION

the optimalized θ	9.8573	10.138	10.242	10.021	10.288	10.25	9.7111	10.186	9.8068	10.396	9.945	9.8076	9.7858	10.174	9.8837	9.9342	9.739	10.103	9.8716	9.5907	10.044	9.6365	9.9159	9.7051	9.3926
the iteration #1	9.8828	10.166	10.202	10.128	10.396	10.208	9.7914	10.426	9.9143	10.396	10.074	9.7422	9.8426	10.377	9.9042	9.8235	9.7824	10.093	9.9813	9.6546	9.9732	9.5551	9.8485	9.7888	9.4828
the iteration #2	9.8753	10.078	10.183	10.049	10.303	10.189	9.7781	10.302	9.8709	10.318	10.028	9.7733	9.8403	10.164	9.8854	9.8714	9.7886	10.079	9.938	9.743	9.9527	9.622	9.8812	9.7665	9.5505
the iteration #3	9.8835	10.183	10.202	10.132	10.38	10.208	9.7896	10.396	9.9084	10.39	10.078	9.7477	9.8423	10.405	9.9029	9.8277	9.7817	10.095	9.9794	9.6366	9.9725	9.557	9.8424	9.7838	9.4704
the iteration #4	9.8745	10.058	10.183	10.045	10.316	10.189	9.7796	10.325	9.876	10.323	10.023	9.7687	9.8406	10.133	9.8865	9.8676	9.7894	10.077	9.9398	9.7646	9.9534	9.6201	9.8886	9.7704	9.5652
the iteration #5	9.8844	10.207	10.203	10.135	10.369	10.207	9.7882	10.379	9.9039	10.385	10.084	9.7515	9.8421	10.44	9.9018	9.8312	9.7808	10.096	9.9776	9.6107	9.9719	9.5589	9.8337	9.7808	9.4531
the iteration #6	9.8735	10.03	10.182	10.042	10.325	10.19	9.7808	10.338	9.8798	10.327	10.018	9.7657	9.8408	10.094	9.8876	9.8644	9.7903	10.075	9.9414	9.796	9.9541	9.6183	9.899	9.7728	9.5859
the iteration #7	9.8855	10.241	10.204	10.139	10.361	10.207	9.7872	10.369	9.9006	10.381	10.09	9.754	9.8419	10.486	9.9009	9.8341	9.7798	10.098	9.976	9.5732	9.9712	9.5606	9.8213	9.7789	9.4288
the iteration #8	9.8724	9.9895	10.182	10.038	10.332	10.19	9.7817	10.345	9.8826	10.331	10.011	9.7636	9.8409	10.044	9.8885	9.8617	9.7914	10.073	9.943	9.8414	9.9547	9.6166	9.9137	9.7742	9.6148
the iteration #9	9.8868	10.29	10.204	10.143	10.356	10.206	9.7865	10.364	9.8982	10.378	10.097	9.7557	9.8418	10.545	9.9	9.8365	9.7785	10.101	9.9746	9.5192	9.9707	9.5622	9.8039	9.7779	9.395
the iteration #10	9.8709	9.9329	10.181	10.033	10.336	10.191	9.7823	10.349	9.8847	10.334	10.003	9.7622	9.841	9.9808	9.8892	9.8595	9.7929	10.07	9.9443	9.9072	9.9552	9.615	9.9346	9.775	9.6553
the iteration #11	9.8883	10.36	10.205	10.148	10.353	10.206	9.7859	10.36	9.8963	10.375	10.106	9.7568	9.8417	10.621	9.8993	9.8385	9.777	10.104	9.9733	9.4413	9.9701	9.5637	9.7794	9.7772	9.3477
the iteration #12	9.8692	9.8535	10.18	10.028	10.339	10.191	9.7828	10.352	9.8863	10.337	9.9927	9.7612	9.8411	9.9017	9.8899	9.8577	9.7946	10.067	9.9456	10.002	9.9557	9.6135	9.9642	9.7755	9.712
the iteration #13	9.8902	10.459	10.206	10.153	10.35	10.205	9.7855	10.358	9.895	10.372	10.117	9.7576	9.8416	10.717	9.8987	9.8402	9.7751	10.108	9.972	9.3299	9.9696	9.5652	9.7447	9.7768	9.2819
the iteration #14	9.8671	9.7426	10.179	10.023	10.341	10.191	9.7832	10.354	9.8875	10.339	9.9809	9.7606	9.8412	9.8027	9.8905	9.8562	9.7967	10.063	9.9468	10.139	9.9562	9.6121	10.006	9.7758	9.7914
the iteration #15	9.8925	10.599	10.207	10.159	10.349	10.205	9.7852	10.357	9.894	10.37	10.13	9.7581	9.8416	10.839	9.8981	9.8416	9.7727	10.112	9.9709	9.1719	9.9691	9.5666	9.6959	9.7766	9.1906
the iteration #16	9.8646	9.59	10.178	10.018	10.342	10.192	9.7834	10.354	9.8884	10.341	9.9668	9.7602	9.8412	9.6803	9.891	9.8549	9.7993	10.058	9.9478	10.333	9.9567	9.6108	10.065	9.776	9.9022

MODIFIED CHI-SQUARE CRITERION IN THE SENSE OF RAO

the optimalized θ	9.8573	10.138	10.242	10.021	10.288	10.25	9.7111	10.186	9.8068	10.396	9.945	9.8076	9.7858	10.174	9.8837	9.9342	9.739	10.103	9.8716	9.5907	10.044	9.6365	9.9159	9.7051	9.3926
the iteration #1	9.8926	10.055	10.13	10.08	10.279	10.127	9.8584	10.368	9.923	10.225	10.109	9.7411	9.8985	10.2	9.8843	9.8093	9.8293	10.059	9.9985	9.8508	9.8454	9.6117	9.8457	9.8284	9.6904
the iteration #2	9.9241	9.9879	10.025	10.133	10.269	10.007	10.048	10.614	10.044	10.047	10.257	9.6476	10.022	10.222	9.8848	9.6678	9.9058	10.02	10.122	10.091	9.6151	9.5873	9.7925	9.9833	10.006
the iteration #3	9.9839	9.8674	9.822	10.235	10.247	9.7729	10.492	11.202	10.292	9.6856	10.545	9.4254	10.284	10.262	9.8859	9.3737	10.049	9.9463	10.368	10.581	9.1195	9.5389	9.6997	10.339	10.702
the iteration #4	10.098	9.6506	9.4363	10.431	10.201	9.321	11.572	12.608	10.81	8.9641	11.126	8.9092	10.861	10.336	9.888	8.792	10.318	9.8082	10.859	11.659	8.0767	9.4436	9.5388	11.177	12.337
the iteration #5	10.316	9.2603	8.7193	10.806	10.104	8.4768	14.106	15.419	11.898	7.6092	12.325	7.7898	12.171	10.472	9.892	7.7697	10.834	9.5544	11.848	14.108	6.1783	9.2573	9.2639	13.124	15.626
the iteration #6	10.74	8.5585	7.4687	11.529	9.901	7.0189	17.474	18.225	14.034	5.5995	14.578	5.8465	14.843	10.721	9.9	6.3638	11.842	9.1026	13.752	17.503	4.2869	8.9	8.8073	16.349	18.524
the iteration #7	11.566	7.3217	5.709	12.894	9.4809	5.1052	19.575	19.991	16.937	4.0189	17.259	4.1159	17.803	11.177	9.9157	4.9941	13.761	8.3447	16.566	19.615	3.3756	8.2416	8.0876	18.842	20.229
the iteration #8	13.125	5.482	4.2316	15.196	8.6517	3.7969	20.902	21.172	19.142	3.2554	19.241	3.297	19.713	11.997	9.9464	3.9858	16.556	7.1908	18.901	20.938	2.9052	7.1275	7.0556	20.396	21.361
the iteration #9	15.572	3.9805	3.4103	17.801	7.2254	3.1469	21.843	22.037	20.58	2.8369	20.596	2.8602	20.973	13.401	10.007	3.3394	18.873	5.6834	20.421	21.875	2.6135	5.5737	5.7772	21.464	22.192
the iteration #10	18.048	3.2367	2.9454	19.67	5.4323	2.7725	22.563	22.711	21.594	2.5687	21.578	2.5838	21.885	15.474	10.127	2.9267	20.394	4.2679	21.479	22.59	2.4109	4.1581	4.5137	22.263	22.842
the iteration #11	19.802	2.8255	2.6486	20.934	4.0459	2.5253	23.141	23.26	22.361	2.3787	22.333	2.3895	22.589	17.708	10.366	2.6471	21.455	3.4169	22.273	23.165	2.2598	3.3497	3.6128	22.894	23.373
the iteration #12	21.013	2.5607	2.4405	21.852	3.2949	2.3472	23.621	23.719	22.972	2.2353	22.939	2.2435	23.158	19.465	10.853	2.4453	22.252	2.9415	22.902	23.643	2.1414	2.8989	3.074	23.412	23.819
the iteration #13	21.905	2.3728	2.2849	22.562	2.8685	2.211	24.031	24.114	23.477	2.122	23.442	2.1284	23.632	20.728	11.857	2.2919	22.883	2.6421	23.418	24.051	2.0454	2.6124	2.7356	23.85	24.203
the iteration #14	22.599	2.2306	2.163	23.135	2.5932	2.1026	24.387	24.459	23.905	2.0295	23.87	2.0347	24.038	21.666	13.854	2.1707	23.402	2.4337	23.855	24.405	1.9654	2.4117	2.5035	24.228	24.539
the iteration #15	23.163	2.1182	2.0642	23.613	2.3983	2.0135	24.701	24.764	24.275	1.952	24.241	1.9564	24.392	22.397	16.689	2.0719	23.841	2.2784	24.232	24.718	1.8973	2.2612	2.333	24.56	24.837
the iteration #16	23.635	2.0263	1.9819	24.022	2.2514	1.9386	24.982	25.038	24.601	1.8859	24.568	1.8896	24.704	22.988	18.934	1.9894	24.219	2.157	24.563	24.997	1.8384	2.1431	2.2011	24.855	25.105

HELLINGER DISTANCE CRITERION

the optimalized θ	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8468	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4512
the iteration #1	9.8747	10.112	10.201	10.057	10.302	10.208	9.7625	10.257	9.8554	10.344	10.024	9.7832	9.8263	10.206	9.8866	9.8745	9.7772	10.086	9.9308	9.6945	9.9704	9.6218	9.8793	9.7471	9.4972
the iteration #2	9.874	10.123	10.209	10.059	10.31	10.218	9.756	10.253	9.8511	10.36	10.02	9.7847	9.8196	10.216	9.8879	9.8738	9.7715	10.089	9.9281	9.6768	9.9796	9.6207	9.8811	9.741	9.4751
the iteration #3	9.8736	10.129	10.213	10.061	10.315	10.223	9.7529	10.252	9.849	10.368	10.018	9.7854	9.8163	10.221	9.8886	9.8735	9.7685	10.09	9.9268	9.6675	9.9841	9.6202	9.882	9.7381	9.4636
the iteration #4	9.8734	10.132	10.216	10.061	10.317	10.225	9.7514	10.251	9.8479	10.372	10.016	9.7857	9.8146	10.224	9.8889	9.8733	9.7669	10.091	9.9261	9.6626	9.9863	9.6199	9.8826	9.7367	9.4577
the iteration #5	9.8733	10.134	10.217	10.062	10.318	10.227	9.7507	10.251	9.8474	10.374	10.016	9.7859	9.8138	10.226	9.8891	9.8732	9.766	10.091	9.9257	9.66	9.9874	9.6197	9.8828	9.736	9.4546
the iteration #6	9.8733	10.134	10.217	10.062	10.318	10.227	9.7503	10.251	9.8471	10.375	10.016	9.7859	9.8134	10.227	9.8892	9.8732	9.7656	10.092	9.9256	9.6586	9.9879	9.6197	9.883	9.7357	9.453
the iteration #7	9.8732	10.135	10.218	10.062	10.319	10.228	9.7502	10.25	9.847	10.375	10.015	9.786	9.8132	10.227	9.8892	9.8731	9.7654	10.092	9.9255	9.6579	9.9882	9.6196	9.8831	9.7356	9.4521
the iteration #8	9.8732	10.135	10.218	10.062	10.319	10.228	9.7501	10.25	9.8469	10.376	10.015	9.786	9.8131	10.227	9.8892	9.8731	9.7653	10.092	9.9254	9.6575	9.9883	9.6196	9.8831	9.7355	9.4517
the iteration #9	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8469	10.376	10.015	9.786	9.8131	10.227	9.8892	9.8731	9.7652	10.092	9.9254	9.6573	9.9884	9.6196	9.8831	9.7355	9.4515
the iteration #10	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8469	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7652	10.092	9.9254	9.6572	9.9884	9.6196	9.8831	9.7354	9.4514
the iteration #11	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8469	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6572	9.9885	9.6196	9.8832	9.7354	9.4513
the iteration #12	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8469	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4513
the iteration #13	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8468	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4512
the iteration #14	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8468	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4512
the iteration #15	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8468	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4512
the iteration #16	9.8732	10.135	10.218	10.062	10.319	10.228	9.75	10.25	9.8468	10.376	10.015	9.786	9.813	10.227	9.8893	9.8731	9.7651	10.092	9.9254	9.6571	9.9885	9.6196	9.8832	9.7354	9.4512

KULLBACK-LEIBLER SEPARATOR CRITERION

the optimalized θ	9.8693	10.137	10.226	10.05	10.309	10.236	9.7371	10.224	9.8331	10.383	9.9964	9.7944	9.8036	10.21	9.8875	9.8898	9.7572	10.095	9.91	9.6371	10.005	9.6268	9.8927	9.7242	9.4298
the iteration #1	9.8754	10.096	10.188	10.052	10.287	10.193	9.7717	10.259	9.8609	10.319	10.029	9.7816	9.8362	10.19	9.8843	9.8774	9.7852	10.081	9.9335	9.7197	9.9575	9.6243	9.8774	9.756	9.5294
the iteration #2	9.8748	10.1	10.192	10.052	10.289	10.197	9.7679	10.255	9.858	10.326	10.026	9.7831	9.8328	10.192	9.8846	9.8787	9.7827	10.083	9.9312	9.7121	9.9625	9.6245	9.8788	9.7525	9.5198
the iteration #3	9.8743	10.103	10.195	10.052	10.291	10.201	9.7646	10.252	9.8555	10.331	10.023	9.7844	9.8298	10.193	9.8849	9.8799	9.7803	10.084	9.9291	9.7052	9.9669	9.6248	9.88	9.7494	9.5111
the iteration #4	9.8738	10.106	10.198	10.051	10.293	10.204	9.7615	10.249	9.8532	10.336	10.021	9.7855	9.8271	10.195	9.8852	9.8809	9.7782	10.085	9.9272	9.699	9.9709	9.625	9.8811	9.7466	9.5033
the iteration #5	9.8734	10.109	10.201	10.051	10.295	10.207	9.7589	10.246	9.8512	10.341	10.018	9.7866	9.8247	10.196	9.8854	9.8819	9.7763	10.086	9.9255	9.6933	9.9744	9.6252	9.8821	9.7441	9.4963
the iteration #6	9.873	10.111	10.203	10.051	10.296	10.21	9.7565	10.243	9.8494	10.345	10.016	9.7875	9.8225	10.197	9.8856	9.8827	9.7746	10.087	9.924	9.6882	9.9776	9.6253	9.883	9.7419	9.4899
the iteration #7	9.8727	10.113	10.205	10.051	10.298	10.213	9.7544	10.241	9.8477	10.349	10.014	9.7883	9.8205	10.198	9.8858	9.8834	9.773	10.087	9.9226	9.6835	9.9805	9.6255	9.8838	9.7399	9.4841
the iteration #8	9.8724	10.116	10.207	10.051	10.299	10.215	9.7525	10.239	9.8462	10.352	10.013	9.789	9.8188	10.199	9.886	9.8841	9.7716	10.088	9.9214	9.6793	9.983	9.6256	9.8846	9.7382	9.4789
the iteration #9	9.8721	10.117	10.209	10.051	10.3	10.217	9.7508	10.237	9.8449	10.355	10.011	9.7896	9.8172	10.2	9.8861	9.8847	9.7703	10.089	9.9203	9.6754	9.9853	9.6257	9.8853	9.7367	9.4742
the iteration #10	9.8718	10.119	10.211	10.051	10.301	10.219	9.7493	10.236	9.8437	10.358	10.01	9.7901	9.8158	10.201	9.8862	9.8852	9.7691	10.089	9.9193	9.6719	9.9874	9.6258	9.8859	9.7353	9.47
the iteration #11	9.8716	10.121	10.212	10.051	10.302	10.221	9.748	10.235	9.8426	10.361	10.009	9.7906	9.8145	10.202	9.8864	9.8857	9.768	10.09	9.9184	9.6687	9.9892	9.6259	9.8865	9.734	9.4661
the iteration #12	9.8714	10.122	10.214	10.051	10.302	10.222	9.7468	10.233	9.8417	10.363	10.008	9.791	9.8134	10.203	9.8865	9.8861	9.767	10.09	9.9175	9.6658	9.9908	9.626	9.887	9.733	9.4627
the iteration #13	9.8712	10.124	10.215	10.051	10.303	10.224	9.7457	10.232	9.8408	10.365	10.006	9.7914	9.8124	10.203	9.8866	9.8865	9.7661	10.091	9.9168	9.6632	9.9923	9.6261	9.8875	9.732	9.4595
the iteration #14	9.871	10.125	10.216	10.051	10.304	10.225	9.7448	10.231	9.84	10.367	10.006	9.7918	9.8115	10.204	9.8867	9.8869	9.7653	10.091	9.9161	9.6608	9.9936	9.6262	9.888	9.7311	9.4567
the iteration #15	9.8708	10.126	10.217	10.051	10.304	10.226	9.7439	10.23	9.8393	10.368	10.005	9.7921	9.8107	10.204	9.8867	9.8872	9.7646	10.092	9.9155	9.6587	9.9948	9.6262	9.8884	9.7304	9.4541
the iteration #16	9.8707	10.127	10.218	10.051	10.305	10.227	9.7432	10.23	9.8387	10.37	10.004	9.7923	9.8099	10.205	9.8868	9.8874	9.7639	10.092	9.915	9.6567	9.9959	9.6263	9.8887	9.7297	9.4518

SAMPLE	#26	#27	#28	#29	#30	#31	#32	#33	#34	#35	#36	#37	#38	#39	#40	#41	#42	#43	#44	#45	#46	#47	#48	#49	#50
the true θ	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
original average	9.892	10.136	10.188	10.084	10.356	10.212	9.784	10.252	9.872	10.372	10.04	9.784	9.768	10.284	9.868	9.824	9.8	10.06	9.956	9.68	9.94	9.636	9.896	9.772	9.488
sorted-out average	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

HALDANE'S DISTANCE CRITERION

the optimalized θ	10.105	10.104	10.169	9.8841	10.379	10.285	9.847	10.261	9.8814	10.546	10.484	9.9781	10.284	10.306	10.374	10.244	9.8146	10.4	9.7205	9.3734	10.199	9.7589	9.9601	9.9201	9.6602
the iteration #1	8.1261	10.081	10.238	11.254	9.2752	9.857	8.8439	10.283	9.7834	9.2099	8.8807	6.0623	9.0058	10.075	9.4708	6.8049	9.606	8.4738	11.043	10.601	6.6975	9.1279	9.3204	8.8692	8.8516
the iteration #2	5.553	10.059	10.473	15.159	5.4835	8.1018	4.3724	10.426	9.3622	4.3732	8.0637	3.8512	7.9182	9.8513	8.5365	4.2518	8.3873	4.8649	14.601	16.175	3.9757	6.7689	7.1428	4.4352	6.5769
the iteration #3	3.7314	10.016	11.423	18.621	3.7134	4.1813	3.3581	11.939	7.4945	3.3584	4.5364	3.1359	4.2727	9.3325	4.2779	3.3104	5.8681	3.5329	18.405	19.056	3.1929	3.9883	4.1391	3.3821	4.0459
the iteration #4	3.0786	9.9184	17.251	20.343	3.0698	3.2815	2.8832	17.565	4.1038	2.8834	3.4198	2.754	3.3188	6.4531	3.3209	2.8563	3.8079	2.978	20.215	20.609	2.7881	3.1985	3.2639	2.8966	3.2239
the iteration #5	2.719	9.6417	19.579	21.461	2.7135	2.8398	2.5944	19.744	3.2489	2.5945	2.9174	2.5074	2.8611	3.9307	2.8623	2.5766	3.1155	2.6558	21.372	21.649	2.5307	2.7914	2.8296	2.6032	2.8064
the iteration #6	2.4832	7.8759	20.944	22.278	2.4794	2.5656	2.3949	21.053	2.821	2.395	2.6167	2.3312	2.5798	3.1726	2.5805	2.382	2.7416	2.4389	22.211	22.422	2.3484	2.533	2.5588	2.4012	2.5431
the iteration #7	2.3131	4.1528	21.892	22.917	2.3103	2.374	2.2463	21.972	2.5529	2.2463	2.411	2.1969	2.3843	2.776	2.3849	2.2363	2.4989	2.2798	22.863	23.033	2.2104	2.3501	2.369	2.2511	2.3575
the iteration #8	2.1828	3.2696	22.611	23.439	2.1806	2.2302	2.1299	22.674	2.3647	2.1299	2.2586	2.0902	2.2381	2.5225	2.2386	2.1219	2.3248	2.1565	23.394	23.535	2.101	2.2117	2.2263	2.1337	2.2175
the iteration #9	2.0787	2.833	23.186	23.878	2.0769	2.117	2.0353	23.238	2.223	2.0354	2.1397	2.0025	2.1233	2.3424	2.1237	2.0288	2.192	2.0573	23.84	23.96	2.0115	2.1021	2.1139	2.0385	2.1067
the iteration #10	1.9929	2.561	23.664	24.256	1.9914	2.0247	1.9565	23.707	2.1112	1.9566	2.0434	1.9287	2.03	2.2057	2.0303	1.951	2.0861	1.975	24.222	24.328	1.9364	2.0124	2.0222	1.9592	2.0162
the iteration #11	1.9206	2.3706	24.071	24.587	1.9193	1.9476	1.8895	24.108	2.02	1.8895	1.9634	1.8655	1.952	2.0972	1.9523	1.8847	1.9991	1.9053	24.558	24.65	1.8721	1.9371	1.9454	1.8918	1.9404
the iteration #12	1.8584	2.2276	24.424	24.881	1.8573	1.8817	1.8314	24.457	1.9435	1.8314	1.8953	1.8104	1.8856	2.0083	1.8858	1.8272	1.9259	1.8451	24.855	24.938	1.8162	1.8727	1.8799	1.8334	1.8756
the iteration #13	1.8043	2.1149	24.736	25.145	1.8033	1.8246	1.7805	24.765	1.8783	1.7805	1.8365	1.7619	1.828	1.9337	1.8282	1.7768	1.863	1.7926	25.122	25.197	1.767	1.8168	1.823	1.7822	1.8193
the iteration #14	1.7564	2.023	25.015	25.385	1.7556	1.7745	1.7353	25.041	1.8216	1.7353	1.785	1.7187	1.7775	1.8698	1.7776	1.732	1.8083	1.7461	25.363	25.432	1.7233	1.7675	1.7731	1.7369	1.7697
the iteration #15	1.7138	1.9461	25.266	25.604	1.713	1.7299	1.6948	25.29	1.7718	1.6948	1.7393	1.6799	1.7326	1.8142	1.7327	1.6918	1.76	1.7045	25.584	25.647	1.684	1.7237	1.7287	1.6962	1.7257
the iteration #16	1.6754	1.8805	25.495	25.805	1.6747	1.69	1.6582	25.517	1.7276	1.6582	1.6984	1.6447	1.6924	1.7653	1.6925	1.6556	1.717	1.667	25.787	25.845	1.6484	1.6844	1.6888	1.6595	1.6862

MODIFIED CHI-SQUARE CRITERION IN THE SENSE OF ANDĚL

the optimalized θ	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #1	9.876	10.132	10.209	10.072	10.327	10.218	9.7625	10.282	9.861	10.368	10.03	9.777	9.8228	10.243	9.8909	9.861	9.7726	10.089	9.9385	9.6744	9.9753	9.6109	9.8752	9.748	9.4741
the iteration #2	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #3	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #4	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #5	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #6	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #7	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #8	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #9	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #10	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #11	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #12	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #13	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #14	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #15	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732
the iteration #16	9.876	10.133	10.21	10.073	10.328	10.219	9.7624	10.282	9.861	10.369	10.03	9.777	9.8226	10.245	9.891	9.8608	9.7724	10.089	9.9385	9.6736	9.9755	9.6108	9.8752	9.7479	9.4732

Analýza vybraných problémov reálnej konvergencie Slovenska k EMÚ[1]

Boďa, Martin - Čižmárik, Pavol – Gavliak, Rudolf

Abstract: The fulfillment of nominal convergence criteria is often used as a criterion of economic policy successfulness in countries, which attempt to adopt the euro currency union (Euro Area). Authors of this article point out the fact, that also countries with underdeveloped economic structure (with low level of labor productivity and parallelly with low rate of its real convergence) are able to fulfill the nominal convergence criteria in a short-term period (also called "Maastricht criteria"), however only by taking risk of both internal and external imbalances. This hypothesis is tested in the case of Slovak republic. The authors anticipate that early accession to Euro Area could cause severe economic problems (real convergence deceleration) in that countries, with impact on life standard level with parallel deepening of both internal and external imbalances. This is especially the case of those countries, which fulfill the Maastricht criteria but do not get to the level, which is necessary for the accession to Euro Area by the theory of optimal currency area (Mundell, McKinnon).

Keywords: *Maastricht criteria, Real convergence, Polak's balance of payment model, internal and external equilibrium condition.*

1. Úvod

Veľmi aktuálnou témou diskutovanou na odbornej i laickej úrovni je plánovaný vstup Slovenska do Európskej menovej únie (EMU) a s tým súvisiaceho zavedenia spoločnej európskej meny v roku 2009. Vstup Slovenska do EMÚ je podmienený splnením Maastrichtských kritérií. Vzniká otázka, či sú tieto kritériá vhodné na posúdenie stupňa a rýchlosti konvergencie národnej ekonomiky k Európskej únii. Existujú názory, že Maastrichtské kritériá zahŕňajú iba finančné a monetárne kritériá. Ani jedno z týchto kritérií nezohľadňuje reálny vývoj ekonomiky – konvergenciu reálneho produktu v parite kúpnej sily, konvergenciu produktivity práce a cenovú konvergenciu.

Vzniká tak otázka, či aj krajiny so zaostalou štruktúrou ekonomiky, nízkou a nekonvergujúcou produktivitou práce, nízkou úrovňou produktu na obyvateľa a cenovou hladinou môžu splniť aspoň krátkodobo Maastrichtské kritériá nominálnej konvergencie. Je aspoň krátkodobo možné, aby takáto krajina udržala stabilnú mieru inflácie, stabilný menový kurz, dlhodobé úrokové sadzby na požadovanej úrovni a rovnováhu verejných financií? Pri vstupe do EMÚ je síce posudzovaná udržateľnosť plnenia Maastrichtských kritérií, toto posúdenie budí dojem skôr politického rozhodnutia. Cena za krátkodobé splnenie týchto kritérií je vnútorná alebo vonkajšia nerovnováha. Takáto krajina by však po vstupe do menovej únie mohla reagovať asymetricky na vonkajšie šoky, čo by tejto zaostávajúcej krajine spôsobilo značné problémy a pravdepodobne prehĺbenie tejto nerovnováhy.

2. Reálna konvergencia pred vstupom do EMÚ

Ekonomická teória sa zaoberá optimálnymi menovými oblasťami (*Optimal Currency Area*) a vymedzuje ďalšie teoretické predpoklady, ktoré by mali spĺňovať krajiny vstupujúce do menovej únie. Kenen (1969) ukázal, že v prípade nízkej diverzifikácie výstupu ekonomiky sa zvyšuje pravdepodobnosť asymetrických šokov. Slovenské zameranie na automobilový priemysel by mohlo predstavovať problém pri integrácii Slovenska do EMÚ.

[1] Tento príspevok bol spracovaný ako súčasť riešenia fakultného grantového projektu FG 75 „Modelovanie dopadov reálneho konvergenčného procesu v EÚ na ekonomickú rovnováhu Slovenska" a projektu VEGA 1/25/94/05 „Variantné metódy predikovania finančného vývoja malých a stredných podnikov po zavedení spoločnej európskej meny v Slovenskej republike".

Je empirickou skúsenosťou, že krajiny, ktoré prechádzajú fázou rýchlej reálnej konvergencie, trpia vysokou elasticitou importu (*Darvas, Sass, 2001*). Znamená to, že zvýšenie domáceho dopytu vedie k výraznému nárastu importu a deficitu bežného účtu platobnej bilancie.

Výška nákladov spojených so začlenením krajiny do menovej únie je spojená s rozdielom vo výške cenovej hladiny a v rozdiele v produkte na obyvateľa. Rozdiely vo výške vyprodukovaného produktu na obyvateľa je možné dosiahnuť iba vyššou mierou reálneho rastu. Naproti tomu zníženie rozdielu cenovej úrovne kandidátskej krajiny a priemeru menovej únie môže prebiehať zhodnocovaním meny (kurzový kanál) resp. vyššou mierou inflácie (inflačný kanál). Podľa teórie parity kúpnej sily by sa mal rozdiel v cenových hladinách v menovej únii a kandidujúcej krajine plne prejaviť v zmene menového kurzu:

$$\frac{CPI_{SKK}}{CPI_{EUR}} \approx \Delta ER_{EUR/SKK} - 1, \tag{1}$$

Potom by malo platiť, že miery rastu reálneho produktu vyjadreného v domácej a zahraničnej mene sú rovnaké. V realite ale platí, že reálna konvergencia prebieha ako rozdielnym tempom rastu cenovej hladiny, tak aj zhodnocovaním menového kurzu. Potom by mal podiel produktu vyjadreného v zahraničnej a domácej mene zodpovedať podielu cenových hladín v menovej únii a kandidátskej krajine a nepriamoúmerne menovému kurzu:

$$\frac{Y_{EUR}^{SC}}{Y_{SKK}^{SC}} = \frac{CPI_{SKK}}{CPI_{EUR} \cdot \left(\Delta ER_{EUR/SKK} - 1\right)}, \tag{2}$$

Budeme predpokladať, že tento vzťah neplatí iba vo svojom úrovňovom tvare, ale aj v dynamickej forme, t.j. prírastok produktu v zahraničnej mene je proporcionálny prírastku pomeru domácej inflácie a inflácie v Eurozóne a nepriamoúmerný zmene menového kurzu. Vzťah (2) sa zmení na:

$$\mathrm{dlog}\left(Y_{EUR}^{SC}\right) - \mathrm{dlog}\left(y_{SKK}^{SC}\right) = \mathrm{dlog}\left(CPI_{SKK}\right) - \mathrm{dlog}\left(CPI_{EUR}\right) - \mathrm{dlog}\left(ER_{EUR/SKK}\right), \tag{3}$$

Zmenu prirodzeného logaritmu je možné považovať za aproximáciu tempa rastu. Potom sa zmení vzťah (3) na identitu:

$$r_{EUR}^{SC} - r_{SKK}^{SC} = p_{SKK} - p_{EUR} - er_{EUR/SKK}, \tag{4}$$

Predpokladáme teda, že rozdiel v tempe rastu produktu vyjadreného v eure a v slovenských korunách je závislý od rozdielu v mierach inflácie a zmeny nominálneho menového kurzu. Predpokladáme, že nie oba faktory pôsobia s rovnakou proporciou. Dá sa ďalej predpokladať, že sú možné krátkodobé odchýlenia od tohto stavu, preto odhadneme nasledujúcu stochastickú špecifikáciu ekonometrického modelu:

$$r_{SR,t}^{SC,EUR} - r_{SR,t}^{SC,SKK} = \lambda_0 + \lambda_1\left(p_{SKK,t} - p_{EUR,t}\right) - \lambda_2 er_{EUR/SKK,t} + \varepsilon_t, \tag{5}$$

3. Asymetrické šoky

Najväčšie obavy z predčasného začlenenia krajiny do menovej únie sú spojené s asymetrickými šokmi pri zmene menového kurzu spoločnej meny. Ak dosahuje daná krajina rozdielne koeficienty elasticity zmeny exportu a importu na zmenu menového kurzu krajiny,

potom zmena kurzu spoločnej meny bude mať rôzny dopad na platobné bilancie jednotlivých členských krajín. Tento stav vedie k tlakom na rôzne nastavenie menovo-politických nástrojov. Výrazne vysoké sú obvykle hlavne elasticity importu, resp. exportu surovín (ropa, zemný plyn) na zmenu menového kurzu. Ak sú súčasťou menovej únie čistí importéri surovín a zároveň čistí exportéri surovín, potom pri zmene menového kurzu spoločnej meny dochádza k pnutiam a nie je jednoduché nastaviť menovo-politické nástroje k spokojnosti všetkých zainteresovaných krajín. Zároveň ale platí, že zmeny úrokovej sadzby na úrovni menovej únie spôsobujú spätne asymetrické reakcie na strane agregátneho dopytu aj agregátnej ponuky. V prípade rôznej úrovne finančných aktív na obyvateľa a rôznej úrovne zadlženosti domácnosti sa kumulujú asymetrické šoky v rámci menovej únie. Asymetrický vplyv sa dá predpokladať takisto aj na strane agregátnej ponuky v prípade rozdielnej miery zadlženosti podnikov. Asymetria šokov teda nezáleží iba od deficitu, alebo prebytku platobnej bilancie, ale aj štruktúry importu a exportu a ich dôchodkovej citlivosti, citlivosti na inflačný diferenciál a zmenu menového kurzu. Na hlbšiu analýzu citlivosti na vznik asymetrických šokov využijeme Polakov monetárny prístup k platobnej bilancii v kombinácii s dôchodkovou verziou rovnice výmeny peňazí. Budeme vychádzať z dynamicky definovanej rovnováhy:

$$\mathrm{dlog}\left(EXP\right)-\mathrm{dlog}\left(IMP\right)=\varepsilon_t\,, \tag{6}$$

Z pohľadu dosahovanej rovnováhy je dôležité, aby sa rozdiely v prírastkoch exportu a prírastkoch importu z dlhodobého hľadiska rovnali (vyrovnaná platobná bilancia). Z krátkodobého hľadiska sa dajú predpokladať krátkodobé odchýlky od tohto stavu. Optimálnym stavom by bolo, keby vektor náhodných chýb zo vzťahu (6) mal nulovú strednú hodnotu a konštantný rozptyl: $\varepsilon_t \sim dist\left(\mathbf{0},\sigma^2\mathbf{I}\right)$. Postačujúcou podmienkou, ktorú v našom prípade budeme klásť na rozdiel v tempe zmeny exportu a importu je konvergencia, resp. trendová stacionarita.

Predpokladáme závislosť exportnej funkcie od dôchodku (domáceho aj zahraničného), produktivity práce a výmenného kurzu. Odhadnuté regresné parametre predstavujú koeficienty dôchodkovej elasticity exportu, elasticity exportu na zmenu úrokovej sadzby a citlivosť exportu na zmenu produktivity práce.

$$\mathrm{dlog}\left(EXP_t\right)=\alpha_0+\alpha_1\mathrm{dlog}\left(\mathrm{Y_t}\right)+\alpha_2\mathrm{dlog}\left(PP_\mathrm{t}\right)+\alpha_3\mathrm{dlog}\left(\mathrm{Y_t^{EU}}\right)+\alpha_4\mathrm{dlog}\left(ER_{EUR/SKK}\right)+\varepsilon_t\,. \tag{7}$$

Pri odhade exportnej funkcie sa stretávame s viacerými problémami. Pri empirickej analýze exportnej funkcie nahradíme rast domáceho a zahraničného dopytu v bežných cenách rastom dopytu v stálych cenách a zmeny rozdielu cenovej hladiny v kandidátskej krajine a v menovej únii. Po využití symboliky zo vzťahu (5) a nahradením aproximácií mierou rastu dostaneme:

$$rex_{SR,t}^{BC}=\alpha_0+\alpha_1 r_{SR,t}^{SC}+\alpha_2\mathrm{dlog}\left(PP_\mathrm{t}\right)+\alpha_3 r_{EU,t}^{SC}+\alpha_4 er_{EUR/SKK,t}+\alpha_5\left(p_{SR,t}-p_{EUR,t}\right)+\varepsilon_t\,. \tag{8}$$

Pre funkciu importu predpokladáme závislosť od domáceho produktu, pričom hodnota odhadnutého regresného koeficientu je koeficientom dôchodkovej elasticity importu. Takisto predpokladáme, že na import vplýva menový kurz (hlavne v prípade krajín, ktoré sú čistými dovozcami surovín) a približovanie cenových hladín (cenová konvergencia):

$$\mathrm{dlog}\left(IMP_t\right)=\beta_0+\beta_1\mathrm{dlog}\left(\mathrm{Y}_t\right)+\beta_2\mathrm{dlog}\left(ER_{EUR/SKK}\right)+\beta_3\mathrm{dlog}\left(\frac{CPI_{SKK}}{CPI_{EUR}}\right)+\nu_t. \tag{9}$$

Po nahradení aproximácií mierami rastu dostávame nasledujúci vzťah pre funkciu importu:

$$rim_{SR,t}^{BC}=\beta_0+\beta_1 r_{SR,t}^{SC}+\beta_2 er_{EUR/SKK,t}+\beta_3\left(p_{SR,t}-p_{ER,t}\right)+\nu_t. \tag{10}$$

Z počiatočnej podmienky vonkajšej rovnováhy (6) predpokladáme, že z dlhodobého pohľadu budú reziduá oboch modelov kolísať okolo nulovej hodnoty $E[\boldsymbol{\varepsilon}]=E[\boldsymbol{\nu}]=0$. Vo vzťahoch (8) a (10) nahradíme bežné hodnoty domáceho a zahraničného produktu hodnotami vyrovnanými pomocou Hodrick-Prescottovho (*H-P*) filtra. Technicky je *H-P* filter dvojstranný lineárny filter, ktorý vypočíta vyrovnaný časový rad *s* pôvodného časového radu *y* minimalizovaním rozptylu *y* okolo *s*, vzhľadom k druhej diferencii *s*. *H-P* filter určí hodnoty hľadaného časového radu s minimalizáciou nasledujúceho výrazu:

$$\sum_{t=1}^{n}\left(y_t-s_t\right)^2+\lambda\sum_{t=2}^{n-1}\left(\left(s_{t+1}-s_t\right)-\left(s_t-s_{t-1}\right)\right)^2, \tag{11}$$

kde

λ je parameter kontrolujúci rozptyl vyrovnaného časového radu.

Hodnota parametra λ závisí od frekvencie dát časového radu. Najčastejšie používaná hodnota tohto parametra sa určí ako (množstvo údajov za rok)$^2\times 100$. Pre štvrťročné časové rady je teda hodnota $\lambda=1600$.

Potom reziduá salda platobnej bilancie $\eta=\varepsilon-\nu$ za predpokladu nezávislosti spĺňajú nasledujúcu podmienku:

$$\begin{aligned}\eta_t=&\alpha_0+\left(\alpha_1-\beta_1\right)r_{SR,t}^{SC,POT}+\alpha_2\mathrm{dlog}\left(PP_t\right)+\alpha_3 r_{EUR,t}^{SC,POT}+\left(\alpha_4-\beta_2\right)er_{EUR/SKK,t}\\&+\left(\alpha_5-\beta_3\right)\left(p_{SR,t}-p_{ER,t}\right)\end{aligned} \tag{12}$$

V empirickej analýze odhadneme vzťah (5) a určíme silu kurzového a inflačného kanála pri reálnej konvergencii Slovenska k priemeru Eurozóny.

V druhej časti odhadneme exportnú a importnú funkciu zo vzťahov (8) a (10), pričom odhadneme vektory parametrov $\boldsymbol{\alpha},\boldsymbol{\beta}$. Následne odhadneme pravú stranu výrazu (12) s vplyvom menového kurzu a bez neho vplyvu. Takto získané časové rady porovnáme a otestujeme na stacionaritu.

4. Dáta a výsledky

Zdrojom údajov pre analýzu sú časové rady z databázy Eurostatu. Hrubý domáci produkt je vyjadrený v trhových cenách roku 1995, pričom údaje pre Slovensko sú v vyjadrené ako v Slovenských korunách, tak aj v Eure. Index spotrebiteľských cien je určený v cenách roku 2005 a vzhľadom na výrazné ovplyvnenie časového radu slovenskej miery inflácie cenovými dereguláciami sme za slovenskú mieru inflácie zvolili mieru čistej inflácie. Zdrojom týchto dát je Eurostat. Z tejto databázy sme čerpali aj dáta pre časový rad hrubého domáceho produktu Eurozóny oceneného v eurách. Informačným zdrojom pre časové rady importu, exportu a produktivity práce je databáza Slovstatu. Takisto časový rad slovenského hrubého domáceho produktu ocenený v slovenských korunách a eurách pochádza z tejto databázy. Údaje o priemernom menovom kurze pochádzajú z databázy

Národnej banky Slovenska (NBS). Všetky údaje majú štvrťročnú frekvenciu a sú za obdobie prvého kvartálu 2000 a druhého kvartálu roku 2006.

Kvantifikovali sme rovnicu zo vzťahu (5), kde sme diagnostikovali silnú autokoreláciu prvého stupňa. Ukázalo sa, že lepšie výsledky sme dosiahli pre medziročné a nie medzikvartálne miery rastu hrubého domáceho produktu. Hodnoty uvádzané v zátvorkách sú *t*-štatistiky zodpovedajúcich regresných parametrov.

$$\begin{aligned} r_{SR,t}^{SC,EUR} - r_{SR,t}^{SC,SKK} &= 41,7 + 0,7\left(p_{SKK,t-1} - p_{EUR,t-1}\right) - er_{EUR/SKK,t-1} + 0,06\left(r_{SR,t-1}^{SC,EUR} - r_{SR,t-1}^{SC,SKK}\right) + \varepsilon_t \\ R^2 = 0,99 \qquad (25,52) \qquad (27,79) \qquad & (-24,83) \qquad (2,94) \qquad DW = 1,88 \end{aligned} \tag{13}$$

V ďalšom kroku sme analyzovali exportné a importné funkcie v podmienkach Slovenskej republiky. Vzhľadom na problém autokorelácie sme do pôvodných špecifikačných tvarov zo vzťahov (8) a (10) upravili o oneskorené hodnoty percentuálneho prírastku exportu a importu. Takisto sa ukázalo, že vzťahy medzi jednotlivými premennými nie sú bezprostredné, ale export a import reaguje na zmeny regresorov až s istým oneskorením. Vo výsledných rovniciach neuvádzame nevýznamné regresory.

$$\begin{aligned} rex_{SR,t}^{BC} &= -0,097 + 4,71r_{EU,t}^{SC} + 1,02r_{SR,t}^{SC} - 0,61er_{EUR/SKK,t-1} + 1,87rex_{SR,t-1}^{BC} + \varepsilon_t \\ R^2 = 0,94 \; (-3,72) \quad (4,21) \qquad & (2,96) \qquad (-1,77) \qquad (6,95) \quad DW = 2,33 \end{aligned} \tag{14}$$

Slovenská importná funkcia má nasledujúci tvar:

$$\begin{aligned} rim_{SR,t}^{BC} = \quad & 0,08 - 0,35r_{SR,t}^{SC} + 1,21er_{EUR/SKK,t} - 0,007\left(p_{SR,t} - p_{ER,t}\right) - 0,79rim_{SR,t}^{BC} + \nu_t . \\ R^2 = 0,99 \; (13,89) \quad & (-4,53) \quad (11,16) \qquad (-13,95) \qquad (-27,46) \quad DW = 2,23 \end{aligned} \tag{15}$$

V nasledujúcej časti príspevku interpretujeme výsledky kvantifikovaných funkcií a zhrnieme výsledky.

5. Diskusia

Rozdielne tempo rastu hrubého domáceho produktu vyjadreného v slovenských korunách a v eurách je aspoň čiastočne vysvetliteľné vyššou mierou inflácie na Slovensku ako v Eurozóne a menovým kurzom. Odhadnuté koeficienty predstavujú relatívne koeficienty elasticity. Predelené výrazom jedna mínus koeficient pri oneskorenej hodnote závislej premennej dostaneme dlhodobé koeficienty elasticity. Dá sa teda povedať, že jednopercentná zmena inflačného diferenciálu medzi Slovenskom a Eurozónou bola sprevádzaná 0,75 % nárastom rozdielu v raste HDP v eurách a slovenských korunách. Zhodnotenie slovenskej koruny o jedno percento bolo v minulosti sprevádzané nárastom rozdielu v raste HDP v eurách a slovenských korunách 1,08 %. Ukazuje sa teda, že kurzový kanál hral v minulosti významnejšiu úlohu pri konvergencii úrovne produktu k priemeru Eurozóny.

Pri odhade funkcie exportu v slovenských podmienkach sa ukázali ako nevýznamné regresory produktivita práce a rozdiel v mierach inflácie na Slovensku a v Eurozóne. Ukazuje sa, že slovenský export je výrazne závislý od externého dopytu, ak by rast HDP Eurozóny poklesol o jedno percento, potom by slovenský export poklesol až o 4,71 %. Zmena domáceho dopytu sa takmer proporcionálne prejavila v minulosti v náraste exportu. Zhodnotenie kurzu bolo paradoxne sprevádzané nárastom exportu.

Na slovenský import vplýva rast produktu opačným smerom, ako sme predpokladali. Nárast produktu má vplyv na pokles importu, čo sa dá vysvetliť tak inováciami a využívaním

technológií s menšou energetickou náročnosťou. Import je takisto citlivý na zmenu menového kurzu. Ak v minulosti koruna poklesla o percento, import poklesol nadproporcionálne. V tomto prípade je významný ako regresor aj rozdiel inflácií na Slovensku a v Eurozóne. Platí, že v období s vyššou infláciou na Slovensku rástol import nižším tempom, čo sa dá vysvetliť saturáciou kúpyschopného dopytu

6. Literatúra

GRANGER, C.W. – NEWBOLD, P. 1974. SPURIOUS REGRESSION IN ECONOMETRICS. IN: JOURNAL OF ECONOMETRICS, Č. 2, 1974, S. 111 – 120.

GUJARATI, N. D. 2003. BASIC ECONOMETRICS. 4TH EDITION. NEW YORK: MCGRAW-HILL, 2003. ISBN 0-07-112342-3.

HATRÁK, M. 2007. EKONOMETRIA.. BRATISLAVA: IURA, 2007. ISBN 978-80-8078-150-7.

HEIJ, CH., DE BOER, P., FRANCES, P. H., KLOEK, T., VAN DIJK, H. K. 2004. ECONOMETRIC METHODS WITH APPLICATION IN BUSINESS AND ECONOMICS. NEW YORK : OXFORD UNIVERSITY PRESS, 2004. ISBN 0-19-926801-0.

KENEN, P. B. 1969. ROUND TABLE ON EXCHANGE RATE POLICY. IN: AMERICAN ECONOMIC REVIEW, AMERICAN ECONOMIC ASSOCIATION, VOL. 59 (2), PP 362-364.

MANDEL, M. – TOMŠÍK, V. 2003. MONETÁRNÍ EKONOMIE V MALÉ OTEVŘENÉ EKONOMICE. PRAHA: MANAGEMENT PRESS, 2003. ISBN 80-7261-094-5.

MCKINNON, R. 1963. THE THOERY OF OPTIMUM CURRENCY AREA. IN: AMERICAN ECONOMIC REVIEW, NO. 4, 1963, PP. 717 – 725.

MUNDELL, R. A. 1961. A THEORY OF OPTIMUM CURRENCY AREAS. THE AMERICAN ECONOMIC REVIEW. 1961, VOL. 51, NO. 4, PP. 657–665.

POLAK, J. J. 2001. THE TWO MONETARYAPPROACHES TO BALANCE OF PAYMENTS: KEYNESIAN AND JOHNSONIAN. IMF WORKING PAPER WP01/100, 2001.

Adresa autorov:

Martin Boďa, Ing. et Bc.
Katedra kvantitatívnych metód a informatiky
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
martin.boda@umb.sk

Rudolf Gavliak, Ing.
Katedra kvantitatívnych metód a informatiky
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
rudolf.gavliak@umb

Pavol Čižmárik, Ing. M.A.
Katedra financií a účtovníctva
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
pavol.cizmarik@umb.sk

Value at risk III. Indexný VCV model a diagnostika modelu value at risk[∞]

Martin Boďa[†] *– Rudolf Gavliak*[§]

Abstract: This article completes the series on value at risk. Being understood that the notion of value at risk and its basic methodologies were introduced in the former two articles, the contents of this continuation are a market index variance-covariance model and some casual approaches to the diagnostics of a value at risk model: concretely via back-testing and stress-testing.

Keywords: market index, variance-covariance method, value at risk diagnostics. backtesting, stresstesting.

1. Obsah a zameranie príspevku

Príspevok nadväzuje na články *Value at risk I. Value at risk ako miera rizika, alternatívy, nedostatky a regulačný aspekt* a *Value at risk II. Základné prístupy k modelovaniu*, ktoré boli publikované vo *Forum Statisticum Slovacum 4/2006* a *5/2006.* Prvý z článkov uviedol koncept value at risk a zdôraznil jeho použiteľnosť pri meraní trhovej formy finančného rizika. Druhý článok načrtol rozličné metodológie aplikovateľné pre zostavenie odhadu value at risk konkrétneho obchodného portfólia. Je celkom prirodzené pokračovať v tejto problematike prezentáciou spôsobov, ktorými je možné ohodnotiť kvalitu zostaveného odhadu. Načrtnutý postup, ktorým sa nadväzuje na druhý článok, je plne konzistentný s chápaním konštrukcie odhadov value at risk ako štvorkrokového procesu podľa Boďa (2006c). Podľa tohto prístupu je možné, bez ujmy na všeobecnosti, vyčleniť štyri etapy pri odhadovaní value at risk: (1.) zobrazenie rizika, (2.) špecifikácia vzoru volatility, (3.) voľba a aplikácia metódy a (4.) ohodnotenie výkonnosti predpovedí. Okrem toho, že nosnou témou príspevku je finálny: štvrtý bod, je obsah príspevku rozšírený i o indexný trhový model, ktorý je rozšírením variančno-kovariančnej metódy.

2. Indexný variančno-kovariančný parametrický model

V teórii finančného investovania dominuje niekoľko postulátov a medzi inými i myšlienka, že celkové riziko finančného inštrumentu (portfólia) je možné dezagregovať na jeho

- *systematické (trhové) riziko* pôsobiace celoplošne na všetky subjekty a
- *nesystematické (špecifické) riziko* príznačné danému finančnému inštrumentu (portfóliu).

Pretože sa za zdroj výnosu v najvšeobecnejšej rovine považuje riziko, je justifikovaná dekompozícia výnosu na systematický a nesystematický výnos. Tvrdeniu v najjednoduchšom prípade zodpovedá jednoduchá indexná formulácia výnosu $R_t^{[i]}$ *i*-teho finančného inštrumentu v čase *t* v podobe

$$R_t^{[i]} = \beta_i R_{market,t} + \varepsilon_t^{[i]},$$

v ktorej prvý člen $\beta_i R_{market,t}$ reprezentuje systematický (trhový) komponent daného inštrumentu tvorený koeficientom proporcionality β_i[1] a výnosom $R_{market,t}$ referenčného trhového indexu

[∞] Príspevok bol spracovaný v rámci riešenie grantovej úlohy VEGA/1/2594/05 Analzýza vybraných otázok finančného a bankového trhu po vstupe SR do EÚ.

[†] Ing. et Bc. Martin Boďa. Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Katedra kvantitatívnych metód a informatiky, Tajovského 10, 975 90 Banská Bystrica. E-mail: martin.boda@umb.sk.

[§] Ing. Rudolf Gavliak. Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Katedra kvantitatívnych metód a informatiky, Tajovského 10, 975 90 Banská Bystrica. E-mail: rudolf.gavliak@umb.sk.

a druhý člen $\varepsilon_t^{[i]}$ zastupuje rušivý nesystematický komponent výnosu daného inštrumentu (porov. napr. Fabozzi a Modigliani, 2003 alebo Härdle, Hlávka, Stahl, 2006).[2]

Na model sa kladú podmienky $\mathbf{E}\varepsilon_t^{[i]} = 0$, $\mathbf{D}\varepsilon_t^{[i]} = \sigma^2_{\varepsilon[i]\,t}$, $\mathbf{cov}\left(\varepsilon_t^{[i]}, \varepsilon_s^{[i]}\right) = 0$, $\mathbf{cov}\left(\varepsilon_t^{[i]}, R_{market,\,t}\right) = 0$ (pre všetky indexy $t, s \in \mathrm{T}$). Z predpokladov potom plynie vzťah pre disperziu $\sigma^2_{R[i]\,t}$ výnosu finančného inštrumentu $\sigma^2_{R[i]t} = \beta_i^2 \sigma^2_{market,\,t} + \sigma^2_{\varepsilon[i]t}$ a kovarianciu medzi výnosom *i*-teho a *j*-teho inštrumentu $\mathbf{cov}\left(R_t^{[i]}, R_t^{[j]}\right) = \beta_i \beta_j \sigma^2_{market,\,t}$, kde $\sigma^2_{market,\,t}$ je disperzia výnosu trhového indexu.

Ak sa zvolí zjednodušenie (zobrazenie) $\sigma^2_{R[i]t} \approx \beta_i^2 \sigma^2_{market,\,t}$, ktoré *de facto* znamená, že sa *drasticky* upúšťa od nesystematickej zložky rizika v prospech výpočtového zjednodušenia, je možné štandardný variančno-kovariančný vzťah pre value at risk na hladine α $VaR_\alpha\left(r_{\underline{\Pi}}\right) = -u_{1-\alpha}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}}$ (v ktorom $u_{1-\alpha}$ je α.100%-ný kvantil normovaného normálneho rozdelenia) modifikovať do tvaru

$$VaR_\alpha\left(r_{\underline{\Pi}}\right) = -u_{1-\alpha}\sigma_{market}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\beta})\boldsymbol{\omega}}\,,$$

kde $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ je pseudokovariančná matica koeficientov β, ktorej diagonálne prvky sú β_i^2 a ostatné prvky sú zmiešané $\beta_i\beta_j$ (pre všetky indexy $i \in I$).

Dôvody pre takýto postup možno hľadať nielen v (zachádzajúcej) obľúbenosti faktorov β pri riadení investícií, ale najmä v tom, že namiesto odhadovania $C_2(\underline{N}) + \underline{N} = .5\underline{N}(\underline{N}+1)$ parametrov sa odhaduje iba $\underline{N}$ koeficientov β a 1 disperzia výnosov trhového indexu. Je teoreticky možné odhadovať koeficienty β vlastným regresným modelom, avšak sa ukazuje byť vhodnejšie preberať ich od investičnej agentúry. Treba ale uvážiť skutočnosť, že zanedbaním nesystematickej zložky výnosov finančného inštrumentu vo všeobecnosti platí

$$-u_{1-\alpha}\sigma_{market}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\beta})\boldsymbol{\omega}} \;\le\; -u_{1-\alpha}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}}\,,$$

teda odhady value at risk získané zjednodušenou indexnou verziou variančno-kovariančnej metódy sú tendenčne nižšie (podhodnotené) ako odhady value at risk získané štandardnou variančno-kovariančnou metódou.

Ak sa zjednodušenie nezvolí a pre výnosy finančného inštrumentu sa uvažuje plný vzťah $\sigma^2_{R[i]t} = \beta_i^2 \sigma^2_{market,\,t} + \sigma^2_{\varepsilon[i]t}$, je potrebné získať odhady nielen $\underline{N}+1$ parametrov, ale taktiež dodatočných $\underline{N}$ disperzií $\sigma^2_{\varepsilon[i]t}$ reziduálneho nesystematického komponentu. Následne sa použije pre odhad value at risk vzťah

$$VaR_\alpha\left(r_{\underline{\Pi}}\right) = -u_{1-\alpha}\sigma_{market}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\beta},\boldsymbol{\varepsilon})\boldsymbol{\omega}}\,,$$

v ktorom symbol $\boldsymbol{\Sigma}(\boldsymbol{\beta},\boldsymbol{\varepsilon})$ nahradzuje pseudokovariančnú maticu, ktorá má diagonálne prvky tvaru $\beta_i^2 + \sigma^2_{\varepsilon[i]t} / \sigma^2_{market,\,t}$ a ostané prvky sú $\beta_i\beta_j$ (pre všetky indexy $i \in I$). V tomto prípade zrejme platí

$$-u_{1-\alpha}\sigma_{market}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\beta})\boldsymbol{\omega}} \;\le -u_{1-\alpha}\sigma_{market}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\beta},\boldsymbol{\varepsilon})\boldsymbol{\omega}} \;\overset{?}{\le} -u_{1-\alpha}\sqrt{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}}\,,$$

čo implikuje, že odhady value at risk sprostredkované plnou verziou indexného modelu sú síce kritickejšie, ale nedá sa zaručiť, že riziko nenadhodnocujú.

---

[1] Koeficient β je mierou systematického rizika a v prípade uvedeného indexného modelu vystupuje aj ako ukazovateľ senzitivity sledovaného finančného inštrumentu vzhľadom na trhový výnos.

[2] Model sa veľmi ponáša na Sharpeho jednoduchý indexný model, nezahŕňa však absolútny člen (pozri Sharpe, 1963).

3. Diagnostika modelu value at risk

Z doteraz uvádzaných informácií je zjavné, že kým koncept value at risk je jednoduchý a ľahko pochopiteľný, jeho praktická aplikácia už nie je celkom ideálne naplniteľná a realizuje sa vo forme výstavby modelov za predpokladov, ktoré sa v niektorých prípadoch vzďaľujú od reality. Divergencia týchto predpokladov a zjednodušovanie reality motivujú pri navrhovaní konkrétneho modelu value at risk (vzhľadom na potreby používateľa) použitie diagnostických postupov zameraných na získanie uistenia o spôsobilosti modelu vysvetľovať a merať trhové riziko. V prípade použitia value at risk na regulačné účely v kontexte kapitálovej primeranosti sa diagnostika používaných postupov obligatórne vyžaduje a samotná Nová bazilejská kapitálová dohoda stanovuje ako podmienku na autorizáciu vnútorného modelu value at risk spätné testovanie (*backtesting*) a stresové testovanie (*stress testing*).

3.1 Spätné testovanie

Spätné testovanie spočíva v porovnávaní odhadov value at risk zostrojených *ex ante* so skutočnými výsledkami držby portfólia, tzn. realizovanými ziskami a stratami. Model sa považuje za kvalifikovaný merať (odhadovať) trhové riziko a za dobre nakalibrovaný, ak spĺňa dve podmienky, a síce

- K prekročeniu[3] strát vymedzených odhadom value at risk dochádza v priemere $100.\alpha$ % prípadov.
- Magnitúda (veľkosť) odchýlok odhadov value at risk od realizovaných strát v prípade ich prekročenia nie je veľká.

Kým formulácia prvej podmienky je jednoduchá, v prípade druhej podmienky je otázne, čo možno považovať za významné prekročenie odhadu. Napr. v prípade plnenia prvej podmienky Christoffersen, Hahn a Inoue (1999) nazývajú odhad value at risk eficientným a ich vymedzenie zodpovedá nasledujúcej definícii.

Definícia *Odhad value at risk* $VaR_\alpha(R_{t+1})$ *na hladine* α *konštruovaný v čase* t *na obdobie* $t+1$ *sa nazýva eficientným s ohľadom na dostupnú informáciu* Ψ_t *práve vtedy, len a len vtedy, keď*

$$\mathbf{E}\left[\mathbf{i}\left\{R_{t+1} < -VaR_\alpha(R_{t+1})\right\}\middle|\Psi_t\right] = \alpha,$$

kde $\mathbf{i}\{\bullet\}$ *je indikátorová funkcia.*

Pre oba prípady bola vyvinutá séria štatistických testov a ukazovatele na posúdenie výkonnosti modelu, spomedzi ktorých možno spomenúť

- *testy* (1.) Kupiecov test času do ďalšieho prekročenia, (2.) Kupiecov test nepodmieneného pokrytia založený na binomickom (alternatívnom) rozdelení, (3.) Christoffersenov test podmieneného pokrytia založený na binomickom (alternatívnom) rozdelení a (4.) test nezávislosti výskytu prekročení založený na autokorelačnej funkcii a
- *ukazovatele* (1.) miera priemerného pokrytia, (2.) miera priemerného pokrytia modifikovaná veľkosťou prekročení, (3.) priemerná absolútna percentuálna chyba a (4.) priemerné kvadratické percentuálne skreslenie.

[3] Anglická terminológia je nejednoznačná. Bežne sa v literatúre používajú pojmy *exceedance* (prekročenie), *exception* (výnimka), *violation* (porušenie), *failure* (zlyhanie) a smernica 2006/49/EC kapitálovej primeranosti používa termín *overshooting* (prestrelenie).

Kupiecov test času do ďalšieho prekročenia Test sa zakladá na úvahe, že pri opakovanom odhadovaní value at risk pri zadanej hladine α je počet obchodných dní $\tilde{T}$, ktoré uplynuli medzi dvoma pozorovanými prekročeniami, náhodnou premennou s geometrickým rozdelením $Geom(\alpha)$. Pri každom pozorovanom prekročení sa testuje, či skutočná miera pokrytia $p = \mathbf{E}\left[\mathbf{i}\{R_{t+1} < -VaR_\alpha(R_{t+1})\}|\Psi_t\right]$ zodpovedá požadovanej miere α, t. j.

$$\mathrm{H}_0 : p = \alpha \qquad vs. \qquad \mathrm{H}_A : p \neq \alpha,$$

čo je ekvivalentné s formuláciou

$$\mathrm{H}_0 : \tilde{T} \sim Geom(\alpha) \qquad vs. \qquad \mathrm{H}_A : non\,\mathrm{H}_0. \qquad [1]$$

Dá sa ukázať, že rovnomerne najsilnejším a nevychýleným kritériom je štatistika založená na vierohodnostnom pomere

$$LR_{**} = -2\ln\left[(1-\alpha)^{\tilde{T}-1}\alpha\right] + 2\ln\left[(1-1/\tilde{T})^{\tilde{T}-1}(1/\tilde{T})\right],$$

ktorá má pri platnosti nulovej hypotézy asymptoticky rozdelenie $\chi^2(1)$.

Test má veľmi nízku silu a dopĺňa sa preto ďalšími testami, najčastejšie testami pokrytia, ktoré sa zakladajú na binomickom (alternatívnom) rozdelení.

V definícii bola použitá indikátorová funkcia $\mathbf{i}\{\bullet\}$, ktorou sa v podstate zavádza (na nejakom $(\Omega, \mathsf{F}, \mathsf{P})$) náhodná premenná

$$\mathbf{i}_t = \begin{cases} 1, & \text{ak } R_t < -VaR_\alpha(R_t), \\ 0, & \text{ak } R_t \geq -VaR_\alpha(R_t). \end{cases}$$

Náhodná premenná $\mathbf{i}_t$ vyjadruje, či v čase (obchodnom dni) t došlo k prekročeniu, alebo nie (a teda v súlade s matematickým zadaním nadobúda hodnotu 1, ak v čase t došlo k prekročeniu, a hodnotu 0, ak v čase t k prekročeniu nedošlo). Tomu pri opakovaní odhadu value at risk počas T obchodných dní a jeho porovnaní so skutočným výsledkom obchodného portfólia zodpovedá postupnosť

$$(\mathbf{i}_t)_{t=1}^{t=T} = \mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_T$$

tvorená iba jednotkami a nulami.

Pri dobrom nakalibrovaní modelu value at risk má zrejme náhodná premenná $\mathbf{i}_t$ rozdelenie $A(\alpha)$, kde α je špecifikovaná hladina odhadu value at risk, čo znamená, že ak počas T obchodných dní došlo k prekročeniu odhadu value at risk X-krát, potom má náhodná premenná

$$X = \sum_{t=1}^{t=T} \mathbf{i}_t$$

rozdelenie $bi(T,\alpha)$. Na tomto sa zakladá **Kupiecov test proporcie zlyhaní,** ktorým sa testuje, či skutočná tendencia modelu value at risk k zlyhaniu $p = \mathbf{E}\left[\mathbf{i}\{R_{t+1} < -VaR_\alpha(R_{t+1})\}|\Psi_t\right]$ je totožná s navolenou hladinou α, tzn.

$$\mathrm{H}_0 : p = \alpha \qquad vs. \qquad \mathrm{H}_A : p \neq \alpha,$$

resp., čo je ekvivalentné s testovaním

$$\mathrm{H}_0 : (\forall t = 1,\ldots,T)(\mathbf{i}_t \sim A(\alpha)) \qquad vs. \qquad \mathrm{H}_A : non\,\mathrm{H}_0. \qquad [2]$$

Testu zodpovedá testovacie kritérium založené na vierohodnostnom pomere

$$LR_{uc} = -2\ln\left[(1-\alpha)^{T-x}\alpha^x\right] + 2\ln\left[(1-x/T)^{T-x}(x/T)^x\right],$$

ktoré má za platnosti H_0 asymptoticky rozdelenie $\chi^2(1)$.

Nedostatkom testu je jeho nízka sila (rastúca s klesajúcou hladinou α odhadu value at risk), ktorú možno zvýšiť zväčšením rozsahu výberovej vzorky: teda opakovaním odhadu value at risk.

Chirstoffersen (1998) upozornil na to, že testovanie výlučne iba predpokladu $X \sim bi(T,\alpha)$ na overenie predikčnej schopnosti predpovedí nepostačuje a že ak model je dobrý, musia byť

premenné $\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_T$ v súčte nielen identicky rozdelené ako $bi(T,\alpha)$, ale musia byť tiež sériovo nezávislé, t. j.

$$(\forall t = 1,\ldots,T)(\mathbf{i}_t \sim_{iid} A(\alpha)) .$$

Testovanie [2] Christoffersen označil ako test nepodmieneného pokrytia (*unconditional coverage test: uc*) a test

$$\mathrm{H}_0 : (\forall t = 1,\ldots,T)(\mathbf{i}_t \sim_{iid} A(\alpha)) \qquad vs. \qquad \mathrm{H}_A : non\,\mathrm{H}_0 , \qquad [3]$$

pomenoval **testom podmieneného pokrytia** (*conditional coverage test: cc*). Na testovanie odporučil použitie štatistiky

$$LR_{cc} = LR_{uc} + LR_{ind} ,$$

ktorá spoločne testuje plnenie nepodmieneného pokrytia (LR_{uc}) a nezávislosť (LR_{ind}). Štatistika na testovanie nezávislosti vierohodnostným pomerom má tvar

$$LR_{ind} = -2\ln\left[\left(1-\hat{p}_{1|0}\right)^{T_{0|0}} \hat{p}_{1|0}^{T_{1|0}} \left(1-\hat{p}_{1|1}\right)^{T_{0|1}} \hat{p}_{1|1}^{T_{1|1}}\right] + 2\ln\left[\left(1-\alpha\right)^{T_{0|0}+T_{0|1}} \alpha^{T_{1|0}+T_{1|1}}\right],$$

kde $T_{i|j}$ pre $i, j \in \{0,1\}$ je počet pozorovaní, v ktorých realizácia $\mathbf{i}_i$ nasledovala po $\mathbf{i}_j$ (napr. $T_{1|0}$ zastupuje počet obchodných dní, pri ktorých došlo k prekročeniu bezprostredne po obchodnom dni, v ktorom k prekročeniu nedošlo) a $\hat{p}_{i|j}$ pre $i, j \in \{0,1\}$ sú zodpovedajúce odhady pravdepodobnosti realizácie $T_{i|j}$ (tzn. $\hat{p}_{1|0} = T_{1|0}/(T_{0|0}+T_{1|0})$ a $\hat{p}_{1|1} = T_{1|1}/(T_{1|1}+T_{0|1})$)). Štatistika LR_{ind} má pri sériovej nekorelovanosti H_0 asymptoticky rozdelenie $\chi^2(1)$, čo implikuje, že

$$LR_{cc} \xrightarrow[T]{\mathrm{H}_0} \chi^2(2) .$$

Ďalší **test sériovej nezávislosti prekročení** je založený na autokorelačnej funkcie konštruovanej z postupnosti $(\mathbf{i}_t)_{t=1}^{t=T} = \mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_T$. Test je formulovaný

$$\mathrm{H}_0 : (\forall t = 1,\ldots,T)(\mathbf{i}_t \sim indep) \qquad vs. \qquad \mathrm{H}_A : non\,\mathrm{H}_0 ,$$

resp., v logike autokorelačnej funkcie

$$\mathrm{H}_0 : (\forall t,s = 1,\ldots,T)(t \neq s)(\rho\{\mathbf{i}_t, \mathbf{i}_s\} = 0) \qquad vs. \qquad \mathrm{H}_A : non\,\mathrm{H}_0 . \qquad [4]$$

Pôvodcovia testu Boudoukh, Richardson a Whitelaw (1998) odporučili overovať autokorelácie až piateho rádu (tzn. oneskorenie zodpovedajúce jednému obchodnému týždňu o piatich obchodných dňoch).Na testovanie sa štandardne používa Ljungova-Boxova Q štatistika.

Testovanie sa dopĺňa ukazovateľmi *neskreslenia*, najčastejšie mierou priemerného pokrytia alebo mierou priemerného pokrytia modifikovanou veľkosťou prekročení, a ukazovateľmi skreslenia, napr. priemernou absolútnou percentuálnou chybou a priemerným kvadratickým percentuálnym skreslením.

Bolo už uvedené, že v prípade optimality modelu value at risk musí byť splnené $\mathbf{E}\left[\mathbf{i}\{R_{t+1} < -VaR_\alpha(R_{t+1})\} \,|\Psi_t\right] = \alpha$. Miesto štatistického testovania je možné intuitívne porovnávať očakávanú (teoretickú) mieru pokrytiu so skutočnou (empirickou) mierou pokrytia. Štandardný **ukazovateľ priemerného pokrytia** (neskreslenia: *unbiasedness*) možno pri odhadovaní value at risk počas T obchodných dní definovať

$$MC = \tfrac{1}{T}\sum_{t=1}^{t=T} \mathbf{i}_t$$

a porovnávať s očakávaním α.

Lopez (1998) navrhuje do indikátorovej funkcie inkorporovať magnitúdu prekročení v tvare

$$\mathbf{i}_t^{\circ} = \begin{cases} 1+\left[|R_t| - VaR_\alpha(R_t)\right]^2, & \text{ak } R_t < -VaR_\alpha(R_t), \\ 0, & \text{ak } R_t \geq -VaR_\alpha(R_t). \end{cases}$$

Ukazovateľ **priemerného pokrytia modifikovaný veľkosťou prekročení** na základe uvedeného potom nadobúda podobu

$$MC_{mag} = \frac{1}{T}\sum_{t=1}^{t=T} \mathbf{i}_t^{\circ}$$

a je porovnateľný s teoretickou vzorovou hodnotou (*benchmark value*) skonštruovateľnou pre konkrétny prípad.

Sinha a Chamú (2000) na hodnotenie kvality predpovedí navrhli ukazovateľ kĺzavej **priemernej absolútnej percentuálnej chyby** (*mean absolute percentage error*), ktorý porovnáva počet pozorovaných prekročení odhadu value at risk s očakávaným počtom za dlhšie obdobie odhadovania T obchodných dní. Zvolí sa referenčná dĺžka kĺzavej časti K obchodných dní ($K < T$), čo znamená že celkovému počtu T obchodných dní zodpovedá $T-K+1$ kĺzavých častí. Ku každej kĺzavej časti $i \in \{1,\ldots,T-K+1\}$ sa zistí počet prekročení $X_i^{(K)} = \sum_{t=i}^{t=i+K-1} \mathbf{i}_t$ a porovná sa s očakávaným počtom prekročení αK. Autori definujú ukazovateľ pre rozsah kĺzavej časti K ako

$$MAPE_k = \frac{1}{T-K+1}\sum_{t=1}^{t=T-K+1}\left|\frac{X_i^{(K)} - \alpha K}{K}\right|.$$

Autori zvolili na účely porovnávania metód value at risk $T = 1187$ a $K = 500$. Na účely spätného testovania by bolo zjavne užitočné voliť $T = 250$ a napr. $K = 10$.

Je možné adaptovať ukazovateľ **priemerného kvadratického relatívneho skreslenia** (*root mean squared relative bias*) použitého v Hendricks (1996) na porovnávanie rôznych metodológii a modifikovať ho napríklad do tvaru

$$MSPB = \sqrt{\frac{1}{\sum_{t=1}^{t=T}\mathbf{i}_t}\sum_{t=1}^{t=T}\mathbf{i}_t\left(\frac{VaR_\alpha(R_t) - |R_t|}{|R_t|}\right)^2},$$

ktorý potom reprezentuje kvadratický priemer odchýlok odhadov value at risk od skutočných strát pri nastaní prekročenia. Zrejme je vhodné klásť $T = 250$.

Žiada sa ešte dodať, že tieto indikátory sú používané nielen v rámci spätného testovania, ale tiež pri porovnávaní a vyhodnocovaní viacerých metodológií.

3.2 Stresové testovanie

Pri stresovom testovaní je skonštruovaný model value at risk prehodnotený cez scenáre zadané používateľom s cieľom zistiť správanie sa trhovej hodnoty portfólia pri neočakávanom a menej pravdepodobnom vývoji rizikových faktorov. Význam stresového testovania pri ohodnocovaní modelu value at risk priamo vyplýva zo skutočnosti, že modely value at risk sú spravidla stavané na očakávaných a (vysoko) pravdepodobných zmenách determinujúcich rizikových faktorov. Evidencia z minulosti[4] ukazuje, že zvažovanie zriedkavých a extrémnych udalostí má svoju relevanciu v riadení finančného trhového rizika, preto je vystavovanie modelu potenciálnym trhovým šokom požadované i pri regulácii finančných inštitúcií. Štandardný model value at risk poskytuje informáciu o správaní sa trhového portfólia za normálnych (nepriaznivých) podmienok a stresové testovanie informáciu rozširuje o jeho reakciu za abnormálnych (ale stále nastateľných) podmienok.

Možno diferencovať tri základné typy scenárov (Aragonés, Blanco, Dowd, 2001; Cipra, 2002): historické simulácie, normované scenáre a mechanické stresové testy.

[4] Napr. akciová kríza na amerických burzách asociovaná s čiernym pondelkom 19. 10. 1987, rozpad európskeho menového mechanizmu v rokoch 1992-1993, kríza amerického dlhopisového trhu a mexická kríza z roku 1994, finančná kríza z roku 1998 spôsobená bankrotom LTCM a i.

- Pri historických simuláciách (*actual extreme events*) sa opakuje vývoj minulých nepriaznivých udalostí (teda sa simuluje vývoj rizikových faktorov počas krízy).
- Pri normovaných scenároch (*stylized scenarios*) sa simuluje vývoj vytypovaných rizikových faktorov, ktorý sa štandardne osvedčil v praxi (napr. paralelný posun výnosovej krivky, pád akciového indexu nad bežné medze, katastrofická zmena devízových kurzov ap.).
- Pri mechanických stresových testoch (*mechanical-search scenarious*) sa v systéme nastavia hranice, v ktorých sa hodnoty rizikových faktorov môžu pohybovať, a kombinujú sa rôzne stavy rizikových faktorov a identifikuje najnepriaznivejšie výsledky.

Nutné je podotknúť, že i stresové testy majú svoje slabé miesta. V zásade sú problémy tri: (1.) Scenáre sú volené subjektívne a ich fruktifikovateľnosť pri riadení rizika závisí od skúseností zadávateľa. (2.) Nie je možné zmysluplne pravdepodobnostne ohodnotiť jednotlivé varianty radikálneho vývoja, následkom čoho je otázne, ktorý scenár je pravdepodobnejší a ktorý scenár by mal byť seriózne zvažovaný. Po ukončení stresového testovanie používateľ disponuje číslami s abstraktnou vypovedateľnosťou. (3.) Obyčajne nemožno štatisticky alebo vedecky vyhodnotiť úplnosť alebo vierohodnosť poskytnutých informácií

Bibliografia

ARAGONÉS, Jose R., BLANCO, Carlos, DOWD, Kevin 2001. *Incorporating Stress Tests into Market Risk Modeling.* In: *Derivatives Quarterly.* 2001, č. 7 (jar). S. 44-49.

BERKOWITZ, Jeremy, O' BRIEN James 2002. *How Accurate Are Value-at-Risk Models at Commercial Banks?* In: *The Journal of Finance.* 2002, č. 3 (jún), roč. 57. S 1093-1111.

BOĎA, Martin 2006a. *Value at risk I. Value at risk ako miera rizika, alternatívy, nedostatky a regulačný aspekt.* In: *Forum Statisticum Slovacum.* 2006, č. 4, roč. 2. S. 15-24.

BOĎA, Martin 2006b. *Value at risk II. Základné prístupy k modelovaniu.* In: *Forum Statisticum Slovacum.* 2006, č. 5, roč. 2. 10 s.

BOĎA, Martin 2006c. *Value at risk: on the naïvete of the correct specification of a volatility model.* In: *Forum Statisticum Slovacum.* 2006, č. 5, roč. 2. S. 176-180.

BOUDOUKH, Jacob, RICHARDSON, Matthew P., WHITELAW, Robert 1997. *The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk.* [Acrobat ® pdf online]. National Bureau of Economic Research 1997. [Cit. 23. 10. 2006]. Dostupné na World Wide Web: <http://papers.ssrn.com/sol3/Delivery.cfm/ 9711202.pdf>.

CASSIDY, Colleen, GIZYCKI, Marianne 1997. *Measuring Traded Market Risk: Value-at-risk and Backtesting Techniques* . In: *Research Discussion Papers of Reserve Bank of Australia.* 1997, č. 9708. 37 s.

CIPRA, Tomáš 2002. *Kapitálová přiměřenost ve financích a solventnost v pojišťovnictví.* Praha: Ekopress 2002. 272 s. ISBN 80-86119-54-8.

FABOZZI, Frank J., MODIGLIANI, Franco 2003. *Capital Markets. Institutions and Instruments.* Tretie vydanie. Upper Saddle River [USA]: Prentice Hall 2003. 644 s. ISBN 0-13-045754-X.

HÄRDLE, Wolfgang, HLÁVKA, Zdeněk, STAHL, Gerhard 2006. *On the appropiateness of inappropriate VaR models.* In: *SFB 649 Economic Risk.* Discussion Paper 2006-003. Berlin: Physica 2006. ISBN 0002-6018.

HENDRICKS, Darryll 1996. *Evaluation of Value-at-Risk Models Using Historical Data.* In: *Economic Policy Review.* Federal Reserve Bank of New York. 1996, č. 2 (apríl), roč. 2. S. 39-69.

CHRISTOFFERSEN, Peter F. 1998. *Evaluating Interval Forecasts.* In: *International Economic Review.* 1998, č. 39. S. 841-862.

CHRISTOFFERSEN, Peter F., HAHN, Jinyong, INOUE, Atsushi 1999. *Testing, Comparing, and Combining Value at Risk Measures.* In: *Center for Financial Institutions Working Papers.* The Wharton School. University of Pennsylvania. 1999, č. 99-44, 25 s.

LOPEZ, Jose A. 1998. *Methods for Evaluating Value-at-Risk Estimates.* In: *Economic Policy Review.* Federal Reserve Bank of New York. 1998, č. 3 (október), roč. 4. S. 119-124.

SINHA, Tapen, CHAMÚ, Francisco 2000. *Comparing Different Methods of Calculating Value at Risk.* [Acrobat ® pdf online]. Nottingham [UK]: Nottingham University Business School 2000. [Cit. 29. 09. 2006]. Dostupné na World Wide Web: <http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID706582_code58705.pdf>.

Oblasti necitlivosti pro parametry střední hodnoty ve smíšeném lineárním regresním modelu s podmínkami typu I a s nimi spojené výpočetní problémy

Hana Boháčová[1], Jana Heckenbergerová[2]

Abstract: The article deals with the fixed effects parameters non-sensivity region in mixed linear regression model with constraints. The non-sensivity region is derived and Moore-Penrose pseudoinverse matrix computation problem is described.

Key words: mixed linear regression model with constraints, non-sensitivity region for fixed effects parameters in mixed linear regression model without constraints and with constraints, Moore-Penrose matrix pseudoinverse computation.

1.Úvod

Cílem je vyšetřit závislost změn odhadů parametrů střední hodnoty $\beta_1, \beta_2, ..., \beta_k$ na malých změnách variančních komponent $\theta_1, \theta_2, ..., \theta_r$ ve smíšeném lineárním regresním modelu s podmínkami typu I

$$Y \sim_n (X\beta, \Sigma_\theta) \tag{1}$$

$$b + B\beta = 0 \tag{2}$$

Parametr β je nepřímo měřitelný a musí navíc vyhovovat podmínce (2). Předpokládáme, že matice *X*, která je typu (n,k) je plné hodnosti ve sloupcích, matice B je typu (q,k) a má plnou řádkovou hodnost. Předpokládáme model s *r* variančními komponentami $\theta_1, ..., \theta_r$, tedy

$$\operatorname{var} Y = \Sigma_\theta = \sum_{i=1}^{r} \theta_i V_i, \tag{3}$$

kde $V_1, ..., V_r$ jsou známé symetrické matice, přičemž musí platit, že Σ_θ je alespoň pozitivně semidefinitní. Souhrnně označíme $\theta = (\theta_1, \theta_2, ..., \theta_r)'$

2. Oblast necitlivosti pro parametry střední hodnoty ve smíšeném lineárním modelu bez podmínek

Oblast necitlivosti odvodíme nejprve pro model (1) bez podmínek (použitý postup je analogický postupu uvedenému v [Kubáček, Kubáčková, 2000]), pak pomocí transformace modelu s podmínkami na model bez podmínek stanovíme oblast necitlivosti pro model s podmínkami.

Je známo, že nejlepší lineární nestranný odhad parametru $\beta = (\beta_1, \beta_2, ..., \beta_k)'$ je v modelu (1) tvaru

$$\hat{\beta}(\theta) = (X' \Sigma_\theta^{-1} X)^{-1} X' \Sigma_\theta^{-1} Y. \tag{4}$$

Změníme-li θ o nějaké malé $\delta\theta$, můžeme pomocí diferenciálu vyjádřit

$$\hat{\beta}(\theta + \delta\theta) \approx \hat{\beta}(\theta) + \frac{\partial \hat{\beta}(\theta)}{\partial \theta'} \delta\theta. \tag{5}$$

Protože

[1] Hana Boháčová, Fakulta ekonomicko-správní, Univerzita Pardubice
[2] Jana Heckenbergerová, Ústav elektrotechniky a informatiky, Univerzita Pardubice

$$\frac{\partial\hat{\beta}(\theta)}{\partial\theta_i} = -\left(X'\Sigma_\theta^{-1}X\right)^{-1}X'\Sigma_\theta^{-1}V_i\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right), \tag{6}$$

podle (5) a (6) tedy

$$\hat{\beta}(\theta+\delta\theta) \approx \hat{\beta}(\theta) - \sum_{i=1}^{r}\left(X'\Sigma_\theta^{-1}X\right)^{-1}X'\Sigma_\theta^{-1}V_i\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right)\delta\theta_i\ . \tag{7}$$

Pro varianční matici odhadu $\hat{\beta}(\theta+\delta\theta)$ platí

$$Var_\theta\left[\hat{\beta}(\theta+\delta\theta)\right] = Var_\theta\hat{\beta}(\theta) + \left(X'\Sigma_\theta^{-1}X\right)^{-1}X'\Sigma_\theta^{-1}\Sigma_{\delta\theta}\left(M_X\Sigma_\theta M_X\right)^{+}\Sigma_{\delta\theta}\Sigma_\varsigma^{-1}X\left(X'\Sigma_\theta^{-1}X\right)^{-1}, \tag{8}$$

kde A^+ značí Moore-Penroseovu pseudoinverzi matice A a M_A je matice ortogonální projekce na vektorový prostor kolmý k vektorovému prostoru generovanému sloupci matice A, pro výpočet se používá vztahu $M_A = I - AA^+$.

Mějme libovolný vektor $h \in R^k$ a označme $L_h' = h'\left(X'\Sigma_\theta^{-1}X\right)^{-1}X'\Sigma_\theta^{-1}$. Pak pro lineární kombinaci složek $\hat{\beta}(\theta+\delta\theta)$ platí

$$h'\hat{\beta}(\theta+\delta\theta) \approx L_h'Y - \left(L_h'V_1\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right),\ldots,L_h'V_r\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right)\right)\delta\theta \tag{9}$$

Označíme-li dále $\xi = \begin{pmatrix} L_h'V_1\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right) \\ \ldots \\ L_h'V_r\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right)\end{pmatrix}$, pak $\xi \sim_r (0, Var(\xi))$. Matici $Var(\xi)$ označme W_h, pak

$$\{W_h\}_{i,j} = \mathrm{cov}\left(L_h'V_i\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right), L_h'V_j\Sigma_\theta^{-1}\left(Y - X\hat{\beta}(\theta)\right)\right) = L_h'V_i\left(M_X\Sigma_\theta M_X\right)^{+}V_jL_h\ . \tag{10}$$

Protože $h'\hat{\beta}(\theta)$ a ξ jsou stochasticky nezávislé,

$$Var_\theta\left[h'\hat{\beta}(\theta+\delta\theta)\right] = Var_\theta\left[h'\hat{\beta}(\theta)\right] + \delta\theta'W_h\delta\theta\ . \tag{11}$$

Oblast necitlivosti budeme stanovovat tak, aby pro směrodatné odchylky platilo

$$\sqrt{Var_\theta\left[h'\hat{\beta}(\theta+\delta\theta)\right]} \le (1+\varepsilon)\sqrt{Var_\theta\left[h'\hat{\beta}(\theta)\right]}, \tag{12}$$

tedy

$$\frac{\delta\theta'W_h\delta\theta}{Var_\theta\left[h'\hat{\beta}(\theta)\right]} \le 2\varepsilon + \varepsilon^2\ . \tag{13}$$

Protože ε můžeme volit libovolně malé, lze $2\varepsilon + \varepsilon^2$ na pravé straně nerovnosti (13) přibližně nahradit 2ε. Pro oblast necitlivosti tedy požadujeme splnění nerovnosti

$$\delta\theta'W_h\delta\theta \le 2\varepsilon Var_\theta\left[h'\hat{\beta}(\theta)\right]. \tag{14}$$

Výsledná oblast necitlivosti pro parametr β je elipsoid

$$N_h = \left\{\delta\theta : \delta\theta \in M(W_h), \delta\theta'W_h\delta\theta \le 2\varepsilon Var_\theta\left[h'\hat{\beta}(\theta)\right]\right\}. \tag{15}$$

$M(W_h)$ značí prostor generovaný sloupci matice W_h.

3. Oblast necitlivosti pro parametry střední hodnoty ve smíšeném lineárním modelu s podmínkami typu I

Model (1) s podmínkami (2) nejprve převedeme na model bez podmínky.(Použitá transformace viz [Kubáček-Kubáčková, 2000]). Zvolíme nějaký počáteční vektor β_0 vyhovující podmínkám (2). Pak pro všechny ostatní vektory β splňující podmínky (2) platí

$$\beta = \beta_0 + K_B\gamma, \gamma \in R^{k-q}, \tag{16}$$

kde K_B je matice typu $(k, k-q)$ taková, že $BK_B = 0$. Z modelu (1) vyplývá, že

$$(Y - X\beta_0) \sim_n (XK_B\gamma, \Sigma_\theta). \tag{17}$$

Model (1) s podmínkami (2) jsme tedy převedli na model (6) bez podmínek.

Označíme-li $\hat{\hat{\beta}}$ odhad parametru β v modelu (1), (2), pak oblast necitlivosti pro parametr β je vzhledem k (15) elipsoid

$$N_h = \left\{\delta\theta : \delta\theta \in M(W_{h,\beta}), \delta\theta' W_{h,\beta}\delta\theta \le 2\varepsilon Var_\theta\left[h'\hat{\hat{\beta}}(\theta)\right]\right\} =$$

$$= \left\{\delta\theta : \delta\theta \in M(W_{h,\beta}), \delta\theta' W_{h,\beta}\delta\theta \le \left(M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right)^+ h\right\}, \tag{18}$$

Kde $\left\{W_{h,\beta}\right\}_{i,j} = h'K_B\left(K_B'X'\Sigma_\theta^{-1}XK_B\right)^{-1}K_B'X'\Sigma_\theta^{-1}V_i\left(M_{XK_B}\Sigma_\theta M_{XK_B}\right)^+ V_j\Sigma_\theta^{-1}XK_B \cdot$

$\cdot\left(K_B'X'\Sigma_\theta^{-1}XK_B\right)^{-1}K_B'h.$

4. Problémy spojené s výpočtem oblastí necitlivosti

Při výpočtu oblasti necitlivosti je největším problémem stanovení pseudoinverzní matice $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^+$ .Maple i Matlab sice mají příkazy pro výpočet Moore-Penroseovy pseudoinverze , při kontrole, zda se skutečně jedná o hledanou Moore-Penroseovu pseudoinverzní matici ovšem narazíme na to, že vypočtenou matici můžeme sice přibližně považovat za pseudoinverzi, ovšem nikoliv Moore-Penroseovu. Je proto třeba hledat jinou cestu, jak Moore-Penroseovu pseudoinverzi získat co nejpřesněji.

Obecně musí pseudoinverzní matice A^- k matici A splňovat $AA^-A = A$, tedy

$$AA^-A - A = 0 \tag{19}$$

Moore-Penroseova pseudoinverze A^+ musí kromě $AA^+A = A$ vyhovovat mj. ještě druhé rovnosti $A^+AA^+ = A^+$, tedy celkově musí platit [viz Anděl, 2005]

$$\begin{aligned} AA^+A - A &= 0 \\ A^+AA^+ - A^+ &= 0 \end{aligned}. \tag{20}$$

Vraťme se k hledání pseudoinverze $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^+$. Nechť například

$$\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right] = \begin{pmatrix} 2{,}168267 & 0{,}046133 & -0{,}138400 & -0{,}461333 \\ 0{,}046133 & 3{,}967467 & -0{,}692000 & -1{,}476267 \\ -0{,}138400 & -0{,}692000 & 4{,}567201 & -3{,}044800 \\ -0{,}461333 & -1{,}476267 & -3{,}044800 & 3{,}137067 \end{pmatrix}$$

Maple 10 určí Moore-Penroseovu pseudoinverzi této matice jako

$$\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^{+}=\begin{pmatrix}-0{,}137\cdot10^{9} & -0{,}274\cdot10^{9} & -0{,}411\cdot10^{9} & -0{,}548\cdot10^{9}\\ -0{,}274\cdot10^{9} & -0{,}548\cdot10^{9} & -0{,}822\cdot10^{9} & -0{,}110\cdot10^{10}\\ -0{,}411\cdot10^{9} & -0{,}822\cdot10^{9} & -0{,}124\cdot10^{10} & -0{,}164\cdot10^{10}\\ -0{,}548\cdot10^{9} & -0{,}110\cdot10^{10} & -0{,}164\cdot10^{10} & -0{,}219\cdot10^{10}\end{pmatrix}.$$

Budeme-li ověřovat vztahy (20), vyjdou nám matice

$$\begin{pmatrix}-0{,}009 & 0{,}328 & 0{,}388 & -0{,}452\\ 0{,}650 & 0{,}014 & -0{,}042 & -0{,}138\\ 1{,}338 & 0{,}461 & 9{,}826 & -7{,}935\\ -1{,}891 & 1.337 & -8{,}996 & 6.551\end{pmatrix}$$

a

$$\begin{pmatrix}0{,}11\cdot10^{9} & 0{,}22\cdot10^{9} & 0{,}33\cdot10^{9} & 0{,}44\cdot10^{9}\\ 0{,}11\cdot10^{9} & 0{,}22\cdot10^{9} & 0{,}33\cdot10^{9} & 0{,}44\cdot10^{9}\\ -0{,}15\cdot10^{9} & -0{,}30\cdot10^{9} & -0{,}45\cdot10^{9} & -0.60\cdot10^{9}\\ -0{,}41\cdot10^{9} & -0{,}82\cdot10^{9} & -0{,}12\cdot10^{10} & -0{,}16\cdot10^{10}\end{pmatrix}.$$

Tedy rozhodně ne nulové matice, které by podle (20) vyjít měly.
Paradoxem je, že přestože má matice $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]$ hodnost 3, tedy není regulární, vypočítá Maple matici k ní inverzní:

$$\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^{-1}=\begin{pmatrix}-0{,}199\cdot10^{9} & -0{,}398\cdot10^{9} & -0{,}596\cdot10^{9} & -0{,}795\cdot10^{9}\\ -0{,}398\cdot10^{9} & -0{,}795\cdot10^{9} & -0{,}119\cdot10^{9} & -0{,}159\cdot10^{10}\\ -0{,}596\cdot10^{9} & -0{,}119\cdot10^{10} & -0{,}179\cdot10^{10} & -0{,}239\cdot10^{10}\\ -0{,}795\cdot10^{9} & -0{,}159\cdot10^{10} & -0{,}239\cdot10^{10} & -0{,}318\cdot10^{10}\end{pmatrix}.$$

Dosadíme-li tuto matici do vztahů (20),dostaneme

$$\begin{pmatrix}-0{,}392 & -2{,}957 & 1{,}647 & 0{,}341\\ -1{,}522 & -0{,}553 & -22{,}006 & 17{,}162\\ -3{,}137 & -29{,}387 & -10{,}242 & 23{,}159\\ -0{,}185 & 17{,}761 & 10{,}241 & -16{,}516\end{pmatrix}$$

a

$$\begin{pmatrix}0{,}22\cdot10^{9} & 0{,}44\cdot10^{9} & 0{,}66\cdot10^{9} & 0{,}87\cdot10^{9}\\ 0{,}26\cdot10^{9} & 0{,}52\cdot10^{9} & 0{,}78\cdot10^{9} & 0{,}10\cdot10^{10}\\ 0{,}40\cdot10^{9} & 0{,}80\cdot10^{9} & 0{,}12\cdot10^{10} & 0{,}16\cdot10^{10}\\ -0{,}24\cdot10^{10} & -0{,}48\cdot10^{10} & -0{,}72\cdot10^{10} & -0{,}95\cdot10^{10}\end{pmatrix}.$$

Řádově se tyto matice liší od nulové podobně jako matice počítané při ověřování matice pseudoinverzní. Pokud použijeme některou z matic $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^{+}$, $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^{-1}$ ve výpočtu oblasti necitlivosti (18), dostaneme nesmyslné výsledky, je proto třeba hledat jinou cestu, jak Moore-Penroseovu pseudoinverzi získat. Jako nejlepší se ukazuje počítat A^{+} následujícím způsobem:

Nechť matice A typu (m,n) má hodnost r. Pak můžeme najít její skeletní rozklad $A=BC$, kde matice B je typu (m,r) a plné sloupcové hodnosti r a matice C je typu (r,n) a má plnou řádkovou hodnost r. Matici A^+ dostaneme jako

$$A^+ = C'(CC')^{-1}(B'B)^{-1}B'.$$

Použijeme-li tento postup pro matici $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]$, dostaneme

$$\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^+ = \begin{pmatrix} 0{,}434 & -0{,}048 & -0{,}048 & -0{,}048 \\ -0{,}048 & 0{,}193 & -0{,}048 & -0{,}048 \\ -0{,}048 & -0{,}048 & 0{,}112 & -0{,}048 \\ -0{,}048 & -0{,}048 & -0{,}048 & 0{,}072 \end{pmatrix}.$$

Ověříme-li pro tuto matici splnění rovností (20), vyjdou matice

$$\begin{pmatrix} -0{,}3\cdot10^{-8} & -0{,}65\cdot10^{-8} & -0{,}82\cdot10^{-8} & 0{,}102\cdot10^{-7} \\ -0{,}9\cdot10^{-9} & 0{,}4\cdot10^{-8} & -0{,}3\cdot10^{-8} & 0{,}1\cdot10^{-8} \\ 0{,}12\cdot10^{-8} & -0{,}14\cdot10^{-8} & 0{,}2\cdot10^{-8} & -0{,}1\cdot10^{-8} \\ 0{,}3\cdot10^{-9} & 0{,}1\cdot10^{-8} & 0{,}2\cdot10^{-8} & -0{,}3\cdot10^{-8} \end{pmatrix}$$

a

$$\begin{pmatrix} -0{,}6\cdot10^{-9} & -0{,}53\cdot10^{-9} & -0{,}42\cdot10^{-9} & 0{,}7\cdot10^{-9} \\ -0{,}6\cdot10^{-9} & 0{,}3\cdot10^{-9} & -0{,}9\cdot10^{-10} & -0{,}3\cdot10^{-10} \\ 0{,}11\cdot10^{-9} & -0{,}6\cdot10^{-10} & 0{,}1\cdot10^{-9} & -0{,}8\cdot10^{-10} \\ -0{,}1\cdot10^{-10} & 0{,}5\cdot10^{-10} & 0{,}8\cdot10^{-10} & -0{,}1\cdot10^{-9} \end{pmatrix}.$$

Tyto matice už mají k nulové matici velmi blízko, proto takto vypočítanou $\left[M_{B'}\left(X'\Sigma_\theta^{-1}X\right)M_{B'}\right]^+$ lze použít ve vztahu (18) pro výpočet oblasti necitlivosti. Nevýhodou je, že tento postup hledání pseudoinverzní matice je poněkud zdlouhavý. To je však plně vyváženo rozumnými výsledky, které ze vztahu (18) dostaneme.

5. Literatura

ANDĚL, J. 2005 Základy matematické statistiky. Praha: Matfyzpress, 2005

KUBÁČEK, L. – KUBÁČKOVÁ, L. 2000 Statistika a metrologie. Olomouc: Univerzita Palackého v Olomouci – vydavatelství, 2000.

Adresa autorů:
Mgr. Hana Boháčová
Fakulta ekonomicko-správní, Univerzita Pardubice
Studentská 84, 532 10 Pardubice
Hana.Bohacova@upce.cz

Mgr. Jana Heckenbergerová
Ústav elektrotechniky a informatiky, Univerzita Pardubice
Studentská 95, 532 10 Pardubice
Jana.Heckenbergerova@upce.cz

Demographic Prognosis of the Future Development of the Number of Information Scientists in the Czech Republic

Fiala Tomáš

Abstract:
This paper brings the prognosis of the development of the number and of the age structure of information scientists in the Czech Republic. The estimate of the initial age structure is given in the first chapter. The second chapter is dedicated to the prognosis on the following few next years.

Key words:
information scientist, age structure, population projection

1. Estimate of the age structure of information scientists in 2006

For carrying out demographic analyses it is necessary to know or to estimate not only the present total number of information scientists, but also their age structure, as accurately as possible, preferably according to age units. It would be ideal to know also sex structure, but data of this type are not available and we shall not therefore differentiate the structure by sex.

The only available data on the composition of information scientists were data from the expert survey for IT professions (see Table 1).

Table 1. Age structure of employees in IT professions in the expert survey

Age category	*Number of firms*	*Physical number of employees*	*Ratios of physical numbers*	*Converted number of employees*	*Ratios of converted numbers*
Less then 20	17	83	0,37%	19,7948	0,10%
20–29	649	5 644	25,18%	4 296,2283	22,06%
30–39	886	7 390	32,97%	6 634,2515	34,06%
40–49	797	4 859	21,68%	4 457,0465	22,88%
50–59	614	3 869	17,26%	3 582,0523	18,39%
60 and more	184	572	2,55%	489,2850	2,51%
Total	x	22 417	100,00%	19 478,6584	100,00%

Source: expert surevey

For demographic analysis it is necessary to have an unequivocally set upper and lower limit for all intervals (including the first and the last). In all the considerations given below we have assumed the usual course of school attendance – in other words beginning at the age of 6, completion of elementary school at 15 years, school-leaving exam (maturita) at 19 years, completion of higher vocational college or a bachelor's degree at university at the age of 22 and finally the completion of a master's degree at university at the age of 24.

We took it that an information scientist has at least a secondary education and therefore enters employment at earliest after the school-leaving exam, i.e. at the age of 19. For this reason we considered all information scientists under the age of 20 to be only 19-year-olds (and not younger). The upper limit of the oldest age-group will depend on retirement age. Information scientists will be more often men than women and the retirement age for men at present is close to 62 years. For the sake of simplicity we have therefore assumed that all information scientists retire at the age of 62, meaning that in the age-group of 60 years and over there are only 60–61-year-olds.

For more accurate demographic analysis it is good to estimate the distribution of information scientists in the wider age-groups into one-year age-groups. We may assume that the distribution of the information scientists of the appropriate group according to units of age is

relative to the distribution of the all citizens of the Czech Republic in this age-group according to age units. In the 20–29-year-old age-group, however, this does not apply; The number of information scientists of the given age will depend to a considerable extent on the level of education already completed at this age with the usual course of school attendance. At each age from 20 years there may be information scientists who entered employment immediately after completing secondary school. But only from the age of 22 years can they also be graduates of higher vocational colleges or bachelor degree courses who are not continuing their studies, and only from the age of 24 will there be "added" graduates with master's degrees. The results of the given estimates are given in Table 2.

Table 2: Estimate of the more detailed age structure of 20–29-year-old employees in IT professions in the expert survey

Age group	*Number of citizens in CR*	*Number of information scientists according to highest completed level of education*			
		Secondary	*Vocational + Bac.*	*Mgr.*	*Total*
19	133 059	83	0	0	83
20–21	274 047	171	0	0	171
22–23	280 333	175	570	0	745
24–29	1 007 473	628	2 050	2 050	4 728
Total 20–29	1 561 853	974	2 620	2 050	5 644

Source: expert survey, Czech Statistical Office, own computations

Now we are able to make an estimate of the distribution of information scientists in the CR in 2006 according to the above-mentioned age-groups. According to estimates there were in all 233 100 information scientists in the Czech Republic in 2006. We divide this number into age-groups (with the above-mentioned division of the 20–29 group into 20–21, 22–23 and 24–29) in the same ratio as the number of 22 417 information scientists is divided in the expert survey. The results are given in Table 3.

Table 3: Estimate of distribution of information scientists in the CR in 2006 according to age

Age group	*Number of citizens of CR*	*Expert survey of information scientists*				*Information scientists of CR in 2006*		
		Physical numbers	*Age structure (in %)*	*Converted numbers*	*Share of converted (in % of physical numbers)*	*Physical numbers*	*Converted numbers*	*Share in number of citizens of CR of given age (in %)*
19	133 059	83	0,37	20	23,85	863	206	0,65
20–21	274 047	171	0,76	41	23,85	1 778	424	0,65
22–23	280 333	745	3,32	567	76,12	7 749	5 898	2,76
24–29	1 007 473	4 728	21,09	3 688	78,01	49 162	38 351	4,88
30–39	1 556 304	7 390	32,97	6 634	89,77	76 844	68 985	4,94
40–49	1 365 783	4 859	21,68	4 457	91,73	50 526	46 346	3,70
50–59	1 557 909	3 869	17,26	3 582	92,58	40 231	37 247	2,58
60–61	256 505	572	2,55	489	85,54	5 948	5 088	2,32
Total	6 431 413	22 417	100,00	19 479	86,89	233 100	202 546	3,62

Source: expert survey, Czech Statistical Office, own computations

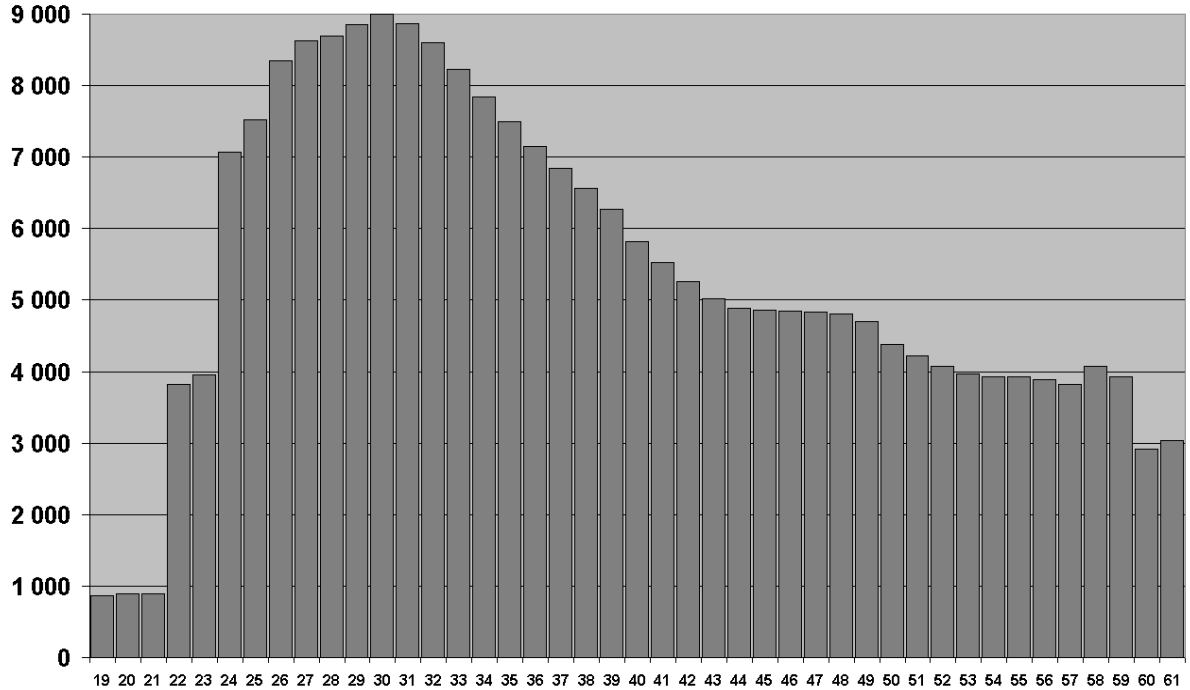
The share of the converted figure in the 6th column of this table was (with the exception of the partial age-groups of the 20–29 year interval) calculated on the basis of Table 1 as the share of the converted and physical numbers for the given age. For the age-group of the 20–21-year-olds the share of the converted numbers was assumed to be the same as for the 19-year-olds, for the 22–23-year-old group it was assumed to be the same as the share for the whole group of 20–29-year-olds from Table 1. For the age-group of the 24–29-year-olds the share of the converted figures was calculated so that the total number of converted figures at the age of 20–29 years was equal to the appropriate value from Table 1 (i.e. 4 296,2283).

The distribution of physical numbers of information scientists in the CR in 2006 according to age was carried out according to the structure for the expert survey, the converted numbers were calculated by multiplying the physical numbers by the shares of converted numbers (again according to the expert survey). Interesting information is provided by the last column (share of information scientists in the number of citizens of the CR of the given age).

The proportion of information scientists declines from the age of 30 with increasing age and it can therefore be seen that the younger the population the larger the proportion of them that are employed in the sphere of informatics. The proportion of 24–29-year-old information scientists is not, however, larger than the proportion of 30–39-year-olds. It is possible that some information scientists complete their studies or begin their working career as an information scientist later than the usual time for completion of studies, i.e. later than the age of 24. This may, however, also be evidence that in the last generation the proportion of information scientists is no longer rising, but stagnating. This must then be seen as a certain "warning" for the future. The generations embarking on economic activity will be fewer and fewer in number. If the proportion of information scientists in the population does not grow, then the number of "new" information scientists will continue to decline in the future.

Table 3 has the age-groups selected in such a way that within each of them we may assume that the distribution of the number of persons according to units of age is proportional to the distribution of the entire population of this age-group according to units of age. On the basis of this assumption an estimate was made of the distribution of information scientists in the CR according to units of age. The calculated values were balanced with the moving average method because the growth (or decline) of the share of information scientists depending on age was certainly rather smoother than jumpy. The result is depicted in Graph 1.

Graph 1. Estimate of the age structure of information scientists in the CR in 2006



Source: own computations

It is clear from the graph at first glance that in the near future one may expect a relatively marked increase in the number of information scientists leaving for retirement. (This will be due to the fact that post-war generations, strong in numbers, will be reaching retirement age.) The following chapter is devoted to this in greater detail. From the graph it is also evident that the numbers of information scientists aged around 30 achieve values of 8–9 thou-

sand for each unit of age. Clearly, then, a number of present information scientists acquired their education in another fashion than by completing the appropriate school (or moved into the country from abroad). Otherwise around the year 2000 the annual number of graduates in subjects concerning informatics would have to have been around 9 000, which probably does not correspond to reality.

2. Projection of development in the next years

We are assuming that every information scientist works in his profession until he retires and that therefore a drop in the number of information scientists occurs only as the result of death or of reaching retirement age at 62. We are not considering the emigration or immigration of information scientists. The proportion of men among information scientists is certainly higher than the proportion of men in the population. On the other side the life expectation of persons with higher education is somewhat higher than the life expectation of the whole population (regardless of education). In particular the life expectation of men with secondary or university education is roughly at the level of the life expectation of women regardless of education. For the projection of the mortality of information scientists were therefore used the life tables for women in the CR (regardless of education) and a further growth of the life expectation was assumed. The results of the projection of the decline in information scientists are shown in the following table and graphs.

Table 4. Expected changes in the number of employees in IT professions in 2007–2010

Year	*2007*	*2008*	*2009*	*2010*
Number on 1st January	233 100	233 208	233 448	233 204
New young graduates	3 495	3 499	3 993	4 206
Deaths	383	386	391	389
Reaching retirement age	3 005	2 873	3 846	3 966
Increase(+)/Decrease(-)	108	240	-244	-148

Source: own computations

We see that whereas the number of information scientists dying would be roughly around 400 a year, the number of information scientists retiring will increase roughly from 3 000 to 4 000. It is probable that the increase will not be so sudden, rather smoother, but in any case the number of information scientists taking up retirement in the next few years will increase.

We assume that all the information scientists who have died or retired should be replaced by young graduates from various schools (we are not considering, then, immigration from abroad or the situation where they are replaced by an older graduate who has not worked in informatics before). What would the numbers of graduates need to be?

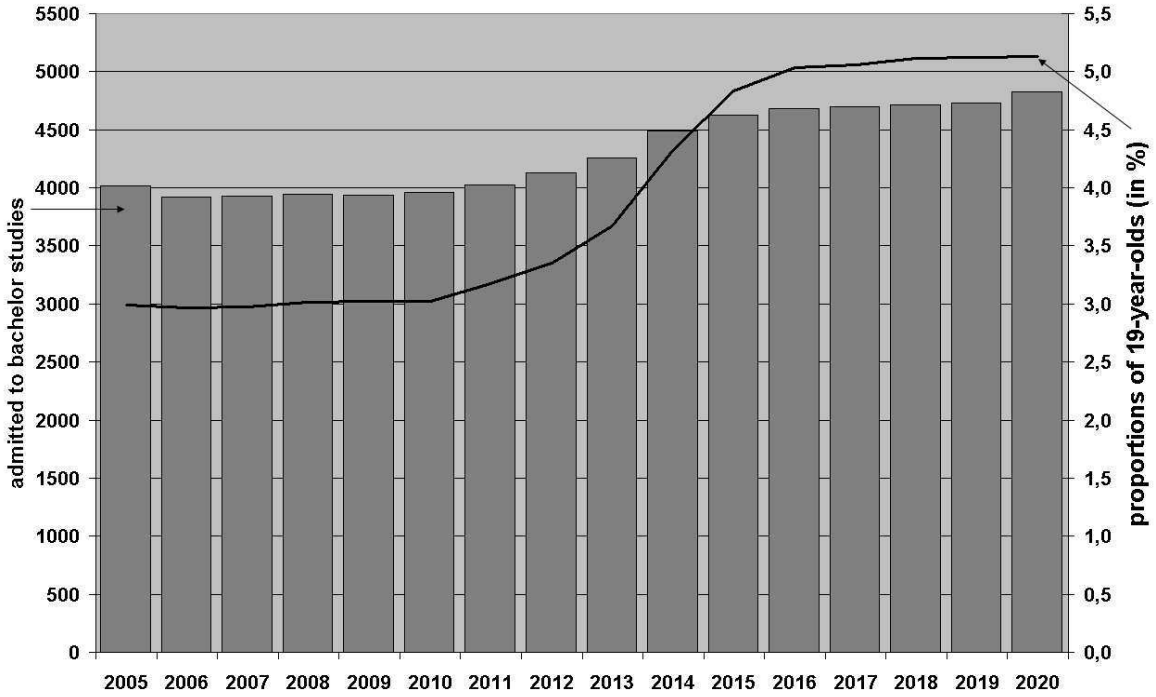
Let us consider that roughly 57% of these young "new" information scientists should be university graduates and the remaining roughly 43% would be information scientists without university education (i.e. graduates of further education colleges, unsuccessful students from bachelor courses, etc.). Experience shows that roughly only 20% of graduates with bachelor's degrees take up employment immediately. The remaining 80% continue with the two years of master's studies, which 90% of them complete successfully and they therefore start working after these two years.

With regard to the expected marked increase in the number of information scientists retiring at the end of this decade the annual number of graduates with bachelor's degrees in informatics subjects should already this year be around 2 200 and in further years around 2 400 of them are needed each year. In the second half of the next decade the annual number of graduates should rise further up to 2 800.

At the same time not every student accepted for bachelor studies will complete the course successfully. Experience shows that only around 60% of those admitted actually com-

plete their bachelor studies. The numbers of students admitted to bachelor studies should therefore be considerably higher (and, of course, grow adequately "in advance"). See Graph 2.

Graph 2. Required numbers of those admitted to bachelor studies in informatics and their share in the total population of the appropriate age



Source: own computations

We can see that at present the number admitted for bachelor study of informatics should be around 4 000 a year. But already in the second half of the next decade this should increase gradually to 4 500 and at the end of the second decade come close to the value of 5 000. With regard to the decline in the number of young people in the population and the competition among universities it will clearly be increasingly difficult to acquire suitable candidates for studies. At present it is enough for around 3% of the 19-year-olds to begin studying informatics. In the first half of the next decade, however, this proportion should rise to over 5%.

3. References

FIALA, T. 2006. Dva přístupy modelování vývoje úmrtnosti v populační projekci a jejich aplikace na populaci ČR. Bratislava 05.10.2006 – 06.10.2006. In: *Forum Statisticum Slovacum 4/2006.* Bratislava : Slovenská štatistická a demografická spoločnosť, 2006, s. 44–55. ISSN 1336-7420.

Address of the author:
RNDr. Tomáš Fiala, CSc.
Department of Demography
Faculty of Informatics and Statistics
University of Economics, Prague
nám. W. Churchilla 4
130 67 Praha 3
Czech Republic
fiala@vse.cz

This article came into being within the framework of the long-term research project 2D06026, "Reproduction of Human Capital", financed by the Ministry of Education, Youth and Sport within the framework of National Research Program II.

Ukazatele vzdělanosti populace: využitelnost odhadů z různých datových zdrojů[1]

Jakub Fischer, Petr Mazouch

Abstract: This paper is focused on comparison of data on education-level structure of population from different types of surveys. Education is one of the basic type of investments in the human capital. For the deep analyses of the impact of these investments on both micro-economic and macroeconomic indicators (wages, life expectancy, health status, labour productivity, multifactor productivity) it is necessary to measure the level of education of the population. Due to the fact that the administrative data sources provide data just on education processes and their results and not on the education level of population, the data are at a disposal only from the Housing and Population Census and from the Labour Force Sample Survey as well. The aim of the article is to compare data from these surveys.

Key words: education level, Housing and Population Census, Labour Force Sample Surveys

1. Úvod

Vzdělání je jedním ze základních druhů investic do lidského kapitálu s významným přínosem jak na úrovni jednotlivce (vyšší mzdy, delší střední délka života, lepší zdravotní stav apod.), tak i na úrovni společnosti (růst produktivity práce a souhrnné produktivity faktorů, nižší náklady na zdravotní péči, vyšší vybrané daně apod.). S ohledem na potřebu tyto mikroekonomické i makroekonomické přínosy kvantifikovat vzniká otázka, jakým způsobem úroveň vzdělanosti zjišťovat (přesněji jakými šetřeními) a nakolik jsou údaje z těchto šetření spolehlivé. Na rozdíl od statistiky zaměstnanosti, kde kromě Sčítání lidu, domů a bytů (SLDB) a Výběrového šetření pracovních sil (VŠPS) máme k dispozici ještě podnikovou statistiku (výkaznictví), v níž se řada ukazatelů zaměstnanosti zjišťuje, u vzdělání jsme odkázáni na data ze SLDB a z VŠPS. Podniková statistika však údaje o vzdělání zaměstnanců nezjišťuje a administrativní zdroje dat o vzdělání se týkají vzdělávacích procesů a jejich výsledků, nikoli úrovně vzdělání. Výsledky vzdělání jsou použitelné maximálně na úrovni primárního a sekundárního vzdělávání, neboť jsou k dispozici nikoli podle bydliště respondenta, ale podle sídla vzdělávací instituce. Zatímco migraci absolventů základních škol lze těžko v masovém měřítku předpokládat, u absolventů středních škol si tím nejsme jisti a u absolventů vysokých škol je předpoklad zjevně neudržitelný. Řada absolventů se vrací zpět do místa svého trvalého bydliště, druhá významná skupina zůstává v sídle školy, kde nachází pracovní uplatnění, třetí skupina se přesouvá za prací na úplně jiné místo. Ukazatele vzdělanosti se objevují i ve strukturálním šetření výdělků, nicméně toto šetření není zaměřeno na zjišťování vzdělanosti, ale na zjišťování mezd. V našem příspěvku se budeme zabývat porovnáním odhadů stejných ukazatelů ze dvou různých šetření provedených ve stejném období – SLDB a VŠPS.

2. Cíl a metodika

V následující části se pokusíme porovnat data o počtu vzdělaných osob ze Sčítání lidu, domů a bytů (SLDB) a z Výběrového šetření pracovních sil (VŠPS). Vzhledem k tomu, že je potřeba kvantifikovat dopady růstu vzdělanosti na řadu ukazatelů, vzniká potřeba pro tyto kvantifikace získat relevantní data. Dosud byly provedeny propočty vztahu vzdělanosti a mezd (Hudrlíková, 2007), vzdělanosti a střední délky života (Mazouch, Fischer, 2007a,

[1] Příspěvek vznikl za podpory Národního programu výzkumu II Ministerstva školství, mládeže a tělovýchovy ČR č. 2D06026 „Reprodukce lidského kapitálu“.

2007b), vzdělanosti a nezaměstnanosti (Fischer, Mazouch, 2007a, 2007b), vzdělanosti a důchodového systému (Mazouch, Fischer, 2007c), vzdělanosti a přidané hodnoty (Fischer, Mazouch, 2007b). Ve všech těchto propočtech byly vzaty v úvahu údaje ze SLDB. Tato data mají tu výhodu, že pocházejí z vyčerpávajícího zjišťování, bohužel jsou k dispozici pouze jednou za deset let (navíc se značným zpožděním po rozhodném okamžiku sčítání). Z tohoto pohledu je otázkou, zdali je možné, a to i pro menší celky, tj. na úrovni regionů či odvětví, použít odhady z VŠPS jako relevantní.

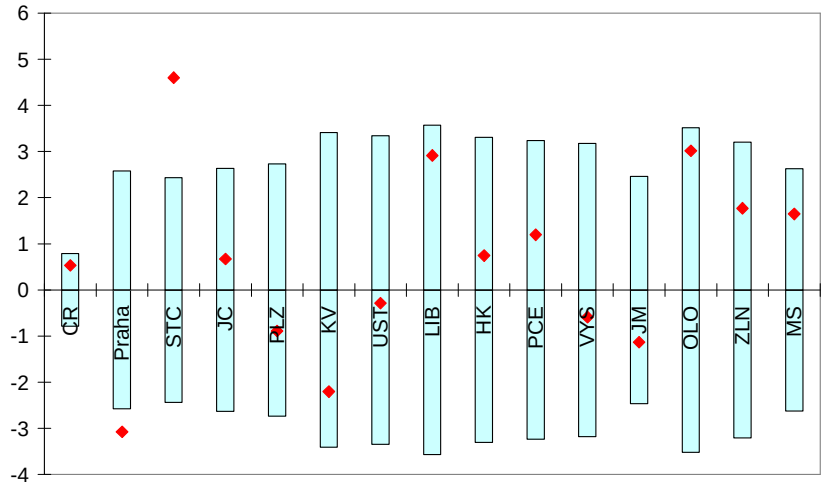
Porovnáváme tedy výsledky ze SLDB a z VŠPS. Vzhledem k tomu, že rozhodným okamžikem při SLDB byla půlnoc z 28. února na 1. března 2001, porovnáváme výsledky ze SLDB s odhady z VŠPS za 1. čtvrtletí 2001. Ukazatele ekonomické aktivity, zaměstnanosti, nezaměstnanosti a vzdělání jsou vymezeny shodně.

Výsledky z VŠPS jsou publikovány nikoli pouze jako bodové odhady, ale současně s nimi jsou k dispozici i 95% intervaly spolehlivosti pro úhrn a podíl. Z toho důvodu můžeme porovnat, zdali údaj ze SLDB leží v 95% intervalu spolehlivosti či nikoli. Pokud ne, naznačuje to buď nastání oné situace s 5% pravděpodobností, kdy výběrový interval leží mimo hodnotu ze základního souboru, nebo nastání nevýběrové chyby při výběrovém šetření, a nebo konečně chybu ve vyčerpávajícím zjišťování.

Nejprve porovnáme odhady počtu zaměstnaných osob (první nás zajímá z důvodu vlivu na přidanou hodnotu a na možnost překlopit další údaje z VŠPS do podnikové statistiky za účelem obohacení produkční funkce o aspekt vzdělání), poté porovnáme odhady míry nezaměstnanosti (z důvodu posuzování vlivu vzdělanosti na míru nezaměstnanosti) a následně se zaměříme na porovnání počtu osob v dané vzdělanostní kategorii (pracujeme se čtyřmi agregovanými skupinami – osoby se základním vzděláním a bez vzdělání; osoby se středním vzděláním bez maturity; osoby se středním vzděláním s maturitou; osoby s vysokoškolským vzděláním). Vždy porovnáváme údaj ze SLDB s 95% intervalem spolehlivosti z VŠPS. Pro možnost srovnání odhadů v jednotlivých regionech jsou všechna porovnání počtu osob převedena na relativní odchylku odhadu z VŠPS od údaje ze SLDB, vyjádřenou v %. Odchylka odhadu míry nezaměstnanosti z VŠPS od údaje ze SLDB je uvedena v procentních bodech. Porovnání odhadu počtu nezaměstnaných osob a počtu ekonomicky aktivních osob je uvedeno v (Fischer, Mazouch, 2007c).

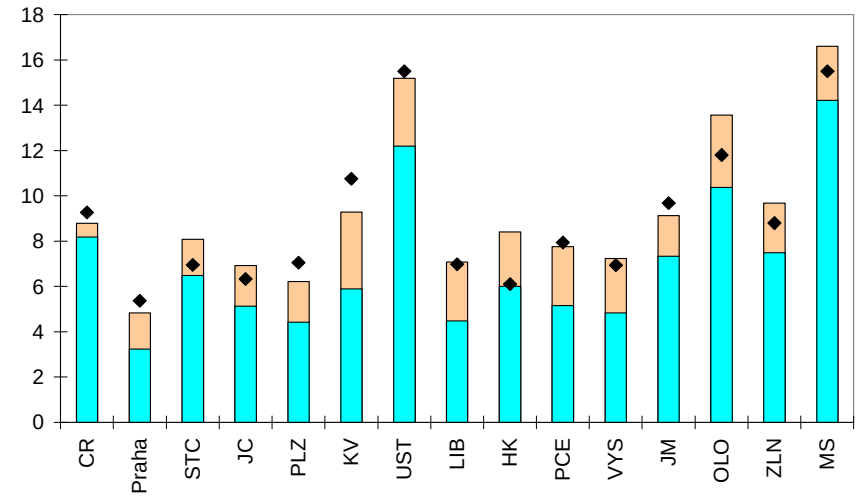
3. Výsledky

Graf 1: Porovnání počtu zaměstnaných osob ze SLDB a z VŠPS (odchylky v %)



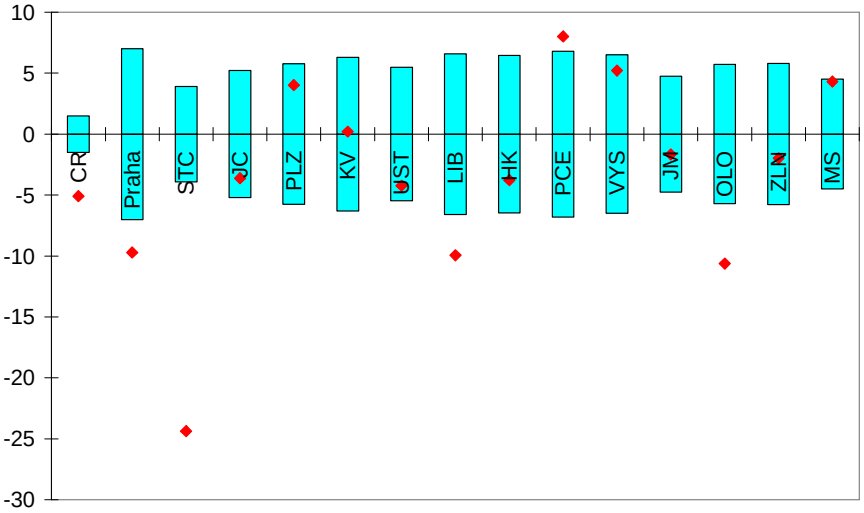
Zdroj: ČSÚ (2001, 2005), propočty autorů.

Graf 2. Porovnání regionálních měr nezaměstnanosti z VŠPS a ze SLDB



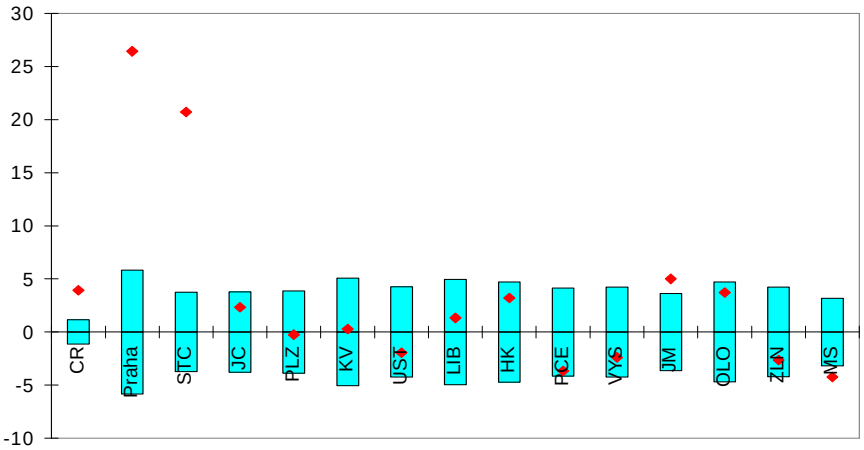
Zdroj: ČSÚ (2001, 2005), propočty autorů

Graf 3. Porovnání počtu osob starších 15 let se základním vzděláním (rozdíly v %)



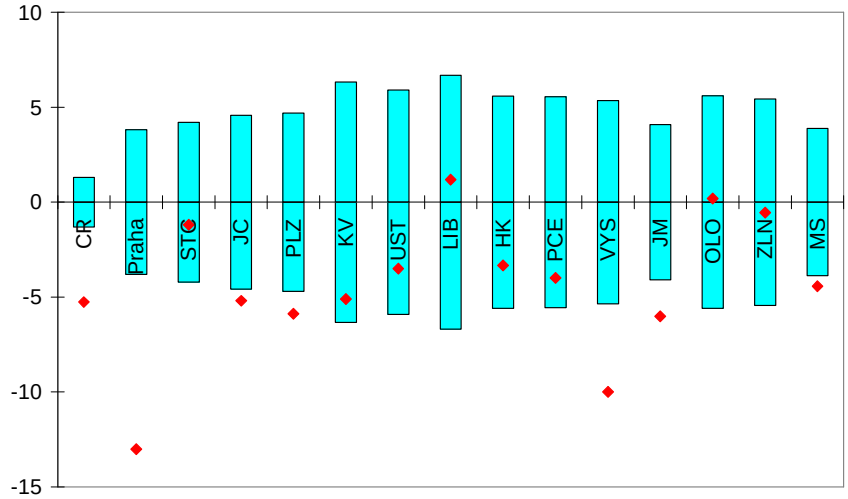
Zdroj: ČSÚ (2001, 2005), propočty autorů

Graf 4. Porovnání počtu osob 15+ let se středním vzděláním bez maturity (rozdíly v %)



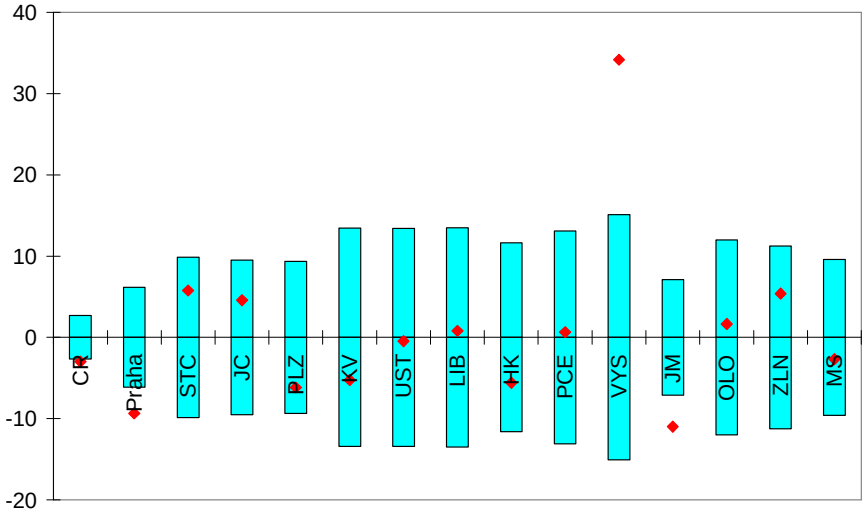
Zdroj: ČSÚ (2001, 2005), propočty autorů

Graf 3.5 Porovnání počtu osob starších 15 let se středním vzděláním s matur. (rozdíly v %)



Zdroj: ČSÚ (2001, 2005), propočty autorů

Graf 3.6 Porovnání počtu osob starších 15 let s vysokoškolským vzděláním (rozdíly v %)



Zdroj: ČSÚ (2001, 2005), propočty autorů

V grafu č. 1 jsou porovnány odhady počty zaměstnaných osob, v grafu č. 2 odhady míry nezaměstnanosti, v grafech č. 3 až 6 odhady počtu osob starších 15 let s příslušným stupněm dosaženého vzdělání. Ve všech grafech sloupce označují 95% intervaly spolehlivosti pro relativní odchylku, v grafu č. 2 je vyznačena přímo regionální míra nezaměstnanosti (bod) včetně intervalu spolehlivosti.

Největší rozdíly mezi VŠPS a SLDB jsou u odhadů pro hl. m. Prahu, pro Středočeský kraj a pro kraj Vysočina. Pro Prahu Výběrové šetření pracovních sil podhodnocuje odhady počtu osob se středním vzděláním bez maturity (rozdíl činí téměř 25 % oproti SLDB), naopak nadhodnocuje odhady počtu osob se středním vzděláním s maturitou (o 13 %). Ve Středočeském kraji se výrazně liší odhady pro počty osob se základním vzděláním (+25 %) a pro počty osob se středním vzděláním bez maturity (-20 %). Na Vysočině je nadhodnocen odhad pro počet osob se středním vzděláním s maturitou (o 10 %) a podhodnocen odhad pro počet osob s vysokoškolským vzděláním (o 35 %). Též je třeba zachytit skutečnost, že odhady pro počty osob pro celou ČR jsou taktéž většinou mimo intervaly spolehlivosti (byť jsou tyto intervaly zřetelně užší).

4. Diskuse a závěr

Rozdíly v odhadech počtu zaměstnaných osob, míry nezaměstnanosti i počtu osob v jednotlivých vzdělanostních skupinách jsou mezi VŠPS a SLDB poměrně výrazné. Kromě velkých rozdílů při odhadech za malé územní celky (kde se údaje ze SLDB „netrefí“ ani do relativně širokých intervalů spolehlivosti) nejsou příliš v souladu ani údaje za celou ČR. Z tohoto důvodu je pro analytické účely nezbytné přistupovat k odhadům z VŠPS s jistou rezervou.

Ukazuje se přitom, že zejména z důvodu odhadů za menší celky (regionálně i odvětvově) je šetření typu cenzu v současné době nezbytné. Přitom analýzy prováděné z regionálního hlediska jsou v současné době velmi poptávané. Bez kvalitních datových podkladů jsou ale tyto analýzy nepříliš cenné. Porovnáním na odvětvové úrovni se budeme zabývat v dalším výzkumu.

5. Literatura

ČSÚ. 2001. Zaměstnanost a nezaměstnanost podle výběr. šetření pracovních sil 1. čtvrtletí 2001.Praha, 2001.

ČSÚ. 2005. Sčítání lidu, domů a bytů 2001. Pramenné dílo. Praha, 2005.

FISCHER, J. – MAZOUCH, P. 2007a. Proč mají regiony investovat do lidského kapitálu? Praha 20.09.2007. In: *MSED na VŠE [CD-ROM]*. Praha : Typograf, 2007, s. 1–5. ISBN 978-80-254-0257-7.

FISCHER, J. – MAZOUCH, Petr. 2007b. Souvislosti vzdělanosti, nezaměstnanosti a ekonomického růstu z regionálního hlediska. Olomouc 23.05.2007 – 24.05.2007. In: *Regionální demografie [CD-ROM]*. Praha : DemoArt pro Českou demografickou společnost, 2007, s. 188–197. ISBN 80-86746-04-6.

FISCHER, J. – MAZOUCH, P. 2007c. Popis stanovanj in prebivalstva: anahronizem ali potreba? Radenci 05.11.2007 – 07.11.2007. In: *17. Statistični dnevi [CD-ROM]*. Ljubljana : Statistični urad Republiky Slovenije, 2007, s. 1–7.

MAZOUCH, P. – FISCHER, J. 2007a. Střední délka života podle nejvyššího ukončeného vzdělání. Brno 08.03.2007 – 09.03.2007. In: Firma a konkurenční prostředí 2007. Brno: MSD, 2007, p. 91–95. ISBN 978-80-86633-86-2.

MAZOUCH, P. – FISCHER, J. 2007b. Arriagova metoda dekompozice střední délky života mezi ČR a SR. Janské Lázně 04.10.2007 – 07.10.2007. In: *Makos 2006.* Olomouc : Univerzita Palackého, 2007, s. 53–59. ISBN 978-80-244-1760-8.

MAZOUCH, P. – FISCHER, J. 2007c. Longer Life Caused by Higher Attained Level of Education: How to Valuate this Advantage. Lisabon 22.08.2007 – 29.08.2007. In: *Bulletin of the International Statistical Institute 56th Session – ISI 2007 [CD-ROM]*. Lisabon : International Statistical Institute, 2007, s. 1–4.

HUDRLÍKOVÁ, L. 2007. Mzdy v České republice v souvislosti s dosaženým vzděláním. Bakalářská práce. VŠE v Praze, 2007.

Adresa autorů:
Ing. Jakub Fischer, Ph.D., Ing. Petr Mazouch
Katedra demografie Fakulty informatiky a statistiky Vysoké škola ekonomické v Praze
nám. W. Churchilla 4
CZ 130 67 Praha 3
Česká republika
fischerj@vse.cz, mazouchp@vse.cz

Testování hypotéz o paralelnosti silnic aneb problémy satelitního mýtného systému

Jana Heckenbergerová[1], Hana Boháčová [2]

Abstract: Today there exists several methods of the position determination. Research in The Global Navigation Satellite Systems (GNSS) area shows, that it is good to use satellite navigation for this purpose. This paper is mainly devoted to the position determination and testing algorithms along the highway. Supposed highway id defined by analytical function. All of described analytical algorithms are based on a statistical model with constraints.

Key words: GNSS, Statistical Model with Constraint, Best Linear Unbiased Estimate (BLUE), Hypotheses Testing, Statistics of Test

1. Úvod

Pod pojmem Globální Navigační Satelitní Systémy (GNSS) rozumíme vlastně několik existujících družicových systémů (např. GPS-USA, GLONASS-Rusko, GALILEO-EU). GNSS receivery (přijímače GNSS signálu) poskytují uživateli informaci o poloze a její přesnosti. Některé GNSS receivery jsou schopné sloučit informace z různých družicových systémů. Více informací o fungování GNSS lze najít ve skriptu [1], kde jsou popsány i algoritmy určení *(x,y)*-polohy daného GNSS receiveru. GNSS se nyní úspěšně používají v automobilovém průmyslu (GPS navigace), geodézii a moderních telekomunikačních technologiích. Propracované metody bezpečného určení polohy pomocí GNSS jsou užívány v letecké dopravě.

V současnosti se v České republice, ale i v rámci celé Evropy, řeší problémy s mýtným systémem. Zdá se, že systém založený na GNSS, by mohl být v českých podmínkách použitelný. Hlavním problémem je však účtování mýtného automobilům, které jedou po paralelní nezpoplatněné silnici a naopak neúčtování mýtného automobilům, které jedou po dálnici, ale jejich GNSS poloha leží na paralelní nezpoplatněné silnici. Pomocí implementace níže popsaných algoritmů do GNSS mýtného systému by měl být tento nedostatek satelitního mýtného systému odstraněn.

2.1 Formulace problému

Pomocí GNSS je odhadnuta poloha n bodů, které určují polohu referenčního bodu automobilu v časech $t_1,\ldots,t_n$. Body jsou dány x-ovou a y-ovou souřadnicí. Informaci o přesnosti GNSS polohy každé souřadnice lze získat přímo z GNSS receiveru, varianční matice potom označme Σ.

Dále předpokládejme, vyšetřovaná část silnice je přímková a tedy skutečné hodnoty referenčních bodů splňují rovnici:

$$y = q + kx, \tag{1}$$

což je rovnice přímky ve směrnicovém tvaru, kde q je posunutí a k je směrnice. Předpokládejme také, že směrnice silnice, po které se pohybuje referenční bod, je známa, posunutí však známo není. Lze ho odhadnout a porovnat s posunutím paralelní silnice, jejíž parametry jsou známé. Uvažujme tedy dvě rovnoběžné silnice, jedna je zpoplatněná, druhá nikoli. Směrnicové rovnice těchto silnic jsou známé. Z GNSS poloh odhadneme hodnotu

[1] Jana Heckenbergerová, Ústav elektrotechniky a informatiky, Univerzita Pardubice
[2] Hana Boháčová, Fakulta ekonomicko-správní, Univerzita Pardubice

posunutí a testujeme, zda automobil jede po zpoplatněné silnici či po paralelní.

Předpokládejme tedy, že observační náhodný vektor naměřených dat

$$Y = \left\{ \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} \right\}_{i=1}^{n}$$

má *2n*-rozměrné normální rozdělení se střední hodnotou μ a kovarianční maticí Σ, kde μ je neznámý *2n*-rozměrný vektor skutečných hodnot a Σ je známá pozitivně definitní matice typu *2n* x *2n*, tedy $Y \sim N_{2n}(\mu, \Sigma)$.

Formulaci problému lze rozložit na dvě části:

- Z n-tice dvojic $\left\{ \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} \right\}_{i=1}^{n}$ naměřených GNSS poloh odhadujeme n-tici dvojic přesných hodnot $\left\{ \begin{pmatrix} \mu_{i,1} \\ \mu_{i,2} \end{pmatrix} \right\}_{i=1}^{n}$, za podmínky že přesné hodnoty leží na přímce, tedy

  $$\mu_{i,2} = \beta + \gamma \mu_{i,1} \qquad i = 1,...,n,$$

  kde γ je známá hodnota směrnice koleje a β je neznámý parametr posunutí.

- Testujeme nulovou hypotézu

  H_0: $\beta = \beta_0$

  proti alternativě

  H_a: $\beta \neq \beta_0$,

  kde $y = \beta_0 + \gamma x$ je rovnice předpokládané (zpoplatněné) silnice, jejíž parametry β_0 a γ jsou známé.

Z formulovaného problému lze vytvořit statistický model

$$Y \sim N_{2n}(\mu, \Sigma) \text{ s podmínkou } \mu_{i,2} = \beta + \gamma \mu_{i,1} \quad i = 1,...,n, \tag{2}$$

který je speciálním případem modelu nepřímého měření s podmínkou typu II na parametr 1. řádu. Vlastnosti tohoto statistického modelu jsou popsány v knize [2].

2.2 Odhady parametrů silnice zaměřené pomocí GNSS

K určení parametru posunutí silnice užijeme větu 1. Důkaz této věty je založen na optimalizaci Lagrangeovy funkce a je uveden v [3].

Věta 1: V uvedeném statistickém modelu (2)

$$Y \sim N_{2n}(\mu, \Sigma) \text{ s podmínkou } \mu_{i,2} = \beta + \gamma\mu_{i,1} \quad i = 1,\ldots,n,$$

je BLUE (nejlepší lineární nestranný odhad) parametrů μ a β dán vztahy

$$\begin{aligned} \hat{\hat{\mu}} = \begin{pmatrix} \hat{\hat{\mu}}_1 \\ \hat{\hat{\mu}}_2 \end{pmatrix} &= \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} - \Sigma \begin{pmatrix} \gamma I \\ -I \end{pmatrix} \left[G^{-1} - G^{-1}1\left(1'G^{-1}1\right)^{-1}1'G^{-1} \right] (\gamma Y_1 - Y_2), \\ \hat{\hat{\beta}} &= -\left(1'G^{-1}1\right)^{-1}1'G^{-1}(\gamma Y_1 - Y_2), \end{aligned} \tag{3}$$

kde

$$G = (\gamma I, -I)\Sigma \begin{pmatrix} \gamma I \\ -I \end{pmatrix}.$$

2.3 Testovací algoritmus

Druhou částí formulovaného problému je testování hypotézy o paralelnosti silnic. K vyřešení tohoto problému využijeme věty 2, jejíž důkaz je založen na vlastnostech statistického modelu s podmínkou typu II. Postup důkaz je popsán v [3].

Věta 2: V uvedeném statistickém modelu (2)

$$Y \sim N_{2n}(\mu, \Sigma) \text{ s podmínkou } \mu_{i,2} = \beta + \gamma\mu_{i,1} \quad i = 1,\ldots,n,$$

testujme nulovou hypotézu

$$\mathrm{H}_0: \quad \beta - \beta_0 = 0,$$

proti alternativní hypotéze

$$\mathrm{H}_a: \quad \beta - \beta_0 \neq 0.$$

Potom pro testovací statistiku T platí

$$T = \frac{\left(\hat{\hat{\beta}} - \beta_0\right)^2}{\left(1'G^{-1}1\right)^{-1}} \sim \begin{cases} \chi_1^2(0), & \mathrm{H}_0 \text{ platí} \\ \chi_1^2(\delta), & \mathrm{H}_0 \text{ neplatí} \end{cases}, \tag{4}$$

kde parametr necentrality δ je ve tvaru

$$\delta = \frac{\left(\beta^* - \beta_0\right)^2}{\left(1'G^{-1}1\right)^{-1}},$$

kde β^* je skutečná hodnota parametru β.

Nulovou hypotézu nelze zamítnout, je-li testovací statistika $T < \chi_1^2(1-\alpha)$, kde α je zvolené riziko (pravděpodobnost chyby 1. druhu), nulovou hypotézu zamítáme ve prospěch alternativy na hladině významnosti α, jestliže $T \geq \chi_1^2(1-\alpha)$.

3. Ilustrační numerický příklad

Mějme naměřené referenční body automobilu pomocí GNSS

x	Y_{11}=-0,0064	Y_{12}=0,5226	Y_{31}=0,9951	Y_{41}=1,6167
y	Y_{21}=4,9952	Y_{22}=6,4912	Y_{32}=8,0131	Y_{42}=9,6317

a mějme zadánu rovnici dálnice ve tvaru

$$y=3x+5,$$

tedy $\gamma_1=5$ $\gamma_2=3$.
Kovarianční matice chyb měření je ve tvaru:

$$\Sigma=\begin{pmatrix} 0.0025 & 0 & 0 & 0 & 0.001 & 0 & 0 & 0 \\ 0 & 0.0025 & 0 & 0 & 0 & 0.001 & 0 & 0 \\ 0 & 0 & 0.0025 & 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0.0025 & 0 & 0 & 0 & 0.001 \\ 0.001 & 0 & 0 & 0 & 0.0025 & 0 & 0 & 0 \\ 0 & 0.001 & 0 & 0 & 0 & 0.0025 & 0 & 0 \\ 0 & 0 & 0.001 & 0 & 0 & 0 & 0.0025 & 0 \\ 0 & 0 & 0 & 0.001 & 0 & 0 & 0 & 0.0025 \end{pmatrix}$$

K určení odhadů $\hat{\hat{\beta}}$ a $\begin{pmatrix}\hat{\hat{\mu}}_1\\ \hat{\hat{\mu}}_2\end{pmatrix}$ je nutno učinit několik předchozích výpočtů

$$G=\left(\gamma_2 I \;\; ; \;\; -I\right)\Sigma\begin{pmatrix}\gamma_2 I\\ -I\end{pmatrix}=\begin{pmatrix} 0.019 & 0 & 0 & 0 \\ 0 & 0.019 & 0 & 0 \\ 0 & 0 & 0.019 & 0 \\ 0 & 0 & 0 & 0.019 \end{pmatrix},$$

$$\left(1'G^{-1}1\right)^{-1}=0.0047.$$

Odhady parametrů modelu jsou tedy ve tvaru

$$\hat{\hat{\beta}}=4.9368 \;;\; \begin{pmatrix}\hat{\hat{\mu}}_1\\ \hat{\hat{\mu}}_2\end{pmatrix}=\begin{pmatrix} 0.0201 \\ 0.5180 \\ 1.0262 \\ 1.5636 \\ 4.9972 \\ 6.4908 \\ 8.0155 \\ 9.6276 \end{pmatrix}.$$

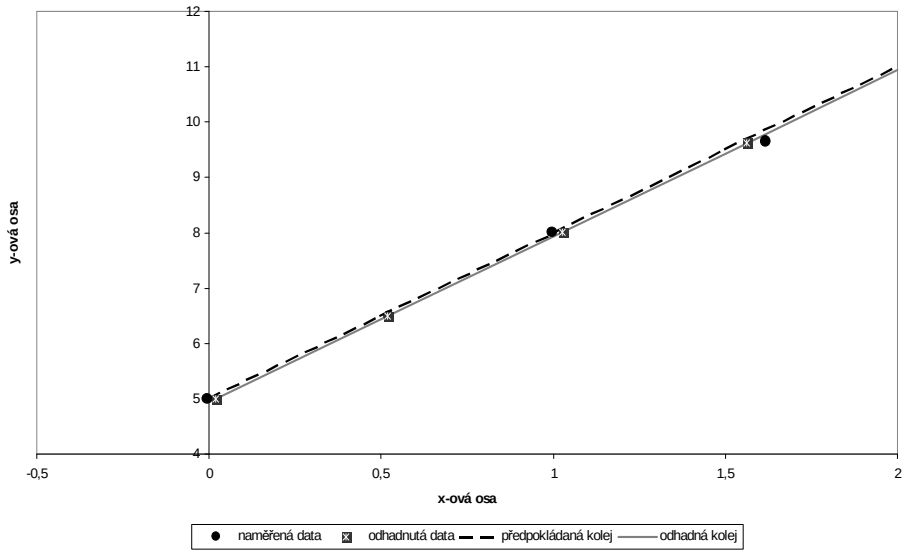
V druhé části problému určíme hodnotu testovacího kritéria

$$T=0.8409\ .$$

Jelikož T < 1.074 nelze hypotézu H_0 zamítnout na hladině významnosti α=0,3.

Řešení problému je graficky znázorněno na obrázku číslo 1.

Obrázek 1



Posuneme-li rovnici dálnice o 1m, budeme-li tedy přepokládat, že rovnice silnice je ve tvaru

$$y=3x+6,$$

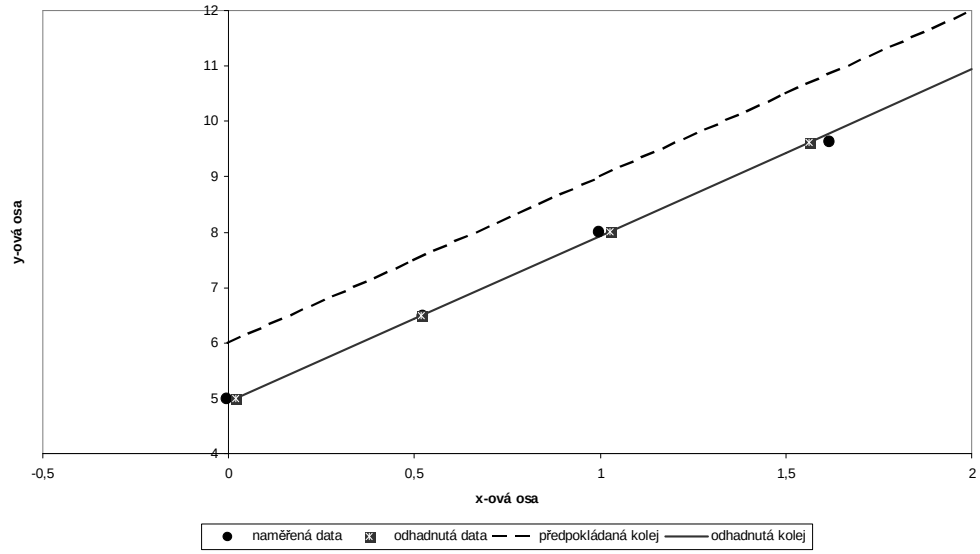
kde parametry mají hodnoty $\gamma 1=6$ a $\gamma 2=3$.
A budeme i nadále přepokládat, že přesnost měření je stejná, tedy kovarianční matice je ve tvaru Σ.

Potom se odhady parametrů $\hat{\hat{\beta}}$ a $\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix}$ nezmění, ale testovací statistika má hodnotu

$$T=237,9777.$$

Neboť $T>12,1$ zamítáme hypotézu H_0 ve prospěch alternativy na hladině významnosti $\alpha=0,0005$.
Řešení pozměněného problému je graficky znázorněno na obrázku číslo 2.

Obrázek 2



3. Závěr

Popsané modely byly úspěšně testovány na reálných datech. Výsledky těchto testů jsou popsány v [3]. Cílem budoucího výzkumu jsou testy algoritmů online a jejich implementace do satelitního mýtného systému. Dále bude potřeba prozkoumat vlastnosti statistických modelů v případě, že silnice není přímková. Tyto úlohy vedou na nelineární statistické modely, jejichž linearizace vede k nesnadným problémům.

V neposlední řadě je potřeba upozornit na univerzálnost popsaných algoritmů. Lze je s úspěchem použít i pro železniční dopravu (např. určení nástupiště, ke kterému vlak přijíždí, dle GNSS poloh referenčního bodu lokomotivy tohoto vlaku). Statistické problémy spojené s bezpečným určením polohy vlaku pomocí GNSS jsou řešeny ve článcích [4] a [5].

4. Literatura

[1] MERVART, L. *Základy GPS*. Vydavatelství ČVUT, Praha, 1993.

[2] KUBÁČEK, L., KUBÁČKOVÁ, L. *Statistika a metrologie*. Univerzita Palackého v Olomouci 2000, ISBN 80-244-0093-6.

[3] DVOŘÁKOVÁ, J. *Statistické metody pro identifikaci polohy vlaku.* Olomouc 2004.

[4] DVOŘÁKOVÁ, J., MOCEK, H., MAIXNER, V., *Statistical Approach to the Train Position Integrity Monitoring*. Sborník „Reliability, safety and diagnostics of transport structures and means 2005“, Pardubice, ISBN 80-7194-769-5.

[5] HECKENBERGEROVÁ, J. *Parametrické algoritmy pro ověření GNSS polohy vlaku.* Sborník Infotrans 2007, Pardubice, ISBN 978-80-7194-989-3.

Adresa autora:
Jana Heckenbergerová, Mgr.
Ústav elektrotechniky a informatiky, Oddělení informatiky
Univerzita Pardubice
Studentská 95
532 10 Pardubice
jana.heckenbergerova@upce.cz

Priame metódy merania úrovne regiónov

Hurbánková Ľubica

Abstract: In this paper we present one of the methods for measurement of regions level – direct methods. We use this methods when we monitor an indicator for efficiency measurement in particular regions. We chose the indicator gross added value per capita. We know these direct methods – comparison method on the ground of growth rate and comparison method on the ground of unit rate. We apply these methods on the data for Bratislava region and for Presov region.

Key words: region, growth rate, unit rate, a gross added value per capita

1. Úvod

Existujú viaceré spôsoby merania úrovne regiónov. Jedným z nich sú priame metódy. Používajú sa v tých prípadoch, keď sledujeme ukazovateľ používaný ako indikátor na meranie efektívnosti jednotlivých krajov. V príspevku sme si zvolili za tento indikátor hrubú pridanú hodnotu per capita (ďalej HPH_{pc}). V literatúre sa môžeme stretnúť s nasledovnými priamymi metódami[1] na meranie úrovne regiónov:

- metóda porovnávania na základe tempa rastu,
- metóda porovnávania na základe mernej rýchlosti.

2. Metóda porovnávania na základe tempa rastu

Cieľom metódy je určiť minimálne tempo rastu zaostávajúceho regiónu (kraja) v porovnaní s vyspelejším, pri docielení ktorého sa nebudú ďalej rozdiely medzi regiónmi (krajmi) prehlbovať.

Predpokladajme dva kraje A a B, ktoré dosahujú výstup Y_A a Y_B, pričom platí:

$$Y_A < Y_B.$$

Tempo rastu v krajoch vyjadríme nasledovne:

$$r_A = \frac{\Delta Y_A}{Y_A} \qquad r_B = \frac{\Delta Y_B}{Y_B} \tag{1}$$

kde

$$\Delta Y_A = Y_{A,k} - Y_{A,j} \qquad \Delta Y_B = Y_{B,k} - Y_{B,j}$$

[1] Glaser-Opitzová, H. – Myslíková, I.: Metodológia a metódy merania sociálno-ekonomickej úrovne regiónov. Bratislava: Infostat, 2001

Vyčíslime teoretické tempo rastu ($r´_A$), ktoré musí menej rozvinutý kraj dosiahnuť, aby sa rozdiely v ekonomickej úrovni neprehlbovali.

$$r´_A = \frac{r_B * Y_B}{Y_A} \tag{2}$$

K vyrovnaniu rozdielov dochádza vtedy, ak skutočne dosahované tempo rastu menej rozvinutého kraja r_A je väčšie ako teoretické tempo rastu $r´_A$.

3. Metóda porovnávania na základe mernej rýchlosti

Uvažujme podobne ako v predchádzajúcom prípade, že máme dva kraje A a B s určitými hodnotami výstupu Y_A a Y_B, z čoho na 1 obyvateľa kraja pripadá hodnota:

$$\frac{Y_A}{N_{0-100,A}}$$

$$\frac{Y_B}{N_{0-100,B}} \tag{3}$$

Predpokladajme, že

$$\frac{Y_A}{N_{0-100,A}} \quad \langle \quad \frac{Y_B}{N_{0-100,B}}$$

Mernú rýchlosť (m) za určité obdobie vypočítame nasledovne:

$$d_A = \frac{Y_{A,k}}{N_{0-100,A,k}} - \frac{Y_{A,j}}{N_{0-100,A,j}} \tag{4}$$

$$d_B = \frac{Y_{B,k}}{N_{0-100,B,k}} - \frac{Y_{B,j}}{N_{0-100,B,j}} \tag{5}$$

$$m = \frac{d_A}{d_B} \tag{6}$$

Ak merná rýchlosť nadobúda hodnotu menšiu ako 1, nedochádza v sledovanom období k odstraňovaniu rozdielov v ekonomickej úrovni krajov. V prípade, že $m=1$, úroveň rozdielov zostáva nemenná, a ak m je väčšie ako 1, dynamika rastu výstupu Y_A je postačujúca na postupné odstraňovanie nižšej úrovne ekonomického rozvoja v kraji A.

4. Aplikačná časť

Z viacerých analýz realizovaných na údajoch za Slovenskú republiku a jej jednotlivé regióny sme zistili, že za najzaostalejší región možno považovať Prešovský kraj. Naopak za najvyspelejší sa považuje Bratislavský kraj. Existuje viacero dôvodov, prečo sú to práve tieto kraje:

- Bratislavský kraj je najviac rozvinutý v porovnaní s ostatnými krajmi SR, čo je spôsobené najmä tým, že sa tu vytvára podstatná časť hrubej pridanej hodnoty, pretože svoje sídlo tu má mnoho veľkých podnikov a neustále do tejto oblasti prúdia zahraničné investície na výstavbu nových zahraničných podnikov. Z toho možno usudzovať, že tento kraj bude i naďalej napredovať.

- Prešovský kraj je najzaostalejší v porovnaní s ostatnými krajmi SR. Dôvodov, prečo práve tento kraj je v takej situácii, je viacero. Jedným z nich je to, že je tu nedostatok pracovných príležitostí, ľudia odchádzajú z tohto kraja do krajov s lepším uplatnením (kde majú viacero možností nájsť si prácu), čo je dôsledkom toho, že sa tu vyprodukuje oveľa nižšia hrubá pridaná hodnota. Tento kraj priťahuje aj menej zahraničných investorov.

Z tohto dôvodu sme sa rozhodli aplikovať spomínané metódy na tieto dva kraje. Údaje sme čerpali z internetovej stránky ŠÚ SR (www.statistics.sk):

Tabuľka 1. Vstupné údaje

Ukazovateľ	Bratislavský kraj		Prešovský kraj	
	1997	**2003**	**1997**	**2003**
Hrubá pridaná hodnota b.c. (mil. Sk)	159 186	277 225	58 454	96 569
Počet obyvateľov – stredný stav (fyzické osoby)	618 958	599 815	775 280	793 898

Zdroj: www.statistics.sk

Metóda porovnávania na základe tempa rastu

Najprv si vypočítame tempo rastu extenzitného indikátora hrubá pridaná hodnota (ďalej *HPH*) v porovnávaných krajoch na základe vzťahu (1). Tempo rastu *HPH* v Prešovskom kraji za najnovšie obdobie 1997 – 2003 je 65,21 % a v Bratislavskom kraji 74,15 %. V ďalšom kroku vyčíslime teoretické tempo rastu, ktoré musí Prešovský kraj dosiahnuť, aby sa rozdiely neprehlbovali. Na základe vzťahu (2) je teoretické tempo rastu rovné 201,93 %. Z nasledovnej analýzy vyplýva, že k vyrovnaniu rozdielov medzi Bratislavským a Prešovským krajom dochádza vtedy, ak skutočne dosahované tempo rastu Prešovského kraja 65,21 % je väčšie ako teoretické tempo rastu 201,93 %.

Metóda porovnávania na základe mernej rýchlosti

Základom tejto metódy je porovnávania nie extenzitnej veličiny výstupu, ale intenzitnej veličiny výstupu, ktorá dáva do pomeru extenzitnú veličinu výstupu a počet obyvateľov, teda v našom prípade ide o porovnávanie pomocou indikátora *HPHpc* v Prešovskom a Bratislavskom kraji.

Mernú rýchlosť za sledované obdobie 1997 – 2003 vypočítame na základe vzťahov (4), (5), (6) a jej hodnota je 0,2256. Z toho vyplýva, že merná rýchlosť nadobúda hodnotu menšiu ako 1, čo znamená, že v sledovanom období nedochádza k odstraňovaniu rozdielov v úrovni analyzovaných krajov.

5. Literatúra

GLASER-OPITZOVÁ, H. – MYSLÍKOVÁ, I. 2001: Metodológia a metódy merania sociálno-ekonomickej úrovne regiónov. Bratislava: Infostat, 2001

HURBÁNKOVÁ, Ľ. – PARDELOVÁ, R.: Meranie a analýza úrovne národnej ekonomiky a jej regiónov pomocou pyramidálneho modelu. In: *Ekonomika a informatika* 2/2005. Bratislava: 2006. S.42-54

Regionálne porovnania v Slovenskej republike 2001. Bratislava: ŠÚ SR, 2002

Regionálny hrubý domáci produkt a ukazovatele regionálnych účtov SR 1995 – 2000. Bratislava: ŠÚ SR, 2002

www.statistics.sk

Adresa autora:

Ľubica Hurbánková, Ing., PhD.
Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave
Dolnozemská cesta 1/b
852 35 Bratislava
lubica.hurbankova@euba.sk

Názory vybraných odborníkov na problematiku imigrácie

Jozef Chajdiak, Ján Luha

Abstact: This article brinks results from public enquiry based on questionnaire for patricipants on 11. Slovak Demographical Conference. Main results are concerned on immigration of certain groups of population to Slovak Republik.

1. Úvod

Pri príprave 11. Slovenskej demografickej konferencie (11. SDK), ktorá sa konala 17. až 19. septembra 2007 v hoteli Čingov v Slovenskom raji, s nosnou témou „Migrácia“, vznikol nápad získať názory odborníkov na migráciu a najmä na imigráciu do SR. Po pripomienkovaní bol pripravený dotazník navrhnutý J. Chajdiakom. Dotazník pôvodne vyplnili účastníci 11. SDK. Naviac boli získané názory účastníkov medzinárodnej konferencie aplikovanej štatistiky FERNSTAT 2007, SCHOLA STATISTICA 2 a 3 a kurzov A a B „Finančné analýzy a metódy na podporu rozhodovania v manežérskom riadení“. Dotazník (s celkovými výsledkami uvádzame v prílohe) obsahuje aj otázky o emigrácii, základom je však batéria 33 otázok hodnotiacich súhlas alebo nesúhlas s imigráciou určitých skupín osôb do SR.

2. Analýza základných výsledkov

Prehľadné výsledky za celý súbor 131 respondentov sú uvedené v prílohe. V tejto kapitole sa iba zmienime o otázkach č. 1, 2, 4 a 5, ktoré sa problematike imigrácie a emigrácie venujú viac všeobecne. Vzhľadom na rozsah súboru sa nebudeme podrobne venovať demografickým charakteristikám – výsledky sú v prílohe. Batériu 33 otázok zaradených ako tretiu otázku bližšie analyzujeme v 3. kapitole.

Prvá meritórna otázka mapovala názor na imigráciu ľudí do SR a druhá naopak názor na emigráciu zo SR. Z výsledkov v prílohe vidno, že pri otázke **o1: Mala by Slovenská republika podporovať imigráciu cudzincov do SR?** je najfrekventovanejšia odpoveď 2=mala by aktívne podporovať imig. u špec.sk. osôb a aktívne brzdiť imigr. pri iných špec. sk.osôb (47,7%) a potom 3=mala by nechať imigráciu na prirodzený vývoj (36,9%). Pri otázke **o2: Mala by Slovenská republika podporovať emigráciu občanov zo SR?** je najpočetnejšia odpoveď 3=mala by nechať emigráciu na prirodzený vývoj (64,8%). **o4: Čo si myslíte o medzikultúrnych manželstvách?**, kde je najfrekventovanejšia odpoveď 2=manželstvo, kde môže, ale nemusí dôjsť k nezhodám (iná kultúra/zvyky) (61,5%).

Posledná otázka dotazníka bola otvorená: **5. Rozveďte, prosím, podrobnejšie váš názor na medzikultúrne manželstvá**: Táto otázka dávala možnosť respondentovi rozviesť kategorizovane odpoveď danú otázkou č. 4. Na túto otázku odpovedalo 38% respondentov, ktorí viac-menej rozviedli otázku č.4, pre nedostatok miesta sa v tomto príspevku bližšie týmito odpoveďami nezaoberáme.

Analýzu názorov odborníkov na batériu otázok č. 3 o „afinite“ k imigrácii do SR vybraných skupín občanov uvádzame v nasledovnej kapitole.

3. Analýza názorov odborníkov na imigráciu do SR

Podrobnejšie sme problematiku imigrácie skúmali pomocou batérie 33 otázok. Respondenti mali vyjadriť mieru súhlasu resp. nesúhlasu s imigráciou konkrétnych skupín populácie. **Otázka: 3. Na stupnici -2,-1, 0, +1, +2 vyjadrite svoj nesúhlas/súhlas s imigráciou nasledovných skupín osôb do Slovenskej republiky.**
Táto batéria otázok jednoducho meria **úroveň afinity** s imigráciou vybraných skupín osôb. Mierou je priemer získaných odpovedí. Kladné hodnoty vyjadrujú určitý prevládajúci súhlas s imigráciou danej skupiny osôb, záporné hodnoty zase prevládajúci nesúhlas a hodnoty blízke nule zase prakticky indiferentný názor na imigráciu danej skupiny osôb. Výsledky sme usporiadali a uvádzame v tabuľke:

Tabuľka afinity s imigráciou

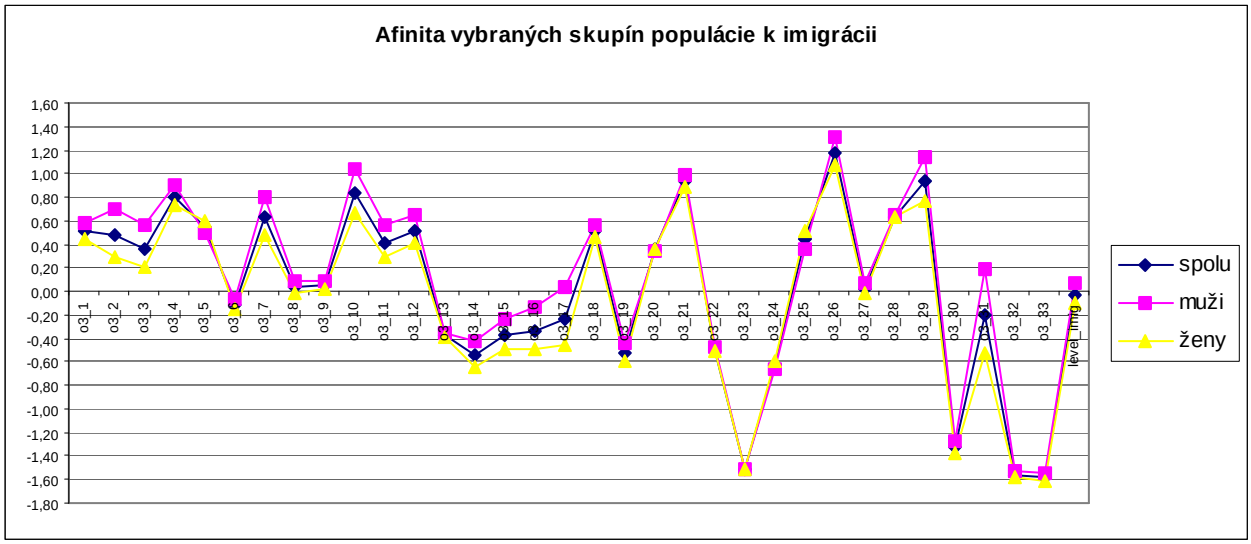
skupina:	**Mean**	**Std. Deviation**
o3_26 osôb s vysokoškolským vzdelaním	1,17	0,94
o3_29 netrestaných osôb	0,94	0,99
o3_21 osôb zdravých	0,93	0,97
o3_10 Slovanov	0,83	0,95
o3_4 osôb vo veku od 16 do 29 rokov	0,80	1,06
o3_28 bohatých osôb	0,64	0,99
o3_7 osôb bielej farby pleti	0,62	0,92
o3_5 osôb vo veku od 30 do 49 rokov	0,55	1,03
o3_12 Chorvátov	0,52	0,94
o3_18 kresťanov	0,51	0,93
o3_1 mužov	0,51	0,96
o3_2 žien	0,47	0,94
o3_25 osôb so stredným vzdelaním	0,44	1,02
o3_11 Srbov	0,41	1,00
o3_3 osôb vo veku do 15 rokov	0,37	1,16
o3_20 osôb bez vyznania	0,35	0,86
o3_9 osôb iných farieb pleti	0,05	1,01
o3_8 osôb čiernej farby pleti	0,04	1,02
o3_27 chudobných osôb	0,02	1,10
o3_6 osôb vo veku viac ako 50 rokov	-0,12	1,15
o3_31 osôb trestaných za politické trestné činy	-0,20	1,25
o3_17 Kórejcov	-0,24	1,11
o3_16 Vietnamcov	-0,34	1,15
o3_13 Albáncov	-0,38	1,15
o3_15 Číňanov	-0,38	1,17
o3_22 osôb chorých	-0,50	1,11
o3_19 osôb s islamským vierovyznaním	-0,53	1,17
o3_14 Arabov	-0,55	1,16
o3_24 osôb bez vzdelania	-0,62	1,09
o3_30 osôb trestaných za nepolitické trestné činy	-1,33	1,01
o3_23 osôb chorých s nákazlivými prenosnými chorobami	-1,51	0,91
o3_32 členov radikálnych politických strán	-1,56	0,78
o3_33 členov nacionalistických politických strán	-1,58	0,76
priemerná celková afinita k imigrácii	**-0,02**	**0,70**

Už jednoduché usporiadanie dáva dobrý pohľad na diferencované názory na imigráciu jednotlivých skupín osôb. Podľa odborníkov, účastníkov 11. Slovenskej demografickej konferencie a vybraných akcií SŠDS približne v rovnakom období ako bola spomínaná konferencia, máme záujem o imigráciu vysokoškolsky vzdelaného, netrestaného, zdravého, slovana, mladého, bohatého, bielej farby pleti. Na opačnom konci spektra je člen nacionalistickej alebo radikálnej politickej strány, chorý s nákazlivou prenosnou chorobou, trestaný, bez vzdelania, arab, s islamským vierovyznaním a chorý. Jednoznačme sme proti imigrácii členov nacionalistických alebo radikálnych politických strán, osôb chorých s nákazlivými, prenosnými chorobami a trestaných osôb. Negatívny, hoci v menšej miere je tiež postoj k imigrácii osôb bez vzdelania, arabov, osôb s islamským vierovyznaním a chorých.
V tabuľke uvádzame aj priemernú afinitu k imigrácii získanú ako priemer za všetky sledované otázky. Vidno, že celkovo je priemerná afinita k imigrácii mierne negatívna – takmer indiferentná. Ako vidno z tabuľky analýzy reliability je vyjadrenie priemernej celkovej afinity k imigrácii vysoko reliabilné:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,952	,951	33

Grafické vyjadrenie za celý súbor ako aj podľa pohlavia je v nasledujúcom grafe.



Signifikantné rozdiely v názoroch podľa pohlavia sme zaznamenali pri:
o3_2 Súhlas/nesúhlas s imigráciou žien - muži boli „afinnejší“ (0,70 / 0,29),
o3_10 Súhlas/nesúhlas s imigráciou Slovanov – taktiež ochotnejšie súhlasili s imigráciou muži (1,03 / 0,66),
o3_17 Súhlas/nesúhlas s imigráciou Kórejcov – muži boli takmer indiferentní, ale ženy boli proti (0,03 / -0,45),
o3_29 Súhlas/nesúhlas s imigráciou netrestaných osôb - ochotnejšie súhlasili s imigráciou muži (1,15 / 0,77),
o3_31 Súhlas/nesúhlas s imigráciou osôb trestaných za politické trestné činy – ochotnejší boli muži, kým ženy boli proti (0,20 / -053).

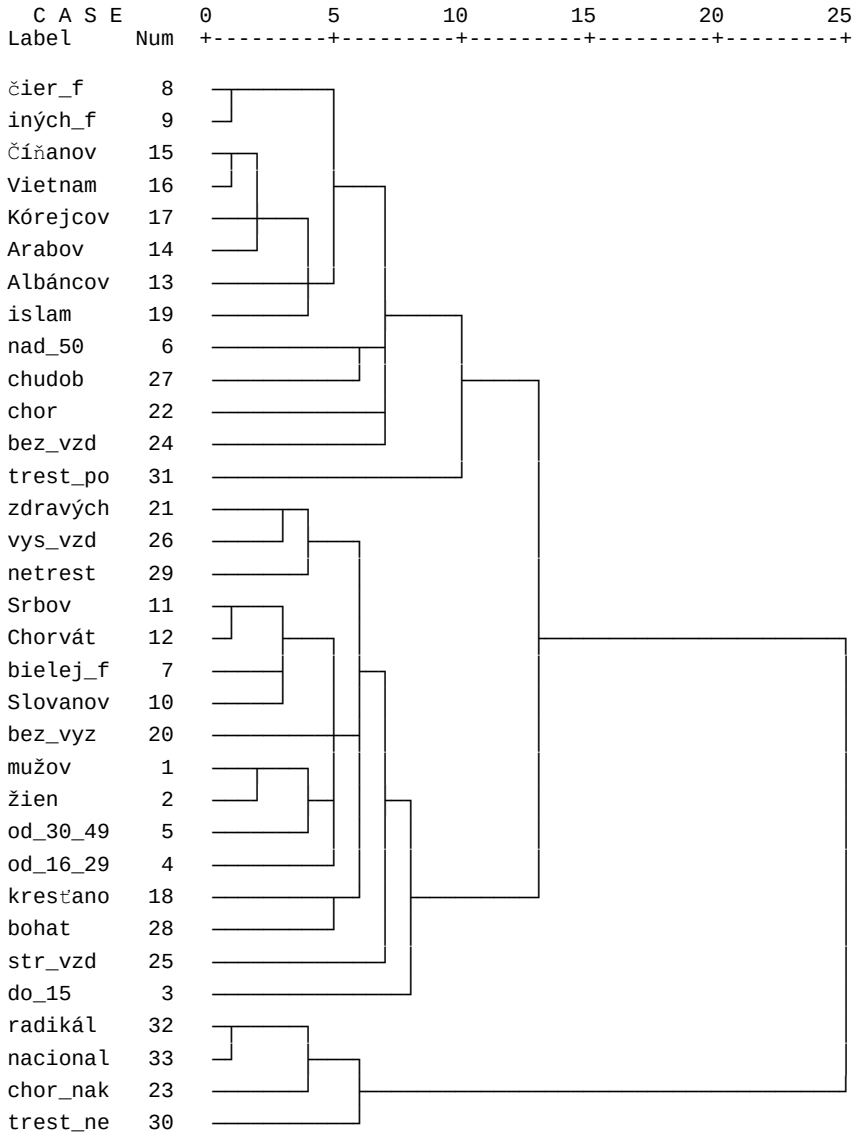
Komplexnejší pohľad na problematiku imigrácie meranú uvedenou batériou otázok získame aplikáciou klastrovej analýzy a faktorovej analýzy.

Najprv uvedieme výsledky klastrovej analýzy, kde vidno zaujímavé vzťahy medzi premennými vyjadrené v zhlukoch blízkych skupín osôb z hľadiska názorov na ich imigráciu. V dendograme sme museli použiť skratky, ktoré však čitateľ ľahko identifikuje z prehľadnej tabuľky uvedenej prv a podľa čísla otázky.

* * * * * * * * * ** * * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine



Na poslednej úrovni zhlukovania sme získali dva zhluky, pričom v jednom sa zoskupili 4 skupiny osôb (o3_32 členov radikálnych politických strán, o3_33 členov nacionalistických politických strán, o3_23 osôb chorých s nákazlivými prenosnými chorobami, o3_30 osôb trestaných za nepolitické trestné činy), čo veľmi korešponduje s rebríčkom *afinity* uvedenom

hore. Druhý zhluk sa ďalej rozkladá na dva stále početne pomerne veľké zhluky, ktoré sa takisto ďalej členia.

Uvedieme stručne výsledky faktorovej analýzy metódou hlavných komponent s Varimax rotáciou. Získali sme až 7 faktorov, ktoré vysvetľujú 74,2% variability. Výsledky sú taktiež veľmi zaujímavé. Pre obmedzený rozsah príspevku ich podrobnejšie nekomentujeme uvádzame iba výslednú rotovanú tabuľku fatorovej analýzy:

Rotated Component Matrix(a)

	Component						
Súhlas/ nesúhlas s imigráciou:	1	2	3	4	5	6	7
o3_15 Číňanov	**,913**						
o3_14 Arabov	**,905**						
o3_16 Vietnamcov	**,898**						
o3_17 Kórejcov	**,810**						
o3_13 Albáncov	**,776**						
o3_19 osôb s islamským vierovyznaním	**,770**						
o3_8 osôb čiernej farby pleti	**,749**						
o3_9 osôb iných farieb pleti	**,739**						
o3_24 osôb bez vzdelania	**,526**						
o3_22 osôb chorých	**,509**						
o3_27 chudobných osôb	**,463**						
o3_26 osôb s vysokoškolským vzdelaním		**,774**					
o3_28 bohatých osôb		**,769**					
o3_21 osôb zdravých		**,759**					
o3_18 kresťanov		**,651**					
o3_29 netrestaných osôb		**,602**					
o3_20 osôb bez vyznania		**,493**					
o3_1 mužov			**,704**				
o3_2 žien			**,671**				
o3_4 sôb vo veku od 16 do 29 rokov			**,656**				
o3_3 osôb vo veku do 15 rokov			**,583**				
o3_7 osôb bielej farby pleti			**,544**				
o3_12 Chorvátov				**,828**			
o3_11 Srbov				**,823**			
o3_10 Slovanov				**,778**			
o3_33 členov nacionalistických politických strán					**,892**		
o3_32 členov radikálnych politických strán					**,890**		
o3_30 osôb trestaných za nepolitické trestné činy					**,665**		
o3_23 osôb chorých s nákazlivými prenosnými chorobami					**,541**		
o3_6 osôb vo veku viac ako 50 rokov						**,629**	
o3_25 osôb so stredným vzdelaním						**,552**	
o3_5 osôb vo veku od 30 do 49 rokov						**,550**	
o3_31 osôb trestaných za politické trestné činy							**,753**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 8 iterations.

4. Závery

Jednoznačme sme proti imigrácii členov nacionalistických alebo radikálnych politických strán, osôb chorých s nákazlivými, prenosnými chorobami a trestaných osôb.

Máme záujem o imigráciu vysokoškolsky vzdelaného, netrestaného, zdravého, slovana, mladého, bohatého, bielej farby pleti.

5. Literatúra

Chajdiak J. (2003): Štatistika jednoducho. Statis Bratislava 2003.
Chajdiak J. (2005): Štatistické úlohy a ich riešenie v Exceli. STATIS, Bratislava.
Kanderová M., Úradníček V. (2005): Štatistika a pravdepodobnosť pre ekonómov, 1. časť. OZ FINANC, Banská Bystrica 2005.
Kanderová M., Úradníček V. (2007): Štatistika a pravdepodobnosť pre ekonómov, 2. časť. OZ FINANC, Banská Bystrica 2007.
Luha, J. (2003): Matematickoštatistické aspekty spracovania dotazníkových výskumov. In: Štatistické metódy vo vedecko-výskumnej práci ´2003. Bratislava : SŠDS, 2003. ISBN 80-88946-32-8
Luha J. (2005): Viacrozmerné štatistické metódy analýzy kvalitatívnych znakov. *EKOMSTAT 2005*, Štatistické metódy v praxi.SŠDS Trenčianske Teplice 22. – 27. 5. 2005.
Luha, J.(2006): Štatistické metódy analýzy kvalitatívnych znakov. FORUM STATISTICUM SLOVACUM 2/2006. SŠDS Bratislava. ISSN 1336-7420
Stankovičová I., Vojtková M. (2007): Viacrozmerné štatistické metódy pre ekonómov a manažérov. IURA, Bratislava 2007.

Príloha – dotazník s výsledkami:

NÁZORY ÚČASTNÍKOV 11. SLOVENSKEJ DEMOGRAFICKEJ KONFERENCIE O IMIGRÁCII A EMIGRÁCII OSÔB DO A ZO SLOVENSKEJ REPUBLIKY

Údaje o respondentovi

Meno (nemusí sa vyplňovať): _____

Pohlavie: muž (45%), žena (55%)

Vek: _____(22 až 62 rokov)_____

Štátna príslušnosť: SR (85,3%), ČR (14,7%), iná (uveďte).....

Vzdelanie: stredoškolské (12,6%), vysokoškolské (74,8%), vedecké (12,6%)

Vierovyznanie: bez vyznania (38,7%), rímskokatolícke (47,8%), iné kresťanské (10,8%),iné (2,7%)

Otázky

1. Mala by Slovenská republika podporovať imigráciu cudzincov do SR?

a) vo všeobecnosti aktívne áno (9,2%)
b) mala by aktívne podporovať imigráciu u špecifických skupín osôb a aktívne brzdiť imigráciu pri iných špecifických skupinách osôb (47,7%)
c) mala by nechať imigráciu na prirodzený vývoj (36,9%)
d) mala by aktívne brzdiť imigráciu osôb vo všeobecnosti (6,2%)

2. Mala by Slovenská republika podporovať emigráciu občanov zo SR?

a) vo všeobecnosti aktívne áno (3,1%)
b) mala by aktívne podporovať emigráciu u špecifických skupín obyvateľov a aktívne brzdiť emigráciu pri iných špecifických skupinách obyvateľov (18,8%)
c) mala by nechať emigráciu na prirodzený vývoj (64,8%)
d) mala by aktívne brzdiť emigráciu obyvateľov vo všeobecnosti (13,3%)

3. Na stupnici -2,-1, 0, +1, +2 vyjadrite svoj nesúhlas/súhlas s imigráciou nasledovných skupín osôb do Slovenskej republiky

|----------|----------|----------|----------|

-2	-1	0	+1	+2
úplne nesúhlasím	skôr nie ako áno		skôr áno ako nie	úplne súhlasím

1. mužov ..... ___0,51___
2. žien ..... ___0,47___
3. osôb vo veku do 15 rokov ..... ___0,37___

4. osôb vo veku od 16 do 29 rokov ..... ___0,80___
5. osôb vo veku od 30 do 49 rokov ..... ___0,55___
6. osôb vo veku viac ako 50 rokov ..... ___-0,12___
7. osôb bielej farby pleti ..... ___0,62___
8. osôb čiernej farby pleti ..... ___0,04___
9. osôb iných farieb pleti ..... ___0,05___
10. Slovanov ..... ___0,83___
11. Srbov ..... ___0,41___
12. Chorvátov ..... ___0,52___
13. Albáncov ..... ___-0,38___
14. Arabov ..... ___-0,55___
15. Číňanov ..... ___-0,38___
16. Vietnamcov ..... ___-0,34___
17. Kórejcov ..... ___-0,24___
18. kresťanov ..... ___0,51___
19. osôb s islamským vierovyznaním ..... ___-0,53___
20. osôb bez vyznania ..... ___0,35___
21. osôb zdravých ..... ___0,93___
22. osôb chorých ..... ___-0,50___
23. osôb chorých s nákazlivými prenosnými chorobami ___-1,51___
24. osôb bez vzdelania ..... ___-0,62___
25. osôb so stredným vzdelaním ..... ___0,44___
26. osôb s vysokoškolským vzdelaním ..... ___1,17___
27. chudobných osôb ..... ___0,02___
28. bohatých osôb ..... ___0,64___
29. netrestaných osôb ..... ___0,94___
30. osôb trestaných za nepolitické trestné činy ..... ___-1,33___
31. osôb trestaných za politické trestné činy ..... ___-0,20___
32. členov radikálnych politických strán ..... ___-1,56___
33. členov nacionalistických politických strán ..... ___-1,58___

4. Čo si myslíte o medzikultúrnych manželstvách?

a) manželstvo ako každé iné (27,7%)
b) manželstvo, kde môže, ale nemusí dôjsť k nezhodám (iná kultúra/zvyky) (61,5%)
c) manželstvo, ktoré prinesie zaručene problémy (iná kultúra/zvyky) (10,8%)

5. Rozveďte, prosím, podrobnejšie váš názor na medzikultúrne manželstvá:

_____

_____

Age structure of the inhabitants of the estates of the Eggenbergs in 1651 according to the List of Serfs according to Faith

Kačerová Eva

Abstract: The List of Inhabitants according to Faith of 1651 captured the situation immediately following the end of the Thirty Years War. On the Eggenberg estates children are recorded systematically only from the age of 12. The basic prerequisite for estimating the number of children under the age of 12 was that women (men) of the same age in various parts of the estates had on average the same or at least a similar number of children in the individual age groups.

Key words: age structure, 17th century,List of Inhabitants according to Fait, 17th century

1. Introduction

The List of Inhabitants according to Faith of 1651 is one of the most valuable sources of historical demography. It captured the situation immediately following the end of the Thirty Years War, covers extensive territory of the Czech Lands and was compiled basically in a uniform manner and on a single date.

2. Text of paper

We shall deal further only with the South Bohemian estates of the Eggenbergs, where 10,595 persons were recorded by the List. The acquisition of as precise a picture as possible of the age structure of the population as it was in the middle of the 17th century, however, is rendered difficult by several factors at once, the most important of these being that not all the population was recorded. Some were not recorded systematically (the clergy and the military) and some might have escaped the list by mistake (farm-labourers, hands, journeymen, beggars and travelling authorities). A further pitfall is the inaccuracy of the ages given, or rather their accumulation to certain values. In the time of the compiling of the List age did not play such a part as it does nowadays and it was not even very important for the purpose of the List. In addition elderly people often did not even know their precise age. Age was rounded up in tens, less often also to numbers ending in five and at a younger age (roughly up to 30 years), under the influence of the sexagesimal system, also to multiples of six and twelve.

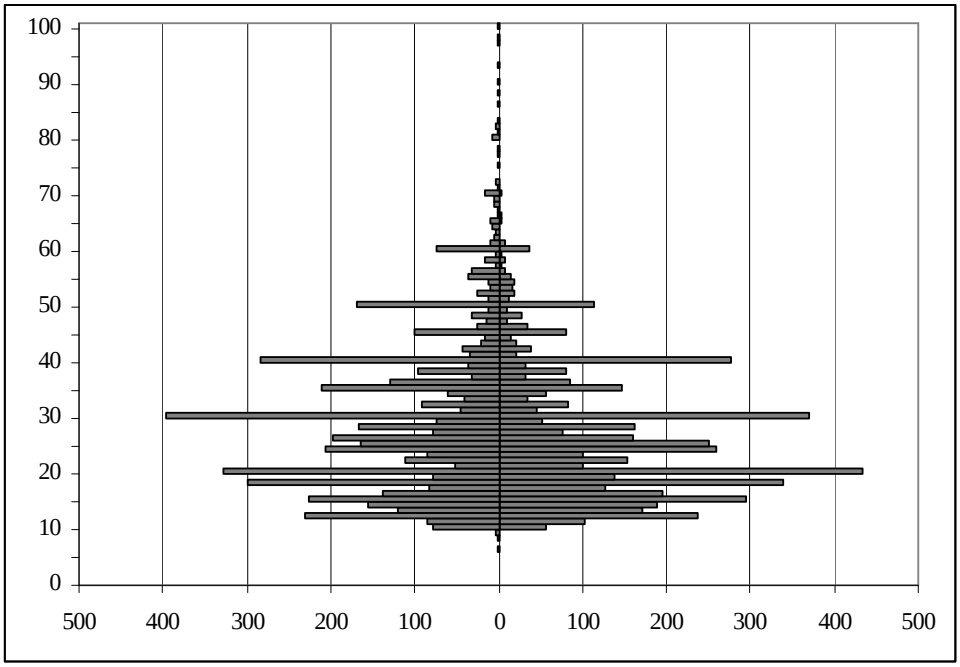
Of the women on the Eggenberg estate in the 30–39 age-group 38% were aged 30 and 15% were aged 35. The age of women from the group of 40–49-year-olds was rounded up somewhat more: women aged 40 accounted for 52% of all those in their forties and women of 45 were 15%. Also interesting is the connection of age with certain social categories. Second-time widows were often ascribed the age of 40 years and if they were older than 60 years. In the case of “lone women” with children, who were probably unmarried mothers, we again often find the age of 30 years and so on. Odd numbers were more popular than even ones.

The level of distortion may be measures by the index of age accumulation *ik:*

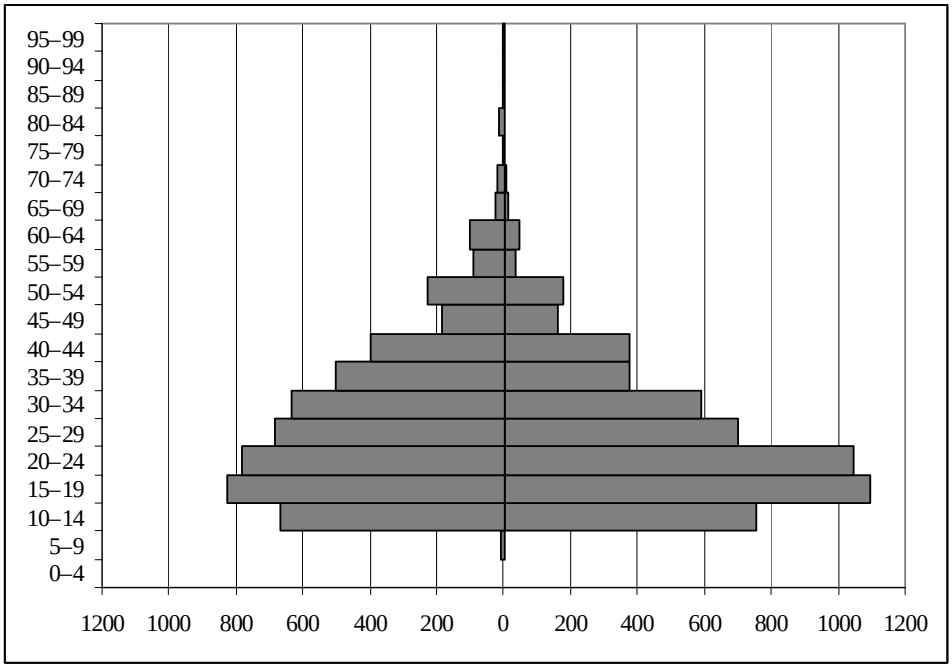
$$ik^{ž} = \frac{5 * \sum_{0}^{7} S^{p}_{25+5x}}{\sum_{23}^{62} S^{p}_{x}} \qquad (1)$$

The rounding-up of ages has a negative impact on the results of the study of the age structure of the population. This distortion can be reduced by the use of ten-year age intervals in which the most frequented value will always be in the middle, i.e. 5–14, 15–24, etc. Unfortunately, in order to be able to compare with present-day statistics and other authors dealing with the analysis of the List of Serfs according to Faith of 1651, it is also necessary to retain the usual age intervals: 0–4, 5–9, 10–14, etc. In some studies we may still find the ten-year intervals 0–9, 10–19, etc. The selection of the age intervals may influence the significance of the rounding-up in the resultant age structure.

Picture 1: Age structure, one year intervals



Picture 2: Age structure, five years intervals



The most complicated pitfall for the researcher studying the age structure of the population on the basis of the List is the sub-registration of children before the age at which they can make confessions on certain estates (or in certain communities). On the Eggenberg estates children are recorded systematically only from the age of 12 and children under the age of 10 were recorded only in exceptional cases.

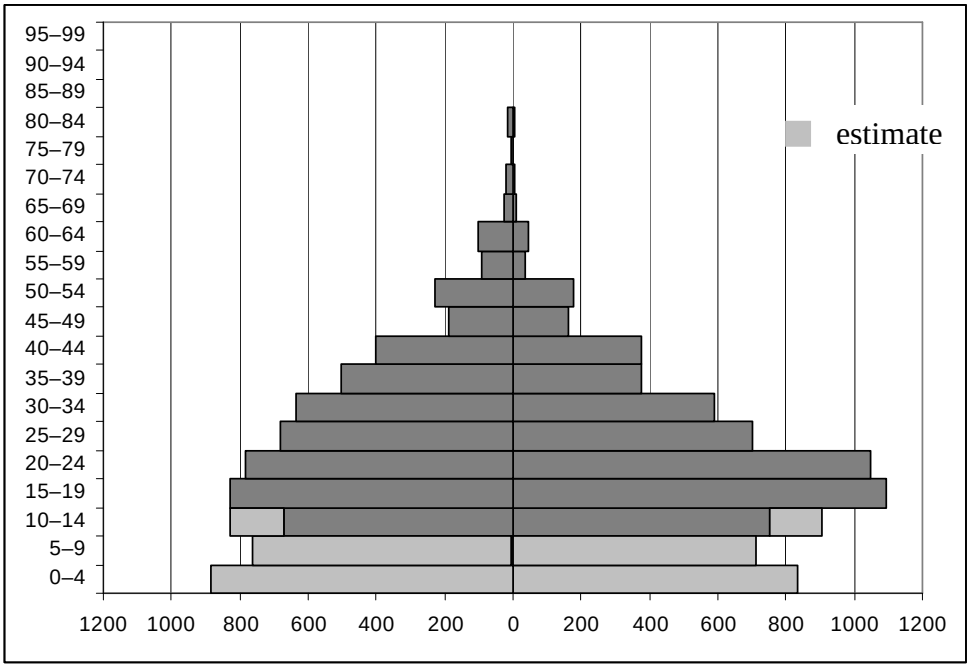
The basic prerequisite for estimating the number of children under the age of 12 was that women (men) of the same age in various parts of the estates had on average the same or at least a similar number of children in the individual age groups. As the initial data for this method per analogiam use was made of the records of the List from the estate of Radenín and Choustník, where children were recorded from birth and which is also part of the Bechyně region.

For checking the described estimate of the number of children under the age of 12 a second estimate was created in which the percentage representation and the average number of children in the various age-groups was related not to the age of the woman (mother), but to the age of the man (father). This was also done because a relatively large number of children lived with the father, who had married a young girl after becoming a widower. If this is really how things were in the family we can only guess from the information available in the List according to the ages of the children and spouses. The two estimates differ only minimally and so the calculated percentage shares would also be almost the same. However, there are relatively large differences in the individual age-groups of children.

Tab. 1. The estimate of the numer of children under the age of 12

Estimate based on age of	Age of child							
	0	1–4	5–9	10–11	0	1–4	5–9	10–11
	numbers				%			
woman	412	1394	1476	369	2,92	9,90	10,48	3,78
man	470	1245	1479	630	3,32	8,80	10,45	4,45

Picture 3: Age structure, five years intervals



3. Conclusions

The accuracy of the results of the estimate of the number of children according to the age of the woman and according to the age of the man cannot be precisely verified, but it is possible to compare them with the estimates of other authors or else with the proportions of children under 10 and under 12 in places where all the children were recorded and which have already been processed in literature. From the comparison it is clear that the estimates of the number of children on the Eggenberg estate generally do not deviate in any way from the estimates of other authors and the proportions of the child element on estates recording children from birth. In spite of this, especially for the children under 10, the estimate for the Eggenberg estate is lower than the other estimates, but nevertheless this difference is in no way extreme in comparison with the ascertained proportions of children on other estates.

4. Sources

Soupis poddaných podle víry - Bechyňsko

Adresa autora:

Eva Kačerová, RNDr.

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra demografie

130 67 Praha 3

kacerova@vse.cz

Optimalizácia portfólia investičných nástrojov v prostredí MS Excel

Kanderová Mária

Abstract: The aim of the paper is to present two different covariance matrix mean estimation approaches. Estimated covariance matrices will be used to portfolio optimization in mean-variance space. The described technique will be applied to selected shares listed on Slovak capital market.

Keywords: *Portfolio Optimization, Exponential Weighted Estimate, Minimum Variance Portfolio Frontier, Efficient Portfolio*

1. Úvod

Pojem portfólio označuje skladbu aktív v držbe individuálneho investora – buď fyzickej, alebo právnickej osoby. Hlavným podnetom na formovanie portfólií je diverzifikácia. Diverzifikácia je alokácia investičných fondov do variety aktív s cieľom znížiť rizikovosť spojenú s jednotlivými cennými papiermi.

Úspešnosť investora pri investovaní sa meria výškou výnosu. Sú dve primárne formy výnosu investora, a to dôchodok a kapitálový zisk. Dôchodok je výnos investora vo forme platieb obdŕžaných ako dôsledok vlastníctva aktíva. Kapitálový zisk je výnos investora spôsobený zmenou ceny aktíva, ktoré má vo vlastníctve.

Nech P_t je cena aktíva v čase t a r_t výnos aktíva v čase t. Potom pre výnos aktíva platí:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \approx ln\left(\frac{P_t}{P_{t-1}}\right). \qquad (1.1)$$

Finančná teória definuje riziko ako disperziu neočakávaných stavov a to na obidve strany od očakávanej (strednej) hodnoty. Z toho vyplýva, že ako kladné, tak aj záporné hodnoty sú považované za zdroj rizika. V súlade s finančným chápaním rizika vhodnou mierou na jeho kvantifikovanie je rozptyl resp. smerodajná odchýlka.

Pri výpočtoch s takto definovaným rizikom sa predpokladá, že skutočné výnosy sú normálne rozdelené. V skutočnosti je rozdelenie výnosov špicatejšie ako normálne rozdelenie, takže nízke kladné a záporné výnosy sa objavujú častejšie ako predpokladá normálne rozdelenie. Okrem toho rozdelenie výnosov má „tučnejšie“ konce ako normálne rozdelenie, to znamená, že pravdepodobnosť výskytu extrémnych výnosov (kladných alebo záporných) je vyššia ako pri normálnom rozdelení. Empirické štúdie dokazujú, že v praktických situáciách s rastúcim počtom pozorovaní sa rozdelenie logaritmov výnosov blíži k normálnemu rozdeleniu. Pri analýze vychádzame z predpokladu normálneho rozdelenia výnosov.

2. Odhad parametrov modelu

Najpoužívanejšie metódy analýzy rizika považujú očakávaný výnos ako indikátor predpokladanej ziskovosti aktíva a rozptyl ako indikátor miery rizika aktíva.

Jeden zo spôsobov ako získať očakávaný výnos aktíva je odhadnúť ho z historických dát.

Nech $r_k(t)$ je výnos aktíva k v čase t, kde k = l,...,n, t = 1,...,T, T je dĺžka časového horizontu. Výnos aktíva r_k je náhodná premenná, očakávaný výnos aktíva sa odhaduje pomocou strednej hodnoty

$$E[r_k] = \frac{1}{T}\sum_{t=1}^{T} r_k(t)\,. \qquad (1.2)$$

Takýto odhad očakávaného výnosu však nie je veľmi reálny. Udalosti, ktoré sa odohrali ku koncu sledovaného obdobia t→T vplývajú na súčasnosť väčšou váhou ako udalosti na začiatku sledovaného obdobia t→0. Toto tvrdenie získava na dôležitosti so zväčšujúcou sa dĺžkou časového horizontu. Odhad výnosov podľa (1.2) priraďuje každému časovému okamihu rovnakú váhu. Presnejšie odhady výnosu získame, ak výnosy vážime parametrom λ, ktorý smerom do minulosti exponenciálne klesá. Potom pre očakávaný výnos aktíva platí:

$$E\left[r_k\right]_E = \frac{\sum_{t=1}^{T}\lambda^{T-t} r_k(t)}{\sum_{t=1}^{T}\lambda^{T-t}} \approx (1-\lambda)\sum_{t=1}^{T}\lambda^{T-t} r_k(t)\,, \qquad (1.3)$$

kde $\lambda \in (0,1)$. Pre λ blízko 1 dávame väčší doraz na novšie informácie. Takto získame reálnejší odhad očakávaného výnosu. Odhad očakávaného výnosu pomocou (1.3) používame pri exponenciálne váženom prístupe.

Riziko aktíva pri klasickom (neváženom) prístupe odhadujeme pomocou disperzie jeho výnosov

$$D[r_k] = \frac{1}{T}\sum_{t=1}^{T}\left(r_k(t) - E[r_k]\right)^2. \qquad (1.4)$$

Pre odhad rizika pri exponenciálne váženom prístupe platí:

$$D\left[r_k\right]_E = \frac{\sum_{t=1}^{T}\lambda^{T-t}\left(r_k(t) - E\left[r_k\right]_E\right)^2}{\sum_{t=1}^{T}\lambda^{T-t}} \approx (1-\lambda)\sum_{t=1}^{T}\lambda^{T-t}\left(r_k(t) - E\left[r_k\right]_E\right)^2. \qquad (1.5)$$

Pre odhad kovariancie dvoch aktív pri klasickom prístupe platí:

$$E\left[r_k, r_l\right] = \frac{1}{T}\sum_{t=1}^{T}\left(r_k(t) - E\left[r_k\right]\right)\times\left(r_l(t) - E\left[r_l\right]\right). \qquad (1.6)$$

Pre odhad kovariancie pri exponenciálne váženom prístupe platí:

$$E\left[r_k r_l\right]_E = (1-\lambda)\sum_{t=1}^{T}\lambda^{T-t}\left(r_k(t) - E\left[r_k\right]_E\right)\left(r_l(t) - E\left[r_l\right]_E\right). \qquad (1.7)$$

Ak portfólio pozostáva z n aktív s váhami $\omega_1, \omega_2, \ldots \omega_n$, potom pre očakávaný výnos a disperziu portfólia platí:

$$E_p = \boldsymbol{\omega}^T \boldsymbol{E} \qquad (1.8)$$

$$D_p = \sigma^2_p = \boldsymbol{\omega}^T \boldsymbol{\Sigma}\, \boldsymbol{\omega} \qquad (1.9)$$

kde $\boldsymbol{\omega}^T$ je 1 x n rozmerný vektor váh (podielov) aktív v portfóliu,

$\boldsymbol{E}$ je n x1 rozmerný vektor očakávaných výnosov aktív

$\boldsymbol{\Sigma}$ je n x n rozmerná kovariančná matica výnosov aktív

3. Optimalizácie v priestore mean - variance

Vzťah medzi očakávaným výnosom a rizikom portfólia ako prvý naformuloval Harry Markowitz. Táto teória sa nazýva moderná teória portfólia (Modern portfolio theory - MPT). Markowitz poukázal na to, že pre investora nie je dôležitý rozptyl výnosov jednotlivých aktív

v portfóliu. To, čo v skutočnosti investora zaujíma, je príspevok jednotlivých aktív k riziku celého portfólia, teda kovariancia medzi jednotlivými aktívami.

Markowitzova úloha výberu portfólia v priestore mean – variance je úlohou kvadratického programovania v tvare:

$$\begin{aligned} &\text{Minimalizovať } \sigma_p^2 = \boldsymbol{\omega}^T \boldsymbol{\Sigma}\, \boldsymbol{\omega} \\ &\text{za podmienok:} \\ &\boldsymbol{\omega}^T \boldsymbol{E} = E_p; \\ &\boldsymbol{\omega}^T \boldsymbol{e} = \boldsymbol{1}; \\ &\omega_k \geq 0, \text{ pre } k=1, 2,...,n. \end{aligned} \tag{1.10}$$

Úloha (1.10) je úlohou minimalizácie rizika pri požadovanej hodnote výnosu pričom sa neuvažuje s krátkym predajom aktív.

Riešením úlohy (1.10) pre rôzne výnosy získame hranicu portfólií s minimálnym rozptylom. Keďže hranica portfólií s minimálnym rozptylom graficky predstavuje kladnú vetvu hyperboly, iba časť portfólií je efektívna. Na overenie, či získané portfólio je efektívne, riešime úlohu maximalizácie výnosu pri riziku, ktoré je výsledkom úlohy na minimalizáciu rizika. Úloha maximalizácie výnosu pri danom riziku:

$$\begin{aligned} &\text{Maximalizovať } \quad E_p = \boldsymbol{\omega}^T \boldsymbol{E} \\ &\text{za podmienok:} \\ &\boldsymbol{\omega}^T \boldsymbol{\Sigma}\, \boldsymbol{\omega} = \sigma_P^2; \\ &\boldsymbol{\omega}^T \boldsymbol{e} = \boldsymbol{1}; \\ &\omega_k \geq 0, \text{ pre } k=1, 2,...,n. \end{aligned} \tag{1.11}$$

Pri výpočtoch budeme pracovať v programe MS Excel, ktorý nám poskytuje všetky základné nástroje a funkcie potrebné pre túto analýzu.

4. Údajová základňa

Pre analýzu boli vybrané akcie podnikov, ktoré sú na BCPB najčastejšie obchodované. Výber obsahuje akcie 9-tich spoločností, ktoré poskytovali dostatočný rozsah dát pre analýzu. Týmito akciovými spoločnosťami sú Slovnaft, VÚB, Plastika, Biotika Slovenská Ľupča, Železiarne Podbrezová, SES Tlmače, OTP Banka Slovensko, Kraft Foods Slovakia, Chemolak. Pre analýzu boli použili mesačné údaje od januára 2003 do marca 2007 vrátane, teda 51 mesačných údajov, po prepočte na výnosy 50 mesačných údajov.

Na základe historických údajov cien akcií uvedených spoločností boli vypočítané základné charakteristiky: kovariančná maticu, smerodajné odchýlky a stredné hodnoty. Pri odhade parametrov boli použité dva prístupy, a to klasický prístup a exponenciálne vážený prístup. Pri exponenciálne váženom prístupe odhadu vstupných parametrov bola zvolená $\lambda = 0{,}97$.

V tabuľkách 1 a 2 sú uvedené odhadnuté parametre očakávaný výnos a smerodajná odchýlka výnosov akcií jednotlivých spoločností. V tabuľke 1 sú parametre odhadované klasickým prístupom a v tabuľke 2 sú parametre odhadované exponenciálne váženým prístupom.

Tabuľka 1: Vstupné parametre – klasický Klasický prístup

	Očakávaný výnos	Smerodajná odchýlka
Slovnaft	2,46%	8,46%
VUB	2,71%	6,32%
Plastika	1,60%	17,51%
Biotika	1,32%	14,11%
Žel.Podb.	3,06%	8,80%
SES Tlmače	3,21%	12,10%
OTP	1,21%	12,39%
Kraft	0,70%	9,85%
Chemolak	2,92%	11,38%

Zdroj: Vlastné spracovanie

Tabuľka 2: Vstupné parametre- exponenciálne vážený prístup

	Očakávaný výnos	Smerodajná odchýlka
Slovnaft	1,45%	6,86%
VUB	1,55%	5,08%
Plastika	1,38%	17,59%
Biotika	0,79%	11,73%
Žel.Podb.	2,44%	7,88%
SES Tlmače	2,34%	10,76%
OTP	0,52%	10,36%
Kraft	0,79%	9,08%
Chemolak	2,52%	10,86%

Zdroj: Vlastné spracovanie

Parametre odhadnuté exponenciálne váženým prístupom vykazujú nižšie hodnoty ako parametre odhadnuté klasickým prístupom, čo znamená, že v súčasnosti vykazuje väčšina spoločnosti nižšie výnosy ako v minulosti a rozdielne odhady parametrov vznikli priradením nižších váh dátam minulým pri exponenciálne váženom prístupe a rovnakým váham pri klasickom prístupe.

Dôležité je analyzovať jednotlivé poklesy individuálne, lebo u jednotlivých spoločností sa intenzita tohto poklesu očakávaní líši. Najväčší pokles očakávaného výnosu zaznamenali spoločnosti Slovnaft, VUB a SES Tlmače, ktorých odhad pri jednotlivých prístupoch líši až o 1%. Tento výrazný pokles bol spôsobený nízkymi výnosmi prípadne stratami v poslednom období, ako aj vysokými výnosmi v minulosti, ktorých vplyv bol zavedením váh oslabený. Výrazné výnosy akcií boli zaznamenané hlavne medzi aprílom 2004 a marcom 2005, kedy slovenský akciový index SAX zaznamenal prudký nárast. Na druhej strane spoločnosť Chemolak zaznamenala len minimálny pokles očakávaného výnosu. Podobne ako odhad očakávaného výnosu poklesla pri exponenciálne váženom prístupe aj odhadovaná smerodajná odchýlka. Tá zaznamenala vo väčšine prípadov výrazný pokles. Výnimkou je Plastika, ktorej výnosy boli ku koncu skúmaného obdobia výrazne rozkolísané, čo spôsobilo mierny nárast smerodajnej odchýlky odhadnutej pomocou exponenciálne váženého prístupu oproti jej odhadu pri klasickom prístupe.

4. Výsledky modelu

V prvom kroku sme riešili úlohu minimalizácie rizika pri požadovanom výnose(1.10). Za požadovaný výnos sme zvolili očakávaný výnos indexu SAX odhadnutý z mesačných údajov za rovnaké obdobie ako skúmaný súbor akcií, ktorý je rovný 2,15%. Pri exponenciálne váženom prístupe je odhad očakávaného výnosu indexu SAX rovný 1,24%. K ohraničeniam sme pridali rozpočtové obmedzenie na maximálny podiel majetku investovaného do jedného cenného papiera. Maximálny podiel bol ohraničený na 20% podľa pravidiel obmedzenia a rozloženia rizika pre uzavretý podielový fond, ktoré sú upravené v zákone o kolektívnom investovaní (§ 66 zákona č. 594/2004 Z. z.).

Úlohu minimalizácie rizika a následne úlohu maximalizácie výnosu sme riešili v MS Excel pomocou nástroja „Solver“. Výsledky sú v tabuľke 3.

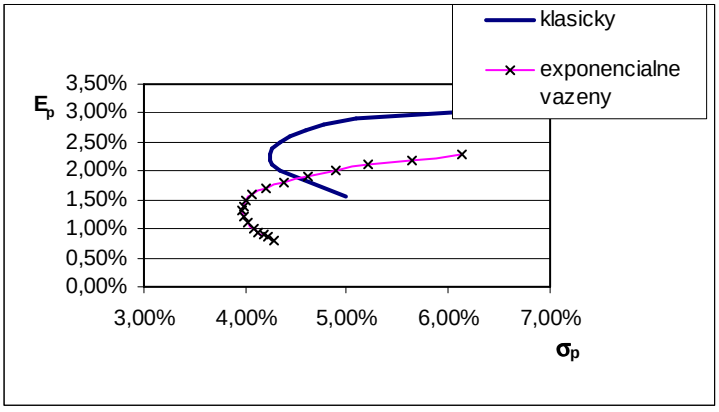
Tabuľka 3: Výsledky optimalizácie portfólia

	Klasický odhad		Exponenciálne vážený odhad	
	minimalizácia rizika	**maximalizácia výnosu**	**minimalizácia rizika**	**maximalizácia výnosu**
Akcia	podiely aktív	podiely aktív	podiely aktív	podiely aktív
Slovnaft	15,93%	16,19%	18,12%	20,00%
VUB	20,00%	20,00%	20,00%	20,00%
Plastika	4,09%	4,00%	3,15%	2,91%
Biotika	8,99%	8,62%	15,61%	6,77%
Zel.Podb.	15,38%	15,85%	7,70%	18,65%
SES Tlm.	0,00%	0,00%	0,00%	0,23%
OTP	2,01%	1,47%	11,73%	0,00%
Kraft	20,00%	20,00%	20,00%	16,62%
Chemolak	13,59%	13,87%	3,70%	14,82%
očakávaný výnos	**2,15%**	**2,17%**	**1,24%**	**1,66%**
smerodajná odchýlka	**4,39%**	**4,39%**	**4,33%**	**4,33%**

Zdroj: Vlastné spracovanie

Portfólia získané riešením úlohy na minimalizáciu rizika pri požadovanom výnose neboli efektívne, pretože v obidvoch prípadoch (pri klasickom odhade aj exponenciálne váženom odhade) pri danom riziku existuje portfólio s vyšším výnosom. Pri porovnaní exponenciálne váženého prístupu s klasickým prístupom zisťujeme, že portfólio pri ktorom vstupné parametre boli odhadnuté klasickým prístupom vykazuje vyšší výnos pri mierne vyššom riziku ako portfólio s exponenciálne váženými odhadmi. Rozdielnosť výsledkov optimalizácie portfólia v závislosti od metódy odhadu je zrejmá aj z grafu 1.

Graf 1: Hranice portfólií s minimálnym rozptylom



Zdroj: Vlastné spracovanie

Z grafu vidíme významný posun v smere osi „výnos portfólia“ v porovnaní s posunom v smere osi „riziko portfólia“.

5. Záver

Cieľom príspevku bolo ukázať dva možné spôsoby odhadu vstupných parametrov pri optimalizácií portfólia. Z výsledkov je zrejmé, že voľba prístupu k odhadu vstupných parametrov ovplyvňuje výsledok optimalizácie portfólia. Neexistuje jednoznačné

odporúčanie, ktorý prístup je vhodnejší. Pri exponenciálne váženom prístupe odhadnuté parametre sa viac približujú k realite, pretože novším údajom je priradená vyššia váha ako starším údajom. Na druhej strane, ak máme akciu, s ktorou sa v poslednom období neobchodovalo, alebo v poslednom období došlo k výraznej jednorázovej zmene, odhad nemusí zodpovedať realite. V takomto prípade by bol vhodnejší klasický odhad. Nevýhodou exponenciálne váženého prístupu je prácnosť pri odhade kovariančnej matice, ktorú nemôžeme odhadnúť pomocou štandardných softvérových nástrojov pre odhad kovariančnej matice.

6. Literatúra

1. BREADLEY, R.A. – MYERS, S.C. 1999. Teorie a praxe firemních financií. Praha: East Publishing, 1999. ISBN 80-85605-24-4.
2. ELTON, E.J. – GRUBER, M.J. 2003. Modern portfolio theory and investment analysis. Willey, 2003. ISBN 978-0-471-23854-6.
3. MLYNAROVIČ, V. 2001. Finančné investovanie. Bratislava: Iura Edition, 2001. ISBN 80-89047-16-5.
4. POPJAK, J. 2007. Kvantifikácia rizika cenných papierov. In: Diplomová práca. Ekonomická fakulta, UMB Banská Bystrica 2007.
5. ZMEŠKAL, Z. a kol. 2004. Finanční modely. Praha: Ekopress, 2004. ISBN 80-86119-87-4

Adresa autora

Mária Kanderová, Ing., PhD.
Katedra kvantitatívnych metód a informatiky
Ekonomická fakulta, UMB
Tajovského 10
974 11 Banská Bystrica
maria.kanderova@umb.sk

Exploračná analýza meraní hĺbky vodnej hladiny v hydrogeologických vrtoch na zosuve v Ľubietovej

Samuel Koróny[1], Roberta Prokešová[1], Alžbeta Medveďová[2]

Abstract: The paper provides basic exploratory data analysis of groundwater level depths in seven hydrogeological drill holes at Ľubietová landslide area during 2007 including graphical presentation in boxplots and sequence plots. From its results it is clear that: depths are normally distributed, have not equal variances, groundwater level depths in hydrogeological drill holes 2, 5A and 6A form homogenous group, water volume in drill holes 1, 2, 5A, 6A and 7 decreased during 2007.

Key words: *Exploratory data analysis, Landslides, Monitoring methods*

1. Úvod

Zosuvy predstavujú veľkú skupinu nie celkom jednotne definovaných svahových pohybov. Najvšeobecnejšiu definíciu zosuvov navrhol Cruden (1991), podľa ktorého možno ako zosuv označiť gravitáciou riadený pohyb masy hornín, pôdy alebo sutín dolu svahom. Vznik zosuvov môže byť podmienený súhrou viacerých faktorov, napríklad morfológiou svahu, priaznivou geologickou stavbou, nevhodnými ľudskými zásahmi do prírodného prostredia apod. K samotnej aktivizácii pohybov dochádza zväčša vplyvom nejakého externého podnetu, tzv. spúšťacieho faktora. Popri seizmickej aktivite najčastejšími spúšťacími faktormi sú faktory klimatické (intenzívne zrážky, rýchle topenie snehu, dlhé zrážkovo bohaté obdobia alebo ich kombinácie), ktoré ovplyvňujú hydrologické pomery na svahoch, náchylných na zosúvanie. Ich priamym dôsledkom je saturácia svahového materiálu (pôda, sedimenty) vodou, resp. zvýšenie hladiny podzemnej vody. Modely niektorých autorov pritom preukázali, že zvýšenie hladiny podzemnej vody nad potenciálnu šmykovú plochu môže až zdvojnásobiť riziko vzniku zosuvu (Fleming & Varnes 1991).

Napriek tomu, že negatívna úloha zrážkových eventov pri aktivizácii zosuvov je známa, nie je jednoduché ju empiricky zhodnotiť a zovšeobecniť. Každý zosuv je totiž výsledkom jedinečnej kombinácie rôznych faktorov (príčinných aj spúšťacích) čo je hlavnou príčinou obtiažnosti v časovo-priestorovom predpovedaní zosuvnej udalosti. Už len samotný prísun vody do svahu v podobe zrážok je výrazne modifikovaný viacerými faktormi, ktoré nie je jednoduché hodnotiť, napr. povrchový odtok, evapotranspirácia, erózne procesy pri mimoriadne intenzívnych zrážkach, kombinácia topenia snehu a tekutých zrážok apod. Najzložitejší problém pritom predstavujú zosuvy s hlbšie založenou šmykovou plochou (> 5 m) a zosuvy, ktoré vznikli remobilizáciou starších zosuvov. V oboch prípadoch sa jedná o územia so zložitou a často oneskorenou reakciou na zrážkové udalosti (napr. Wasowski 1998, Iverson 2000).

Jedným z najznámejších (negatívne) zosuvov na Slovensku je tzv. katastrofálny ľubietovský zosuv (pri obci Ľubietová na strednom Slovensku). Prívlastok „katastrofálny“ si vyslúžil vďaka materiálnym škodám, ktoré v čase svoje aktivizácie na jar 1977 spôsobil. Za hlavný spúšťací faktor ľubietovského zosuvu býva považovaná zrážková anomália v zimnom období 1976/1977. Prekročenie mesačných zrážkových úhrnov o 50 – 120% viedlo

[1] Samuel Koróny, Roberta Prokešová, Ústav vedy a výskumu UMB
[2] Alžbeta Medveďová, Katedra geografie a krajinnej ekológie, Fakulta prírodných vied UMB
Príspevok bol spracovaný v rámci grantovej úlohy VEGA 1/4052/07.

k aktivizácii pohybu o rýchlosti 2,5 m za deň (Nemčok 1982). Vytvoril sa prúdový zosuv s dĺžkou 1300 a maximálnou šírkou 400 m. Hĺbka šmykovej plochy bola určená na 6 – 8 m v čelnej a na 30 m v odlučnej časti zosuvu (Fussgänger et al. 1978). Zosuv bol urýchlene stabilizovaný sieťou odvodňovacích rigolov a šikmých vrtov. Mnohé z týchto opatrení však už stratili svoju funkciu. Ľubietovský zosuv je pritom klasickým príkladom hlbokého zosuvu so zložitou stavbou, kde sa dá očakávať komplexná a oneskorená reakcia na spúšťacie impulzy. Akékoľvek poznanie jeho režimu môže priniesť nové zovšeobecňujúce závery. Príspevok predstavuje prvé štatistické zhodnotenie niektorých našich meraní za rok 2007.

2. Dáta

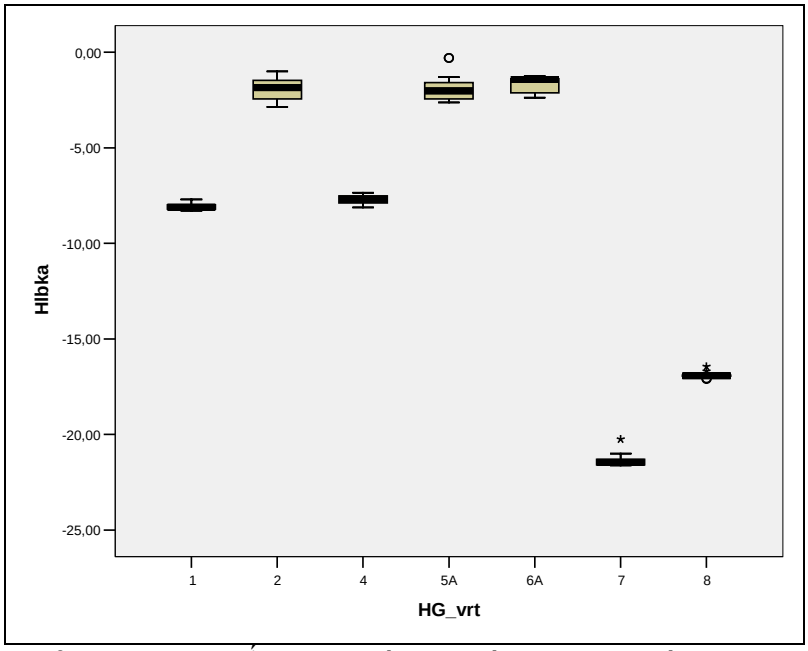
Počas roku 2007 sme urobili spolu 14 meraní hĺbky vodnej hladiny v siedmich hydrogeologických vrtoch na danom zosuve. Podľa možností a počasia to bolo zhruba každých 14 dní od marca do októbra. Hodnoty hĺbok v jednotlivých vrtoch sú ďalej uvedené v metroch. Pre lepšiu interpretáciu sú v záporných číslach (ak sa zvýši objem vody, tak jeho hĺbka sa zmenší a naopak).

3. Metódy

Pre prvotnú orientáciu sme použili boxplot a základné štatistické charakteristiky skúmaného súboru. Normalitu sme testovali presným jednovýberovým Kolmogorovovým-Smirnovovým testom. Stredné hodnoty sme porovnali ANOVA procedúrou, ktorej súčasťou je aj Levenov test rovnosti rozptylov. Pre zistenie trendu bol použitý Spearmanov korelačný koeficient. Všetky výpočty a grafy boli urobené v systéme SPSS verzia 13.

4. Výsledky

Z boxplotov (graf 1) je zrejmé, že hĺbky vodnej hladiny vo vrtoch sa dajú vizuálne rozdeliť do 4 skupín: najväčšia hĺbka je vo vrte 7, druhá najväčšia vo vrte číslo 8, vrty 1 a 4 sú z tohto hľadiska podobné a vrty z najnižšou hĺbkou (2, 5A, 6A) tiež. Hodnoty vo vrtoch 5A, 7 a 8 obsahujú extrémne hodnoty.



Graf 1. Boxploty hĺbky vodných hladín v sledovaných hydrogeologických vrtoch za rok 2007 (v metroch)

V tabuľke 1 sú uvedené základné charakteristiky opisnej štatistiky. Z nich je presnejšie vidieť rozdiely v aritmetických priemeroch ako aj v rozptyloch. Pre urobenie záverov je vhodné otestovať rovnosť priemerov a rozptylov. Z veľkostí priemerov a smerodajných odchýlok tiež vyplýva možná normalita hodnôt.

Tabuľka 1. Základné štatistické charakteristiky hĺbky vodnej hladiny vo vrtoch

HG_vrt	Mean	N	Std. Deviation	Minimum	Maximum
1	-8.0793	14	.16973	-8.29	-7.70
2	-1.9679	14	.56237	-2.86	-1.00
4	-7.7136	14	.25327	-8.12	-7.35
5A	-1.8386	14	.77145	-2.63	-.30
6A	-1.6950	14	.43641	-2.38	-1.24
7	-21.3507	14	.35877	-21.62	-20.25
8	-16.9007	14	.16098	-17.10	-16.45
Total	-8.5065	98	7.31865	-21.62	-.30

Pre testovanie normality hĺbok vodnej hladiny sme použili presný (vzhľadom na malý rozsah výberu) jednovýberový Kolmogorovov-Smirnovov test, ktorý je súčasťou SPSS modulu Exact tests. Pre ušetrenie priestoru jeho výsledky vo forme tabuliek neuvádzame. Výsledok bol rovnaký pre všetky vrty: nebola zamietnutá normalita hodnôt hĺbky vodných hladín v sledovaných vrtoch. Je nutné dodať, že sila jednovýberového testu pre malé rozsahy nie je dostatočná pre jeho zamietnutie.

Priemery hĺbok vodnej hladiny v jednotlivých vrtoch boli porovnané ANOVA procedúrou. Jej výsledky sú v tabuľke 2. Nulová hypotéza je zamietnutá ($p < 0{,}001$), medzi priemermi hĺbok vodnej hladiny v sledovaných vrtoch sú signifikantné rozdiely.

Tabuľka 2. Základná tabuľka výstupu ANOVA v SPSS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5178.034	6	863.006	4476.692	.000
Within Groups	17.543	91	.193		
Total	5195.577	97			

Ďalej sme testovali rovnosť rozptylov. V tabuľke 3 je výsledok testovania Levenovým testom. Nulová hypotéza je zamietnutá ($p < 0{,}001$), preto sme vybrali adekvátny porovnávací postup – post hoc Tamhanov test. Pre stručnosť tu neuvádzame jeho tabuľku.

Tabuľka 3. Testovanie rovnosti rozptylov ako súčasť výstupu ANOVA v SPSS

Levene Statistic	df1	df2	Sig.
8.001	6	91	.000

Tamhanov test (rovnako aj ďalšie tri pre post hoc porovnania priemerov pri nerovnosti rozptylov) nezamietol rovnosť pre hĺbky vo vrtoch 2, 5A a 6A. Tie tvoria homogénnu skupinu. Ostatné štyri vrty tvoria samostatné skupiny.

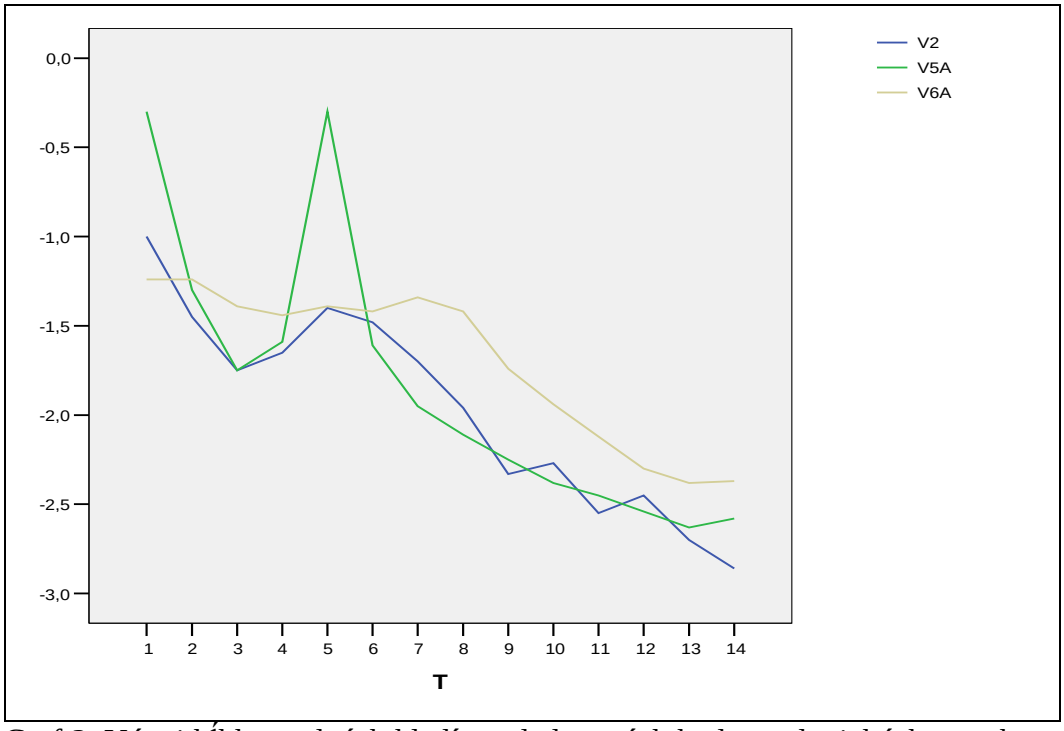
Užitočnou informáciou pri analýze dát je aj existencia trendu v časových radoch. Vzhľadom na to, že časová premenná je poradová a samotné určenie hladín má apriórnu chybu, sme pre trend použili Spearmanov koeficient korelácie ako robustný analóg Pearsonovho koeficientu korelácie.

Tabuľka 4. Testovanie trendu hĺbky hladín Spearmanovým korelačným koeficientom

V1	V2	V4	V5A	V6A	V7	V8
-.924(**)	-.921(**)	-.095	-.944(**)	-.913(**)	-.895(**)	-.080
.000	.000	.748	.000	.000	.000	.787
14	14	14	14	14	14	14

** Correlation is significant at the 0.01 level (2-tailed).

Výsledky sú v tabuľke 4. Z nich vyplýva, že objem vody klesá (hĺbka stúpa) vo vrtoch 1, 2, 5A, 6A a 7 ($p < 0{,}001$). Vo vrtoch 4 a 8 stagnuje. Na grafe 2 je to zobrazené pre vrty 2, 5A a 6A. Hodnoty z piateho merania signalizujú určitú anomáliu v podobe lokálneho maxima, to bude tiež súčasť ďalších analýz.



Graf 2. Vývoj hĺbky vodných hladín v sledovaných hydrogeologických vrtoch za rok 2007 (v metroch)

5. Záver

Príspevok uvádza prvé výsledky exploračnej analýzy dát časti meraní, ktoré sa uskutočnili v roku 2007 na zosuve v Ľubietovej. Z nich pre hodnoty hĺbok hladín podzemnej vody v siedmich sledovaných vrtoch vyplýva:

- hodnoty sú normálne rozdelené a nemajú rovnaké rozptyly ($p < 0,001$),
- hodnoty vo vrtoch 2, 5A a 6A tvoria homogénnu skupinu,
- objem vody vo vrtoch 1, 2, 5A, 6A a 7 v rámci roka 2007 klesá ($p < 0,001$),
- objem vody vo vrtoch 4 a 8 stagnuje.

6. Literatúra

CRUDEN D.M. 1991: A simple definition of a landslide. In: Bulletin of the International Association of Engeneering Geology, roč. 43, 1991, s. 27–29

FLEMING R.W. – VARNES D.J. 1991: Slope movements. In: Kierch G.A. (ed.): The Heritage of Engineering Geology. The First Hundred Years. Boulder, Colorado, Geological Society of America, Centencial Special Volume 3.

FUSSGÄNGER F., JADROŇ D., BANSKÝ M. & TYLEČEK B.: Ľubietová – prúdový zosun. Záverečná správa z predbežného inžiniersko-geologického prieskumu zosunu. Manuskript – archív Geofond Bratislava, 1978, 79s

IVERSON R.M. 2000: Landslide triggering by rain infiltration. In: Water Resources Research, roč. 36, č.7, 2000, s. 1897 – 1910

NEMČOK A. 1982: Zosuvy v Slovenských Karpatoch. Veda, Bratislava, 1982, 319 s

SPSS 13.0 USER'S GUIDE. Chicago : SPSS Inc., 2004. ISBN 0-13-185723-1

WASOWSKI J. 1998: Understanding rainfall-landslide relationships in man-modified environments: a case-history from Caramanico Terme, Italy. In: Environmental Geology, roč.35, č. 2–3, 1998, s. 197 –209

Adresa autora:

RNDr. Samuel Koróny
Ústav vedy a výskumu UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
Email: samuel.korony@umb.sk

RNDr. Roberta Prokešová, PhD.
Ústav vedy a výskumu UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
Email: roberta.prokesova@umb.sk

Mgr. Alžbeta Medveďová
Katedra geografie a krajinnej ekológie
Fakulta prírodných vied UMB
Tajovského 40
974 01 Banská Bystrica
Email: medvedov@fpv.umb.sk

The Ageing of the Population and its Consequences

Langhamrová Jitka, Fiala Tomáš

Abstract:
The Czech population is ageing. The proportion of younger people in the population is declining and the number of older people, on the contrary, is increasing. This is caused chiefly by the fact that the mortality rate is dropping, people are reaching a greater age and at the same time the birth rate is declining. One of the most frequently mentioned consequences of the ageing of the population is the impact on the field of pension security.

Key words:
population ageing, population projection, pension system

1. Introduction

The Czech Republic and its population are ageing at the present time. A process is occurring that is called the second demographic revolution in demographic terms and we also observed similar changes in the composition of the population and in demographic behaviour in advanced populations in the period after the Second World War. The proportion of younger people in the population is declining and the number of older people, on the contrary, is increasing. This is caused chiefly by the fact that the mortality rate is dropping, people are reaching a greater age and at the same time the birth rate is declining, less children are born and fertility has dropped below the so-called preservation limit of simple reproduction. In the future, too, it is anticipated that the death rate will decline still further and people will die at a later age and also the birth rate will continue to drop or stagnate. The ageing of the population is taking place in all advanced countries.

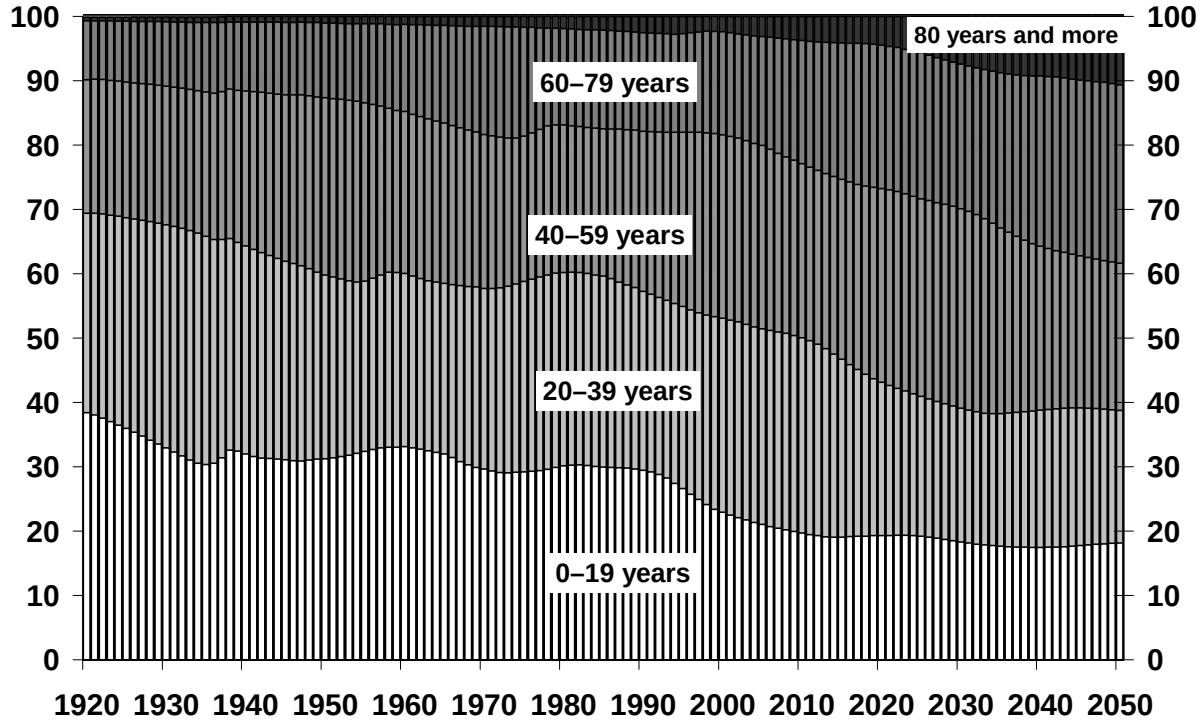
2. Main part

The basic characteristics of the age structure of the population are changing, the average age of the population is rising, and there is a reduction in the proportion of younger people and an increase in the proportion of older people in the population.

In 1920 there were still almost 40% of the population under the age of 20 and only 10% older than 60. At present the proportion of people younger than 20 is only 20% and it will no longer increase in the future, and the proportion of people over the age of 60 has exceeded 20% and will further increase in the future.

According to the projection of the Czech Statistical Office the proportion of people of productive age, i.e. the age of 20–59 years, will drop from the present 60% to almost 40% in the year 2050. There will therefore be a considerable decline in workforces. At the same time the proportion of people aged 60 and over will increase from the present 20% to almost 40%. According to the projection every tenth inhabitant will be over the age of 80.

Graph 1. Development of the age structure of the Czech Republic (in %)



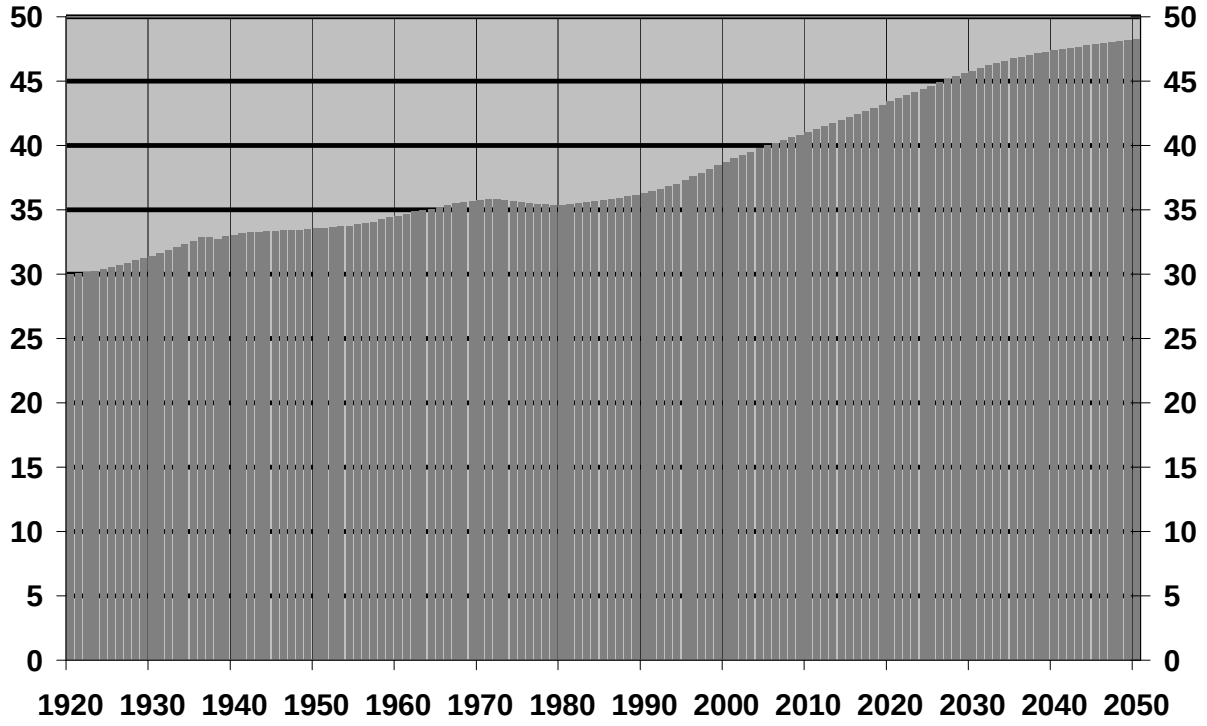
Source: Czech Statistical Office data, own computations

The population will age considerably. The average age of the inhabitants of the Czech Republic in 1920 was only 30 years; today we are almost 40 years old and around the year 2050 the average age will be close to 50 years. The Czech population will be among the oldest in the world.

The age structure of the population of the Czech Republic is similar to the structure of the population of the EU. The age structure of the CR is, however, more irregular. It differs from the EU chiefly in the higher proportion of those aged between 50–64 years (due to the increase in the number of births during World War Two, which was caused by the specific situation compared with the other European countries – Bohemia and Moravia were occupied, the army was dissolved and the starting of a family became one of the ways to avoid what was known as total service – enforced labour). The Czech Republic also has a lower proportion of 35–49-year-olds in comparison with the EU (this was caused by the drop in the birth rate in the sixties and seventies when the weak population years born in the thirties at the time of economic crisis were giving birth). There is also a strikingly higher proportion here of 25–34-year-olds (those giving birth were from the strong population years following WW2 and the pro-population measures of the seventies also had an impact). From the point of view of demography these were unsuitably timed pro-population measures that resulted in further increasing the already great irregularities in age structure. In the CR there is also a lower proportion of children under the age of 10, caused by the decline in the birth rate as a result of the political, economic and social changes after 1989. Fertility is dropping far below the so-called preservation line of mere reproduction and young people are postponing the birth of a

child until a later age as is known to be the case in the majority of the countries of Western Europe. New opportunities exist for personal and professional self-realisation, greater application to work is required and the role of education is increasing.

Graph 2. Development of the average age in the Czech Republic



Source: Czech Statistical Office data, own computations

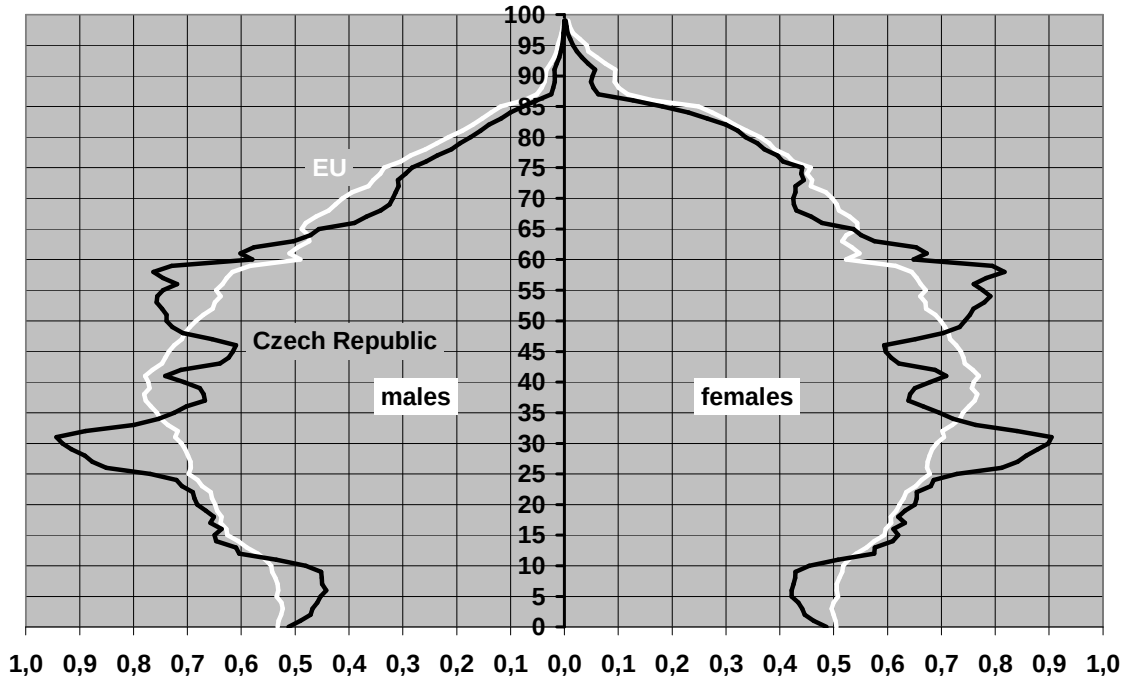
The ageing of the population will have a number of consequences in all areas of the life of society.

One of the most frequently mentioned consequences of the ageing of the population is the impact on the field of pension security. At present pension security is also based on the so-called pay-as-you-go system. Economically active people pay contributions to pensions into the system of pension security and this is immediately re-allocated and paid out to present pensioners.

In this system, then, the degree of burden on the system is the ratio of the number of persons of retirement age to the number of persons of productive age.

At present for 100 people of productive age there are on average around 37 persons of pensionable age. As the population ages there will also be an increase in the number of persons thus dependent. This will not be prevented even by the present law, which envisages a gradual raising of the retirement age of men and childless women to 63 years (for women with children the expected retirement age is 1–4 years earlier, depending on the number of children). Around 2030 there may already be about 50 pensioners for every 100 persons of productive age and around 2050 it may be as many as 75 pensioners.

Graph 3. Demographic Structure of the Czech Republic and the EU (31. 12. 2005)

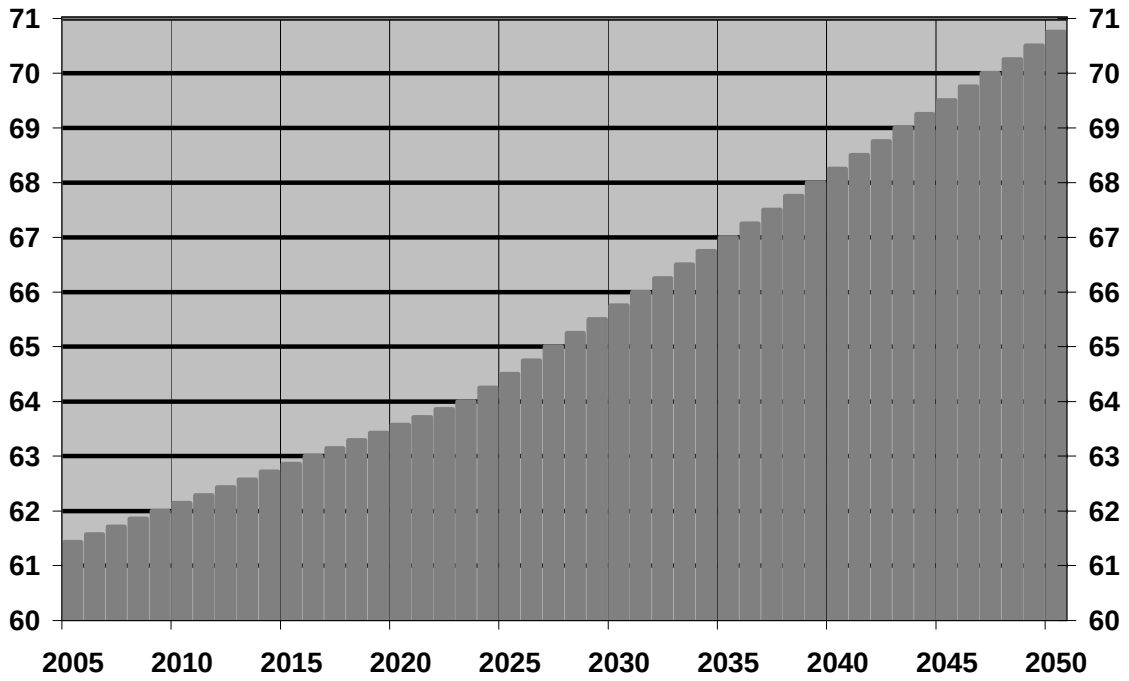


Source: Czech Statistical Office data, own computations

3. Conclusions

One of the possibilities for reducing the burden on the pension system is the constant raising of retirement age.

Graph 4. Growth of the retirement age the Czech Republic if the ratio of pensioneers to people in working age would stay at the level 40 %.



Source: Czech Statistical Office data, own computations

Nevertheless, in order for the number of persons of retirement age around the year 2050 per 100 persons of productive age to be only about 40, retirement age at that time would have to be over 70 years. (The question is, of course, whether with such an increase there would be sufficient suitable jobs for seniors.)

A number of further solutions also appear, among them raising the contributions to the fund of pension security, additional pension insurance, etc. For the future it may also be assumed that as the education structure of the population improves, so labour productivity will also increase and a smaller number of economically active people will be capable of supporting more dependent persons.

4. References

LANGHAMROVÁ, J. – FIALA, T. 2007. Stárnutí obyvatelstva a jeho důsledky v oblasti sociální a zdravotní péče. Poster pro Den vědy na pražských vysokých školách 30. 11. 2007. Praha

LANGHAMROVÁ, J. – FIALA, T. 2007. Ageing of the population and its impact on the financing of health care in the Czech Republic t. Lisboa 22.08.2007 – 29.08.2007. In: *ISI 2007*. Lisboa : International Statistical Institute, 2007, s. 627. ISBN 978-972-8859-71-8.

FIALA, T. 2006. Dva přístupy modelování vývoje úmrtnosti v populační projekci a jejich aplikace na populaci ČR. Bratislava 05.10.2006 – 06.10.2006. In: *Forum Statisticum Slovacum 4/2006.* Bratislava : Slovenská štatistická a demografická spoločnosť, 2006, s. 44–55. ISSN 1336-7420.

Address of authors:

Ing. Jitka Langhamrová, CSc., RNDr. Tomáš Fiala, CSc.
Department of Demography
Faculty of Informatics and Statistics
University of Economics, Prague
nám. W. Churchilla 4
130 67 Praha 3
Czech Republic
langhamj@vse.cz, fiala@vse.cz

This article came into being within the framework of the long-term research project 2D06026, "Reproduction of Human Capital", financed by the Ministry of Education, Youth and Sport within the framework of National Research Program II.

Vliv znečištěného životního prostředí na vrozené vady a samovolné potraty v České republice

Löster Tomáš

Abstract: The environment is being polluted as a result of developed industrial production and energy industry. Every region of the Czech Republic has been affected by industrial activities on different level and therefore the quality of the environment is different in each of 14 regions of the Czech Republic. The objective of this article is to prove that pollution of the environment certainly affects, among others, the spontaneous abortions and congenital defects. One way analysis of variance (ANOVA) has been used to analyze the spontaneous abortion rate. Tukey method has been used to assess statistical significance of differences of the spontaneous abortion rates in the Czech regions. Statistical system STATGRAPHICS has been used for analysis of variance. Development of the spontaneous abortions in all Czech regions in the period 2000 – 2005 is part of this analysis (of spontaneous abortions).

Key words: congenital defects, spontaneous abortion, environment, one-way analysis of variance.

1. Úvod

Vrozené vady a samovolné potraty mohou být důsledkem špatného životního prostředí, mj. znečištěného ovzduší. Při snaze prokázat vliv stavu životního prostředí na samovolné potraty a vrozené vady je vhodné použít rozdělení území České republiky do jednotlivých územních celků – krajů. V závislosti na stavu životního prostřední v jednotlivých krajích, který je dán mimo jiné celkovou úrovní průmyslové výroby a energetického průmyslu, lze zkoumat důsledky na sledované veličiny. Pro analýzu důsledků stavu životního prostřední na sledované veličiny je vhodné určit míru samovolných potratů jako podíl počtu samovolných potratů a počtu těhotných, kde počet těhotných je možné určit jako součet počtu samovolných potratů a živě narozených. Míra samovolných potratů je dále označována jako *míra_SP*. Kvalita ovzduší je v jednotlivých krajích zkoumána pomocí obsahu různých látek v ovzduší. Jedná se například o následující látky: CO, O_3, SO_2 atd.

Cílem tohoto příspěvku je prokázat, že znečištěné prostředí ovlivňuje vrozené vady i samovolné potraty a tedy kraje s vyšší průmyslovou výrobou a energetickým průmyslem (jako je např. Ústecký a Moravskoslezský kraj) mají vyšší míru samovolných potratů a vyšší počet vrozených vad.

2. Vrozené vady

Dá se předpokládat, že vrozené vady budou nepříznivě ovlivňovány znečištěným životním prostředím. Ke kvantifikaci důsledků znečištěného životního prostředí na vrozené vady budou konstruovány různé míry vrozených vad, které jsou zachyceny v tabulce č. 1, v členění podle okresů.

První sloupec tabulky č. 1 obsahuje jednotlivé územní celky ČR – kraje. Druhý sloupec obsahuje podíly jednotlivých krajů na celkovém počtu vybraných vrozených vad (např. rozštěp rtu, cystická nemoc ledvin, vrozená brániční kýla). Třetí sloupec obsahuje podíly jednotlivých krajů na celkovém počtu všech vrozených vad. Čtvrtý sloupec obsahuje podíl každého z krajů na celkovém počtu živě narozených. Pátý sloupec obsahuje rozdíl mezi podílem příslušného kraje na celku z hlediska všech vrozených vad a živě narozených dětí (tj.

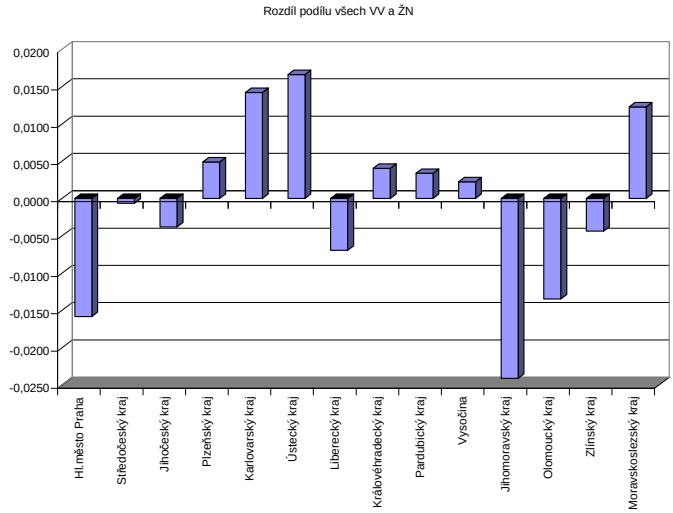
rozdíl třetího a čtvrtého sloupce). Z pátého tohoto sloupce (a následně také z grafu č. 1) je možné zjistit, které kraje mají vyšší podíl vrozených vad než je jejich podíl na živě narozených. Je například patrné, že kraj Praha má nižší podíl vrozených vad než je její podíl na celkovém počtu narozených dětí. Naopak Ústecký kraj má větší podíl na vrozených vadách než podíl na živě narozených. Z toho vyplývá, že vrozené vady jsou závislé na územním celku (kraji). Vzhledem k tomu, že územní celky mají rozdílné hodnoty ve zkoumaných látkách obsažených v ovzduší, je tedy zřejmé, že ovzduší (životní prostředí) má vliv na míru a samozřejmě tedy i na počet vrozených vad.

Tabulka č. 1: Podíly vrozených vad (VV) a živě narozených na celkovém počtu narozených (ŽN)

Kraj	Podíl vybraných VV na celku	Podíl všech VV na celku	Podíl ŽN na celku	Rozdíl podílu všech VV a ŽN
Hl.město Praha	*0,0858*	*0,0948*	*0,1106*	-0,0158
Středočeský kraj	*0,1049*	*0,1129*	*0,1136*	-0,0006
Jihočeský kraj	*0,0650*	*0,0567*	*0,0604*	-0,0038
Plzeňský kraj	*0,0527*	*0,0581*	*0,0532*	0,0049
Karlovarský kraj	*0,0422*	*0,0453*	*0,0311*	0,0142
Ústecký kraj	*0,0915*	*0,1044*	*0,0877*	0,0166
Liberecký kraj	*0,0451*	*0,0366*	*0,0436*	-0,0070
Královéhradecký kraj	*0,0536*	*0,0577*	*0,0537*	0,0040
Pardubický kraj	*0,0533*	*0,0527*	*0,0494*	0,0034
Vysočina	*0,0606*	*0,0525*	*0,0503*	0,0022
Jihomoravský kraj	*0,0966*	*0,0838*	*0,1079*	-0,0241
Olomoucký kraj	*0,0569*	*0,0474*	*0,0609*	-0,0135
Zlínský kraj	*0,0496*	*0,0514*	*0,0557*	-0,0044
Moravskoslezský kraj	*0,1356*	0,1343	0,1220	0,0123

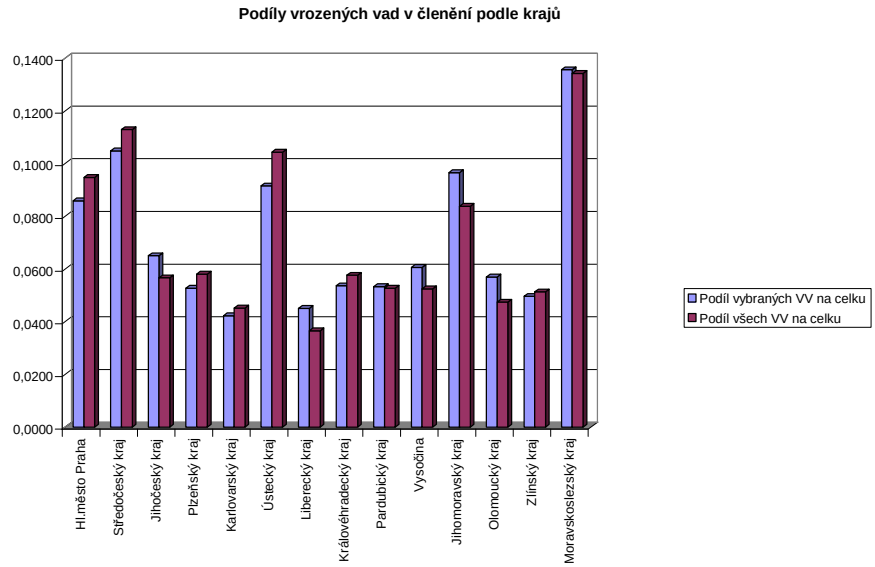
Graf 1 představuje znázornění posledního sloupce tabulky č. 1. Z tohoto grafu je na první pohled patrné, které kraje České republiky mají vyšší podíl vrozených vad než podíl živě narozených na celku. Jedná se o následující kraje: Plzeňský, **Karlovarský**, **Ústecký**, Královéhradecký, Pardubický, Vysočina a **Moravskoslezský**. Tučně vyznačené kraje se pohybují o více než + 1% bod v rozdílu podílů vrozených vad a živě narozených na celku. Vzhledem k výčtu krajů je vliv znečištěného ovzduší na vrozené vady zřejmý.

Graf 1: Rozdíl podílu všech vrozených vad a živě narozených



Graf 2 znázorňuje podíly jednotlivých krajů na celkovém počtu vrozených vad (vybraně sledovaných i všech vrozených vad). Jedná se tedy o druhý a třetí sloupec tabulky č. 1. Potvrzuje mj. závěr z grafu č. 1 a je zde vidět, že např. Moravskoslezský kraj je z hlediska podílu (tedy i počtu) vrozených vad na jednom z předních míst. Je to dáno právě mj. znečištěním životního prostředí. Z grafu je zřejmé, že Moravskoslezský kraj se na celkovém počtu vrozených vad podílí cca 14 %.

Graf 2: Podíly vrozených vad v jednotlivých krajích



Závěrem k oblasti vrozených vad lze konstatovat, že se z příslušných výpočtů jednotlivých měr zobrazených v tabulce č. 1 a grafech 1 a 2 dá předpokládat závislost vrozených vad na jednotlivých krajích. Vzhledem k rozdílnosti jednotlivých krajů z hlediska kvality ovzduší a životního prostředí lze říci, že vrozené vady jsou ovlivňovány kvalitou ovzduší.

3. Samovolné potraty

Stejně jako u vrozených vad se i u samovolných potratů dá očekávat, že budou ovlivňovány nepříznivým životním prostředím (znečištěným ovzduším). K analýze tohoto předpokladu a vzhledem k povaze dat může být využita analýza rozptylu. Pro účely této analýzy je vytvořena veličina označená jako *míra_SP*, která představuje podíl samovolných potratů na počtu „těhotných“ (počet živě narozených plus počet samovolných potratů).

Cílem je prokázat, že znečištěné prostředí a vyšší průmyslová výroba (jako je např. v Ústeckém a Moravskoslezském kraji) ovlivňuje míru samovolných potratů. K analýze závislosti míry samovolných potratů na jednotlivých krajích je využita jednofaktorová analýza rozptylu, která je řešena v systému STATGRAPHICS_Plus. Ověření předpokladů užití analýzy rozptylu je provedeno pomocí Bartlettova testu (test o shodě rozptylů). Z grafu 3 vyplývá, že lze předpoklad o rovnosti skupinových rozptylů považovat za splněný.

Graf 3: Výstup analýzy rozptylu ze systému STATGRAPHICS_Plus

```
Variance Check

Cochran's C test: 0,225241   P-Value = 0,0640827
Bartlett's test: 1,36287   P-Value = 0,0897009
Hartley's test: 18,0919
```

Z vypočtených hodnot ve výstupu ze systému STATGRAPHICS_Plus v grafu 4, ve kterém je ověřována závislost *míry_SP* na jednotlivých krajích je patrné, že na všech obvykle volených hladinách významnosti (5 % nebo 1 %) je testovaná hypotéza (o rovnosti středních hodnot) zamítnuta a tedy existuje alespoň jedna významná dvojice krajů, která se významně liší z hlediska *míry_SP*. Na základě tohoto výpočtu je tedy prokázána závislost mezi mírou samovolných potratů a kvalitou ovzduší.

Graf 4: Výstup analýzy rozptylu ze systému STATGRAPHICS_Plus

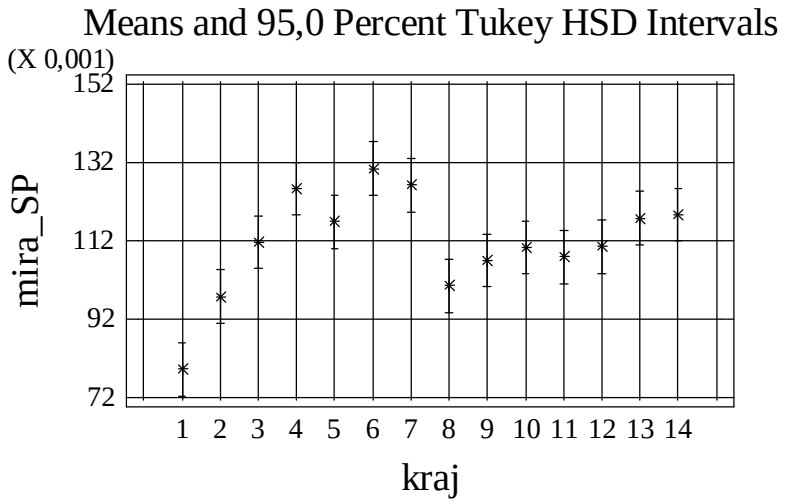
ANOVA Table for mira_SP by kraj

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	0,0136653	13	0,00105117	22,99	0,0000
Within groups	0,00320073	70	0,0000457247		
Total (Corr.)	0,016866	83			

Vzhledem k prokázání závislosti míry_SP a jednotlivých krajů je vhodné dále zkoumat, které dvojice krajů jsou z hlediska míry_SP statisticky rozdílné (na 5 % hladině významnosti) a kolik % tento rozdíl činí. K tomuto srovnání je využita Tukeyho metoda, která je opět vypočtena pomocí systému STATGRAPHICS_Plus. Lze zjistit, že za statisticky významnou dvojici je považována taková dvojice, jejíž rozdíl v míře_SP je více než +/–1,35%. Průměrná úroveň *míry_SP* pro jednotlivé kraje za celé sledované období (rok 2000 – 2005) je znázorněna v grafu 5.

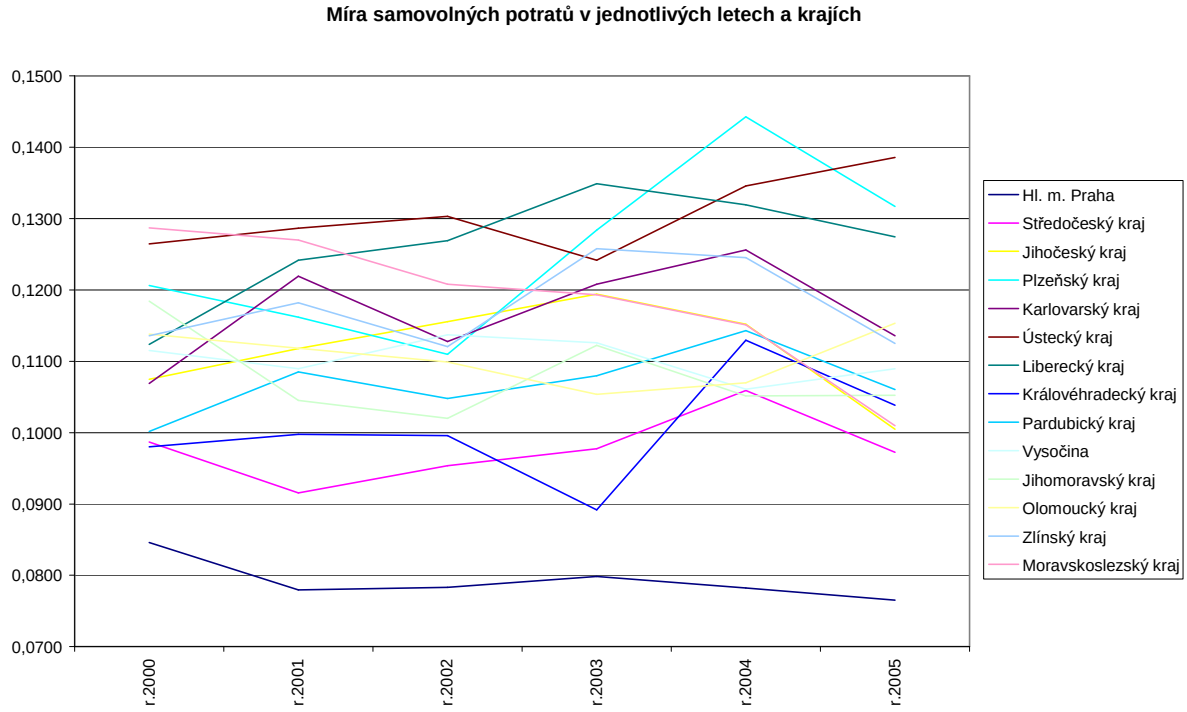
Graf 5: Průměrná úroveň míry_SP za celé období v členění podle jednotlivých krajů



Z grafu 5 je patrné, že nejvyšší míru samovolných potratů během sledovaného období měl kraj 6 (Ústecký kraj). Moravskoslezský kraj, označený číslem 14, je z hlediska míry samovolných potratů nadprůměrný (stejně jako v případě vrozených vad).

Graf 6 zachycuje vývoj míry samovolných potratů v jednotlivých letech v členění podle jednotlivých krajů. Průběh v letech 2000 – 2005 je patrný při pohybu po spojnici. Například je patrný pokles míry samovolných potratů ve Středočeském kraji během sledovaného období.

Graf 6: Míry samovolných potratů v jednotlivých krajích (vývoj v letech)



4. Závěr

Závěrem lze říci, že na základě provedených výpočtů a grafických výstupů je zřejmé, že míra samovolných potratů i vrozených vad (tedy i jejich počty) jsou ovlivněny úrovní znečištění ovzduší. Jednotlivé kraje se mezi sebou liší úrovní znečištění ovzduší a to se, jak bylo ukázáno, mimo jiné projevuje i na míře samovolných potratů a vrozených vad. Jak již bylo uvedeno výše, Ústecký kraj je z hlediska míry samovolných potratů jako jeden z nejhorších krajů v České republice.

5. Literatura

HEBÁK, P. a kol.: Vícerozměrné statistické metody (1), Informatorium, 2004.
JAROŠOVÁ, E., PECÁKOVÁ, I.: Příklady k předmětu statistika B, VŠE v Praze, 2000.
LÖSTER, T.: Závislost samovolných potratů na životním prostředí, Typograf, 2007.
MAREK, L. a kol.: Statistika pro ekonomy aplikace, Profesional Publishing, 2005.

Kontaktní adresa:
Ing. Tomáš Löster
Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
nám. W. Churchilla 4
130 67 Praha 3
Česká republika
losterto@vse.cz

FDH DEA model ako alternatívna metóda benchmarkingu
FDH DEA model as an alternative benchmarking method

MAJOROVÁ Martina

ABSTRACT

In the paper we deal with computing the measures of technical efficiency of FDH DEA models as an alternative benchmarking method. The analysis was carried out on the sample size of 142 German banks in the year 2005. The measures of technical effciency of FDH DEA models were compared with the measures of technical efficiency of other radial DEA models (CCR, BCC, NIRS and NDRS). Those banks or their branches that were identified as effective ones were ordered by the measures of technical efficiency produced by using the Super-Efficiency DEA model.

KEY WORDS
DEA, technical efficiency, benchmarking, FDH

ÚVOD

Všeobecne možno benchmarking definovať ako kontinuálnu, systematickú metodológiu pre identifikáciu, meranie a porovnávanie pracovných procesov alebo funkcií vlastnej organizácie počas určitého časového horizontu s jedným alebo viacerými organizáciami s cieľom interných zlepšení (Mard et al., 2004). Z uvedeného vyplýva, že benchmarking je dôležitým nástrojom pre evaluáciu efektívnosti a relatívnej výkonnosti podnikov v rámci sektoru.

Klasifikácie benchmarkingu sa rôznia, avšak podľa OECD sa benchmarking člení predovšetkým na podnikový (interný), branžový (odvetvový) a benchmarking rámcových podmienok (zameraný na kľúčové faktory ovplyvňujúce operačné prostredie podnikov a tým majúci vplyv na konkurencieschopnosť; [8]).

Z hľadiska analýzy dátových obalov (**D**ata **E**nvelopment **A**nalysis – DEA) je benchmarking proces definovania validných mier porovnávania výkonnosti rozhodovacích jednotiek referenčnej množiny, ktoré sa využívajú na stanovenie relatívnych pozícií podnikov referenčnej množiny, vytvárajúc tak určitý výkonnostný štandard (štandard kvality). Týmto spôsobom sa DEA dá považovať za nástroj benchmarkingu, pretože konštruovaná produkčná hranica prezentuje empirický výkonnostný štandard (Zhu, 2003).

MATERIÁL A METÓDY

Metodológia FDH analýzy bola sformulovaná autormi Deprins, Simar a Tulkens (1984), neskôr rozšírená aj o využitie vo finančnom sektore a bankovníctve (Tulkens, 1993). FDH modely sú známe dvoma charakteristikami: platí pre ne podmienka *silnej disponibility zdrojov*[1] (free disposability), no na rozdiel od ostatných základných DEA modelov je konštruovaný *obal nekonvexný* (t.j. na FDH modely sa nevzťahuje predpoklad konvexnosti). Ďalšou odchýlkou oproti klasickým DEA modelom je spôsob výpočtu mier technickej

[1] Silná disponibilita zdrojov (free, strong disposability) znamená, že napr. prebytočných vstupov sa možno zbaviť bez dodatočných nákladov.

efektívnosti. FDH model je totiž upravený DEA model, v ktorom sa predpokladá, že hodnotené rozhodovacie jednotky sa *porovnávajú iba so skutočne jestvujúcimi efektívnymi podnikmi* (nie s virtuálnymi, ako to praktizuje DEA). Pri výpočte technickej efektívnosti sa uplatňuje prístup, v ktorom kombinácie vstupov, resp. výstupov môžu nadobúdať len také hodnoty, aké sú nám známe v skutočnosti. Matematický zápis inputovo-orientovaného **FDH modelu**[2] uvádza model (1):

$$\begin{aligned} &\min \theta \\ &\quad Y\lambda \geq y_0 \\ &-\theta x_0 + X\lambda \leq 0 \\ &\quad 1^T \lambda = 1 \\ &\quad \lambda = \begin{matrix}0\\1\end{matrix}\} \end{aligned} \qquad (1)$$

Ako vyplýva z modelu (1), *obmedzujúce podmienky* pre parameter lambda sú definované ako *bivalentné* (binárne), t.j. nadobúdajú buď hodnotu 0 alebo 1. Keďže hranica produkčných možností pri FDH analýze je zhodná s „vnútorným obsahom“ obalu pri DEA modeloch, bude produkovať oveľa vyššie priemerné miery technickej efektívnosti ako DEA. Vo všeobecnosti platí, že modely za predpokladu konštantných výnosov z rozsahu sú prísnejšie ako modely za predpokladu variabilných výnosov z rozsahu a najmenej prísne sú práve FDH modely. Výsledkom tvrdenia je počet podnikov, ktoré sú pri jednotlivých modeloch identifikované ako efektívne: najmenej ich je pri KVR, viac pri VVR a najviac efektívnych podnikov dosiahneme použitím metodológie FDH analýzy.

Navyše, pri FDH sa nevyžaduje žiadny predpoklad o forme rozdelenia neefektívnosti v pozorovaných rozhodovacích jednotkách. Podniky, ktoré sa nachádzajú na FDH fronte sú *nedominované* [3] (100% efektívne), tie ostatné sú dominované nimi[4] (sú neefektívne).

Postup pri FDH analýze je nasledovný: najprv sa preskúmajú dáta o rozhodovacích jednotkách, aby sa identifikovali tie efektívne a v druhej etape sa z týchto podnikov pomocou parametrickej regresnej analýzy aproximuje FDH hranica (Thiry, Tulkens, 1992).

VÝSLEDKY A DISKUSIA

Údajovú základňu pre výpočet mier technickej efektívnosti DEA modelov sme získali z nemeckého spolku bánk. Výberový súbor pozostával z 142 bánk (prípadne ich pobočiek), ktoré sme analyzovali za rok 2005. V dôsledku konzistentnosti údajov, nebolo potrebné pre aplikáciu DEA metodológie vylúčiť žiadnu banku.

Ako vstupné premenné sme si zvolili plochu banky v m^2, počet poskytnutých úverov a predaných obligácií (vyjadrených počtom účtov v konkrétnej banke) a výstupnou premennou sa stali osobné depozity na účty (taktiež vyjadrené počtom účtov v konkrétnej banke). Vstupná databáza vykazovala pomerne veľkú variabilitu v údajoch (pozri Tabuľka 1).

---

[2] FDH je odvodené z anglického názvu **F**ree **D**isposal **H**ull, do slovenského jazyka sa neprekladá.

[3] Nedominovaný variant znamená, že neexistuje žiaden iný variant, ktorý by bol vo všetkých hodnotených kritériách lepší ako ostatné varianty (pri minimalizačných kritériách nadobúda nižšie hodnoty a pri maximalizačných kritériách zase vyššie hodnoty kritéria).

[4] Opakom nedominovaného variantu je variant dominovaný, t.j. existuje aspoň jeden variant, ktorý je v akomkoľvek hodnotenom kritériu lepší ako daný variant.

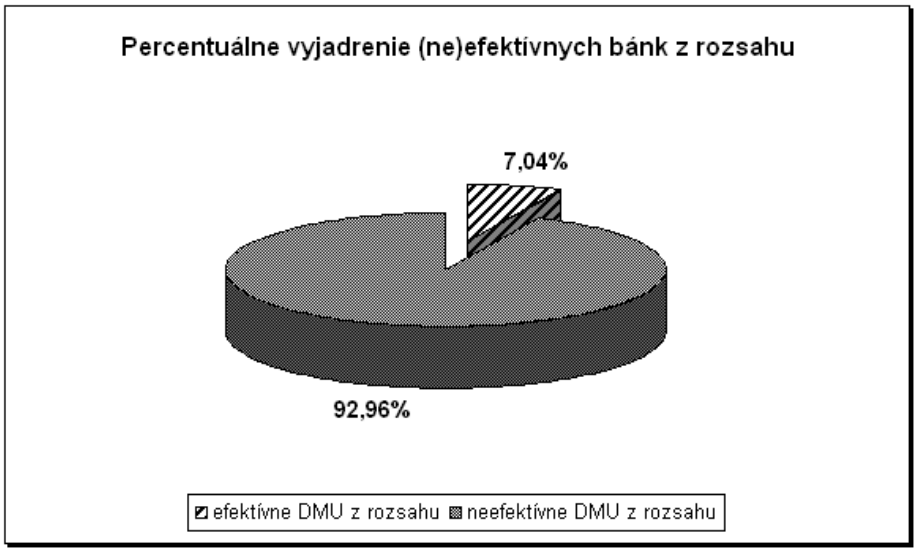
Tabuľka 1 Základné štatistické charakteristiky premenných vstupov a výstupu

Premenné	Minimum	Maximum	Priemer	Smer.odch.
Plocha banky [m^2]	64	1228	297	213
Úvery [počet účtov]	6	499	124	100
Obligácie [počet účtov]	7	503	96	86
Osobné vklady [počet účtov]	432	7851	1847	1456

Z tabuľky 1 vyplýva, že na základe všetkých hodnotených premenných by bolo možné klasifikovať dostupný súbor bánk aj vzhľadom k veľkosti plochy banky na nasledovné kategórie: malé (menšie) banky, stredne veľké banky a veľké banky[5]. Na základe tohto predpokladu je variabilita v rámci konkrétneho ukazovateľa považovaná za prirodzenú.

Pri hodnotení efektívnosti jednotlivých bánk je vhodné vedieť určité doplnkové informácie o konkrétnych rozhodovacích jednotkách[6]. Medzi tie podstatnejšie môžeme zaradiť výpočet efektívnosti z rozsahu[7] a definovanie oblasti výnosov z rozsahu, v ktorom banka operuje. V danej databáze údajov prevažujú banky neefektívne z rozsahu (v relatívnom vyjadrení 92,96%) nad bankami efektívnymi z rozsahu (7,04%), pozri Graf 1.

Graf 1 Percentuálne vyjadrenie (ne)efektívnych bánk z rozsahu



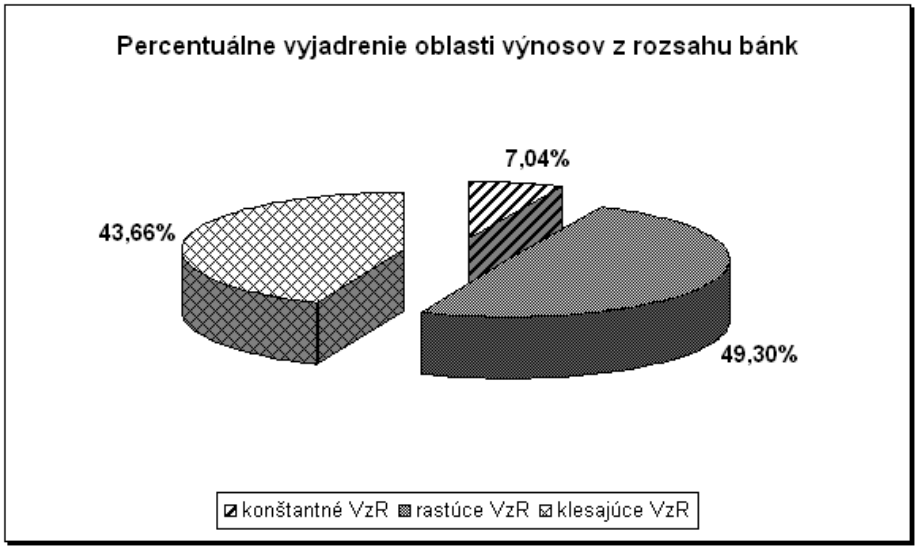
Priaznivá skutočnosť sa ale potvrdila pri definovaní oblasti výnosov z rozsahu, kedy majoritná časť bánk pôsobí v oblasti rastúcich výnosov z rozsahu (49,30%). Podstatné zastúpenie majú aj banky operujúce v oblasti klesajúcich výnosov z rozsahu (43,66%); podiel bánk spadajúcich do oblasti konštantných výnosov z rozsahu je viac-menej zanedbateľný (pozri Graf 2).

[5] Keďže nie sú k dispozícii presné kritéria na takto definovanú klasifikáciu bánk, bolo by uvedené rozdelenie značne subjektívne. Z tohto dôvodu nebolo aplikované pri ďalších výpočtoch.

[6] Rozhodovacia jednotka je slovenský preklad anglického pojmu **D**ecision **M**aking **U**nit (**DMU**), ktorá je štandardným základným pojmom v teórii efektívnosti a produktivity.

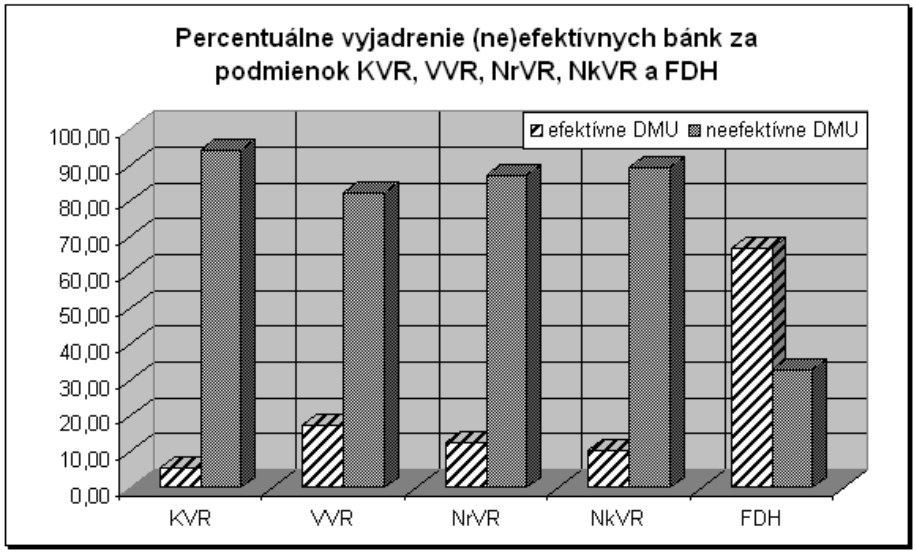
[7] Banka môže byť buď efektívna z rozsahu (t.j. má ešte „priestor“ na zväčšovanie svojho výstupu vzhľadom k danej množine vstupov) alebo je rozsahom neefektívna, pričom jej neefektívnosť vyplýva z príliš veľkého výstupu.

Graf 2 Percentuálne vyjadrenie oblasti výnosov z rozsahu bánk



Charakter metodológie FDH analýzy (spôsobený predovšetkým bivalentnosťou podmienok nezápornosti modelu) „produkuje“ v priemere najvyššie miery technickej efektívnosti v porovnaní so základnými radiálnymi DEA modelmi (KVR, VVR, NrVR a NkVR), čo môžeme symbolicky zapísať $TE_{KVR}<TE_{NkVR}<TE_{NrVR}<TE_{VVR}<TE_{FDH}$. Táto podmienka bola vo väčšej miere splnená pri výpočte mier technickej efektívnosti uvedených modelov (s výnimkou TE_{NrVR} a TE_{NkVR}). Dosiahnuté priemerné miery technickej efektívnosti spomínaných DEA modelov a percentuálne vyjadrenie počtu (ne)efektívnych bánk sú uvedené v grafe 3 a tabuľke 2.

Graf 3 Percentuálne vyjadrenie (ne)efektívnych bánk pri aplikácii uvedených DEA modelov



Legenda
KVR – konštantné výnosy z rozsahu
VVR – variabilné výnosy z rozsahu
NrVR – nerastúce výnosy z rozsahu (kombinácia klesajúcich a konštantných výnosov z rozsahu)
NkVR – neklesajúce výnosy z rozsahu (kombinácia rastúcich a konštantných výnosov z rozsahu)

Tabuľka 2 Miery technickej efektívnosti pre jednotlivé DEA modely

Premenné	**TE KVR**	**TE VVR**	**TE NrVR**	**TE NkVR**	**TE FDH**
Efektívne banky [%]	5,63	17,61	12,68	10,56	66,90
Priemerná efektívnosť	0,7109	0,7682	0,7337	0,7454	0,9350

Tabuľka 3 Poradie bánk na základe SuperDEA modelu

Poradové číslo banky	**TE SuperDEA**	**Poradie**
DMU140	**inf.**	1.
DMU10	**3,3873**	2.
DMU138	**2,0958**	3.
DMU97	**1,6551**	4.
⋮ ⋮ ⋮	⋮ ⋮ ⋮	⋮ ⋮ ⋮
DMU117	**1,0418**	22.
DMU139	**1,0313**	23.
DMU129	**1,0284**	24.
DMU82	**1,0149**	25.

V prípade, že by bolo žiadúce usporiadať efektívne banky, resp. pobočky do určitého poradia, je možné využiť DEA model pre výpočet superefektívnosti[8], ktorého špecifickou črtou je, že *hodnotený podnik* (jeho vstupy a výstupy) *sa* pri výpočte mier technickej efektívnosti *vylúči* z konvexnej kombinácie vstupov/výstupov ostatných podnikov, čím sa dosiahnu miery technickej efektívnosti inputovo-orientovaných modelov väčšie ako 1. Výsledné usporiadanie bánk podľa ich poradového čísla uvádza tabuľka 3.

Najzaujímavejšia banka z pohľadu superefektívnosti je banka pod poradovým číslom 140. Jej miera technickej efektívnosti nie je špecifikovaná ako konkrétne číslo, ale ako kladné nekonečno. Pre manažéra tejto banky by to znamenalo, že môže ľubovoľne zvyšovať vstupy tejto banky, pričom jej efektívnosť sa nezmení, t.j. táto banka bude efektívna pri akomkoľvek zvyšovaní definovaných vstupných premenných.

SÚHRN

Príspevok sa zaoberá výpočtom mier technickej efektívnosti FDH modelu ako alternatívnej metódy benchmarkingu. Analýza bola realizovaná na výberovom súbore 142 nemeckých bánk za rok 2005. Vypočtané miery technickej efektívnosti FDH DEA modelov boli komparované s mierami technickej efektívnosti ostatných radiálnych DEA modelov (KVR, VVR, NrVR a NkVR). Banky, resp. pobočky, ktoré boli identifikované ako efektívne, boli následnou aplikáciou DEA modelu pre výpočet superefektívnosti zoradené do poradia.

KĽÚČOVÉ SLOVÁ

DEA, technická efektívnosť, benchmarking, FDH

[8] DEA model pre výpočet efektívnosti sa označuje aj ako Andersen-Petersenov model.

LITERATÚRA

[1] COELLI, T. – RAO, D.S. Prasada – O'DONNELL, CH.J. – BATTESE, G.E.: An introduction to efficiency and productivity analysis, 2nd edition. New York: SpringerBusiness, 2005, 341 s. ISBN 978-0387-24265-1

[2] DEPRINS, D. – SIMAR, L. – TULKENS, H.: Measuring labor ineffciency in post offices, in Marchand, M., Pestieau, P. and Tulkens, H. (eds.). In *The Performance of Public Enterprizes*: Concepts and Measurements, 1984, s.243-267

[3] FANDEL, P.: FDH analýza ako metóda benchmarkingu. In *Acta oeconomica et informatica*, zv.1, 2005, s.6-9

[4] MARD, J.M. – DUNNE, R.R. – OSBORNE, E. – RIGBY, J.S.: Driving Your Company's Value: Strategic Benchmarking for Value. New Jersey: John Wiley & Sons, Inc., 2004, 208 s. ISBN 0471648558

[5] THIRY, B. – TULKENS, H.: Allowing for inefficiency in parametric estimation of production functions for urban transit firms. In *Journal of Productivity Analysis*, zv.3:1/2, 1992, s.45-65

[6] TULKENS, H.: On FDH efficiency analysis: Some methodological issues and application to retail banking, courts and urban transit. In *Journal of Productivity Analysis*, zv.4, 1993, s.183-210

[7] ZHU, J.: Quantitative models for performance evaluation and benchmarking: Data envelopment analysis with spreadsheets and DEA Excel Solver. Massachusetts: Kluwer Academic Publishers, 2003, 328 s. ISBN 1402070829

[8] Slovenské informačné centrum benchmarkingu. [online]. [cit. 2007-11-23]. Dostupné na Internete: <http://www.sbic.sk/2bench2.htm>

KONTAKTNÁ ADRESA

Ing. Martina Majorová, Katedra štatistiky a operačného výskumu
Fakulta ekonomiky a manažmentu, Slovenská poľnohospodárska univerzita v Nitre
Trieda A. Hlinku 2, 949 76 Nitra, tel. 037/6414 813
e-mail: Martina.Majorova@fem.uniag.sk

Confidence interval for relative frequency

Lubos Marek, Michal Vrabec

Introduction: We often solve the problem of estimation of relative frequency (i.e., the parameter value π of an alternative distribution) in the parent population of size N on the basis of a random sample of size n. There are two way how to build this estimation. We can compute a point estimate (where proportion p is used as an estimate of the unknown parameter value π) or an interval one. In the latter instance, the confidence interval must be determined. Endpoints of the confidence interval are usually calculated with the aid of an approximation by normal distribution. More accurate values can be looked up in tables. Let us consider another method, based on the use of MS Excel software.

Key words: *Alternative distribution, binomial distribution, relative frequency, confidence interval, approximations, MS Excel*

1. Confidence interval for relative frequency

When determining the confidence intervals for the relative frequency values (or the basic absolute frequency) we have to base our method on the binomial distribution, which is also used for point estimates because this distribution governs the occurrence of the observed sampling-with-replacement sign.

Considering the sampling *with replacement*, the $(1 - \alpha)\%$ confidence interval for the relative frequency value π is given as:

$$p' < \pi < p'', \tag{1}$$

where the lower bound p' is obtained by solving the following equation (for given m and α_1; m is the number of "successes" in n trials)

$$\sum_{x=m}^{n} \binom{n}{x} p'^{x} (1 - p')^{n-x} = \alpha_1 ; \tag{2}$$

similarly, the upper bound p'' is obtained by solving the following equation (for given m and α_2)

$$\sum_{x=0}^{m} \binom{n}{x} p''^{x} (1 - p'')^{n-x} = \alpha_2 , \tag{3}$$

where $\alpha_1 + \alpha_2 = \alpha$.

Putting $\alpha_1 = 0$ (implying $\alpha_2 = \alpha$), we get the *right-hand-sided* interval $0 \le \pi < p''$. It should be noted that this is the only option if the sampling provides $m = 0$, because in this case (2) holds identically for all $p' \in \langle 0;1 \rangle$. Putting $\alpha_2 = 0$ (implying $\alpha_1 = \alpha$), we get the *left-hand-sided* interval $p' < \pi \le 1$; and again, this is the only option if $m = n$, because in this case (3) holds identically for all $p'' \in \langle 0;1 \rangle$.

In sampling surveys, however, the most frequently employed approach puts $\alpha_1 = \alpha_2 = \alpha/2$ (cf. also the Figure below) thus utilising *two-sided* confidence intervals. In order to simplify practical considerations, values p' and p'' have been calculated for different values of m and n and tabled. These tables are easily available on the Internet and published in most textbooks on sampling surveys – e.g. Čermák, Vrabec.

Larger tables (and also nomographic charts made by CLOPPER & PEARSON (1934) on the basis of the calculated values) are published in most collections of statistical tables. Among domestic collections, let us refer to JANKO (1958) and LIKEŠ & LAGA (1978). Among foreign ones, BOLŠEV & SMIRNOV (1965) and OWEN (1962). Tables with shorter (narrower) intervals published by BUNKE (1960) are also worth mentioning.

If the sample is sufficiently large, and the estimated relative frequency value π is neither too large nor too small, say, $0.1 < \pi < 0.9$, the confidence interval can be calculated from normal distribution, which binomial distribution asymptotically approaches for $n \to \infty$ and fixed π. Some authors only require $n > 50$, but others take the value of π into consideration and say that the sample size should be such that $n\,\pi\,(1-\pi) > 9$. Textbook COCHRAN (1977) reflects the fact that we only have p at our disposal, not π, and recommends the following table of size samples n for the corresponding sampled relative frequency values p:

p	0.5	0.4	0.3	0.2	0.1	0.05
n	30	50	80	200	600	1400

Let us present two ways of utilising this approximation.

In practice, binomial distribution is often replaced with normal distribution (with suitably chosen parameters).

Approximation 1

For sufficiently large n, the statistics p is approximately normally distributed with parameters $E(p) = \pi$ and $D^2(p) = \pi\,(1-\pi)/n$. This implies that the approximated $(1 - \alpha)$% two-sided confidence interval for π according to upper equation is given as follows:

$$P\left\{-u_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} < p-\pi < u_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right\} = 1-\alpha \qquad (4)$$

or

$$P\left\{|p-\pi| < u_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right\} = 1-\alpha\,. \qquad (5)$$

Solving the equation

$$|p-\pi| = u_{1-\alpha/2}\sqrt{\pi(1-\pi)/n}\,,$$

or equivalently

$$(p-\pi)^2 = u^2_{1-\alpha/2}\,\pi(1-\pi)/n$$

in the unknown variable π we get the endpoints p' and p'' of the interval (1)

$$p', p'' = \frac{np + \frac{1}{2}u^2_{1-\alpha/2} \mp u_{1-\alpha/2}\sqrt{\frac{1}{4}u^2_{1-\alpha/2} + np(1-p)}}{n + \frac{1}{2}u^2_{1-\alpha/2}}\,. \qquad (6)$$

The approximation of binomial distribution by normal distribution can be improved by adding to the admissible error $u_{1-\alpha/2}\sqrt{\pi(1-\pi)/n}$ a correction term, equal to $1/(2n)$; this value takes into account discontinuity of binomial distribution. A discontinuous random variable m with binomial distribution can take on integer values 0, 1, 2, ..., n only. That is why, when calculating the probability value of $P\{x_1<m<x_2\}$ with the aid of normal distribution, the values of the distribution function are considered at points $x_1 - 1/2$ and $x_2 + 1/2$ instead of x_1 and x_2; in other words, the equality

$$P(x_1 < m < x_2) \doteq F\left(\frac{x_2 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - F\left(\frac{x_1 - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

is replaced with the corrected equality

$$P(x_1 < m < x_2) \doteq F\left(\frac{x_2 + 1/2 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - F\left(\frac{x_1 - 1/2 - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

Going from random variable m to $p = m/n$, the correction is correspondingly changed from 1/2 to $1/(2n)$.

Inequalities in (1.1) take on the form

$$P\left\{-u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} - \frac{1}{2n} < p - \pi < u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \frac{1}{2n}\right\} = 1-\alpha \tag{7}$$

which is the same as

$$P\left\{|p-\pi| < u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} + \frac{1}{2n}\right\} = 1-\alpha \tag{8}$$

The endpoints p' and p'' of the confidence interval (1) are then determined by solving quadratic equations

$$\left(p - \pi - \frac{1}{2n}\right)^2 = u_{1-\alpha/2}^2 \frac{p(1-p)}{n}$$

and

$$\left(p - \pi + \frac{1}{2n}\right)^2 = u_{1-\alpha/2}^2 \frac{p(1-p)}{n}$$

for the unknown π. The result is

$$p', p'' = \frac{np \mp \frac{1}{2} + \frac{u_{1-\alpha/2}^2}{2} \mp u_{1-\alpha/2}\sqrt{\frac{u_{1-\alpha/2}^2}{4} + n\left(p \mp \frac{1}{2n}\right)\left(1 - p \pm \frac{1}{2n}\right)}}{n + u_{1-\alpha/2}^2} \tag{9}$$

Approximation 2

The second method for determining the approximate confidence interval for π is based on the statements cited at the beginning of the paragraph; namely, on the assumed approximate normal distribution of random variable *p*.. Let us suppose that the statistics

$$\frac{p-\pi}{\sqrt{p(1-p)}}\sqrt{n-1} \tag{10}$$

is asymptotically governed by normed normal distribution; consequently, the following equation holds for large *n* at the approximate level of confidence $1 - \alpha$

$$p - u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}} < \pi < p + u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}} \tag{11}$$

This interval can also be corrected with respect to discontinuity:

$$p - \frac{1}{2n} - u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}} < \pi < p + \frac{1}{2n} + u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n-1}} \tag{12}$$

Both of the above-mentioned methods for determining the confidence interval (whether corrected for the discontinuity or not) provide more or less good results for values of π around 0.5; if the value of π is too small ($\pi < 0.1$) or too large ($\pi > 0.9$), the approximation by normal distribution is not very suitable. The calculated endpoints of the confidence interval then significantly differ from those tabled by exact solution of equations (2) and (3). This fact is easy to understand, because normal distribution is symmetric, hence it provides a symmetric two-sided confidence interval. Binomial distribution is symmetric if $\pi = 0.5$, which is far from the reality for $\pi < 0.1$ or $\pi > 0.9$ – in such situations, skewness of binomial distribution is very high. In such situations the above-mentioned approximations should not be used and equations (1.2) and (1.3) should be solved directly. The values can also be looked up in the above-mentioned tables. Another option is to calculate your own solutions; MS Excel provides an ideal tool for such calculations in its **Solutions** procedure. First of all, let us have a look at binomial distribution within MS Excel.

Binomial distribution within MS Excel

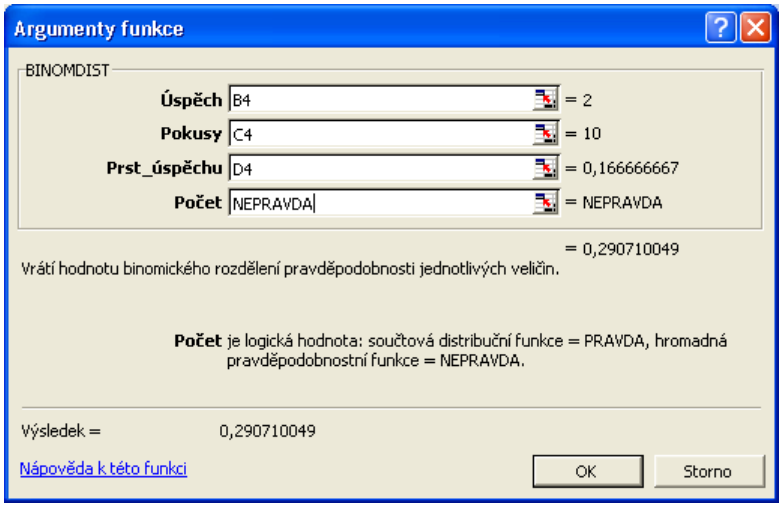
Function BINOMDIST is used in MS Excel for the distribution function and probability values of binomial distribution. Its arguments have the following syntax:

Úspěch – x (number of successes in trials). It is the value at which F(x) or P(x) is calculated.

Pokusy – n (number of independent trials).

Prs_úspěchu – π. Probability of success in each trial.

Počet – NEPRAVDA for the probability mass function, $P(x)$; PRAVDA for the distribution function, $F(x)$.



The Solutions procedure in Excel

The following relationships are easily derived from equations (2) and (3):

$$\alpha_1 = 1 - F(m-1) \tag{13}$$

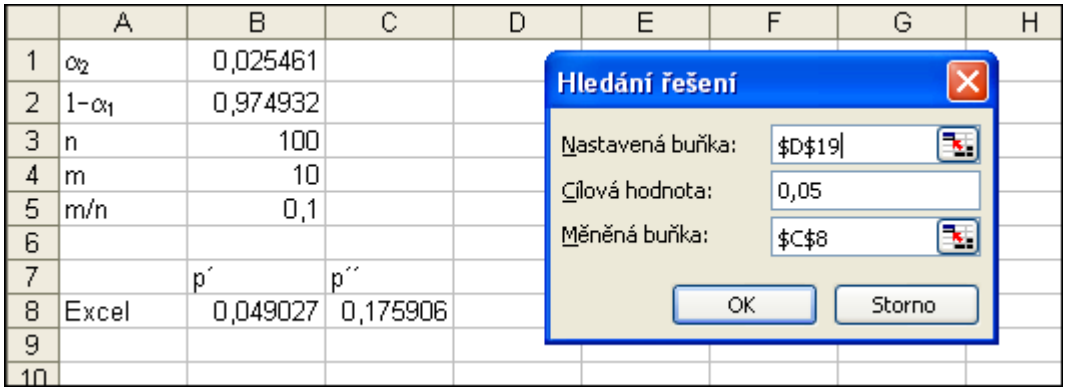
equivalently

$$1 - \alpha_1 = F(m-1) \tag{14}$$

$$\alpha_2 = F(m) \tag{15}$$

These relationships can be utilised when directly seeking the solutions of equations (2) and (3) in Excel. If looking for the 90% confidence interval for parameter π with the aid of the Excel Solver, we employ the following procedure:

Set Target Cell B1 (in which the value of distribution function of binomial distribution is calculated =BINOMDIST(B4;B3;C8;1)) to value 0.05, By Changing Cell C8, and hit the Solve button.



Similarly, find the solution for B2 (=BINOMDIST(B4-1;B3;B8;1)) while changing cell B8.

Practical calculation

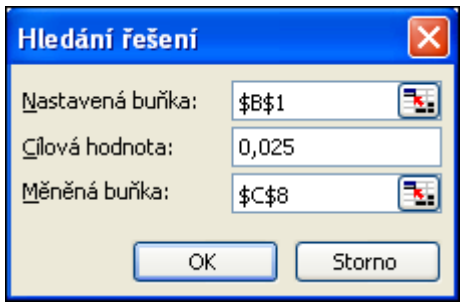
Example 1: Out of 180 sampled elements (let us recall we consider independent observations, that is, sampling with replacement or from a very large population), 80 elements have the monitored sign; in other words, $n = 180$, $m = 80$; consequently $p = 0.444$.

The exact confidence interval (p', p''), covering the unknown value of the basic relative frequency P with confidence value 0.95, will be determined with the aid of tables. According to tables JANKO (1958), it is

$$p' = 0.370 \text{ and } p'' = 0.520.$$

Looking for a solution in Excel:

	A	B	C	
1	α_2	0,0253246		
2	$1-\alpha_1$	0,9746204		
3	n	180		
4	m	80		
5	m/n	0,4444444		
6				
7		p´	p´´	
8	Excel	0,370762	0,520011	
9				



When using the first method of normal approximation, we get

according to (6), $p' = 0.374$ and $p'' = 0.517$,

and according to (7), $p' = 0.371$ and $p'' = 0.520$.

When using the second method of normal approximation, we get

according to (11), $p' = 0.372$ and $p'' = 0.517$,

and according to (12), $p' = 0.369$ and $p'' = 0.520$.

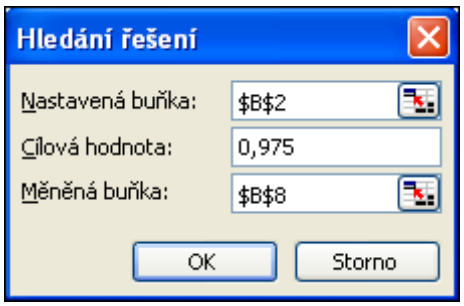
In this case, the normal approximation of binomial distribution is very good for determining the confidence interval, especially if the correction term $1/(2n)$ is applied. The result of Excel Solver is also very good.

Example 2: Let us verify the normal approximation under less favourable conditions; e.g. $n = 90$ and $m = 10$, with $p = 0.111$.

The endpoints of the 95% confidence interval looked up in tables are

$$p' = 0.055 \text{ and } p'' = 0.196.$$

Finding the solution in Excel:



	A	B	C
1	α_2	0,0258866	
2	$1-\alpha_1$	0,9758043	
3	n	90	
4	m	10	
5	m/n	0,1111111	
6			
7		p´	p´´
8	Excel	0,0542907	0,194199

When using the first method of normal approximation, we get

according to (6), $p' = 0.061$ and $p'' = 0.193$,

and according to (7), $p' = 0.057$ and $p'' = 0.199$.

When using the second method of normal approximation, we get

according to (11), $p' = 0.046$ and $p'' = 0.176$,

and according to (12), $p' = 0.040$ and $p'' = 0.182$.

The differences are considerably larger now. Even if the discontinuity corrections are included, for example for the second method, the interval is not sufficiently extended to get the upper endpoint to the tabled value of 0.196. On the other hand, the results obtained with the aid of Excel Solver is very good, close to the tabled values.

3. Conclusions

When determining the confidence interval for parameter value π of an alternative distribution, approximation by normal distribution can be successfully applied (with possible correction for discontinuity) if the actual value of π lies between 0.1 and 0.9. However, if $\pi < 0.1$ or $\pi > 0.9$, tabled values must be utilised. In both instances, MS Excel Solution can be used.

4. References – books

ARKIN, H.: Handbook of Sampling for Auditing and Accounting. New York, McGraw-Hill 1963

COCHRAN, W.G.: Sampling Techniques, 3rd ed. New York, Wiley 1977, Chap. 2.

HÁJEK, J.: Theory of probabilistic sampling with applications to sampling surveys (in Czech). Prague, ČSAV 1960

HÁTLE, J., LIKEŠ, J.: Elements of the probability theory and mathematical statistics (in Czech). Prague, SNTL/Alfa 1972

CHUNG, J.H., DELURY, D.B.: Confidence Limits for the Hypergeometric Distribution. Toronto, University of Toronto Press, 1950.

JANKO, J.: Elements of statistical induction (in Czech). Prague, State Statistical Office 1937

JANKO, J.: Statistical tables (in Czech). Prague, ČSAV 1958

KORDOS, J.: Tablice minimalnej liczebności próbki i wielkości przedzialów ufności. Warsaw, GUS 1967

References – articles

BUNKE, O.: Neue Konfidenzintervalle für den Parameter der Binomialverteilung. Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, mathem.-naturwissenschaftliche Reihe, 9, 1960, 3, 335 - 363.

BUONACORSI, J.P.: A note on confidence intervals for proportions in finite populations. American Statistician, 41, 1987, 3, 215 – 218.

BURSTEIN, H.: Finite population correction for binomial confidence limits. JASA, 70, 1975, 349, 67 – 69.

CLOPPER, J., PEARSON, E.S.: The use of confidence or fiducial limits ilustrated in the case of the binomial. Biometrika, 26, 1934, 1-2, 404 – 413.

ERDÖS, P., RÉNYI, A.: On the central limit theorem for samples from a finite population. MTA-MKIK, A 4, 1959, 1, 49 – 61.

HÁJEK, J.: Limiting distributions in simple random sampling from a finite population. MTA-MKIK, A 5, 1960, 3, 361 – 374.

KOŻNIEWSKA, I.: Porównanie efektywności losowania ze zwracaniem i bez zwracania przy nieznanej wariancji populacji generalnej. Zastosowania matematyki, 2, 1955, 3, 297 – 303.

QUESENBERRY, C.P., HURST, D.C.: Large sample simultaneous confidence intervals for multinomial proportions. Technometrics, 6, 1964, 1, 191 – 195.

Address of authors:

Luboš Marek, Doc., RNDr., CSc.

VŠE Praha

Vysoká škola ekonomická v Praze

nám. W. Churchilla 4

130 67 Praha 3

marek@vse.cz

Michal Vrabec, Mgr., CSc.

VŠE Praha

Vysoká škola ekonomická v Praze

nám. W. Churchilla 4

130 67 Praha 3

vrabec@vse.cz

Prognóza struktury vzdělanosti v České republice do roku 2050[1]

Petr Mazouch, Tomáš Fiala, Jakub Fischer

Abstract: This paper is focused on the population prognosis made by level of education for the Czech Republic, till 2050 year. Formal education is one of the basic type how the human capital could be measured and the prognosis of this structure is one of the most important factors for prediction of other socio-economic indicators. The prognosis is based on the classical population prognosis and on the prognosis of number of school leavers of each level of education in future. Joining of these two things we can predict the education structure of the population in the future.

Key words: education level, population prognosis, human capital

1. Úvod a cíl

S růstem významu lidského kapitálu a jeho vztahu k ekonomickým veličinám je od počátku zmiňována otázka měření lidského kapitálu. Existuje několik možností, jak lidský kapitál změřit, ale podstatným nedostatkem většiny z nich jsou chybějící data nutná pro kvantifikaci. Nejčastěji se k měření lidského kapitálu používá nejvyšší úroveň dosaženého vzdělání, nebo lépe řečeno počet let strávených studiem a případně i účast na dalším vzdělávání.

Vycházíme–li z předpokladu, že vzdělávání by mělo plnit funkci selektivní a funkci produktivní, pak lze říci, že selektivní funkce vybírá a „třídí“ jedince na základě jejich vrozených či v rodině získaných schopností, talentů a zejména motivace, a funkce produktivní předpokládá, že škola k tomuto základnímu kapitálu přidává další hodnotu – rozvíjí a zdokonaluje základní schopnosti a znalosti. V případě ideálního fungování tohoto modelu by tedy pro zaměstnavatele bylo spolehlivým indikátorem úrovně lidského kapitálu předložení diplomu o absolvování některého ze vzdělávacích stupňů.

K uskutečňování formálního vzdělání slouží vzdělávací instituce (základní školy, střední školy, odborné školy, odborné učiliště, vyšší odborné školy a vysoké školy). V České republice je většina škol zřízena státem, regionálními a obecními úřady a studium na nich je bezplatné. Kromě toho existují soukromé školy, které jsou částečně financovány státem a částečně z plateb školného od studentů. Stát pomocí Ministerstva školství a tělovýchovy (MŠMT) provádí kontrolu nad oběma typy těchto školních institucí.

Na úrovni lidského kapitálu přitom větší či menší měrou závisejí hodnoty řady dalších socio-ekonomických ukazatelů, jako jsou produktivita práce, nezaměstnanost nebo střední délka života. Závislost mezi úrovní vzdělanosti, tedy velikostí lidského kapitálu a ostatními socio-ekonomickými ukazateli je možné měřit několika způsoby. Například lze porovnat různé úrovně vzdělanosti a ostatních ukazatelů v čase nebo v prostoru a z těchto hodnot se pak pokusit modelovat vztah mezi jednotlivými proměnnými.

Pro takové srovnání jednotlivých ukazatelů a zejména pro jejich predikci je nutné mít delší časovou řadu právě hodnot úrovně lidského kapitálu, na základě které se bude predikovat jejich vývoj.

Protože se v poslední době autorský tým věnuje zejména popisu a zkoumání právě takových vztahů, bylo jen otázkou času, kdy začne být nevyhnutelné pokusit se predikovat

[1] Příspěvek vznikl za podpory Národního programu výzkumu II Ministerstva školství, mládeže a tělovýchovy ČR č. 2D06026 „Reprodukce lidského kapitálu“.

úroveň vzdělanosti, která by v budoucnu byla základem pro předpovědi dalších socio-ekonomických ukazatelů.

Předmětem předkládaného článku je tedy predikce úrovně vzdělanosti populace osob starších 15 let na území České republiky do roku 2050. Základem takové prognózy musí být prognóza populační a také nutně prognóza změny vzdělanostního systému, který bude v příštích obdobích realizovaný právě v České republice.

2. Metodika

Jak již bylo řečeno výše, základem celé projekce vzdělanostní struktury je klasická populační prognóza. Tato populační prognóza vychází ze základního předpokladu, že demografické chování české populace bude (s jistým zpožděním) kopírovat demografické chování populace Nizozemska. Nizozemsko bylo vybráno z toho důvodu, že se jedná o populaci, kde již byl dokončen přesun plodnosti do vyššího věku a plodnost je zde poměrně stabilní, rovněž úmrtnost v Nizozemsku se zdá být poměrně stabilní. Navíc se jedná o populaci geograficky nepříliš vzdálenou a co do velikosti v jistém smyslu srovnatelnou s ČR. Jako výchozí demografická struktura bylo pro obě varianty prognózy použito složení obyvatelstva ČR podle pohlaví a jednotek věku k 1. 1. 2007 dle údajů Českého statistického úřadu (ČSÚ).

Odhad vývoje plodnosti byl prováděn na základě odhadu vývoje plodnosti jednotlivých „pseudokohort“, tj. vzájemně se překrývajících kohort žen vždy dvou sousedních ročníků narození, jejichž plodnost lze odhadovat pomocí průřezových specifických měr plodnosti. Předpokládalo se, že plodnost kohort českých žen bude s určitým zpožděním „kopírovat“ plodnost žen Nizozemska, kde již byl ukončen přesun plodnosti do vyššího věku a kohortní plodnost se zde zdá být poměrně stabilní.

Podobně jako v případě plodnosti se i v případě úmrtnosti předpokládalo, že úmrtnost mužů i žen ČR se bude blížit úmrtnosti v Nizozemsku. Během posledních 10 let rostla střední délka života žen v Nizozemsku v průměru o 0,13 roku ročně. Byl přijat předpoklad, že takto poroste střední délka života nizozemských žen i nadále až do roku 2050.

Růst střední délky života mužů v Nizozemsku byl v posledních letech zhruba dvojnásobný – střední délka života rostla průměrně o 0,26 roku ročně. Přitom však střední délka života žen v Nizozemsku v roce 2005 byla pouze o 4,42 roku vyšší než střední délka života mužů. Podle současných poznatků bývá střední délka života žen bývá vždy minimálně o zhruba 4 roky vyšší než střední délka života mužů. Proto jsme přijali předpoklad, že růst střední délky života mužů v Nizozemsku do roku 2010 poklesne na úroveň růstu střední délky života žen a na této úrovni setrvá.

Ze složek populačního vývoj je migrace tou nejobtížněji prognózovatelnou. I nadále lze předpokládat, že Česká republika zůstane zemí imigrační. Zdrojem kladného migračního salda budou zejména přistěhovalí ze Slovenska, Ukrajiny, Ruska a Vietnamu.

V letech 2007 a 2008 předpokládáme migrační saldo 30 000 resp. 32 500 osob. V letech 2009 – 2016 nadále uvažujeme zejména pracovní migraci mužů, přičemž o ženách se lze domnívat, že budou muže následovat s odstupem 2 let. V tomto období předpokládáme migrační saldo 35 000 osob. V letech 2017 – 2021 uvažujeme stejné věkové a pohlavní složení migračního salda jako v předchozím období, každý rok tohoto pětiletého období se saldo zvýší o 1 000 osob. Od roku 2022 uvažujeme migrační saldo 40 000 osob ročně.

Takto získaná populační prognóza byla základem pro projekci výkonů vzdělávací soustavy, kterou vypracoval Bc. Vladimír Hulík z Ústavu pro informace ve vzdělávání.

Projekce osob je asi nejjednodušší částí prognózy, protože zde existuje velmi podrobná evidence. Podobně jako v dalších vzdělanostních stupních, i zde se používá metoda propustnosti mezi jednotlivými ročníky. Počet absolventů je tedy ovlivněn počtem přijatých – tzv. mírou účasti a prostupností mezi ročníky.

V případě středoškolského vzdělávání je základním faktorem pro počet absolventů opět počet nově přijatých. Ten je zde odhadován na základě podílu absolventů předcházejícího vzdělávacího stupně. Počet žáků je kalkulován na základě prostupnosti mezi ročníky a počet absolventů je vypočten jako podíl žáků z posledního ročníku. Data o středoškolském vzdělávání však nedosahují kvalit dat o základním vzdělávání.

Asi nejsložitější situace je v souvislosti s vysokoškolským stupněm vzdělávání. Opět se zde vychází z počtu absolventů předchozího stupně studia – zde z počtu maturantů, z prostupnosti mezi jednotlivými ročníky a také se uvažuje, že studenti studují standardní dobu určenou pro absolvování studia.

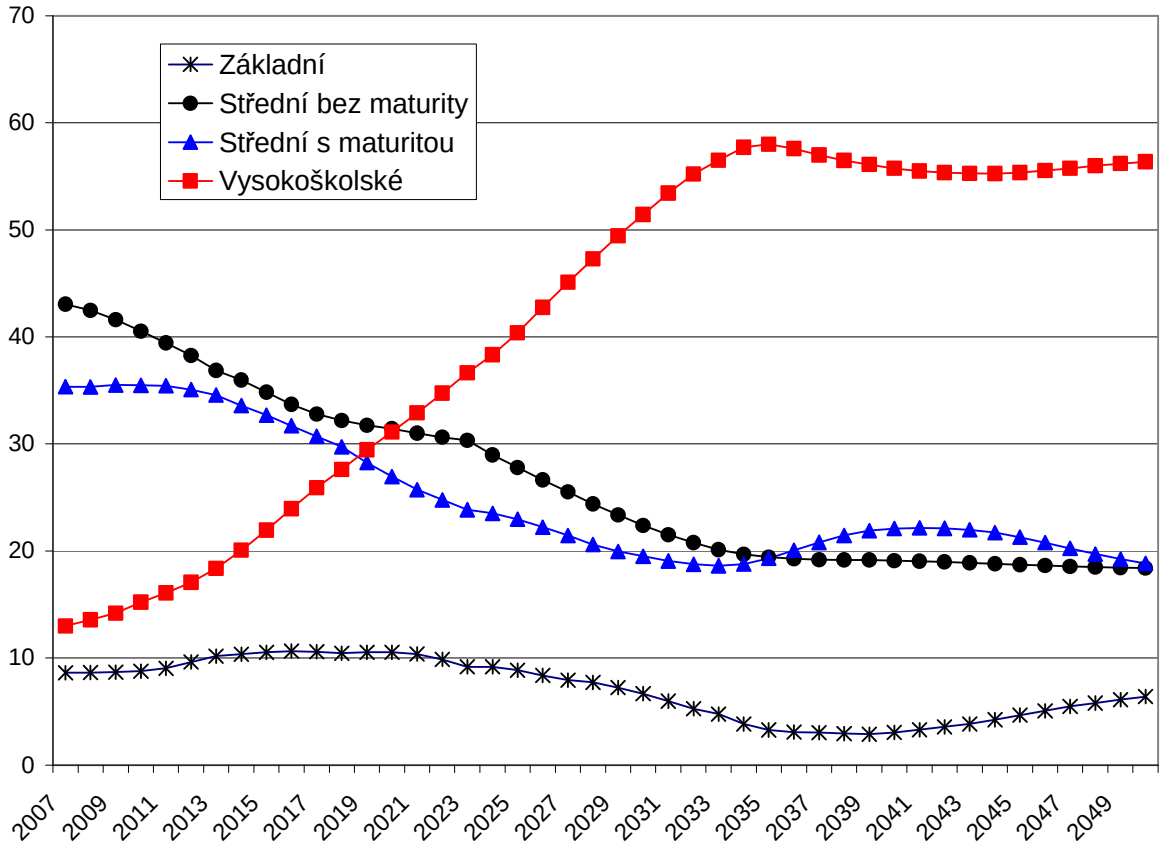
Výsledkem této projekce je počet absolventů jednotlivých stupňů studia v jednotlivých letech, pro které je kalkulována prognóza.

Na základě znalosti vzdělanostní struktury populace, kterou máme ze Sčítání lidu, domů a bytů (SLDB), znalosti rozdílné úmrtnosti obyvatelstva podle pohlaví a vzdělání, základní populační prognózy a prognózy počtů absolventů jednotlivých vzdělanostních stupňů můžeme zkonstruovat prognózu obyvatelstva podle vzdělání, jejímž výsledkem bude vzdělanostní struktura populace České republiky až do roku 2050.

Při takové prognóze uvažujeme, že osoby s určitým stupněm vzdělání se vlastně rodí až v okamžiku absolvování tohoto stupně a umírají buď přirozenou smrtí nebo absolvováním navazující úrovně vzdělání. Realizace takové prognózy je tedy modelováním dalších čtyř podpopulací v rámci každého roku.

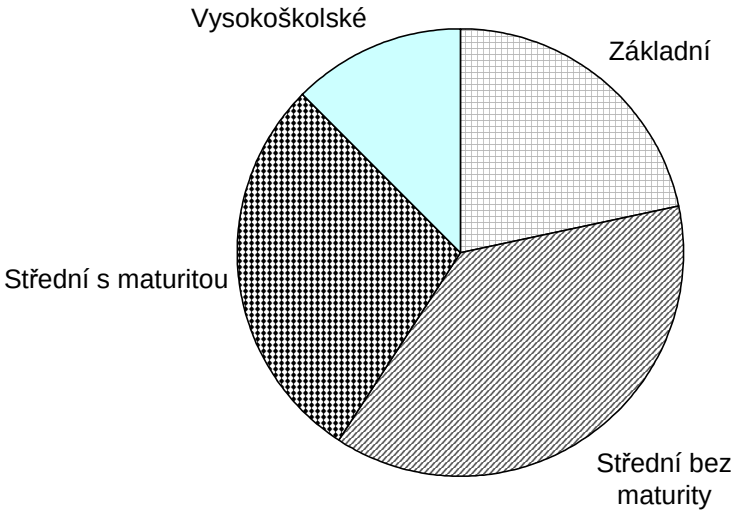
3. Výsledky

Graf 1: Vývoj relativního podílu osob podle vzdělání ve věkové skupině 30 – 39 let (v %)



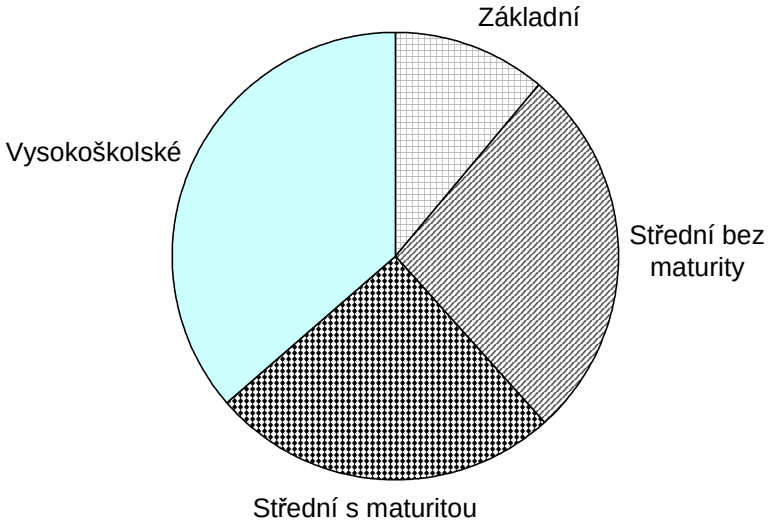
Zdroj: propočty autorů.

Graf 2. Relativní zastoupení jednotlivých vzdělanostních skupin v populaci 15+ v roce 2007



Zdroj: propočty autorů

Graf 3. Relativní zastoupení jednotlivých vzdělanostních skupin v populaci 15+ v roce 2050



Zdroj: propočty autorů

Vzhledem k predikci vývoje školského systému v České republice bude docházet k značnému nárůstu podílu osob s vysokoškolským vzděláním. V grafu č. 1 je zobrazena

pouze skupina osob ve věku 30 – 39 let, ale již na první pohled je patrný dopad politiky vzdělanosti, aplikované v České republice v současné době, tedy rapidní nárůst počtu absolventů vysokých škol, který se v současné době blíží 50 % osob z populačního ročníku.

Tyto silné „absolventské ročníky“ se dostanou do skupiny 30 –39 let přibližně za deset let, což je patrné z růstu podílu této skupiny, až přibližně do roku 2030, kdy se tento růst zastaví přibližně na hodnotě 55 %. Tento silný růst má dopad na vzdělanostní skupinu středoškolsky vzdělaných osob, ať už s maturitou, či bez maturity. Obě tyto skupiny během celé projekční doby klesnou přibližně na polovinu své hodnoty z roku 2007. Skupina osob se základním vzděláním je již v této době velmi malá a vývoj jejího počtu je podobný jako u předchozích dvou skupin, tedy pokles přibližně na polovinu současné hodnoty.

V grafu č. 2 a 3 je srovnání populace 15+ v roce 2007 a 2050 z hlediska vzdělanostního složení. Je zde zřejmý nárůst podílu vysokoškolsky vzdělaných osob na úkor skupin se základním vzděláním a se středoškolským vzděláním bez maturity. Skupina osob se středoškolským vzděláním s maturitou je přibližně stejná v obou obdobích.

4. Diskuse a závěr

Cílem příspěvku byla konstrukce prognózy skupin obyvatelstva podle vzdělání, která by byla možná využít v dalších analýzách vztahů s ostatními socio-ekonomickými ukazateli. Výsledkem je zjištění, že v dalších letech, zejména vlivem posilování vysokého školství, bude rychlým tempem narůstat podíl osob s vysokoškolským vzděláním, zejména v produktivní generaci. Podíl osob s nižším vzděláním bude postupem času v této skupině menší, protože se tyto osoby budou vlivem času (stárnutí) přesouvat do vyšších – poproduktivních věkových skupin.

Tyto závěry dávají možnost realizovat predikci dopadů takových změn ve vzdělanostní struktuře, např. v oblasti důchodových reforem, kde se většinou předpoklad změny vzdělanostní struktury, který má zásadní vliv na schopnost populace zvyšovat produkci, opomíjí.

V dalších projektech se tedy budeme snažit ještě lépe rozpracovat tuto prognózu, a to zejména pro menší územní celky, a také na jejím základě realizovat některé již zmíněné analýzy vývoje dalších socio-ekonomických ukazatelů.

5. Literatura

ČSÚ . Pohyb obyvatelstva v České republice v roce 2001. Praha: Český statistický úřad 2002.
ČSÚ . Sčítání lidu, domů a bytů v roce 2001. Pramenné dílo. Praha: Český statistický úřad 2002.

Adresa autorů:
Ing. Petr Mazouch, RNDr. Tomáš Fiala, CSc., Ing. Jakub Fischer, Ph.D.
Katedra demografie Fakulty informatiky a statistiky Vysoké škola ekonomické v Praze
nám. W. Churchilla 4
CZ 130 67 Praha 3
Česká republika
mazouchp@vse.cz, fiala@vse.cz, fischerj@vse.cz

Gender equality indikátory a databázy vedomostnej spoločnosti

Anna Nemcová[1]

Abstract: Gender equality indicators are source of databases of knowledge economy. In praxis informal education occures in the role of knowledge effectiveness indicator. Hierarchy of indicators from gender equality is discussed as

(1) statistics data (man and female files)
(2) gender indicies
(3) indicators of individual gender oriented politicies
(4) implementation gender to all proceedings as whole

For gender equality are necessary statistics approaches to

- databases
- monitoring
- proceedings and management mechanisms

SR long, middle and short time gender gap pay prognoses shows convergency of female to man unemployment rate

Key words: Gender equality, monitoring, prognoses, proceedings

1. Úvod

Prognostické tímy (experti rôzneho zamerania vrátane ekonómov za pozornosti Svetovej Banky) pre tretie milenium vytypovali okolo tridsať ukazovateľov trvaloudržateĺného rozvoja, v prvej polovici ktorých sa umiestnilo kritérium rovnosti mužov a žien (gender equality).

I keď protagonisti holistických postupov a uplatnenia ďalších transformačných teórií na problematiku súčasnej ekonómie považujú uspokojovanie potrieb súčasnej populácie za riešiteľné, dnes stojíme pred mnohými prozaickými otázkami praktických riešení, akými je napríklad aj získavanie údajov pre rozvoj vedomostnej spoločnosti, založenej na poznatkovej ekonomike a profitujúcej z inovácií v informačných a nových technológiách.

Najbližším cieľom pre naše teritórium napríklad je, že do roku 2010 si Európska únia stanovila ambiciózny cieľ rozvoja vedomostnej spoločnosti (Lisabonskou stratégiou) porovnateĺný s USA, a to aj zvýšením zamestnanosti žien na 60 % celkovej úrovne zamestnanosi a 70 % celkovej populácie žien.

Pre presadenie takéhoto rozvoja sa prijal úzus rodovej rovnosti transhormovaný do rovnosti príležitostí pre ženy a mužov, ktorý sa má uplatniť vo všetkých postupoch. V zamestnávaní to znamená odstrániť rodovú segregáciu na trhu práce. V údajovej základni ide o pokračovanie smernice OSN z roku 1997, že všetky štatistické dáta sa majú členiť na údaje o mužoch a ženách.

Treba teda vyriešiť otázku, do akej miery je takéto delenie dát zmysluplné: Ide o spätnú väzbu dát s reálnou praxou tak, aby aj samotný vývoj databáz bol v neposlednom rade ekonomicky efektívny.

[1] Anna Nemcová, Centrum pre výskum regiónov

2. Gender indikátory a vedomostná spoločnosť (Pilotný výskum SR a východiská vo svete)

Na základe pilotných skúseností v oblasti vývoja a tvorby indikátorov rodovej rovnosti na trhu práce v regióne Pezinok- Modra, ako sa uvádza v NEMCOVÁ, A. 2007 (1), v ktorom sme skúmali

- Teoretické východiská
- Štatistické spracovanie údajov
- Ne/riadený experiment
- Predpoklady pre prácu informálneho interkultúrneho informatického intergeneračného vzdelávania

možno hierarchizovať gender equality indikátory, ak vychádzame z premisy, že vedomostnú spoločnosť reprezentuje úroveň teórie poznania samotnej vedomostnej spoločnosti a praktické kroky na jej dosiahnutie.

Otázka: Do akej miery majú získavané dáta vypovedaciu schopnosť? zostáva naďalej vecou každodenného skúmania možností štatistiky v interdisciplinárnom prístupe. Ekonomický rozmer časopriestoru reprezentuje regionálny prístup, ktorý akcentuje EU, pričom obsahový rámec rodového rozpočtovania (gender budgeting) pre vedomostnú spoločnosť v zásade rovnako ako miléniové koncepty prešiel vo svete vývojom od separovaných ukazovateľov k holistickému prístupu. (Ide o vývoj od delenia dát za mužov a ženy podľa inštrukcie OSN 1997 k celostnému uplatňovanu rodovej rovnosti v politikách Európskej únie na základe Iniciatívy spoločenstva EQUAL II 2004, ale aj o uplatnenie ukazovateľov štatistík ako akceptovateľných verenosťou – napr. medián miezd, Giniho koeficient, atď. na rozdiel od štatisticky využívanej terminógie v podobe smerodatnej odchýľky, R a i.)

Teóriu poznania štádií vedomostnej spoločnosti možno oprieť o poznanie míľnikov pri rozvoji – podľa autora RIFKIN, J. 2007, postindustriálneho sociológa a poradcu EU pre obnoviteľné energie míľnikom/novou priemyselnou revolúciou je diverzifikácia zdrojov analogická diverzifikácii IT: Každý, kto chce a môže pracovať na PC, dnes efektívne pracuje bez udržiavania nákladnej centrály (obdobný trend - diverzifikácia bez centrály - sa už postupne presadzuje osobitne v trvalo obnoviteľných zdrojoch pre životné prostredie).

Diverzifikácia zdrojov v tejto stati obmedzíme na informáciu z oblasti dopytu a ponuky vzdelávania v IT a na základné zhrnutie, ktoré vyplýva z výskumu a diskusií (napr. (1) (2)):
Pre vedomostnú spoločnosť je nutné vytvoriť system signalizácie efektívnosti vzdelávania:

> Databázy a štatistiky formálnych inštitúcií SR neevidujú úspešnosť absolventov v ekonomickom živote (2) - z toho vyplýva, že monitoring doplňujúceho vzdelávania predstavujú vo všeobecnosti databázy inštitútov informálneho vzdelávania, ktoré sa vyvíjajú podľa reálnej potreby trhu a sebauvedomenia pracovnej sily (vrátane gender aspektov z hľadiska uplatnenia slobody individua a potreby vypočuť hlasy tých, ktorí chcú a môžu pracovať - v zmysle štúdie Svetovej Banky z roku 2005, podľa ktorej v Strednej a východnej Európe trh nesplnil očakávania ohľadom odstránenia nezamestnanosti a je treba hľadať doplňujúce riešenia).

Vzdelávanie u nás možno skúmať aj z pozície

1. ÚPSVaR – respondenti sú však pod ekonomickým tlakom a majú predestinovaný výber vzdelania: databázy sú jednorazové a obmedzené na obdobie nezamestnanosti
2. obce - respondenti sú vyberaní len náhodne a výber sa aj pre nedostatok zdrojov spravidla neopakuje

Databázy zo všetkých troch inštitútov však majú nezastupiteľnú výhodu a tou je spravidla 100% návratnosť dotazníkov, t. j. nevychýlenosť pravdepodobnostných odhadov(LUHA, J. 2005).

Pozn.: Pre konštrukciu dotaznikov je zaujímavé, že kým na otázky dotazníkov v ÚPSVAR a dotazníkov pre zisťovanie situácie gender equality vrátane gender equality vo vzdelávaní v obci, ktoré sme v rámci projektu (1) pripravili, odpovedajú respondenti bez zábran, otázky experimenálneho dotazníkového skúmania, konformného so skúmaním *AEO (Adult Education Observing) EUROSTATU* respondenti z radov uchádzačov o zamestnanie odmietajú v niektorých prípadoch zodpovedať, čo zdôvodňujú obťažnosťou otázok a individuálne chápaným prekračovaním intimity.

3. Gender indikátory a postupy (Hierarchické usporiadanie)

Z výskumu (1) NEMCOVÁ, A. 2007 vyplýva hierarchizácia štatisticky sledovateľných gender charakteristík nasledovne:

1. Databázy EUROSTATU: ide o vyčlenenie súborov dát s potenciálom deľby údajov viazaných na mužov a ženy osobitne (v SR ide komplex dostupných údajov, v ČR sa publikuje ako osobitný systém; špeciálne v zmysle dodžiavania smerníc EÚ sú analyzované databázy miezd mužov a žien SR)
2. Koeficienty a kvóty ako kvantifikácie jednotlivých izolovaných javov: znamenajú výpočty z existujúcich údajov a získavanie nových údajov a charakteristík
    a) Zo štatistických ukazovateľov, ako je stredná odchýľka atď. sa využívajú tie, ktoré majú bezprostredne interpretovateľný obsah. Ide nielen o ukazovatele počtu zamestnaných mužov a žien ako základ odstránenia segregácie v zamestnaní, ale aj o vyrovnanie miezd elimináciou rozdielov miezd mužov a žien *(gender pay gap).* Popri používanom *indexe disimilarity pre zamestnanosť* tak môžme napr. *vyvodiť index disimilarity miezd* ako

$$I_{DM} = ½ \sum (M_{ii}/M - F_{ii}/F) \qquad (1)$$

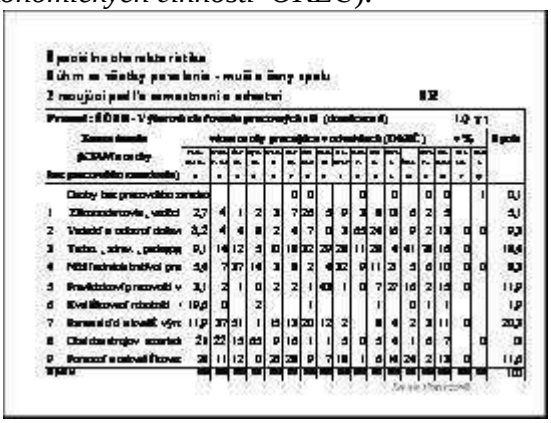
    pre $i=1, \ldots, n$; M = premenná (objem skúmaného mzdového pásma i z celkového objemu miezd mužov) prislúchajúca mužom, F = premenná (objem skúmaného mzdového pásma i z celkového objemu miezd žien) prislúchajúca ženám).
    b) Kým v roku 2006 vzorku na získavanie údajov o zamestnanosti a mzdách v SR - Informačný system o cene práce (ISCP) s 500 000 respondentmi - možno považovať za konsolidovanú, v roku 1996 sa použitím nevážených a vážených priemerov získali údaje s pomerne variabilnými výsledkami o smerovaní rodovej mzdovej medzery - *gender pay gap.* Skúmala sa disimilarita zamestnanosti a miezd, vážené a nevážené priemery a ich vplyv na zmeny v rode, odvetví, kvalifikácii, veku, vlastníckej štruktúre (NEMCOVÁ, A. 1996).
    c) Extirpáciou databázy ISCP z 300 000 respondentov sa zistil ukazovateľ

$$\alpha = 0{,}000\,3\,\% \quad \text{a} \quad \beta = 0{,}1 \qquad (2),$$

    ktorý znamená, že v celom sledovanom súbore ISCP sa našlo 1 pracovné miesto s podmienkami pre ženu rovnakými ako pre 9 mužov, teda pre rovnosť v zamestnávaní by sa mohli aspoň štyri takéto miesta obsadiť ženami. Pre dôkaz rovnakej odmeny za rovnakú prácu mužov a žien, ktoré EU požadovala, sa tieto štatistické zistenia ukázali dostatočné (obdobné zistenia ako sa uvádzajú v NEMCOVÁ, A. 1996 sa robia i pre rok 2006).
3. Jednotlivé politiky presadzovania rovnosti mužov a žien uplatňované v spoločnosti: znamenajú skúmanie a hodnotenie komplexov súborov dát izolovaných politík z hľadiska progresu jednotlivých politík pre ženy a mužov.

Úspešnosť jednotlivých politík – ako politiky koncipované gender indikátory (napr. presadzovanie politiky zamestnanosti EÚ 70 % zamestnanosťou žien pri celkovej 60 % zamestnanosti) sa považujú za závislé od presadzovania politickými skupinami, čo bráni ekonomickému efektu presadiť sa celostne (stráca sa synergický efekt).

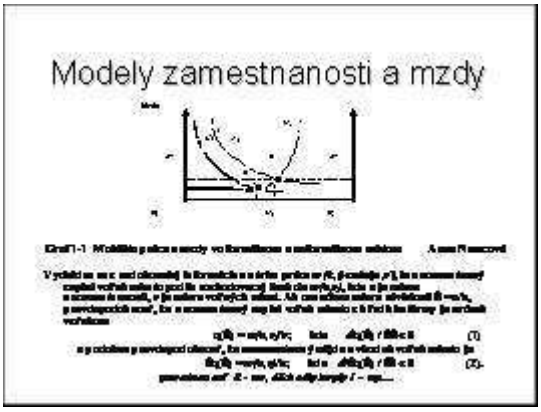
Z hľadiska úspešnosti politík možno oceniť napr. že v SR ukazovatele pre nezamestnanosť ÚPSVaR možno monitorovať detailne, pričom je napr. priestor na zdokonalenie kontinuity prechodov v ne/zamestnanosti a prepojenie jednotlivých politík v HDP. (*Obrázok 1.* ilustruje zber dát o ne/zamestnanosti do úhrnu všetkých povolaní ako Súhrnný informačný list odvetvových politík podĺa *Katalógu zamestnanosti KZAM* a *Odvetvového katalógu ekonomických činností OKEČ*).



Obrázok 1. Súhrnný informačný list odvetvových politík

4. Celostné ukazovatele: Ekonomicky sa dokazuje, že nie je možné presadiť rovnosť príležitostí žien a mužov využitím súhrnu jednotlivých politík, ukazovateľov či dát (predchádzajúce body 1, 2, 3), ale len ako celostný (holistický) prístup ku všetkým postupom v živote spoločnosti (príkladom prijatia komplexého gender opatrenia je Iniciatíva Spoločenstva EÚ EQUAL). *Obrázok 2. Systém monitorovania equilibria modelmi práce ako equilibria zamestnanosti a mzdy* ilustruje systém pre vyhodnotenie a evaluáciu zhody equilibria vedomostnej spoločnosti a sveta práce.

Monitorovací mechanizmus tohto procesu predstavuje prienik equilibria rozvoja spoločnosti do bodu rovnováhy ne/zamestnanosti. Sancou pre vedomostnú spoločnosť je teda orientácia equilibria miezd na vedomostnú bázu mužov a žien s ich špecifikami.



Obrázok 2. Systém monitorovania equilibria modelmi práce ako equilibria zamestnanosti a mzdy

4. Záver

Pre presadenie rovnosti príležitostí (gender equality) vo vedomostnej spoločnosti existujú vo svete a sú v SR aj experimentálne overené pilotné verzie štatistických súborov pre

° databázy

° monitorovanie
° rozhodovacie mechanizmy

ktoré možno zdokonaliť

° prienikom riešení výskumu z hľadiska
- interdisciplinárnej filozofie
- kvantitatívneho postihnutia skutočnosti (finančné banky skúmajú vyjadrenie šťastia)
- legislatívnych návrhov s tlakom na decizívne orgány a akceptáciu verejnosťou

° implementáciou dobrej praxe

Ako príklad dobrej praxe využitia statistik pre gender postupy na uplatnenie rovnosti príležitostí mužov a žien vo vedomostnej spoločnosti možno uviesť dlhodobú, strednedobú (*Obrázok 3. Strednedobá gender prognóza CVR znázorňuje vypočítané hodnoty pre SR)* a krátkodobú prognózu miezd (KVETAN, V. 2006), ktorá už dnes ukazuje na znižovanie nezamestnanosti žien a zmenšovanie rozdielu v rovnosti príležitostí medzi mužmi a ženami.



Obrázok 3. Strednedobá gender prognóza CVR

Využitím štatistiky je v tomto zmysle odhad (gender) zmien, ktoré sa môžu zrealizovať na dosiahnutie predpokladaného cieľa (vedomostnej spoločnosti).

4. Literatúra

(1) NEMCOVÁ, A. 2007. Výskum a tvorba indikátorov rodovej rovnosti. Centrum pre výskum regiónov. In: Rodová rovnosť. Zborník z konferencie 18.10.2007. Trnavská univerzita 2007. 6 s.
(2) NEMCOVÁ, A. 2007. Informálne vzdelávanie. Seminár vedomostná spoločnosť. Bratislava: FES - UK, 8.11.2007. 1 s.
(3) LUHA, J. 2005. Reprezentatívnosť vo výskumoch verejnej mienky. In: Forum statisticum Slovacum, č. 2, 2005. Bratislava: SŠDS, 2005, s. 95-99.
(4) RIFKIN, J. 2007. Hard Talk. TV: BBC, November 15th, 2007.
(5) NEMCOVÁ, A. 1996. Relácie miezd mužov a žien. Výskumná práca. Bratislava: VÚPSVR, 1996. 27 s.
(6) KVETAN, V. 2006. Prognóza miezd I-III. Výskumná práca. Modra: Centrum pre výskum regiónov – CVR, 2006. 60 s.

Adresa autora:
Anna Nemcová, Ing. PhD.
Pod vinicami 17
900 01 Modra
ane.cvr@gmail.com Tel. : 421 905 882681

Prognóza populačného potenciálu Slovenska v roku 2025 (Aplikácia modelu populačného potenciálu na územie Slovenska)

Gabriela Nováková

Abstract: *In geographical terminology the term "potential" can be understood as an indicator or index of existing status (at present) or as a capacity for coming into being (in future). In this paper the term "potential" is used in context of population potential model. The population potential model has appeared in the human geography literature as an index of the intensity of possible interaction between social or economic groups at different locations. It our paper we present the application of this model on spatial system of the Slovak Republic to forecast the population potential surface of Slovakia in 2025.*

Úvod

Potenciálový model patrí do skupiny matematických modelov, ktoré sa v posledných desaťročiach udomácnili v geografických vedných disciplínach. Pod trend stále častejšieho využívania matematických modelov v tejto vednej disciplíne sa podpísal fakt, že spracovanie mnohých štatistických údajov, ktoré model vyžaduje, je značne uľahčené vďaka modernej výpočtovej technike.

Model populačného potenciálu sa používa ako ukazovateľ intenzity možných interakcií medzi rôzne lokalizovanými sociálnymi a ekonomickými skupinami (Rich 1980). Má blízky konceptuálny, empirický a historický vzťah s gravitačným modelom, z ktorého bol odvodený.

Interpretácia populačného potenciálu v čase jeho prvej formulácie Stewartom nebola úplne jasná (Isard 1966). Široké možnosti využitia potenciálového modelu však ponúkli postupne rôzne formy jeho interpretácie v závislosti od charakteru sledovaného problému. Model populačného potenciálu v našom príspevku budeme chápať ako objektívny ukazovateľ priestorových vzťahov medzi geografickými objektmi v kontexte priestoru Slovenskej republiky, resp. ako ukazovateľ možných (pravdepodobných) interakcií medzi jednotlivými územnými jednotkami Slovenska.

Metodika výpočtu

Model populačného potenciálu sa vyznačuje tromi základnými dimenziami, ktorými sú priestor, masa (hmotnosť) a vzdialenosť. Teória modelu potenciálu predpokladá teoretickú možnosť výpočtu potenciálu pre každý bod zemského povrchu. To znamená, že každý bod zemského povrchu sa určitou mierou podieľa na hodnote potenciálu všetkých zostávajúcich bodov na Zemi (Kusendová 1996). Takýto výpočet v praxi však nemožno zrealizovať, preto sme nútení **ohraničiť priestor**, pre ktorý sa potenciál vypočítava a zároveň **určiť body výpočtu**, pre ktoré budeme hodnotu populačného potenciálu stanovovať. Priestorom, pre ktorý sme určovali hodnoty populačného potenciálu, bolo pre nás územie Slovenskej republiky rozdelené na 68 častí, ktoré boli reprezentované 68 bodmi výpočtu. Týmito priestorovými jednotkami boli prevažne okresy Slovenska (v počte 66), centrá výpočtu boli okresné mestá a body výpočtu predstavovali „stredy" týchto centier určené kartografickými súradnicami v systéme S-JTSK (Systém jednotnej trigonometrickej siete katastrálnej). Špecifickými územnými jednotkami boli pre nás Bratislava a Košice (tzv. mestské bratislavské a košické okresy), u ktorých sme ignorovali ich členenie na päť, resp. štyri časti, a do výpočtov vstupovali ako celky, pričom sme k nim priradili ešte územia okolitých okresov vo funkcii ich zázemia (keďže svojím zázemím vo výpočtoch disponujú všetky ostatné okresy Slovenska). Tak vnikol tzv. „bratislavský región" tvorený Bratislavou-mestom a územiami okresov Malacky, Pezinok a Senec a tzv. „košický región" tvorený Košicami-mestom a okresom Košice-okolie. V príspevku ich budeme nazývať „Bratislava" a „Košice". Bod výpočtu (centroid) bol v oboch prípadoch lokalizovaný do mestskej časti Staré Mesto.

Interakčná schopnosť priestorových jednotiek je výsledkom ich spoločensko-ekonomických aktivít. **Masa** (hmotnosť) by mala byť určitým meradlom tejto interakčnej schopnosti - schopnosti danej hmoty vytvárať kontakty. Čím je hmotnosť väčšia, tým je interakčná schopnosť priestorových jednotiek silnejšia. V procese empirických výskumov bola masa meraná rôznymi spôsobmi. Najčastejším a najjednoduchším vyjadrením hmotnosti v modeli populačného potenciálu je veľkosť populácie (resp. jej presne definovanej podmnožiny) - počet obyvateľov prislúchajúci k danej priestorovej jednotke (sídlu, regiónu). Masu v našej aplikácii modelu populačného potenciálu predstavoval počet obyvateľov jednotlivých územných jednotiek prognózovaný pre rok 2025 (stredný variant) (Infostat 2004). Pre porovnanie sme určili hodnoty populačného potenciálu pre rok 2001. Masu v tomto prípade predstavovali počty obyvateľov príslušných priestorových jednotiek ku dňu sčítania (25.5.2001).

V potenciálových modeloch **vzdialenosť** indikuje negatívny vplyv geografických podmienok na veľkosť interakcie populácií v danom geografickom priestore. Predstavuje *„odpor, ktorý kladie prostredie pri uskutočňovaní interakcie“* (Hlavička 1993, str. 35), resp. pôsobí ako bariéra na interakčné schopnosti skúmaných populácií (Kusendová 1996).
Do funkcie vzdialenosti d_{ij} bola v našom modeli dosadená priama vzdušná vzdialenosť medzi jednotlivými bodmi výpočtu.

Na výpočet prognózy populačného potenciálu Slovenska sme použili model **celkového populačného potenciálu**, ktorý môžeme matematicky zapísať vzťahom:

$$CP_i = IP_i + VP_i ,$$

kde CP_i je celkový populačný potenciál priestorovej jednotky *i*,

IP_i je indukovaný populačný potenciál priestorovej jednotky *i*,

VP_i je vlastný potenciál priestorovej jednotky *i*.

Indukovaný populačný potenciál priestorovej jednotky *i* odráža vplyv okolitých územných jednotiek na jednotku *i* bez započítania tzv. vlastného potenciálu priestorovej jednotky *i*. Určili sme ho ako súčet podielov masy P_j ostatných bodov výpočtu *j* a ich priamej vzdušnej vzdialenosti d_{ij} od jednotky *i* (v km) podľa vzťahu:

$$IP_i = \sum_{\substack{j=1 \\ j \neq i}}^{68} \frac{P_j}{d_{ij}}$$

Vlastný potenciál územnej jednotky *i* vyjadruje interakčnú aktivitu prislúchajúcu vlastnej priestorovej jednotke *i*. Na určenie jeho hodnoty sme použili výpočet podľa vzťahu:

$$VP_i = \frac{P_i^c}{1\ km} + \frac{P^z{}_i}{A_i},$$

kde P_i^c je počet obyvateľov žijúcich v centre územnej jednotky *i* (v okresnom meste, v Bratislave-meste alebo v Košiciach-meste),

P_i^z je počet obyvateľov žijúcich v zázemí (mimo centrum) územnej jednotky *i*,

A_i je redukčný koeficient - polomer kruhu s rovnakým obsahom ako je plocha územnej jednotky *i*.

Keďže sme nemali k dispozícii údaje o prognózovanom počte obyvateľov centier výpočtu, populáciu prognózovanú pre rok 2025 sme rozdelili medzi centrum a zázemie každej priestorovej jednotky v pomere, ktorý bol aktuálny v roku 2001. Tým pádom sa musíme vzdať ambície hodnotiť zmeny v populačnom potenciáli spôsobené presunom obyvateľov

medzi centrami jednotlivých územných jednotiek a ich zázemiami a obmedziť sa len na možný vplyv zmeny počtu obyvateľov územných jednotiek ako celku.

Výsledky

Východiskom pre porovnanie boli pre nás hodnoty populačného potenciálu aktuálne pre rok 2001.

Z hľadiska populačného potenciálu prognózovaného pre rok 2025 predpokladáme, že najvyšší potenciál si zachová bratislavský región a nezmení sa ani postavenie Košíc na druhom mieste (tab. 1). Vplyvom zvýšenej hodnoty populačného potenciálu Košíc a zmenšenia hodnoty populačného potenciálu Bratislavy sa rozdiel medzi nimi zmenší (populačný potenciál Košíc bude predstavovať 63,3% z maximálneho populačného potenciálu v súbore územných jednotiek). Zmena však nastane v prípade Nitry, ktorá zostúpi z tretej priečky na štvrtú vplyvom výrazného nárastu populačného potenciálu Prešova (zo 150395 obyv./km na 163980 obyv./km). Prešov odsunie aj okres Žilinu zo 4. na 5. miesto. Poradie na ďalších troch miestach v slede Banská Bystrica, Trnava, Martin sa nezmení, na ďalších priečkach si Prievidza a Trenčín vymenia svoje miesto. Kým na začiatku poradia došlo k niekoľkým vyššie uvedeným zmenám, na konci tabuľky sa poradie nezmení. Najnižšie hodnoty potenciálu preukážu (podobne ako v roku 2001) regióny Snina, Stropkov, Svidník, Sobrance a Medzilaborce (menej ako 14% z maximálnej hodnoty populačného potenciálu v súbore), aj keď pre hodnotu ich populačného potenciálu predpokladáme neveľké zvýšenie (najviac o 2213 obyv./km v okrese Stropkov). V porovnaní s rokom 2001 sa Prešov a Žilina zaradia do oblasti najvyšších hodnôt populačného potenciálu k Bratislave, Košiciam a Nitre; do kategórie vysokých hodnôt sa pravdepodobne presunie Zvolen; do intervalu nízkych hodnôt populačného potenciálu by mali postúpiť z oblasti najnižších hodnôt okresy Stará Ľubovňa a Tvrdošín a naopak, pokles z kategórie stredne veľkých hodnôt do nižšej by mali zaznamenať územné jednotky Komárno a Nové Mesto nad Váhom.

Ak porovnáme prognózované hodnoty populačného potenciálu v roku 2025 s rokom 2001 prostredníctvom koeficientu rastu (tab. 1), zistíme, že územie Slovenska sa rozdelí na dva takmer kompaktné celky: severné (Orava, Kysuce + Žilina a Bytča) a východné Slovensko môžu bez výnimky očakávať zvýšenie populačného potenciálu s najvyššími hodnotami koeficientu rastu (nad 1,05) v oblasti sústredenej okolo okresu Sabinov (Kežmarok, Stará Ľubovňa, Bardejov, Prešov, Vranov nad Topľou, Spišská Nová Ves, Levoča); západná časť Slovenska s výnimkou Dunajskej Stredy a Zvolena má podľa prognózy zaznamenať pokles hodnoty populačného potenciálu s najmenšími hodnotami koeficientu rastu v okresoch Prievidza, Nové Zámky a Komárno (menej ako 0,96) (obr. 1).

Zmeny pozícií v usporiadaní podľa veľkostí hodnôt populačného potenciálu v roku 2025 v porovnaní s rokom 2001 (tab. 2) vyjadruje Kruegerov - Spearmanov koeficient poradovej korelácie, ktorý svojou pomerne vysokou hodnotou (0,9893) naznačuje, že môžeme očakávať vysokú podobnosť usporiadania priestorových jednotiek so stavom v roku 2001. Najväčšia zmena by sa mala prejaviť v poklese Komárna (o 8 miest) a v postupe okresov Kežmarok, Sabinov (o 7 miest), Bardejov, Levoča, Spišská Nová Ves a Vranov nad Topľou (o 6 miest). Ostatné priestorové jednotky zmenili svoju prognózovanú pozíciu maximálne o 5 priečok.

Záver

Porovnaním prognózovaných hodnôt populačného potenciálu pre rok 2025 s hodnotami tohto ukazovateľa v roku 2001 sme zistili, že v populačnom povrchu Slovenska nedôjde k výrazným zmenám. Zvýšenie hodnôt populačného potenciálu, a tým aj pravdepodobnosti interakcie, môže očakávať región severného a východného Slovenska, južné a západné Slovensko, naopak, svoju schopnosť interagovať oslabia. Jednoznačnú dominanciu si zachová bratislavský región a to aj napriek svojej excentrickej polohe. Košice

v hodnotách populačného potenciálu budú tvoriť jeho protipól na opačnom konci republiky. Najmenšie hodnoty populačného potenciálu si zachovajú okrajové, málo ľudnaté okresy na severovýchode a východe Slovenska.

Model populačného potenciálu je praktický a v súvislosti s progresiou výpočtovej techniky v súčasnosti aj ľahko použiteľný nástroj, ktorý sa môže pre geografov stať vhodným pomocníkom v procese poznávania zákonov a zákonitostí organizácie socio-ekonomického geografického priestoru. Veríme, že teoretické a empirické aplikácie modelu populačného potenciálu nachádzajú a naďalej budú nachádzať svoje využitie vo výskume socio-ekonomických a regionálnych štruktúr Slovenska.

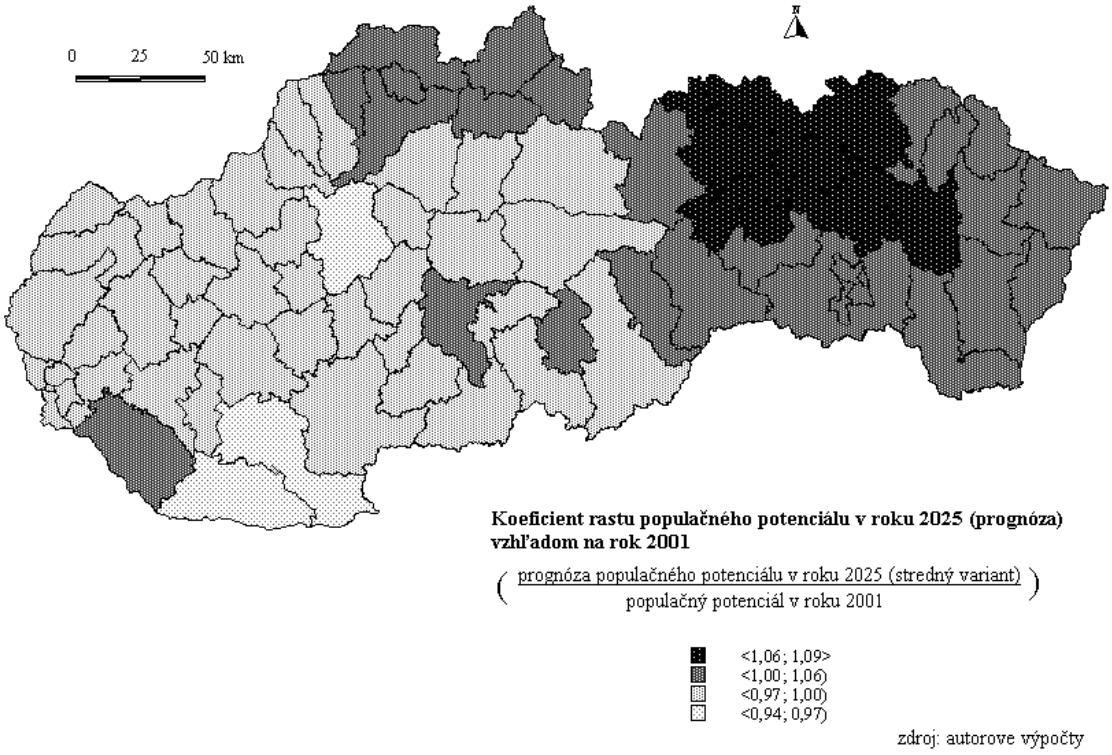
Gabriela Nováková
Univerzita Komenského v Bratislave
Prírodovedecká fakulta
Katedra humánnej geografie a demogeografie
novakova@fns.uniba.sk

Použitá literatúra a štatistické pramene

Hlavička, V. (1993). *Teoretická východiska a souvislosti konstrukce gavitačních modelù v geografii*. Sborník ČSG, 1, pp. 34-43.
Isard, D. W. (1966). *Metody regionaľnovo analiza*. Progress, Moskva, pp. 659.
Kusendová, D. (1996). *Kartografické modelovanie v prostredí geoinformačných systémov: Teoretická báza a geografické aplikácie.* Kandidátska dizertačná práca. Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Bratislava, pp. 117.
Rich, D. C. (1980). *Potential Models in Human Geography. Concepts and Techniques in Modern Geography*, 26, Macquarie University, Australia, pp. 38.

Prognóza vývoja obyvateľov v okresoch SR do roku 2025. (2004). Bratislava, Infostat.
Sčítanie obyvateľov, domov a bytov z roku 2001.

Obr. 1: Koeficient rastu hodnoty populačného potenciálu územných jednotiek Slovenska v roku 2025 (prognóza) vzhľadom na hodnotu populačného potenciálu v roku 2001



Tab. 1: Prognóza populačného potenciálu územných jednotiek Slovenska v roku 2025

názov	1	2	3	4
Bratislava	473041	467911	100,0	0,99
Košice	285019	296191	63,3	1,04
Prešov	150395	163980	35,0	1,09
Nitra	158095	157160	33,6	0,99
Žilina	152046	153924	32,9	1,01
Banská Bystrica	148016	145539	31,1	0,98
Trnava	139973	135998	29,1	0,97
Martin	127394	125073	26,7	0,98
Trenčín	123502	121535	26,0	0,98
Prievidza	123647	118890	25,4	0,96
Poprad	113791	118587	25,3	1,04
Zvolen	110221	110782	23,7	1,01
Považská Bystrica	112385	110484	23,6	0,98
Spišská Nová Ves	100539	108752	23,2	1,08
Topoľčany	103703	101373	21,7	0,98
Piešťany	102654	101276	21,6	0,99
Šaľa	101258	99533	21,3	0,98
Nové Zámky	103790	98915	21,1	0,95
Hlohovec	100632	98822	21,1	0,98
Partizánske	100750	98713	21,1	0,98
Levice	98017	95152	20,3	0,97
Bánovce nad Bebravou	95863	94555	20,2	0,99
Ružomberok	94192	93554	20,0	0,99
Michalovce	90143	92541	19,8	1,03
Galanta	93054	92043	19,7	0,99
Kysucké Nové Mesto	90710	91769	19,6	1,01
Liptovský Mikuláš	90880	90414	19,3	0,99
Žiar nad Hronom	91339	89050	19,0	0,97
Dunajská Streda	87595	88160	18,8	1,01
Púchov	88734	87457	18,7	0,99
Nové Mesto nad Váhom	89625	87452	18,7	0,98
Zlaté Moravce	88082	86962	18,6	0,99
Humenné	85798	86768	18,5	1,01
Bardejov	81182	86707	18,5	1,07

názov	1	2	3	4
Levoča	80865	86043	18,4	1,06
Vranov nad	79870	84833	18,1	1,06
Bytča	83866	84519	18,1	1,01
Komárno	89412	84481	18,1	0,94
Kežmarok	77871	84022	18,0	1,08
Dolný Kubín	81995	84001	18,0	1,02
Čadca	83154	83530	17,9	1,00
Lučenec	82867	82065	17,5	0,99
Brezno	82102	80521	17,2	0,98
Turčianske Teplice	79744	79327	17,0	0,99
Banská Štiavnica	79458	78014	16,7	0,98
Senica	79374	77879	16,6	0,98
Sabinov	72545	77526	16,6	1,07
Detva	77605	76446	16,3	0,99
Trebišov	73677	76336	16,3	1,04
Žarnovica	77580	76167	16,3	0,98
Rimavská Sobota	76137	75989	16,2	1,00
Ilava	77072	75648	16,2	0,98
Rožňava	72601	73548	15,7	1,01
Myjava	75620	73453	15,7	0,97
Gelnica	68404	71084	15,2	1,04
Stará Ľubovňa	66372	70931	15,2	1,07
Tvrdošín	67200	70391	15,0	1,05
Revúca	68889	69953	15,0	1,02
Krupina	69745	68784	14,7	0,99
Veľký Krtíš	67208	66041	14,1	0,98
Námestovo	62464	65517	14,0	1,05
Poltár	64726	64964	13,9	1,00
Skalica	65355	64901	13,9	0,99
Snina	62224	63208	13,5	1,02
Stropkov	58679	60892	13,0	1,04
Svidník	58039	59753	12,8	1,03
Sobrance	48759	49511	10,6	1,02
Medzilaborce	46408	47253	10,1	1,02

zdroj: autor

1 populačný potenciál (obyv./km) v roku 2001
2 prognóza hodnoty populačného potenciálu (obyv./km) v roku 2025 (stredný variant)
3 podiel (%) z maximálneho prognózovaného populačného potenciálu v súbore (maximálna hodnota = prognóza populačného potenciálu Bratislavy)
4 koeficient rastu prognózy populačného potenciálu v roku 2025 vzhľadom na hodnotu populačného potenciálu v roku 2001

Tab. 2: Zmeny poradia územných jednotiek Slovenska podľa hodnôt populačného potenciálu rokoch 2001-2025*

názov	1	2	3
Komárno	30	38	-8
Kežmarok	46	39	7
Sabinov	54	47	7
Bardejov	40	34	6
Levoča	41	35	6
Spišská Nová Ves	20	14	6
Vranov nad Topľou	42	36	6
Brezno	38	43	-5
Čadca	36	41	-5
Lučenec	37	42	-5
Dunajská Streda	33	29	4
Krupina	55	59	-4
Michalovce	28	24	4
Nové Zámky	14	18	-4
Stará Ľubovňa	60	56	4
Ilava	49	52	-3

názov	1	2	3
Myjava	51	54	-3
Trebišov	52	49	3
Žiar nad Hronom	25	28	-3
Bytča	35	37	-2
Gelnica	57	55	2
Námestovo	63	61	2
Nové Mesto nad Váhom	29	31	-2
Partizánske	18	20	-2
Prešov	5	3	2
Revúca	56	58	-2
Skalica	61	63	-2
Tvrdošín	59	57	2
Veľký Krtíš	58	60	-2
Žarnovica	48	50	-2
Banská Štiavnica	44	45	-1
Detva	47	48	-1

názov	1	2	3
Dolný Kubín	39	40	-1
Galanta	24	25	-1
Humenné	34	33	1
Kysucké Nové Mesto	27	26	1
Liptovský Mikuláš	26	27	-1
Nitra	3	4	-1
Považská Bystrica	12	13	-1
Prievidza	9	10	-1
Púchov	31	30	1
Rimavská Sobota	50	51	-1
Senica	45	46	-1
Trenčín	10	9	1
Turčianske Teplice	43	44	-1
Zvolen	13	12	1
Žilina	4	5	-1

zdroj: autor

1 poradie územnej jednotky podľa hodnoty populačného potenciálu v roku 2001(zostupné usporiadanie)
2 poradie územnej jednotky podľa hodnoty populačného potenciálu v roku 2025 (zostupné usporiadanie)
3 diferencie v poradí podľa prognózy hodnôt populačného potenciálu v roku 2025 v porovnaní s rokom 2001

zhoršenie pozície zlepšenie pozície

*v tabuľke nie sú uvedené územné jednotky, ktorých pozícia v súbore sa nezmenila

Konkurenceschopnost vybraných regionů EU a jejich klasifikace

Jakub Odehnal[1], Jaroslav Michálek[2]

Abstract: The goal of the contributions is the classification of the competitiveness of selected regions in European Union. Used variables were obtained from the set of variables of Statistical Office of the European Communities. The statistical techniques are based on factor analysis, cluster analysis and classification trees.

Key words: Cluster Analysis, classification trees, factor analysis, competitiveness of regions .

1. Úvod

Kvalita podnikatelských subjektů v regionech a jejich ekonomická výkonnost tvoří nedílnou součást podmínek ekonomického růstu země a faktorů zvyšování životní úrovně obyvatelstva. Regiony s kvalitním podnikatelským prostředím tak sehrávají klíčovou úlohu při rozhodování subjektů domácích i zahraničních, o lokalizaci nových či rozšiřování stávajících podnikatelských aktivit. Vzniklé meziregionální rozdíly pramenící jednak z přirozeného historického vývoje regionů, z rozdílné vybavenosti výrobními faktory, z geografického umístění vytváří tlaky na zvyšování konkurenceschopnosti jednotlivých regionů vedoucí k ekonomickému růstu a prosperitě. Aktuální důraz je kladen především na proces regionálního rozvoje založeného na znalostech, vědomostech a inovacích vedoucích k přechodu na znalostně založenou ekonomiku umožňující realizaci vzniklé konkurenční výhody do podnikové praxe.

Hodnocení regionální konkurenceschopnosti vybraných regionů EU je problematika velmi aktuální, viz např. [2]. Cílem příspěvku je klasifikace vybraných regionů EU provedená pomocí charakteristik konkurenceschopnosti na základě dat získaných z databáze Eurostatu [4]. V práci je provedena vzájemná komparace vzniklých meziregionálních rozdílů s důrazem na posouzení vhodnosti a aktuálnosti nástrojů hospodářské politiky užitých v problémových regionech jednotlivých států EU.

2. Zdroje dat

K hodnocení konkurenceschopnosti bylo z aktuální veřejné databáze Eurostatu vybráno 78 regionů NUTS2 tvořících společně 5 členských států EU (Česká republika – 8 regionů, Německo – 41 regionů, Rakousko – 9 regionů, Slovensko – 4 regiony, Polsko – 16 regionů). Tyto regiony byly popsány pomocí třinácti regionálních charakteristik konkurenceschopnosti. Pro zajištění vzájemné porovnatelnosti mezi regiony byly vybrané proměnné statisticky upraveny (zejména absolutní počty byly transformovány na relativní) a vhodně standardizovány. Detailní popis vybraných charakteristik je uveden v odstavci 3.

3. Použité metody klasifikace

V první části příspěvku byla použita faktorová analýza k redukci počtu proměnných a pro konstrukci ekonomických faktorů umožňujících posouzení regionální konkurenceschop-

[1] Ing. Jakub Odehnal, Ekonomicko správní fakulta, Masarykova univerzita, Brno e-mail: odehnal@mail.muni.cz

[2] Doc. RNDr. Jaroslav Michálek, CSc., Faculty of Mechanical Engineering, Brno UT e-mail: michalek@fme.vutbr.cz

Příspěvek vznikl za podpory VZMSM0021622418.

nosti. Dále pomocí hierarchických shlukovacích postupů byly jednotlivé regiony rozděleny do skupin s podobnými, respektive blízkými hodnotami vypočtených faktorů a vytvořené skupiny (shluky) regionů byly charakterizovány s ohledem na konkurenceschopnost.

Ve druhé části příspěvku je užita technika „data mining" a pomocí ní byl vytvořen optimální klasifikační strom pro provedení alternativní nezávislé klasifikace jednotlivých regionů. V závěru práce jsou obě metody porovnány a jsou zhodnoceny jejich přínosy pro ekonomickou klasifikaci regionů z hlediska konkurenceschopnosti.

4. Charakteristiky regionální konkurenceschopnosti a jejich syntéza faktorovou analýzou

Sledované charakteristiky regionální konkurenceschopnosti zahrnující oblasti makroekonomického vývoje regionů, ukazatelů trhu práce, inovačního potenciálu regionů, kvalifikační struktury lidských zdrojů a infrastruktury cestovního ruchu jsou uvedeny v tabulce 1. Součástí tabulky je i matice faktorových zátěží získaná faktorovou analýzou (byla použita metoda hlavních komponent s ortogonální rotací varimax). Pomocí faktorové analýzy byly vytvořeny 3 výsledné faktory regionální konkurenceschopnosti:

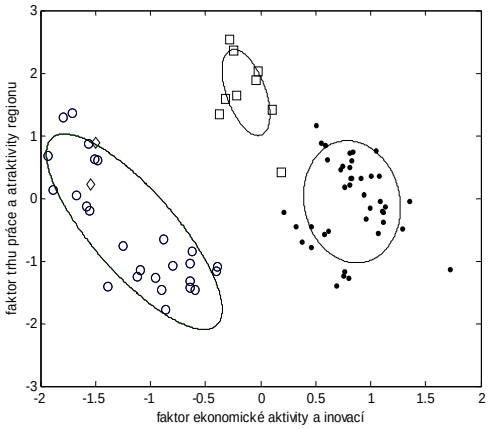
F1. Faktor ekonomické aktivity a inovací,

F2. Faktor trhu práce a atraktivity regionu,

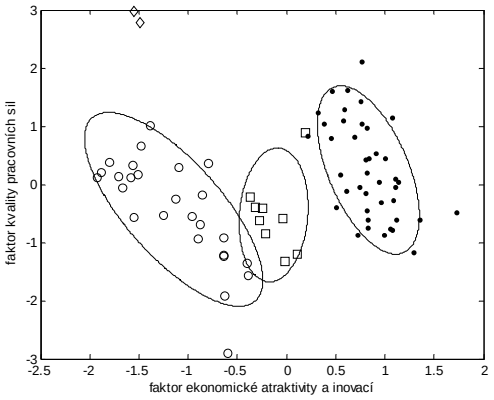
F3. Faktor kvality pracovních sil.

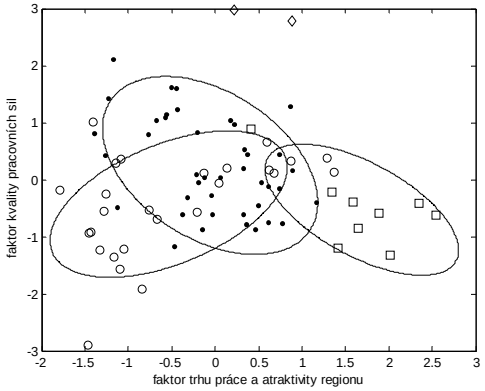
Tabulka 1. Matice faktorových zátěží. Faktorové zátěže rovné aspoň 0,6 jsou vytištěny tučně.

Vybrané charakteristiky konkurenceschopnosti	*F1*	*F2*	*F3*
reálný ekonomický růst	**-0,630**	-0,173	-0,323
příjem domácností	**0,869**	0,402	0,155
HDP na obyvatele	**0,798**	0,477	0,281
kapacita reg. ubyt. zařízení	0,099	**0,807**	-0,006
počet reg. ubyt. zařízení	0,326	**0,864**	0,090
počet odprac. hodin (týden)	**-0,917**	-0,205	0,015
míra zaměstnanosti	0,135	**0,862**	0,341
míra nezaměstnanosti	-0,274	**-0,836**	-0,060
podíl obyvatelstva v reg.	**-0,764**	0,239	0,007
ekonomicky aktivní obyv.	-0,055	0,303	**0,847**
prac. síla s terciál. vzděl.	0,592	-0,098	**0,704**
počet ICT patentů v regionu	**0,775**	0,407	0,218
prac. síla ve vědě a výzk.	0,543	0,222	**0,600**
podíl na celkové variabilitě	*0,359*	*0,283*	*0,150*



Obr. 1. Faktorové skóre (F1 vs. F2)
- *regiony Německa,*
- ○ *regiony ČR, SR a Polska*
- ◊ *Praha, Bratislava*
- □ *regiony Rakouska*



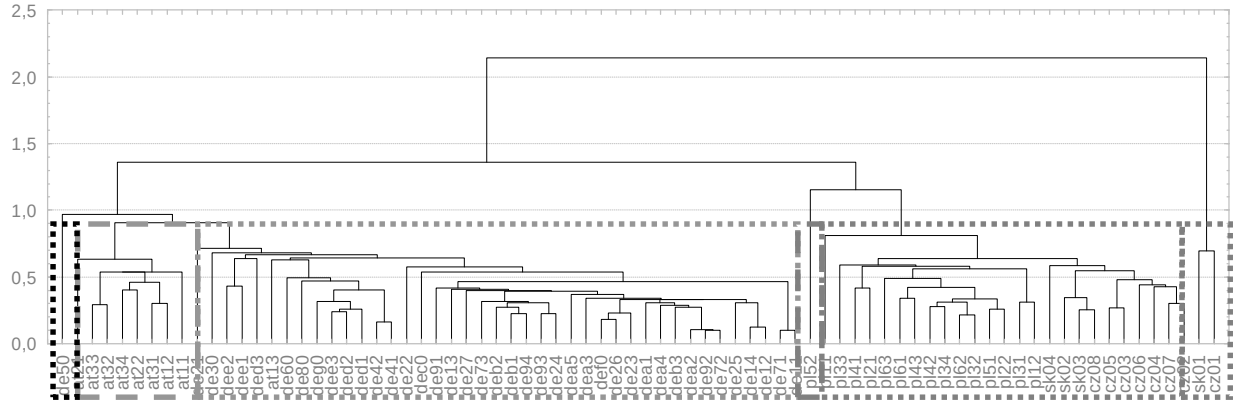


Obr. 2. Faktorové skóre (F1 vs. F3) *Obr. 3. Faktorové skóre (F3 vs. F2)*
• regiony Německa, ○ regiony ČR, SR a Polska, ◊ Praha, Bratislava, □ regiony Rakouska

Dále byly jednotlivé regiony charakterizovány jejich faktorovým skórem v závislosti na všech třech nalezených faktorech. Na obrázcích 1, 2 a 3 je potom znázorněna poloha jednotlivých regionů z hlediska hodnot jejich faktorových skórů. Každý z těchto tří obrázků představuje polohu každého regionu v závislosti na dvou vybraných faktorech. Elipsy znázorňují shluky regionů podle dvou hodnot v obrázku prezentovaných faktorů. Obrázky tedy dokumentují vzniklé meziregionální rozdíly s ohledem na specifikované faktory konkurenceschopnosti. Regionální podobnost zjištěnou mezi regiony ČR, SR a Polska potvrzuje následně i shluková analýza. Dominantní postavení metropolitních regionů Prahy a Bratislavy patrné z obr. 2 a 3 naznačuje výrazné rozdíly mezi těmito regiony a zbylými regiony ČR a SR. Historický původ těchto center, kvalitní pracovní síla, vysoký podíl podnikatelských subjektů, infrastruktura tvoří z těchto regionů ekonomicky nejvýkonnější centra sledovaných zemí.

5. Klasifikace vybraných regionů EU

Faktorová skóre, která byla popsána v předchozím odstavci, byla dále použita ke klasifikaci regionů pomocí shlukové analýzy. Hierarchická konstrukce shluků vycházela z metody jednoduchého spojení objektů a za míru nepodobnosti byla zvolena euklidovská vzdálenost (viz. [1]). Z výsledného dendrogramu jsou dobře patrné podobnosti sledovaných regionů podle jednotlivých charakteristik konkurenceschopnosti. 78 hodnocených regionů se přirozeně rozpadá do 6 regionálních skupin. Na hladině spojení 0.85 pozorujeme klasifikaci regionů téměř odpovídající jejich geografickému uspořádání podle jednotlivých zemí.

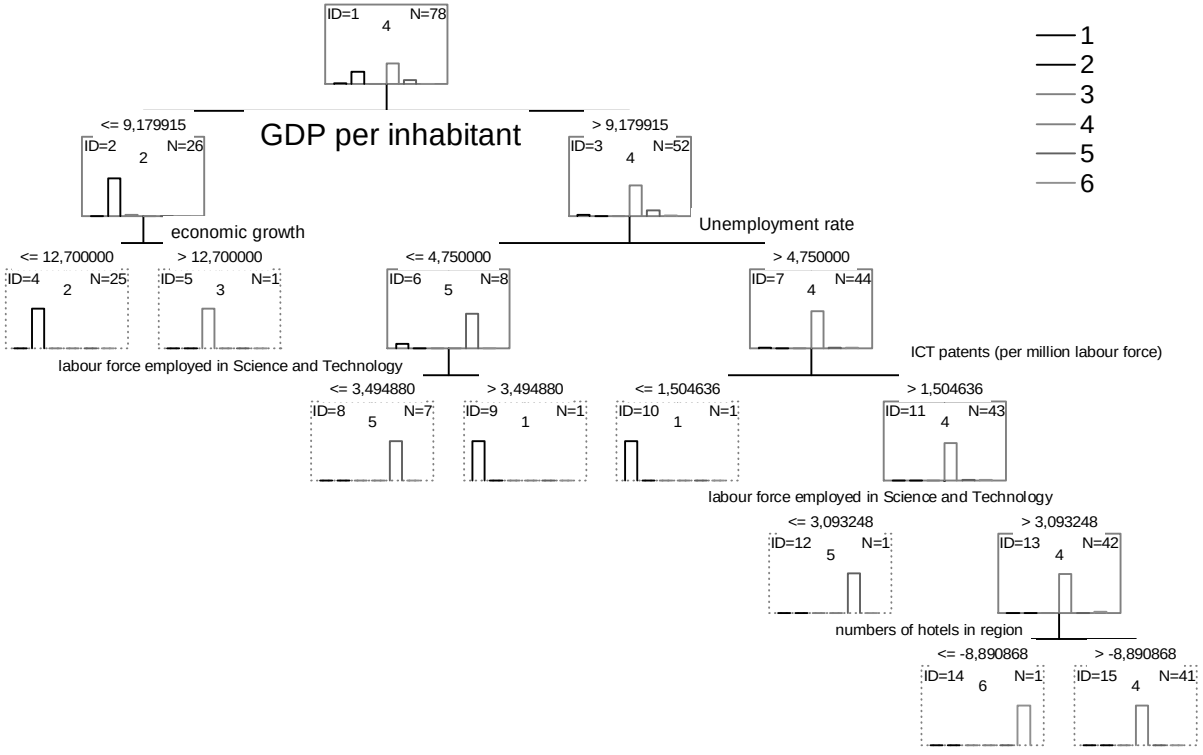


Obr. 4. Výsledný dendrogram

Významné samostatné postavení pozorujeme u regionů Praha a Bratislava, jejichž nadprůměrné hodnoty u většiny sledovaných proměnných dokazují jejich dominantní postavení v rámci středo-východních regionů EU. Ve prospěch ojedinělé pozice polského regionu Opolskie (PL52) svědčí zejména charakteristika ekonomického růstu, jehož tempo bylo v meziročním srovnání nejvyšší ze všech hodnocených regionů. Zvýšením shlukovací hladiny na úroveň 1.3 pozorujeme klasifikaci regionů na 3 samostatné skupiny, skupiny tradičních regionů EU, skupiny nových regionů v EU a ostatní (Praha, Bratislava).

6. Hodnocení regionální konkurenceschopnosti dle významnosti proměnných

Popsané výsledky klasifikace použijeme dále pro alternativní hodnocení konkurenceschopnosti za pomoci klasifikačního stromu. Abychom získali nezávislou klasifikaci na předchozí metodě, budeme klasifikační strom konstruovat přímo pomocí všech sledovaných proměnných. Vzhledem k asymetrii rozdělení některých ekonomických proměnných byly tyto proměnné předem logaritmovány (HDP na obyvatele, míra zaměstnanosti, počet ICT patentů v regionu, počet a kapacita regionálních ubytovacích zařízení, pracovní síla ve VV). 78 regionů bylo při prvním štěpení podle HDP na obyvatele klasifikováno do 2 různorodých skupin. Štěpícímu pravidlu o hodnotě 9,179915 (proměnná byla zlogaritmována) odpovídá hodnota HDP na obyvatele 9700 v PPS. Regiony s hodnotou nižší (26 regionů) jsou klasifikovány do společné skupiny ekonomicky slabších regionů a regiony s hodnotou vyšší (52 regionů) do skupiny ekonomicky silnějších. Slabších 26 regionů je následně klasifikováno dle další ekonomicky významné proměnné, dle meziročního ekonomického růstu. Regiony s hodnotami odpovídajícími růstu nižšímu než 12,7 % za rok (regiony ČR, SR a Polska – bez CZ01, SK01 a PL52) klasifikujeme do výsledné skupiny s označením ID4. Region s ekonomickým růstem vyšším (Opolskie – PL52) do skupiny ID5. Obdobně probíhá další štěpení 52 ekonomicky silnějších regionů na počátku klasifikovaných do skupiny s vyššími hodnotami regionálního produktu na obyvatele.



Obr. 5. Výsledný klasifikační strom

Výsledky klasifikačního stromu můžeme pozorovat z jednotlivých konečných uzlů: (ID4 – regiony ČR, SR, Polska, ID5 – Opolskie, ID8 – regiony Rakouska, ID9 – Praha, ID10- Bratislava, ID12 – Burgenland, ID 14 – Bremen, ID 15 – regiony Německa a Wien).

7. Závěr

Analýza regionální konkurenceschopnosti prostřednictvím mnohorozměrných statistických metod prokázala v obou případech existenci významných meziregionálních rozdílů mezi regiony „nových“ členských států EU a tradičními zeměmi EU. Hlavní odlišnosti mezi sledovanými regiony spatřujeme zejména v ukazateli HDP na obyvatele, kdy zejména regiony postižené strukturálními změnami v národních ekonomikách během transformačního procesu dosahují nižších hodnot regionální HDP. Důraz nástrojů hospodářské politiky EU ke zmírnění těchto nerovností by měl být kladen zejména na zvýšení kvality podnikatelského prostředí v postižených regionech, vedoucích ke snížení nezaměstnanosti a k následnému růstu životní úrovně obyvatel.

8. Literatura

[1] Lukasová, A., Šarmanová, J.Metody shlukové analýzy. Vyd. 1. Praha: SNTL, 1985. 210 s.

[2] Ronald L. M., A Study on the Factors of Regional Competitiveness: A draft final report for the European Commission Directorate-General Regional Policy. Cambridge Econometrics. ECORYS-NEI. 2005.

[3] Johnson R.A., Wichern D.W. Applied Multivariate Statistical Analysis. 3rd edition. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1992.

[4] http://epp.eurostat.ec.europa.eu/

Adresy autorů

Ing. Jakub Odehnal
Katedra aplikované matematiky a informatiky
Ekonomicko správní fakulta Masarykovy univerzity
Lipová 41a, 600 00 Brno
odehnal@mail.muni.cz

Doc. RNDr. Jaroslav Michálek, CSc.
Ústav matematiky
Fakulta strojního inženýrství
Vysoké učení technické v Brně
Technická 2896/2, 616 69 Brno
michalek@fme.vutbr.cz

Počítačové zpracování a statistická analýza rozsáhlých souborů dat

Miroslav Otáhal

Abstract

A description of the computer processing of extensive data files is presented. Some problems in data collection, in the database processing and in statistical analysis of such data are discussed. An example of algorithm that can be used to make a systematic analysis of insurance data is presented.

Příklad pořizování rozsáhlých souborů dat v pojišťovnictví

Předkládaný článek demonstruje použití statistických a databázových softwarů při analýze rozsáhlých souborů dat. Jako příklad je použit soubor dat o pojistných událostech (PU). Podnět pro analýzu byla snaha o nalezení vhodných algoritmů pro stanovení optimální výše rezerv na pojistné nároky.

Procesy v moderním pojišťovnictví lze popsat následovně:

- Klientovi vznikne škoda, která je předmětem pojištění (vznik PU).
- Klient tuto událost nahlásí na pojišťovně (hlášení PU).
- Pojišťovna PU přezkoumá a rozhodne o výši pojistného plnění a toto plnění klientovi vyplatí (výplata PU).
- Jednotlivé pobočky pojišťoven mají (v dnešní době online) přístup přes terminály nebo PC do provozního systému (SAP, Golem) mateřské pojišťovny. Tento systém běží na serveru v sídle pojišťovny a je reálné předpokládat, že během pracovních dní jsou připojeny u velkých pojišťoven i tisíce terminálů současně. Data jsou v provozním systému ukládána do databáze, která je vytvořená v některém databázovém prostředí, např. Oracle, Informix nebo FoxPro.
- Na serveru se prakticky neustále aktualizují provozní databáze novými a novými údaji. Ve většině pojišťoven se pro řadu analýz nepoužívají data přímo z ostré provozní databáze, ale využívá se tzv. datového skladu (DS). Jde o speciální databázi, kde nedochází k neustálé aktualizaci (její naplnění probíhá mimo špičku výkonu hlavního serveru tj. např. v noci nebo o víkendech) a která též neobsahuje všechny údaje potřebné pro běžný provoz pojišťovny. Datasklad má tu výhodu, že zpravidla běží na jiném počítači a neovlivňuje a ani není ovlivněn ostrým provozem pojišťovny. Je ideálním místem pro spouštění různých sestav a je zdrojem dat pro různé analýzy. DS může sloužit jako místo, kde se připravují vstupní data do specializovaných analytických softwarů jako např. Profet, SAS, SPSS nebo MATLAB.

Nástin matematického modelu popisující pojišťovnictví a účel použití statistické analýzy v něm

Důležitým úkolem pojistných matematiků je analýza procesu generujícího PU (PGPU). Tento proces je složen ze čtyř součástí:

- Doba vzniku PU (DVPU). Často se vychází z předpokladu, že PU jsou generované Poissonovým procesem (s proměnnou intenzitou).
- Doba do nahlášení PU (DNPU). Má smysl analyzovat, zda je doba do nahlášení PU v čase konstantní, nebo zda podléhá změnám.

- Doba do (vy)placení PU (DPPU). Pro praktickou aplikaci v pojišťovnictví je vhodnější pracovat místo s údaji o nahlášení škod s údaji o výplatě škod. Pro ukázku matematického modelu a počítačového zpracování ale rozdíl mezi nahlášením škody a výplatou škody můžeme zanedbat a v celém článku mluvit pouze o nahlášení škody.
- Výše škody (VS). Identifikace nejvýznamnějších faktorů ovlivňujících výši škody může být klíčovým nástrojem při strategickém rozhodování pojišťovny.

V počátečních fázích analýzy pojistného kmene je důležité provádět příslušné testy, zda jednotlivé fáze PGPU splňují zvolené apriorní předpoklady.

Popis rezerv v pojišťovnictví je možné nalézt v [Ca]. Předkládaný text je zaměřen na stanovování IBNR rezervy. Tento typ rezerv je používán v neživotním pojištění ke krytí vzniklých, ale doposud nenahlášených PU. Bližší popis této rezervy je možné nalézt v [Ol]. Jedná se vlastně o odhad součtu budoucích nároků, které bude muset pojišťovna uhradit na základě současných smluv. Uvažují se škody již vzniklé, o kterých ale pojišťovna ještě neví, neboť nebyly nahlášené. IBNR je stanovována pojistně-matematickými metodami na základě minulých dat.

Z provozního systému máme zpravidla k dispozici matici čísel, kde jednotlivé sloupce tvoří identifikace PU, DVPU, DNPU, DPPU a VS. Snadným způsobem lze vypočítat dobu do nahlášení každé PU.

V pojišťovnictví se často pracuje s diskrétními časovými obdobími jako je rok, kvartál nebo měsíc. V tomto textu bude bez újmy na obecnosti jako časové období zvolen kvartál.

V oblasti pojistných nároků pracujeme s tzv. pojistnými (PK) a vývojovými (VK) kvartály. O PK mluvíme v souvislosti DVPU a počítáme je od počátku působení pojišťovny na trhu. O VK mluvíme v souvislosti s DNPU a zjednodušeně řečeno je určujeme od okamžiku vzniku PU.

Pro stanovování IBNR se v praxi standardně používá tzv. run-off analýza založená na metodě chain-ladder (viz. [Ma]) a run-off trojúhelníku, který vznikne sumací škod vzniklých v jednom PK a hlášených ve stejném VK (viz. tabulka - Run-off trojúhelník).

Tabulka 1 - Run-off trojúhelník

	1	2	3	…	I
1	S_{11}	S_{12}	S_{13}		S_{1I}
2	S_{21}	S_{22}	S_{23}		$\mathbf{S_{2I}}$
3	S_{31}	S_{32}	S_{33}		$\mathbf{S_{3I}}$
⋮	⋮		…		⋮
I	S_{I1}	$\mathbf{S_{I2}}$	$\mathbf{S_{I3}}$		$\mathbf{S_{II}}$

Náhodná veličina S_{ij} vyjadřuje součet výší škod vzniklých v i-tém PK a hlášených v j-tém VK.

V daném čase má pojišťovna k dispozici horní trojúhelníkovou matici (v celém textu uvažujeme vzhledem k vedlejší diagonále) zaznamenaných hodnot. Je nutné odhadnout doposud neznámé prvky v dolní trojúhelníkové matici a z dopočtených prvků následně stanovit IBNR.

Tento článek se zabývá alternativním postupem, kdy se neznámé prvky run-off trojúhelníka doplní na základě zobecněného lineárního modelu (GLM - General Linear Model). Se zobecněným lineárním modelem je možné se seznámit např. v [Ne], nebo v [Mi].

Vektor náhodných veličin $S = (S_{11}, S_{21}, \ldots, S_{I1}, S_{12}, S_{22}, \ldots, S_{I2}, \ldots, S_{I1}, S_{I2}, \ldots, S_{II})$ se snažíme vysvětlit použitím GLM v závislosti na řádkovém a sloupcovém indexu v příslušném

run-off trojúhelníku. Splnění předpokladů GLM při vhodné volbě předpokladů kladených na PGPU je ukázáno v [Ol].

Základní lineární model pro naše účely má tvar

$$S_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad (1)$$

Kde

- μ vyjadřuje úrovňovou konstantu obsaženou ve všech pozorováních S_{ij}.
- α_i vyjadřuje úroveň trendu v pojistných kvartálech společnou pro všechny S_{ij} z i-tého řádku run-off trojúhelníku
- β_j vyjadřuje úroveň trendu ve vývojových kvartálech společnou pro všechny S_{ij} z j-tého sloupce run-off trojúhelníku
- e_{ij} vyjadřuje náhodnou chybu [i,j]-tého pozorování S_{ij}.

S modelem tvaru (1) je možné se detailně seznámit v [An]. Zde je též diskutovaná nutnost zavedení dodatečné reparametrizační podmínky, která odstraní přeparametrizování modelu. Zavedením dodatečných předpokladů na trendy v pojistných a vývojových kvartálech dojde k vytvoření submodelu s menším počtem odhadovaných parametrů, tím dojde ke zmírnění přeparametrizace.

Výchozím modelem při stanovování IBNR je model tvaru

$$S_{ij} = \mu + \alpha_1 + (i - 1)a + \beta_1 + (j - 1)\, b + e_{ij} \qquad (1)$$

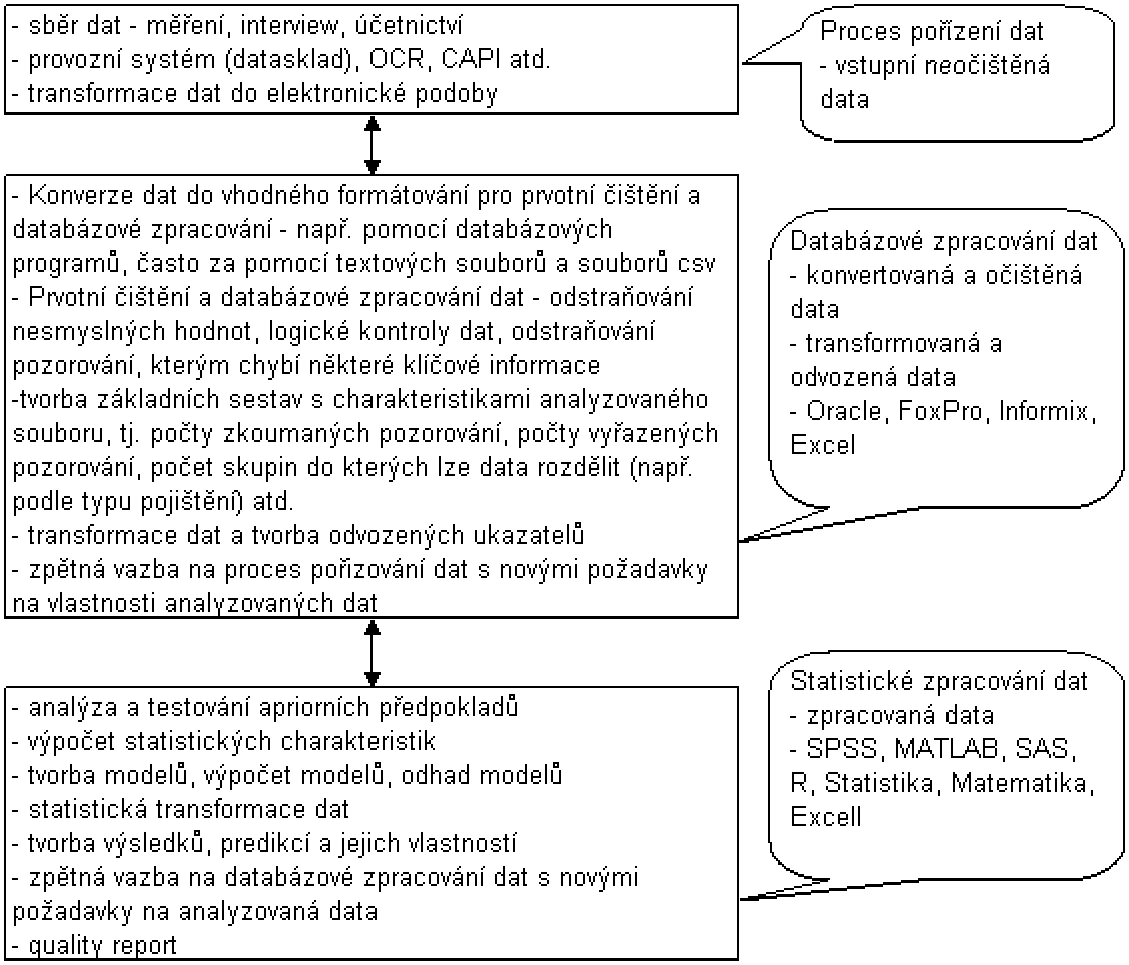
který předpokládá stejné lineární trendy v pojistných i vývojových kvartálech po celou dobu. Daný model odhadneme metodou maximální věrohodnosti a spočteme rezidua. Tato rezidua vykreslíme postupně proti indexu vývojových kvartálů, pojistných kvartálů a indexu kvartálů placení. Pokud rezidua v některém směru vykazují nějaký zbytkový trend, je vhodné zavést dodatečný parametr, pomocí něhož budeme schopni modelovat příslušnou změnu trendu. Takto získáme nový model, který opět odhadneme a stejným způsobem analyzujeme jeho rezidua. Tímto způsobem postupujeme tak dlouho, až se z reziduí odstraní všechny zbytkové trendy. Získáváme tak posloupnost modelů, kde původní model je submodelem modelu nového.

Na závěr přistoupíme k testování hypotéz o vhodnosti výsledného modelu s největším počtem parametrů a o možnosti redukce tohoto modelu na některý z dílčích submodelů. Popis testování je popsán v [Do].

Schéma postup při analýze rozsáhlých souborů dat

Konkrétní příklad výše popsaného postupu je znázorněn na následujícím schématu.

Obrázek 1 - obecné schéma zpracování dat



Konkrétní postup statistické analýzy pojistných dat v systému MATLAB.

Jako příklad komplexního statistického zpracování dat je možné uvést následující postup statistického zpracování dat o PU v systému MATLAB.

1. Prostřednictvím příkazu *querybuilder* vyvoláme okno, v němž je možné spouštět SQL příkazy. Snadným *selectem* je možné naplnit hodnotami uloženými v databázových tabulkách lokální proměnné. Pro další zpracování je zpravidla nutné tyto lokální proměnné konvertovat do datových typů vhodných pro další statistické zpracování.
2. Je možné provést další očištění zpracovávaných dat. Hrubé čištění dat je vhodnější provádět již na úrovni databázových zpracování, ale některé podezřelé nebo chybné hodnoty se objeví až ve fázi statistického zpracování (např. některá odlehlá pozorování). Je možné spouštět stejné zpracování na soubor dat, kde provádíme dodatečné korekce vstupních hodnot. Dobře se tímto způsobem může odhalit vliv chybných nebo odlehlých pozorování na konečné výsledky. Stejné zpracování lze provádět pro jednotlivé podskupiny původních dat. V pojišťovnictví je téměř nemožné dospět k smysluplným výsledkům v případě, že analyzujeme data v celku bez ohledu na typ pojištění. Je nutné data rozdělit na skupiny podle typu pojištění a analýzu provádět separátně pro každý typ pojištění. Někdy ani toto dělení nestačí a je nutné přistoupit k jemnějšímu dělení dat. V této fázi analýzy se projeví zkušenosti, intuice a cit pro analýzu dat.
3. Vlastní statistická analýza dat je spouštěna pomocí parametrizovaných funkcí. Parametricky určujeme např. části analýzy, kterou máme provádět nebo v datech apriori předpokládané rozdělení pravděpodobností. Parametricky rovněž ovlivňujeme chování GLM a χ^2 testů dobré shody. Pro usnadnění práce je vhodné zavést globální vektor parametrů, který v celém programu řídí obecné chování programu - například vypisování mezivýsledků, jazyk výpisů, hladinu významnosti apod. Zavedení parametrizace se vyplácí zejména při opakovaném spouštění analýz.
    a. Dochází ke zjednodušení pravidelné analýzy dat po zahrnutí nových škod do datových souborů. Zpravidla postupujeme tak, že při původní volbě parametrů spustíme program pro větší soubor dat.
    b. Dochází ke značnému usnadnění práce v situaci, kdy zkoumáme jeden datový vektor, ale opakovaně se zaměřujeme na různé části prováděné analýzy. Jednoduchou změnou parametrizace můžeme změnit chování programu tak, aby se prováděla pouze zamýšlená část analýzy (např. pouze analýza výše škody).
    c. Parametrizace též značně usnadňuje situace, kdy při prvním spuštění programu je nutné rekurzivním způsobem odvodit parametry výpočtu (např. odvození tříd, do kterých rozdělujeme zaznamenané realizace náhodných veličin při χ^2 testech dobré shody) a při opakovaném spouštění analýzy dat již nechceme být tímto rozdělováním zdržováni.
4. Další parametrizaci je možné zavést do funkce, která spouští výpočty v GLM. Tyto výpočty je vhodné spouštět opakovaně při různých volbách linkovací funkce. Velmi se osvědčuje rekurzivní volání GLM v situaci, kdy se stávajícím modelem nejsme spokojeni a chceme dodatečným zavedením parametru přejít ke složitějšímu modelu. Tento postup je možné provádět díky rekurzivnímu volání dokud nejsme s výsledným modelem spokojeni. Elegantním způsobem tak můžeme do celkového výpisu výsledků předat výsledky získané ze zvoleného nejvhodnějšímu modelu. Je praktické zaznamenávat postupně výsledky všech

odhadnutých modelů a dodatečně provádět analýzu těchto výsledků. Touto komparativní analýzou lépe vyniknou rozdíly mezi jednotlivými modely.

Celkově lze říci, že komplexní analýza rozsáhlých často nehomogenních souborů dat je složitý a mnohastupňový proces. Je potřeba vnímat celou analýzu jako jeden celek a uvědomovat si provázanost a vzájemnou závislost jednotlivých stupňů analýz. Praktická interpretace odvozených výsledků je seriózní až po začlenění všech stupňů zpracování.

Literatura

[Ol] Otáhal, M.: Stanovování IBNR rezervy s využitím zobecněného lineárního modelu.

[Ca] Cipra, T.: Pojistná matematika - Teorie a praxe, EKOPRESS, s. r. o., Praha 1999

[Ne] McGullagh, P. and Nelder, J. A., Generalized Linear Models, Chapman and Hall, London, 1997

[Mi] Michálek J., Lineární a zobecněný lineární model, In Proceedings ANALÝZA DAT 2003/II, Pardubice, Trilobyte, Ltd., 2004

[Ma] Mack T., Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates, ASTIN Bulletin, Vol 23, 1993

[Ze] Barnett G., Zehnwirth B. Best Estimates for Reserves, Proceedings of the CAS Volume LXXXVII, Numbers 166 and 167, 2000

[An] Anděl J., Statistické metody, Matfyzpress, Praha 2007

[Do] Dobson J., An Introduction to Generalized Linear Models, Chapman end hall 2002

Kontakt

Mgr. Miroslav Otáhal
Na padesátém 81
100 82, Praha 10
Strašnice
Tel.: +420274054376
Email: miroslav.otahal@czso.cz, otahalm@email.cz

Nabídka a poptávka po IT odbornících v ČR 2007

Iva Pecáková[1]

Tento příspěvek vznikl za podpory MŠMT ČR v rámci rozvojového projektu pro vytváření společných struktur mezi vysokými školami a praxí.

Abstract: Between the years 2006 and 2007, a study regarding labour market with IT specialists in the Czech Republic was carried out at the University of Economics in Prague. To confront knowledge requirements of practice with knowledge obtained by the graduates in the Czech Republic, data on the majority of IT study programs offered by Czech universities and data from the sample survey of active economic entities were used. The data was also used for the estimation of the number of IT specialists in the Czech Republic.

Key words: labour market with IT specialists in the Czech Republic

1. Metodologie

Za podpory České společnosti pro systémovou integraci (ČSSI), Sdružení pro informační společnost (SPIO), České asociace manažerů úseků informačních technologií (CACIO), Českého statistického úřadu a MŠMT ČR realizovala VŠE v Praze v letech 2006 a 2007 projekt, jehož cílem bylo získat představu o aktuálním pracovním trhu s IT odborníky v ČR a přispět k hledání odpovědí na jeho významné otázky, a to především:

- Jaký je celkový počet IT odborníků v ČR a jak se bude vyvíjet v nejbližších 3-5 letech?
- Jaká je současná struktura IT odborníků dle profesí a jak se bude vyvíjet v nejbližších 3-5 letech?
- Jaká je požadovaná struktura znalostí základních IT profesí?
- Jaké jsou počty studentů a absolventů IT oborů VŠ a VOŠ?
- Jaká je struktura znalostí těchto absolventů?
- Je počet absolventů IT oborů a struktura jejich znalostí vyhovující z hlediska potřeb globální ekonomiky?
- Jsou VŠ a VOŠ připraveny reagovat na nadcházející změny na trhu práce?

I) V rámci výzkumu bylo provedeno výběrové šetření ekonomických subjektů s cílem získat ze strany poptávky po IT pracovnících originální primární data.

Základní soubor

Základním souborem pro účely šetření byl soubor všech ekonomicky aktivních subjektů v České republice k 31. 12. 2005. Na základě údajů v Registru ekonomických subjektů Českého statistického úřadu bylo k tomuto datu v ČR celkem 2 388 490 registrovaných subjektů, z toho 1 266 336 subjektů ekonomicky aktivních. ČSÚ zveřejňuje soubory ekonomických subjektů tříděné podle různých kritérií. Z hlediska zamýšleného šetření jsme považovali za rozhodující jejich strukturu podle velikosti a podle odvětví převažující činnosti.

Velikost ekonomického subjektu je dána počtem zaměstnanců. V registru je rozlišeno celkem 20 velikostních kategorií, což bylo pro účely výzkumu příliš mnoho. Použili jsme jednodušší strukturu používanou ve statistických šetřeních, jež je

[1] Iva Pecáková, Vysoká škola ekonomická Praha

tvořena šesti velikostními kategoriemi: 0/[2], 1 – 9, 10 – 49, 50 – 249, 250 – 999, 1000 a více.

Odvětví činnosti jsou podle používané odvětvové klasifikace ekonomických činností (OKEČ) tříděna do 17 sekcí. Pro výzkum podstatná byla náročnost jednotlivých odvětví na využívání informačních technologií. Na základě podílu objemu investic do IT na obratu firmy[3] byla odvětví rozdělena do tří kategorií: odvětví s nejnižší náročností (MIT), odvětví středně náročná (SIT) a odvětví s nejvyšší náročností na využívání informačních technologií (VIT):

MIT – zemědělství, myslivost, lesnictví, rybolov a chov ryb, těžba nerostných surovin, stavebnictví, doprava a skladování, vzdělávání, zdravotní a sociální péče, veterinární činnosti, činnosti domácností;

SIT – zpracovatelský průmysl, výroba a rozvod elektřiny, plynu a vody, obchod, opravy motor. vozidel a výrobků, ubytování a stravování, činnosti v oblasti nemovitostí a pronájmu, ostatní veřejné, sociální a osobní služby, exteritoriální organizace a instituce;

VIT – spoje, finanční zprostředkování, činnosti v oblasti výpočetní techniky, veřejná správa a obrana, povinné soc. zabezpečení.

Výběrový soubor; rozsah výběru

Jelikož se dalo se přirozeně očekávat, že počet informatiků ve skupinách firem získaných uvedeným tříděním bude kolísat méně, než v celém souboru ekonomických subjektů, byl základní soubor aktivních ekonomických subjektů na základě šesti velikostních kategorií a tří kategorií podle náročnosti odvětví jejich působnosti na využití informačních technologií rozdělen do 18 strat. Z jednotlivých strat byl proveden pravděpodobnostní výběr bez vracení.

Při rozhodování o velikosti výběrového souboru jsme vycházeli z úvahy, že jeho nejjednodušší, tj. proporcionální rozvržení do oblastí, kdy jsou výběrové podíly ve všech oblastech stejné, nebude v tomto případě výhodné. Oblasti v základním souboru se totiž velikostí dosti liší. Výsledkem by tak byla ve vzorku značná převaha malých firem z málo či středně náročných odvětví na úkor větších firem v odvětvích na informační technologie náročnějších. S malým počtem zaměstnanců bude objektivně souviset menší variabilita v počtu IT pracovníků, která bude nižší také u firem v méně náročných odvětvích. U velkých firem a také v odvětvích náročných na informační technologie lze naopak očekávat podstatně vyšší variabilitu počtu IT pracovníků.

Použití proporcionálního oblastního výběru by tedy negativně ovlivnilo prováděné odhady a jejich přesnost, která by tak v jednotlivých skupinách byla velmi rozdílná.. S ohledem na nevyrovnanou variabilitu by vhodnějším postupem bylo optimální rozvržení, které preferuje variabilnější oblasti před oblastmi méně variabilními. Jeho využití však brání chybějící základní předpoklad – alespoň přibližná představa o variabilitě v oblastech.

Velmi problematickou je v této souvislosti skupina subjektů bez zaměstnanců. Při odhadu počtu IT pracovníků jde totiž vlastně o odhad relativní četnosti jedné hodnoty alternativní proměnné, která, jak známo, pro dosažení žádoucí přesnosti nárokuje značné rozsahy vzorků.

Všechny uvedené skutečnosti jsme se rozhodli zohlednit provedením co nejrozsáhlejšího možného zjišťování ve skupině největších subjektů (nad 1000 pracovníků) a v odvětvích náročných na informační technologie u subjektů nad 50

[2] do této kategorie jsou započítány také subjekty, které počet pracovníků neuvedly
[3] www.gartner.com

pracovníků. V ostatních oblastech stanovením rozsahů výběru stejně velkých vzhledem k velikosti firmy a zvyšujících se vzhledem k náročnosti odvětví, které lze v uvedené souvislosti považovat za významnější. S ohledem na kalkulované náklady šetření pak byl rozsah výběrového souboru stanoven na cca 1000 jednotek. Výběrové šetření bylo realizováno ve druhém pololetí roku 2006 metodou CATI v kombinaci s on-line dotazováním[4]. Rozsah a struktura realizovaného výběru je patrná z tabulky 2.

Tabulka 1. Složení výběrového souboru ekonomických subjektů

	0	**1 – 9**	**10 - 49**	**50 - 249**	**250 - 999**	**více**	**Celkem**
MIT	56	28	28	28	37	16	**193**
SIT	56	56	56	56	71	36	**331**
VIT	56	110	160	122	26	4	**478**
Celkem	**168**	**194**	**244**	**206**	**134**	**56**	**1 002**

Předmět a metoda zjišťování

Jedním ze základních cílů projektu bylo získání představy o počtu IT pracovníků v národním hospodářství ČR a o jejich struktuře. V šetření byl proto zjišťován počet IT pracovníků firmy v členění podle šesti profesních rolí: byznys analytik, manažer IT, obchodník s IT, IS architekt – vývojář, správce sítě, pokročilý uživatel IT – metodik. Kromě aktuálního počtu (rok 2006) byly zjišťovány také údaje pro nejbližší minulost (rok 2005), resp. budoucnost (rok 2007) a dále představa firem o budoucím vývoji v krátkém horizontu (odhad pro rok 2010).

Dalším cílem projektu bylo získání představy o tom, jaké jsou požadavky praxe na speciální znalosti potřebné pro vykonávání jednotlivých profesních rolí IT pracovníků (obligatorními znalostmi, jako je například flexibilita, schopnost komunikace, znalost jazyků atd. se výzkum nezabýval). Definováno bylo celkem 17 různých znalostních domén. Pro vyjádření požadovaného stupně znalosti byla použita pořadová stupnice s hodnotami 0 – „žádné znalosti a dovednosti“ až 5 – „hluboké aktuální znalosti a pokročilé praktické dovednosti“.

S ohledem na skutečnost, že projekt hledal rovněž odpověď na otázku, nakolik vyhovují požadavkům praxe znalosti absolventů informatických oborů na vysokých školách, bylo ovšem nutné definovat hodnoty na stupnici pregnantněji. Úroveň předmětů na vysokých školách je kvantifikována prostřednictvím tzv. kreditů, které zohledňují mj. především čas nutný k jejich zvládnutí. Pro praxi je význam kreditů méně srozumitelný, proto byl nastaven jejich ekvivalent tak, že jeden kredit odpovídá zhruba jednomu dni intenzivního školení. Pro jednotlivé stupně znalosti potom:

1 = 1-2 dny školení/kredity výuky,
2 = 3-5 dnů školení/kreditů výuky,
3 = 6-20 dnů školení/kreditů výuky,
4 = 21-40 dnů školení/výuky,
5 = nad 40 dnů školení/kreditů výuky.

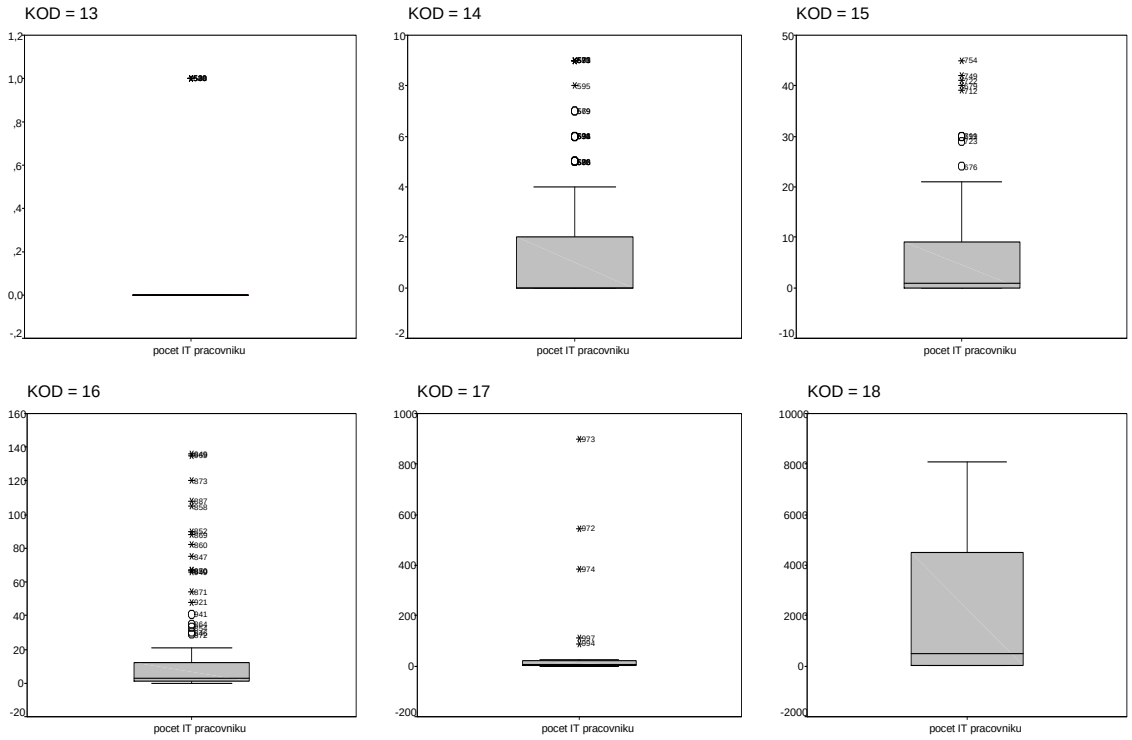
Dotazování ve firmách i ve školách probíhalo na základě standardizovaného dotazníku.

2. Analýza a odhad

Počet IT pracovníků v roce 2006

[4] firmou GfK, jež vzešla z výběrového řízení

Souhrnně lze říci, že (ostatně podle očekávání) jsou rozdělení počtu IT pracovníků v jednotlivých stratech[5] velmi nesymetrická, sešikmená kladně (vpravo). Významnou příčinou sešikmení výběrových rozdělení je (v různých stratech různě vysoký) výskyt firem bez IT pracovníků na straně jedné a výskyt extrémů na straně druhé. V důsledku toho je značný rozdíl mezi variačním rozpětím a kvartilovým rozpětím a samozřejmě také mezi průměrem a mediánem. Vše je patrno z boxplotů (obrázek 1).



Obrázek 1. Boxploty rozdělení počtu IT pracovníků ve stratech 13 – 18 (VIT)

Zjištěná rozdílnost rozdělení je argumentem podporujícím členění základního souboru a vytvoření vzorku výběrem ze strat. Zdá se, že se celkem dobře podařilo zatřídit jednotlivá odvětví OKEČ podle náročnosti na IT. Zvážen by však mohl být počet velikostních skupin subjektů.

Vzhledem k charakteru výběrových rozdělení počtu IT pracovníků je zřejmé, že použití standardních postupů odhadů by vedlo ke značnému nadhodnocení skutečnosti. Odhad byl proto proveden na základě výběrového podílu subjektů uvádějících alespoň jednoho IT pracovníka, mediánu vyjadřujícího úroveň počtu IT pracovníků v takových subjektech v příslušném stratu po eliminování extrémů a počtu subjektů ve stratu v základním souboru. Pro velikostní skupinu subjektů bez zaměstnanců byl s ohledem na příliš nízký rozsah výběru ve stratu použit publikovaný odhad (například [3]) cca 3 %. Dosažený odhad počtu IT pracovníků v ČR obsahuje tabulka 2.

[5] Kódování 18 strat v textu a grafech:

	0	1 – 9	10 - 49	50 - 249	250 - 999	více
MIT	1	2	3	4	5	6
SIT	7	8	9	10	11	12
VIT	13	14	15	16	17	18

Tabulka 2. Odhad počtu IT pracovníků v ČR v r. 2006 (zaokrouhleno)

	0	1 – 9	10 - 49	50 - 249	250 - 999	více	Celkem
MIT	7 900	30 000	6 900	3 100	300	600	48 800
SIT	20 900	75 800	48 100	7 600	900	500	153 800
VIT	1 500	4 000	3 600	600	500	20 500	30 700
Celkem	30 300	109 800	58 600	11 300	1 700	21 600	**233 300**

Struktura počtu IT pracovníků v roce 2006

Pro zkoumání struktury IT pracovníků jsme v zájmu dosažení většího obsazení jednotlivých strat ve výběru použili pouze tři velikostní skupiny firem (do 50 zaměstnanců, 50 – 249, 250 a více zaměstnanců)[6]. V průběhu šetření byly zjišťovány počty IT pracovníků podle šesti profesních rolí: byznys analytik (BA), manažer IT (MA), obchodník s IT (OB), IS architekt – vývojář (VÝ), správce sítě (SP), pokročilý uživatel IT – metodik (PU). Profesní struktura IT pracovníků zjištěná v jednotlivých stratech ve výběru pak byla přepočtena na odhadované počty IT pracovníků v základním souboru.

Tabulka 3. Struktura IT pracovníků v roce 2006

		BA	MA	OB	VÝ	SP	PU
MIT	**0 - 49**	0,07	0,31	0,08	0,11	0,23	0,20
	50 - 249	0,17	0,21	0,10	0,15	0,24	0,13
	250 +	0,02	0,11	0,09	0,12	0,23	0,43
SIT	**0 - 49**	0,18	0,13	0,20	0,11	0,15	0,23
	50 - 249	0,11	0,23	0,32	0,12	0,15	0,07
	250 +	0,37	0,08	0,08	0,15	0,16	0,16
VIT	**0 – 49**	0,11	0,11	0,15	0,32	0,15	0,16
	50 - 249	0,10	0,03	0,09	0,33	0,11	0,34
	250 +	0,02	0,01	0,01	0,11	0,02	0,83
Celkem		0,14	0,16	0,16	0,12	0,16	0,26

Vývoj počtu IT pracovníků

V rámci šetření byl v subjektech zjišťován počet IT pracovníků v letech 2005, 2006, představa pro rok 2007 a odhad pro rok 2010. Odpovědi v jednotlivých letech jsme konfrontovali s cílem získat představu o tom, zda se počty IT pracovníků nějak zásadně změnily v uplynulých letech, a především představu o jejich budoucích záměrech.

Tabulka 4. Očekávaný vývoj počtu IT pracovníků ve firmách (%)

Vývoj	2006/5	2007/6	2010/7
pokles	4,1	16,3	15,6
stejně	85,5	62,3	69,2
růst	10,4	21,5	15,3

[6] Kódování devíti strat v textu a grafech:

	0 - 49	50 – 249	250 a více
MIT	I	II	III
SIT	IV	V	VI
VIT	VII	VIII	IX

Rozhodující část firem se počet IT pracovníků nijak zásadně nemění. Pokud dochází ke změně, jde v první části období spíš o nárůst. V případě prognózy do roku 2010 se však podíl firem uvažujících o zvýšení s podílem firem uvažujících o snížení počtu IT pracovníků vyrovnává (tabulka 4). Lze přitom říci, že v odvětvích náročných na IT je nadále předpokládán nárůst počtu IT pracovníků spíše v menších firmách.

Požadavky firem na znalosti IT pracovníků

Požadavky na znalosti IT pracovníků ze strany firem byly zjišťovány pro každou profesní roli v 17 znalostních doménách na šestibodové pořadové stupnici od 0 do 5 (viz výše). Vzhledem k charakteru používané proměnné jsme zvolili pro vyjádření úrovně požadavků za firmy medián.

Souhrnně vyjádřeno, největší požadavky klade podle našich zjištění praxe na znalosti IT manažerů a byznys analytiků, nejnižší na znalosti pokročilých uživatelů – metodiků, rozdíly jsou také podle očekávání v požadované struktuře znalostí.

Požadavky na znalosti IT pracovníků se liší podle velikosti firem. Velké firmy vyslovily největší požadavky v případě byznys analytiků a správců sítě (obrázek 16 - graf pro BA), u ostatních rolí byly v požadavcích náročnější menší firmy (obrázek 17 - graf pro MA). Lze to zřejmě zdůvodnit možnou zastupitelností IT pracovníků v případě, že jich je ve firmě zaměstnáno více a také většími možnostmi velkých firem IT pracovníky si vyškolit podle svých představ.

Lze očekávat, že vyšší nároky na znalosti IT pracovníků budou vyslovovat firmy působící v odvětvích z tohoto hlediska náročnějších. Tento předpoklad však průzkum potvrdil pouze v případě managerů, prostřední skupina odvětví je nejnáročnější na obchodníky a správce, u ostatních rolí nebyly mezi odvětvími zjištěny podstatnější rozdíly.

II) V rámci výzkumu bylo dále provedeno vyčerpávající zjišťování na vysokých a vyšších odborných školách, jež nabízely ve svém studijním programu IT obory, s cílem získat představu o nabídce IT odborníků a o jejich kvalitě.

Rozsah a složení základního souboru

Osloveny byly všechny vysoké školy (fakulty) a vyšší odborné školy v České republice nabízející ve školním roce 2006/2007 ke studiu IT obory. Informace byly získávány písemným dotazováním, ale i z dalších veřejných zdrojů, z výročních zpráv apod.

Zjišťovány byly počty studentů IT oborů v České republice, počty ke studiu přijímaných uchazečů, počty absolventů, a také představa o vývoji těchto veličin v nejbližších letech. Pro vytvoření představy o kvalitě absolventů byl zjišťován rozsah výuky předmětů vytvářejících znalosti v sedmnácti znalostních doménách definovaných pro výběrové šetření prováděné ve firmách.

Datový soubor tvoří údaje o celkem 203 IT studijních oborech existujících ve studijních plánech vysokých a vyšších odborných škol v České republice – počty vyučovaných oborů a počty studentů v roce 2005/6 obsahuje tabulka 5.

Tabulka 5. Vyučované obory a počty studentů VOŠ a VŠ v ČR v roce 2005/6

	VOŠ	Bc	Mgr (2 roky)	Mgr (5 let)	Celkem
Vyuč. obory	9	80	56	26	171
Počet studentů	707	16 252	2 535	3 304	22 798

Dále z údajů škol dále vyplynulo, že v dohledné době by mělo být otevřeno dalších cca 10 IT oborů pro více než 500 studentů. Počet absolventů se v roce 2006 přiblížil 3 500 studentů, z toho však značnou část představují studenti bakalářského studia, z nichž většina pokračuje ve studiu. Ze srovnatelných údajů o počtu přijímaných uchazečů o studium (tj. na IT oborech, které poskytly informace za celé období) vyplývá v nejbližších letech tendence k mírnému nárůstu počtu absolventů, který by se tak mohl pohybovat kolem 4 000.

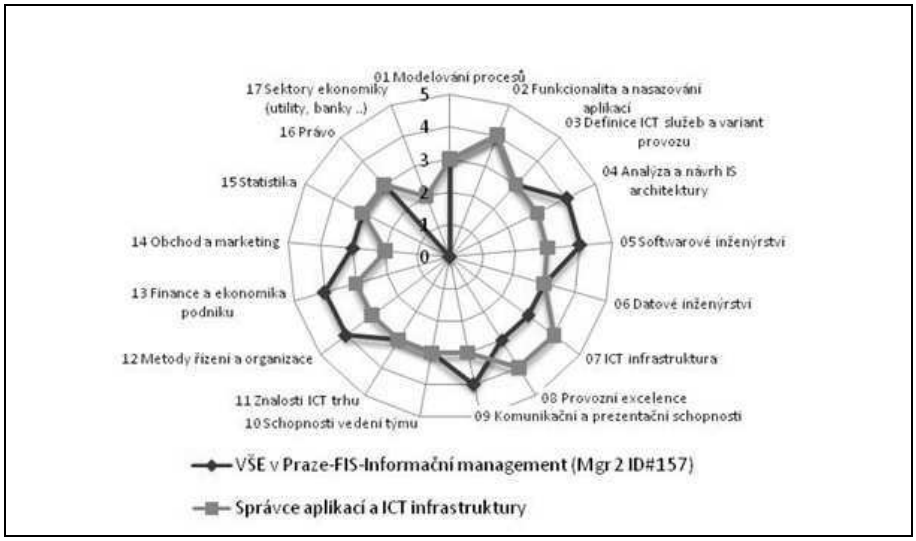
Znalostní profil absolventů IT oborů

Pro vytvoření představy o kvalitě absolventů a pro hledání odpovědi na otázku, nakolik studijní programy IT oborů vyhovují praxi, byly využity zjištěné údaje o rozsahu výuky předmětů vytvářejících znalosti absolventů v sedmnácti definovaných znalostních doménách (tedy 17 proměnných). Nejprve bylo třeba na jejich základě zjistit, které vyučované IT obory jsou si více a které méně podobné, a vytvořit tak jejich určité typické skupiny. Pro tento účel byl použit shlukovací EM algoritmus.

Na rozdíl od algoritmů založených na vzdálenostech, EM (Expectation-Maximization) algoritmus je metodou odhadu parametrů tzv. smíšeného modelu. Ten je založen na představě, že data jsou směsí pozorování pocházejících z různých pravděpodobnostních rozdělení. Cílem je tato rozdělení od sebe oddělit a modelovat je. Vzhledem k pravděpodobnostní podstatě smíšeného modelu lze rozdělení spojitých i nespojitých proměnných různého typu charakterizovat odpovídající hustotou či pravděpodobnostní funkcí. Jejich parametry jsou pak odhadovány tak, aby smíšený model co nejlépe vyhovoval datům. Kritériem je maximum věrohodnostní funkce. Výhodou smíšeného modelu je, že nevylučuje překrývání shluků a poskytuje pravděpodobnosti příslušnosti jednotlivých pozorování k vytvořeným shlukům. Matematické vlastnosti EM algoritmu pak zaručují, že věrohodnost v poskytované posloupnosti odhadů parametrů roste.

Shlukovací algoritmus byl použit zvlášť na bakalářské a zvlášť na magisterské IT obory (obory vyučované na VOŠ a několik oborů s nevyhovujícími údaji zahrnuto nebylo). Výsledkem shlukování byly v obou případech čtyři skupiny studijních oborů.

Pro konfrontace požadavků praxe na znalosti v jednotlivých znalostních doménách a úrovně znalostí poskytovaných vyučovanými IT obory v každé definované profesní roli byly použity paprskové grafy (příklad viz obr. 2). Pro vyjádření úrovně požadavků i nabízených znalostí byl opět používán medián.



Obrázek 2. Příklad paprskového grafu

Výzkum vedl k závěru, že ačkoliv počty přijímaných studentů na informatické obory neustále rostou, VŠ a VOŠ v České republice v nejbližších letech nebudou poskytovat dostatek absolventů pro pokrytí poptávky praxe. Také rozsah výstupních znalostí absolventů je a nadále bude velmi rozdílný. Absolventů, které není třeba ve firmách dále nákladně doškolovat, je a bude i v příštích letech nedostatek.

3. Literatura:

[1] Bradley, P. S. – Fayyad, U. M. – Reina, C. A.: Scaling EM Clustering to Large Databases, Microsoft Research, Microsoft Corp. 1999
[2] Doucek, P. – Novotný, O. – Pecáková, I. – Voříšek, J.: Analýza nabídky a poptávky po IT odbornících v ČR, VŠE Praha 2007 (v tisku)
[3] Frinking, E., Ligtvoet, a., Lundin, P., Oortwijn, W.: The supply and demand of e-skills in Europe, Draft final report, Prepared for DG Enterprise and Industry of the European Commission, June 2005
[4] Statistická ročenka České republiky 2006 (elektronická verze, dostupná na www.czso.cz/csu/2006edicniplan.nsf/publ/10n1-06-_2006)

Doc. Ing. Iva Pecáková, CSc.
nám. W. Churchilla 4
130 67 Praha 3
e-mail: pecakova@vse.cz

Ověřovací studie přístupu hodnocení finanční výkonnosti stavebních spořitelen v ČR

Poměnková Jitka, Kašparovská Vlasta

Abstract:
Using correlation analysis during process of financial efficiency evaluation of the Czech building and loan associations arise problem of credibility in the case of small sample size. Even using knowledge from theory with information character it is impossible to consider correlation analysis results as stable without additional analysis. Suitable step in validation process is using partial correlation analysis. In our case we could not deploy partial correlation due to small sample size. Alternative, validation, approach, is an additive index partition of synthetic indicator. The aim of presented paper is to validate results of correlation analysis used for the financial efficiency evaluation of the Czech building and loan associations by additive index partition method.

Key words: *building and lone association, aditive index partition, correlation analysis, financial performance*

1. Úvod

Při posuzování finanční výkonnosti českých stavebních spořitelen pomocí korelační analýzy narážíme na problém dostupnosti zveřejňovaných dat, který má za následek relativně malý rozsah posuzovaného souboru. Studie vychází z pyramidálního rozkladu syntetického ukazatele finanční výkonnosti ROAA jak byl uveden v Kašparovská, Poměnková, 2007. I přes zapojení znalosti teorie a z ní plynoucího popisu charakteru údajů nelze výsledky korelační analýzy považovat za dostatečně stabilní bez další podrobnější analýzy. Vzhledem ke stejnému problému, tj. malému rozsahu datového souboru, nelze použít ani parciální korelační analýzu.

Alternativním přístupem je aditivní indexový rozklad syntetického ukazatele, který ve své teorii již zohledňuje ekonomický charakter posuzovaných dat a jejich vzájemný vztah. Proto byla za ověřovací metodu zvolena metoda aditivního indexového rozkladu.

Cílem předkládaného příspěvku je ověřit výsledky korelační analýzy při hodnocení finanční výkonnosti českých stavebních spořitelen metodou aditivního pyramidálního rozkladu. Přínos příspěvku je v metodickém spojení standardně využívaných přístupů. Uvedené přístupy bývají využívány izolovaně, bez záměru podpořit výsledky použité metodiky.

2. Metodika

Pro empirickou analýzy bylo využito ročních hodnot dílčích marží v sektoru stavebních spořitelen pro období 1999 – 2005. Korelační analýza využívá hodnot změn ROAA a marží. Druhý přístup aditivní indexový rozklad pyramidálního rozkladu zahrnuje znalost vzájemného vztahu jednotlivých marží (výpočet jejich hodnot jakož i hodnot změn ROAA tuto znalost zohledňuje). Poznamenejme, že hodnota ROAA byla odvozena na základě zmiňovaného aditivního indexového rozkladu a že posouzení probíhá na změnách jak ukazatele ROAA, tak jednotlivých marží.

Z pohledu statistické analýzy závislosti a vlivu posuzovaných ukazatelů byla uvažována korelační analýza, a to výběrový korelační koeficient, který poukazuje na přímou nebo nepřímou závislost ROAA na konkrétní marži. Pro testování statistické významnosti bylo využito oboustranného t-testu (blíže ANDĚL (1978).

Při aditivním rozkladu jsme vycházeli ze znalosti pyramidálního rozkladu (viz. blíže Kašparovská, Poměnková, 2007), a tedy z dekompozice podle vztahu

$$\Delta ROAA = \Delta UM + \Delta PM + \Delta MFO + \Delta MOV - \Delta MSN - \Delta MVN - \Delta RM \quad (1)$$

kde UM - úroková marže, PM - provizní marže, MFO - marže finančních operací, MOV - marže ostatních výnosů, MSN - marže správních nákladů, MVN - marže věcných nákladů a RM - riziková marže. Ekonomické vazby mezi dílčími ukazateli v pyramidovém rozkladu ROAA lze matematicky analyzovat pomocí tzv. aditivního indexního rozkladu (Dluhošová, 2004), kdy je celková změna vrcholového ukazatele (ROAA) vyjádřena jako součet změn dílčích ukazatelů, tedy

$$\Delta ROAA = \sum_{j=1}^{k} \Delta M_j , \quad (2)$$

kde ΔM_j je změna příslušné marže za celé sledované období. Je možné analyzovat jak odchylku absolutní, tak i relativní. V našem případě zaměříme pozornost na odchylku relativní. Vyčíslení jednotlivých vlivů pro aditivní vazbu je určeno poměrem změny ukazatele a celkové změny ukazatelů

$$\Delta ROAA_i = \frac{ROAAA_{i+1} - ROAA_i}{\sum_i \left(ROAA_{i+1} - ROAA_i \right)}, \; i = 1,..., n-1 \quad (3)$$

Pro jednotlivé marže platí

$$\Delta M_j = \frac{M_{i+1} - M_i}{\sum_i \left(M_{i+1} - M_i \right)} \Delta ROAA_i , \; i = 1,..., n-1 \quad (4)$$

kde M_i označuje příslušnou marži.

3. Empirická část

Datovou základnu pro analýzu tvoří údaje získané z výkazů zisku a ztráty a bilancí jednotlivých českých stavebních spořitelen za období 1999 - 2005. Hodnoty průměrných aktiv byly pro jednotlivá období stanoveny z údajů k 31.12. dvou po sobě následujících období. Pro výpočty průměrných hodnot finančních ukazatelů odvětví byly použity údaje všech stavebních spořitelen na českém trhu. Pro získané hodnoty byla nejprve na základě vztahu (1) dopočtena hodnota syntetického ukazatele ROAA. Výpočet syntetického ukazatele, změn syntetického ukazatele i změn jednotlivých marží je zaznamenán v tabulce 1.

Tabulka 1. Výsledky aditivního indexového rozkladu za sektor stavebních spořitelen (globální hodnocení).

	Δ ROAA	ΔUM	Δ PM	Δ MFO	Δ MOV	Δ MSN	Δ MVN	Δ RM
2005 - 2004	0,000087	0,000516	-0,000208	0,000048	-0,000386	-0,000229	0,000035	0,000076
2004 - 2003	-0,005263	0,000009	-0,008568	-0,000217	0,000425	-0,002132	-0,000670	-0,000285
2003 - 2002	0,003538	-0,007571	0,009871	-0,000537	0,000003	-0,002934	0,000673	0,000489
2002 - 2001	0,001815	-0,003543	0,000152	0,000362	-0,000309	-0,004120	-0,000999	-0,000034
2001 - 2000	0,000134	-0,004215	0,001463	0,000181	-0,000082	-0,003517	0,001301	-0,000572
2000 - 1999	0,018783	0,015510	0,002222	-0,000198	-0,000310	-0,002178	0,000739	-0,000119
SUMA	**0,019094**	0,000706	**0,004932**	-0,000360	-0,000660	**-0,015111**	0,001080	-0,000445

Pozn. Tučně vyznačená čísla jsou významnými hodnotami komentovanými v textu
Zdroj:vlastní výpočet

Na získané hodnoty přírůstků marží a ROAA (tabulka 1) byla aplikována korelační analýza s cílem identifikovat významnost vlivu přírůstku marže za celé sledované období na celkový přírůstek syntetického ukazatele. Byly zjištěny níže uvedené hodnoty výběrových korelačních koeficientů.

Tabulka 2. Výsledky korelační analýzy přírůstků ROAA a jednotlivých přírůstků marží pro spořitelní sektor.

	ΔUM	Δ PM	Δ MFO	Δ MOV	Δ MSN	Δ MVN	Δ RM
Výběrový korelační koeficient	0,76903	0,44092	-0,20016	-0,53032	0,00078	0,40433	0,15594
p-hodnota	0,0739	0,3815	0,7038	0,2791	0,9988	0,4266	0,7680
Pořadí stupně závislosti	1	3	5	2	7	4	6

Pozn. p-hodnota < ♕ indikuje statisticky významnou korelaci na (1-♕)% hladině významnosti.
Zdroj:vlastní výpočet

Jednotlivé hodnoty korelačních koeficientů byly testovány na statistickou významnost. Z tabulky 2. vyplývá, že ani jedna hodnota zjištěného výběrového korelačního koeficientu není statisticky významná. Pouze hodnota změny úrokové marže vykazovala statistickou významnost na 90% hladině významnosti. Poznamenejme, že vzhledem k velmi malému rozsahu souboru dat jsou výsledky a jejich vypovídací schopnost ověřovány analýzou změn marží aditivního indexového rozkladu z pohledu nejvýznamnější stanovené změny marže.

Protože se jednotlivé zjištěné marže podílí vlivem svých hodnot rozdílnou váhou na výsledku ROAA, je pro sofistikovanější zjištění vlivu jednotlivých marží provedeno vážení výběrových korelačních koeficientů podílem sumy změn příslušené marže za období 1999-2005 na celkové sumě změn ROAA.

Tabulka 3. Výsledky vážené korelační analýzy přírůstků ROAAA a jednotlivých přírůstků marží pro spořitelní sektor.

	ΔUM	Δ PM	Δ MFO	Δ MOV	Δ MSN	Δ MVN	Δ RM
Váha	0,03699	0,25831	-0,01888	-0,03456	-0,79139	0,05655	-0,02329
Vážený výběrový korelační koeficient	0,02844	0,11390	0,00378	0,01833	-0,00062	0,02287	-0,00363
Pořadí stupně závislosti	2	1	5	4	7	3	6

Zdroj:vlastní výpočet

Z uvedené korelační analýzy vyplývá, že největší vliv na změnu ukazatele ROAA v období 1999 - 2005 vykazuje změna provizní marže následována změnou marže úrokové. Nejmenší vliv pak vykazuje změna marže správních nákladů.

Nyní se zaměříme na ověření dosažených výsledků metodou aditivního indexového rozkladu. Pozornost bude soustředěna primárně na globální hodnocení, které bude doplněno poznatky z horizontálního hodnocení.

Analýza vývoje rentability aktiv v odvětví českých stavebních spořitelen s využitím metody aditivního indexového vyplývá za období 1999-2005 poskytuje několik závěrů. Globální hodnocení aditivního indexového rozkladu je založeno na zkoumání vlivu dílčích ukazatelů na syntetický ukazatel ve zvoleném období jako celku, horizontální hodnocení zkoumá vliv dílčích ukazatelů (dílčích marží) na změny syntetického ukazatele (Δ ROAA) v jednotlivých obdobích jako celku a vertikální hodnocení analyzuje vliv dílčích složek v jednotlivých letech zvoleného období na výslednou souhrnnou hodnotu dílčího ukazatele (Kašparovská, Poměnková; 2007). Pro potřeby této studie je využito výsledků

globálního a horizontálního hodnocení. Poznamenejme, že na rozdíl od korelační analýzy, která zkoumá (jen) oboustrannou závislost, umožní analýza aditivního indexového rozkladu zkoumat problematiku z různých úhlů pohledu. Výsledky aditivního indexového rozkladu za sektor stavebních spořitelen uvádí tabulka 1 (globální hodnocení) a tabulka 4 (horizontální hodnocení).

Z globálního hodnocení vyplývá, že celková změna ROAA ve sledovaném období činila 1,91 %. Z ekonomického hlediska můžeme uvedený nárůst označit jako nárůst podprůměrný, neboť jeho hodnoty bylo dosaženo za období 6 let. Hodnotíme-li vliv globálních změn jednotlivých marží na celkovou změnu ROAA můžeme učinit závěr, že relativně výrazný pokles marže správních nákladů o 1,51 %, která dosáhla v globálním hodnocení nejvyšší změny, se významně nepromítl do změn v ROAA a tudíž i do jeho globální hodnoty. Dále lze vysledovat nárůst provizní marže o 0,49 %. Významně, podle našeho názoru, ovlivnily konečnou relativně nízkou hodnotu ROAA změny v oblasti úrokové marže, jejíž de facto nepříznivý vývoj kompenzovaly stavební spořitelny částečně tlakem na růst provizní marže.

Tabulka 4. Výsledky aditivního indexového rozkladu pro sektor stavebních spořitelen (horizontální hodnocení).

	Δ ROA	ΔUM	Δ PM	Δ MFO	Δ MOV	Δ MSN	Δ MVN	Δ RM
2005 - 2004	0,000087	**0,000516**	-0,000208	0,000048	-0,000386	-0,000229	0,000035	0,000076
2004 - 2003	**-0,005263**	0,000009	**-0,008568**	-0,000217	0,000425	-0,002132	-0,000670	-0,000285
2003 - 2002	0,003538	-0,007571	**0,009871**	-0,000537	0,000003	-0,002934	0,000673	0,000489
2002 - 2001	0,001815	-0,003543	0,000152	0,000362	-0,000309	**-0,004120**	-0,000999	-0,000034
2001 - 2000	0,000134	-0,004215	**0,001463**	0,000181	-0,000082	-0,003517	0,001301	-0,000572
2000 - 1999	**0,018783**	**0,015510**	0,002222	-0,000198	-0,000310	-0,002178	0,000739	-0,000119
SUMA	0,019094	0,000706	0,004932	-0,000360	-0,000660	-0,015111	0,001080	-0,000445

Pozn. Tučně vyznačená čísla jsou významnými hodnotami komentovanými v textu
Zdroj:vlastní výpočet

Z horizontálního hodnocení tabulka 4 vyplývá, že ve třech obdobích (2001-2000, 2003-2002, 2004-2003) měly nejvyšší vliv změny provizní marže, ve dvou obdobích (2005-2004, 2000-1999) měl nejvyšší vliv na Δ ROAA změna úrokové marže a v jednom období (2002-2001) měl nejvýznamnější vliv pokles marže správních nákladů.

4. Závěr

Předkládaný článek se zabývá ověřením výsledků korelační analýzy při hodnocení finanční výkonnosti českých stavebních spořitelen metodou aditivního pyramidálního rozkladu v období 1999-2005. Při hodnocení vlivu jednotlivých složek je posuzován jejich vliv na celkový ukazatel ROAA, přičemž je využito vyčíslení ROAA na základě pyramidálního rozkladu metodou aditivního indexového rozkladu.

Z uvedené korelační analýzy (vážené hodnoty výběrových korelačních koeficientů) vyplývá, že největší vliv na změna ukazatele ROAA v období 1999-2005 vykazuje změna provizní marže následován změnou marže úrokové. Nejmenší vliv pak vykazuje změna marže správních nákladů.

Hodnotíme-li vliv globálních změn jednotlivých marží na celkovou změnu ROAA můžeme učinit závěr, že relativně výrazný pokles marže správních nákladů, která dosáhla v globálním hodnocení nejvyšší změny, se významně nepromítl do změn v ROAA. Dále došlo k nárůstu provizní marže. Významně podle našeho názoru ovlivnily konečnou relativně

nízkou hodnotu ROAA změny v oblasti úrokové marže, jejíž de facto nepříznivý vývoj kompenzovaly stavební spořitelny částečně tlakem na růst provizní marže. Z horizontálního hodnocení aditivního indexového rozkladu vyplývá, že ve třech obdobích (2001-2000, 2003-2002, 2004-2003) měly nejvyšší vliv změny provizní marže, ve dvou obdobích (2005-2004, 2000-1999) měl nejvyšší vliv na změnu ROAA změna úrokové marže a v jednom období (2002-2001) měl nejvýznamnější vliv pokles marže správních nákladů.

Provedeme-li srovnání výsledků korelační analýzy s výsledky globální a horizontální analýzy aditivního indexového rozkladu, můžeme říci, že analýza potvrdila výsledek významnosti (přírůstku) změn provizní marže a (přírůstku) změn úrokové marže na celkovou změnu ROAA. Rovněž v případě změny správní marže, jak vyplynulo z korelační analýzy, byla potvrzena nízká významnost ve smyslu vlivu na změnu syntetického ukazatele ROAA metodou adaptivního indexového rozkladu, jak uváděla studie Kašparovská, Poměnková 2007.

Příspěvek vznikl v rámci výzkumného záměru PEF MZLU číslo MSM 6215648904.

4. Literatura

ANDĚL, J. Matematická statistika. SNTL/ALFA, Praha 1978.

DLUHOŠOVA, D. Přístupy k analýze finanční výkonnosti firem a odvětví na bázi metody

EVA. In: Finance úvěr 11-12/2004, roč. 54, str. 541-559, ISSN: 0015-1920.

KAŠPAROVSKÁ, V. Analýza vývoje trhu stavebního spoření v ČR. In Acta Univerzitatis, MZLU Brno, ročník LIII, 2005, 3, str. 69-77. ISSN: 1211-8516.

V. KAŠPAROVSKÁ, J. POMĚNKOVA. Uplatnění aditivního indexového rozkladu při hodnocení finanční výkonnosti odvětví českých stavebních spořitelen. Acta Univerzitatis, MZLU Brno, 2007 (přijato do tisku)

Adresa autora:
Jitka Poměnková, RNDr., Ph.D
Ústav statistiky a operačního výzkumu, PEF MZLU
Zemědělská l,
613 00 Brno
pomenka@mendelu.cz

Vlasta Kašparovská, Ing., Ph.D
Ústav financí, PEF MZLU
Zemědělská l,
613 00 Brno
kasparov@mendelu.cz

Marketingový průzkum zájmu o produkty poskytovatelů služeb v oblasti cestovního ruchu

Semerák Pavel, Kubanová Jana

Abstract: The paper deals with the problem of marketing strategy in the field of tourism. The basic steps of the investigation are described as pilot study and advance research. The methodology of questionary formation and strategy of inquiry is analysed.

Key words: marketing strategy, pilot study, questionary formation

1. Úvod

Prioritním zájmem každého výrobce nebo poskytovatele služeb je odbyt jeho produktů, neboť s tímto odbytem jsou úzce spjaty i jeho zisky. K odhalení perspektiv budoucího vývoje mohou přispět marketingové průzkumy zájmu o jednotlivé produkty.
Pod marketingovým průzkumem rozumíme soubor činností, které slouží ke zkoumání externích a interních podmínek pro obchodní rozvoj firmy a její publicity.

Jednou z metod, užívaných při takovýchto šetřeních, je metoda dotazníkového šetření. Prvním krokem je sestavení vhodného dotazníku, který splňuje podmínky kvalitního průzkumu a na jehož základě je možné vyjadřovat kvalitní závěry. Teorii tvorby dotazníku a nástinu prováděných postupů a operací je věnován následující příspěvek.

2. Teorie tvorby dotazníku

Výzkum můžeme podle podstaty zkoumání rozlišit na kvantitativní a kvalitativní. V obou případech je zapotřebí nasbírat jisté informace, nebo chceme-li se vyjádřit konkrétněji, jistá data, na jejichž základě vytváříme statistické analýzy a praktické závěry. V případech, kdy se zabýváme sociologickým, marketingovým či jiným výzkumem, kde se data týkají lidí, zákazníků, jejich nálad, názorů, atd., je prvním krokem nashromáždění potřebného množství dat. Nejjednodušším způsobem získání dat je zřejmě dotazování. Nejužívanějšími způsoby dotazování jsou přímý rozhovor, telefonické interview a dotazník. Každá z těchto forem má své výhody a nevýhody. Věnujme se zde pouze dotazníku.

V porovnání s ostatními způsoby dotazování jsou výhody dotazníku následující. Je poměrně snadno distribuovatelný. Umožňuje získání poměrně velkého množství dat při relativně krátkém čase a malé námaze tazatele. Respondentovi zajišťuje poměrně silný pocit anonymity, což je velmi důležité z hlediska pravdivé odpovědi na otázky například intimního rázu či otázky, týkající se témat, jež jsou ve společnosti tabu. Dotazník má i stinné stránky. Nikde nemáme zaručeno, že respondent odpoví na všechny otázky. Při distribuci dotazníků např. do domácností je možné, že dotazník vyplní úplně jiná osoba, než jakou jsme zamýšleli. A konečně, respondent nakonec vůbec nemusí vyplněný dotazník vrátit. Návratnost dotazníků je obecně poměrně nízká, je ji ovšem možné jistými opatřeními zvýšit. Od přiložení ofrankované obálky ke zpětnému zaslání k tazateli až po slib dárku nebo zařazení do soutěže o lákavé ceny.

Kvantitativní výzkum spočívá v ověřování hypotéz. To se provádí tak, že se získaná data statisticky zpracují. Při samotném sestavování dotazníku pro kvantitativní výzkum je samozřejmostí zahrnutí všech takových otázek, které nám získají odpovědi, z nichž budeme schopni otestovat dané hypotézy. Tyto otázky by měly být formulovány tak, aby respondent neměl žádné pochybnosti ohledně pochopení dané otázky. Dále by otázky neměly být sugestivního charakteru. Možná, že ne každý volič ČSSD by na otázku „Volil jste loni socialisty buldozera Paroubka," odpověděl kladně.

Ne na každou otázku se lze ptát přímo. Specielně například v případě motivace zákazníka k nákupu jistého zboží či služby, mohou rozhodovat jisté podprahové faktory, které si ani sám kupující neuvědomuje. V tom případě je lepší zjistit dotazovanou skutečnost spíše sérií nepřímých otázek, jejichž odpovědi nám nakonec jistou představu podají.

Dalšími ze skupiny zásad při sestavování dotazníku jsou např.

- Neredundance otázek. Každá otázka má svůj význam. Každého respondenta by pravděpodobně pětistránkový, otázkami hustě zaplněný dotazník, jistě odradil. Neplýtvejme tedy zbytečnými otázkami a řiďme se nepsaným zákonem, jenž praví: „Každý dotazník je příliš dlouhý“.
- Dotazník by měl mít jistou logickou strukturu a návaznost otázek.
- Na osobní data respondenta či jiná, pro respondenta nepříjemná témata, je lepší ptát se až na konci dotazníku, přece ho hned neodradíme vlezlými otázkami na jeho věk, pohlaví nebo např. výši jeho příjmu.
- Dotazník je vhodné začínat otázkou, která vzbudí respondentův zájem a která ho příjemně naladí k dalšímu vyplňování dotazníku. Tato otázka může být z hlediska významnosti odpovědi i naprosto zbytečná (Vraťme se ale k prvnímu bodu, takových otázek nesmí být moc.) .

Kvalitativní výzkum probíhá tak, že na základě získaných dat výzkumník pátrá po pravidelnostech, objevujících se v těchto datech, po významu těchto dat, formuluje předběžné závěry. Výstupem mohou být nově formulované hypotézy nebo nová teorie.(DISMAN, 2002). Při sestavování dotazníku ja pak cílem nasbírat co nejvíce dat týkajících se zkoumané problematiky.

3. Pilotní studie a předvýzkum

V našem případě jsme byli osloveni cestovní kanceláří (dále CK), nabízející poznávací zájezdy do celého světa. Cílovou populací této CK (a následně i dotazníku) jsou osoby, které jezdí s cestovními kancelářemi za poznáním. Manažeři CK měli jisté představy o tom, na co se ve výzkumu zaměřit, dokonce naformulovali i některé otázky, které by v dotazníku rádi měli. Z toho následně vyplynulo, že zamýšlený výzkum bude jak kvantitativní, tak i kvalitativní. Literatura nabádá, abychom před vytvořením finální podoby dotazníku, kterou budeme dále předkládat tazatelům, provedli pilotní studii a předvýzkum.

Pilotní studie slouží ke zjištění toho, zda informace, jež chceme získat, vůbec existují a zda jsou zjistitelné. Je doporučováno provádět pilotní studii formou skupinového rozhovoru s malým vzorkem cílové populace.

Předvýzkum spočívá v tom, že poté co jsme vymysleli jisté nástroje, v našem případě otázky v dotazníku ke zjištění požadovaných informací, tyto nástroje testujeme opět na malé skupině reprezentující cílovou populaci. Cílem je ověřit, zda jsou všechny otázky srozumitelné a jednoznačné nebo zda respondenti nemají problémy s formulováním odpovědí.

Z důvodu nedostatku času na provedení výzkumu jsme provedli pilotní studii a předvýzkum společně a to formou podání předběžné podoby dotazníku a následné besedy se skupinou 32 klientů CK, kteří se zúčastnili posezení spojeného s promítáním diapozitivů z nedávno uspořádaných zájezdů.

4. Výsledky předvýzkumu

Podle formulace odpovědí jsme zjistili, že respondenti neměli problémy s vyplňováním dotazníku. Všechny otázky byly zřejmě srozumitelné a všichni na všechny otázky odpověděli.

Zdá se tedy, že dotazníkové šetření obsahující tyto otázky by nám mohlo jisté výsledky poskytnout.

Téměř na každou otázku v dotazníku lze aplikovat metody popisné statistiky v rámci skupiny respondentů. Např. v otázce č. 7 nám jako modální odpověď vyšla (podle našeho očekávání) Albánie, do nedávné doby pro turisty uzavřená země. U otázek na délku zájezdu nám dokonce i při malém počtu respondentů vyšly námi očekávané pravidelnosti, totiž že u zájezdu do evropských zemí nikdo nevolil dobu přesahující 3 týdny, naopak u mimoevropských zemí nikdo nebral v potaz délku dvou a méně týdnů.

Dále počítáme s tím, že při větším počtu respondentů bychom testovali námi uvažované hypotézy, zmiňme následující:

- existence závislosti mezi věkem a přípustnou cenou zájezdu
- existence závislosti mezi věkem a typem dopravy při zájezdech do evropských zemí
- existence závislosti mezi věkem a optimální délkou pobytu
- existence závislosti mezi věkem a způsobem, jak se respondent dověděl o CK.

Všechny tyto závislosti bychom testovali pomocí testu nezávislosti v kontingenčních tabulkách.

Podívejme se například na poslední uvedenou hypotézu a otestujme ji pro náš zatím malý soubor respondentů. Tabulka 1 udává kontingenční tabulku sestavenou na základě odpovědí na otázky číslo 15, 16 a 20.

Tabulka1. Kontingenční tabulka

Věk	Dozvěděl se z		∑
	internetu	jiné	
Méně než 50 let	$n_{11}=16$	$n_{12}=6$	$n_{1\cdot}=22$
50 let a více	$n_{21}=3$	$n_{22}=7$	$n_{2\cdot}=10$
∑	$n_{\cdot 1}=19$	$n_{\cdot 2}=13$	32

Podle našich předpokladů by mělo platit, že starší klienti se o CK dozvědí spíše jinak než z internetu, na rozdíl od mladších klientů. Jelikož jsou četnosti v tabulce poměrně malé, nebudeme využívat postupu založeného na asymptotickém χ^2 - rozdělení o 1 stupni volnosti, ale použijeme tzv. Fischerův faktoriálový test, specielně jeho jednostrannou verzi.

Za předpokladu neexistence závislosti mezi věkem a způsobem, jakým se klienti dozvěděli o CK je pravděpodobnost toho, že při daných marginálních četnostech $n_{i\cdot}$ a $n_{\cdot j}$ pro $i, j = 1, 2$ budou hodnoty v tabulce právě n_{ij}, rovna

$$P = \frac{n_{1\cdot}!\; n_{2\cdot}!\; n_{\cdot 1}!\; n_{\cdot 2}!}{n!\; n_{11}!\; n_{12}!\; n_{21}!\; n_{22}!} \quad . \tag{1}$$

Pro naši tabulku je tato pravděpodobnost rovna číslu 0,02577. Neskončíme ovšem výpočtem této hodnoty. Budeme počítat pravděpodobnosti dalších, ještě více proti hypotéze nezávislosti

svědčících skutečností. Jsou to ty, kdy by při zachování všech marginálních četností, byla četnost n_{21} rovna hodnotám 2, 1 a 0. Po vypočítání příslušných pravděpodobností nám jejich součet dává hodnotu
$P_{\sum} = 0{,}02577 + 0{,}003411 + 2{,}106 \cdot 10^{-4} + 4{,}4332 \cdot 10^{-6} = 0{,}03129$. Vidíme, že $P_{\sum} < 0{,}05$, zamítáme tedy hypotézu o nezávislosti na hladině 5%.

5. Závěr

Pilotní studia a předvýzkum prokázaly opodstatněnost sestaveného dotazníku i jeho dostatečnou validitu. V současné době pracujeme na distribuci dotazníku většímu počtu klientů. Cílem bude jak testování hypotéz, tak i na základě kvalitativního výzkumu, vytváření nových pracovních hypotéz, které budou opět podrobeny testování.

6. Literatura:

ANDĚL, J.: *„Matematická statistika“*, SNTL, Praha, 1985
BÁRTOVÁ, H., BÁRTA, V.: *„Marketingový výzkum trhu“*, ECONOMIA, a.s., Praha, 1991, ISBN 80-85378-09-4
DISMAN, M.: *„Jak se vyrábí sociologická znalost“*, Karolinum, Praha, 2002, ISBN 80-246-0139-7

Adresa autorů:

Mgr.Pavel Semerák, doc. PaedDr.Jana Kubanová,CSc
Univerzita Pardubice
Studentská 95
53210 Pardubice
pavel.semerak@upce.cz, jana.kubanova@upce.cz

Příloha: Dotazník zájmu o produkty poskytovatelů služeb v oblasti cestovního ruchu

1) V jaké zahraniční zemi jste byl(a) naposledy?
2) Ze všech Vašich pobytů v zahraničí, která země se Vám nejvíce líbila?
3) Kolik poznávacích zájezdů jste s naší cestovní kanceláří dosud absolvoval(a) ?
4) Pokud jste se zúčastnil(a) více než jednoho poznávacího zájezdu s naší CK, který z dosud absolvovaných zájezdů považujete za nejvydařenější? Uveďte destinaci, případně název zájezdu.
5) Kolik zájezdů do **evropských zemí** jste absolvoval(a)?
6) Jaká je pro Vás přijatelná délka zájezdu do evropské země? □ do 7 dní □ 8-15 dní □ 16-23 dní □ 24-31 dní □ déle než 31 dní
7) Kterou evropskou zemi zamýšlíte v budoucnu navštívit? (uveďte maximálně 3 nejvíce preferované země)
8) Jaký druh dopravy na evropských zájezdech preferujete? □ letadlo □ autobus □ kolo □ jiný, jaký?.....
9) Jaký druh ubytování na evropských zájezdech preferujete? □ hotel □ penzión □ stan □ jiný, jaký?.....
10) Jaká je pro Vás přijatelná cena zájezdu do evropské země? □ do 10 tis. □ 10-15 tis. □ 15-20 tis. □ nad 20 tis.(zaplatím komfort) □ nerozhoduje
11) Kolik zájezdů do **mimoevropských zemí** jste absolvoval(a)?
12) Jaká je podle Vás optimální délka zájezdu do mimoevropské země? □ do 7 dní □ 8-15 dní □ 16-23 dní □ 24-31 dní □ déle než 31 dní
13) Kterou mimoevropskou zemi zamýšlíte v budoucnu navštívit? (uveďte prosím maximálně 3 nejvíce preferované země)
14) Jaká je podle Vás přijatelná cena zájezdu do mimoevropské země (exotika)? □ do 30 tis. □ 30-50 tis. □ 50-80 tis. □ nad 80 tis.(zaplatím komfort) □ nerozhoduje
15) Jak jste se o naší CK dozvěděl(a)? □ od přátel □ z katalogu □ z internetu □ ze sdělovacích prostředků jiné: ……
16) Pokud jste v předchozí otázce odpověděl(a) „z internetu“, vzpomenete si, na jaké internetové adrese to bylo? □ seznam.cz □ centrum.cz □ atlas.cz □ google.cz □ azet.sk □ karavela.cz □ jiný, jaký?.....

17) Pohlaví: □ M □ Ž
18) Vzdělání: □ ZŠ □ SŠ □ VŠ
19) Jazykové znalosti: □ ruský □ německý □ anglický □ francouzský □ jiný.....
20) Věk: □ do 30 □ 30-39 □ 40-49 □ 50-59 □ 60 a více

Nový prístup – nové možnosti štatistického modelovania, alebo ako „ušiť pravdepodobnostný model na mieru“

Ľubica Sipková

Abstract

In the article is presented a useful supplement to the classical statistical approaches to modeling, which gives a new perspective for statistical models and brings acceptable results in cases, when selection of a PDF or CDF does not succeed. Quantile functions, as a random component of statistical model, can be combined and modified with the mathematical kid altogether with a deterministic part of a model to derive new functions with good fit of the data.

Kvantilové metódy pravdepodobnostného modelovania prinášajú dodatočné možnosti tam, kde sa nepodarilo úspešne modelovať tvar klasickým spôsobom, t.j. definovaním pravdepodobnostnej funkcie niektorým známym tvarom funkcie hustoty, prípadne distribučnej funkcie. Kombinovaním známych tvarov pravdepodobnostných rozdelení pri modelovaní náhodnej zložky stochastických modelov, definovaných pomocou inverznej distribučnej funkcie, tzv. kvantilovej funkcie – QF [1], umožňuje pomocou matematického aparátu dosiahnuť vysokú elasticitu tvaru a tak vhodne „napasovať“ teoretické rozdelenie na empirický tvar rozdelenia.

Veľké využitie metód modelovania na kvantilovom základe vidíme aj v ich aplikácii u regresných modelov v prípade keď veľká časť variability ostáva nevysvetlená regresnou zložkou. Pôsobenie iných – neznámych, tzv. náhodných vplyvov, možno priamo zakomponovať do regresnej funkcie pomocou pravdepodobnostnej zložky definovanej QF. Podstatou nových možností je, že výstupom z pravdepodobnostného modelu v kvantilovom tvare nie je pravdepodobnosť, ale tak ako u regresných modelov – hodnota znaku, t. j. kvantil.

Nový prístup k pravdepodobnostnému modelovaniu si vyžiadal aj rozvoj nových metód a postupov vo všetkých fázach štatistického modelovania. Z dôvodu obmedzeného rozsahu príspevku ponúkame v ňom prehľad východiskovej literatúry k štúdiu danej problematiky a upozorňujeme na monografiu pod názvom Modelovanie kvantilovými

[1] Kvantilová distribučná funkcia (tiež kvantilová funkcia - QF) spojitej náhodnej premennej X je vyjadrená vzťahom $Q(p) = x$, pre každé reálne p, pre ktoré platí $0 \leq p \leq 1$. Vyjadruje x, p-kvantil premennej X, ako funkciu p, t.j. pravdepodobnosti, že náhodná premenná X má hodnoty menšie, nanajvýš rovné reálnemu číslu x. $Q(p)$ je neklesajúcou funkciou. V literatúre je definovaná aj ako inverzná funkcia k distribučnej funkcii: $Q(p) = F^{-1}(p)$, pre $0 \leq p \leq 1$.

funkciami [3], ktorá je analýzou východísk, prístupov, metód a aplikácií kvantilového modelovania.

Aplikované boli viaceré kladne zošikmené základné tvary QF známych pravdepodobnostných rozdelení, ako Weibullov, gamma, logistický, power, Paretov, lognormálny tvar (viď. napr. [2], str. 155, tiež [4], str. 63) a hlavne ich viaceré kombinácie, ktoré vznikli skladaním jednoduchých tvarov pomocou matematického aparátu, a to na údajoch o príjmoch domácnosti v SR, ktoré získal ŠÚ SR v rámci zisťovania o príjmoch a životných podmienkach EU SILC 2005 podľa metodiky Eurostat-u (premenná HY020_F – celkový ročný disponibilný príjem domácnosti, v príspevku *CDPD*). Rôznym vážením jednotlivých jednoduchých rozdelení v základnom tvare QF v jednotlivých častiach zloženého tvaru príjmového rozdelenia sa dosahuje plynulosť prechodu z jedného tvaru na ich kombináciu a následne na druhý tvar. Váhy sú pre dosiahnutie primeranej rýchlosti prechodu jedného tvaru na iný a ich vhodným splynutím v strednej časti kvantilového modelu volené ako rôzne funkcie pravdepodobnosti.

Aplikované boli aj viaceré metódy identifikácie tvaru, ktoré sú dostupné v rôznych štatistických programových balíkoch (najrozsiahlejšia zbierka jednoduchých teoretických tvarov k aplikácii bola k dispozícii v Statgraphics Centurion), ktoré však neumožnili nájsť štatisticky významnú – dobrú zhodu ani s jedným z jednoduchých ani viacparametrických teoretických tvarov.

Na hľadanie vhodných východiskových tvarov pre ich kombinovanie do zloženého kvantilového tvaru je vhodné použiť menej známe Q-Q grafické zobrazenia rôznych funkcií empirického a teoretického tvaru, grafické znázornenie hodnôt ich koeficientov korelácie a tabuľkové sumarizácie výpočtov charakteristík osobitne pre pravý a ľavý koniec rozdelenia (viď. [2], str. 173, niektoré v praxi menej známe grafické a kvantitatívne identifikačné analýzy pre inú aplikáciu sú napr. v [3], Aplikačná časť 3.2.4). Záverom identifikačných analýz je, že hľadaný kvantilový tvar má mať rýchly vzrast ľavého konca od hodnoty po tretí percentil, následne ostrejší ohyb k pomalšiemu rastu až po horný decil a opätovný, však plynulejší vzrast od 95. percentilu do dlhého tenkého konca pre rýchlo rastúce horné hodnoty príjmov. Do úvahy pre dolný koniec preto treba brať Weibullov, lognormálny, ganmma, prípadne exponenciálny tvar a pre ľavý koniec Paretov prípadne iné extrémne tvary (pozri napr. [2], str. 122) , pričom v strednej časti treba metematicky zabezpečiť vhodnú kombináciu oboch tvarov.

Teoretickým základom kvantilového modelovania je teória známa pod názvom Order statistics (viď. napr. [1]), ktorú prekladáme ako teória usporiadaných štatistík. Dodržiavaním

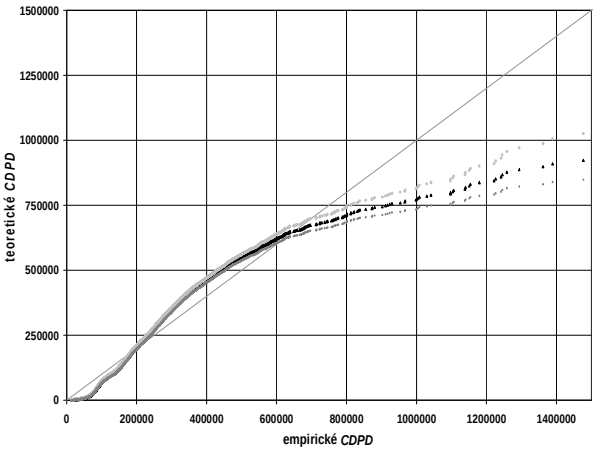
základných pravidiel modifikácie kvantilových funkcií a konštrukcie zložitých kvantilových tvarov (pravidlo reflexie, o sčítaní, o stredovom umiestnení, o násobení, U-transformácii, p-transformácii (pozri [3], kap. 3, alebo [2], str.131) do fázy estimácie boli vybrané tvary Weibullov-Paretov tvar, Gamma-Paretov tvar, Power-Paretov sučinový tvar, záporný Power-Pareto súčtový tvar a logistický zošikmený tvar (tvary kvantilových distribučných modelov pozri [3], kapitola 3.2.2, alebo [2], str. 130).

V súčtových tvaroch boli v procese estimácií použité jednoduché váhy (1-*p*) pre dolný koviec a *p* pre horný koniec, kde *p* je pravdepodobnosť rozdelenia, ako aj tvary váh pre rýchlejší prechod v tvare $p^2(3-2p)$ a ešte rýchlejší v tvare $p^3(10-15p+6p^2)$.

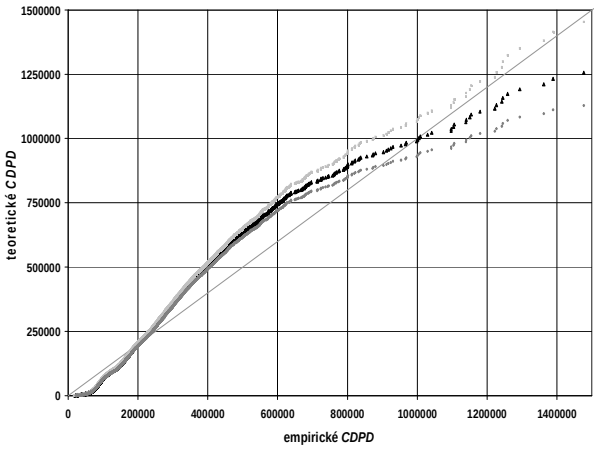
Parametre sme odhadli pre všetky uvedené tvary s rôzne volenými váhami špecifickými metódami estimácie kvantilových tvarov, metódou minimalizácie štvorcov distribučných rezíduí a metódou minimalizácie absolútnych distribučných rezíduí. Obe (viď. [2], str. 198, alebo [3], kap. 4.3) boli volené preto, že sa vhodne dopĺňajú, každá z nich zohľadňuje iné vlastnosti rozdelenia v estimácii. Výpočtovo jednoduchšie metódy ako metóda momentov a lebo kvantilov neprinášali vhodné tvary zložitejších asymetrických rozdelení. Prvá aplikovaná, s umocňovaním rezíduí definovaných cez rankit (viď teória usporiadaných štatistík v prílohe B v [3]), konkrétne ich aproximáciou podľa Öztürkovej-Daleovej metódy estimácie (viď. [3], kap. 4.3), v minimalizačných procedúrach zabezpečila lepšiu zhodu v dlhších koncoch. Príkladom je gamma-Paretov tvar s jednoduchými váhami (1-*p*) pre gamma a *p* pre Paretovo rozdelenie, ktoré neumožnili dobrú zhodu s empirickým rozdelením ale dokumentujú dôsledok použitia dvoch rôznych metód estimácie (viď. graf 1 a graf 2).

Graf 1: Q-Q graf gamma-Paretovho kvantilového rozdelenia *CDPD* (v Sk) odhadnutého metódou minimalizácie absolútnych distribučných rezíduí s hranicami 0,95 intervalu spoľahlivosti

Graf 1



Graf 2



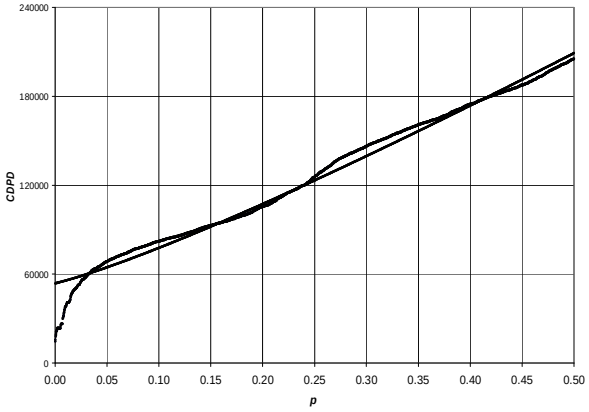
Graf 2: Q-Q graf gamma-Paretovho kvantilového rozdelenia CDPD (v Sk) odhadnutého

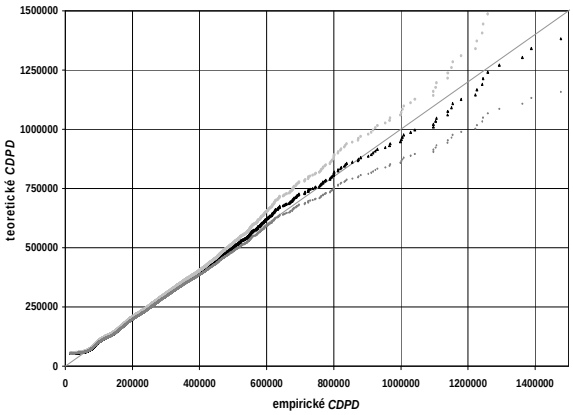
metódou absolútnych distribučných rezíduí s hranicami 0,95 intervalu spoľahlivosti

Druhá aplikovaná metóda je róbustnejšia a univerzálnejšia v kvantilovom modelovaní, nebola však prínosom. Vo všeobecnosti prináša lepšiu zhodu strednej časti rozdelenia a preto veľká asymetria, spôsobená dlhým pravým koncom empirického rozdelenia, nemôže byť vhodne vystihnutá (viď. graf 2).

Pretože váhy volené ako konštantné funkcie *p* nepriniesli akceptovateľný tvar, zvolený bol tvar ω*(1-*p*) a κ**p*, kde ω a κ vstúpili do modelov ako ich parametre. Získaný bol tak šesťparametrický kvantilový gamma-Paretov tvar (viď. graf 5).

Graf 3: Dolná časť kvantilovej funkcie a Q-Q graf päť-parametrického gamma-Paretovho kvantilového rozdelenia CDPD (v Sk), obidva s hranicami 0,95 intervalu spoľahlivosti.





Estimácie sme urobili pomocou procedúr SOLVER-u, tabuľkového programového balíka EXCEL. Použitie EXCEL-u vyplynulo aj z toho, že doterajšie štatistické programové balíky zatiaľ neobsahujú potrebné procedúry pre kvantilové modelovanie a iné simplexové minimalizačné procedúry neboli dostupné. Predpokladáme však, že apikáciou optimalizačných metód na iných (napr. Newtonových, Nelderovo-Meadových,...) základoch by mohla byť dosiahnutá ešte lepšia zhoda.

Rôzne grafické a výpočtové verifikácie odhadnutých tvarov potvrdili, že Weibullov, a lognormálny tvar nie sú vhodným riešením pre rozdelenie *CDPD*, väčšia špicatosť gamma rozdelenia vystihla dolnú časť rozdelenia príjmov vhodnejšie. Ani váhy volené konštantnou funkciou pravdepodobnosti neprinášali uspokojujúce výsledky.

Najvhodnejší odhadnutý tvar rozdelenia *CDPD* je nasledovný:

$$Q(p) = \alpha + \left(\omega (1-p)\,\mathrm{GAMAINV}\,(p;\beta.\gamma) + \kappa p \Big/ (1-p)^{\delta} \right),\ 0 < p < 1|$$

kde vektor parametrov (α, ω, β, γ, κ, σ) = (53876.471015681200; 94517.301607597000; 0.689668672671; 3.189465646509; 148533.621042; 0.362093)

Model vhodne vystihuje celkový tvar, ako aj horný koniec rozdelenia príjmov, v dolnej časti sa nim však nepodarilo vhodne modelovať rýchly pokles príjmov najchudobnejších 3 % domácností. Zvlnenie empirického rozdelenia v dolnej časti by mohlo byť napr. dôsledkom vlasností rozdelení v podsúboroch, stanovených hraníc pre odvod daní či hraníc výšky sociálnych dávok a pod. Nemyslíme si, že by bolo užitočné hľadať tvar modelu, ktorý by vystihoval aj tieto zvlnenia, skôr by malo nadväzovať modelovanie vo vhodne volených štruktúrach súboru.

Odhadnuté poskladané tvary kvantilových modelov majú lepšiu zhodu v odlišných častiach pravdepodobnostného rozdelenia v závislosti od použitých metód odhadu. Vhodný tvar a metódu jeho estimácie je treba voliť aj podľa účelu použitia modelu. Naviac možno niektoré časti odhadnúť s väčšou presnosťou metódami váženia vstupných údajov. Dobrý model je model účelne jednoduchý s dostatočnou zhodou podľa jeho použitia. Za súčasných technických možností a hlavne komplexných zložitých ekonomicko-sociálnych javov už viacparametrický, t. j. model s štyrmi až šiestimi parametrami, nie je zvláštnosťou.

Dovolíme si tvrdiť, že je možné kombinovaním vhodných jednoduchých tvarov pre časti rozdelenia a ich vhodným vážením nájsť viacvrcholové komplikované kvantilové tvary modelov, ale s nedostatkom veľkého počtu parametrov ich konečného tvaru. Otázkou ostáva kedy nájdený tvar je ešte modelom populácie a kedy začína byť už modelom náhodného výberu s prispôsobovaním sa možno náhodným výkyvom empirického tvaru. Nájdený štatisticky významný kvantilový model rozdelenia príjmov domácností SR umožňuje úplne charakterizovať rozdelenie celkových disponibilných príjmov domácností za celú SR v roku 2004 a môže byť východiskom k modelovaniu v štruktúrach a v čase.

Literatúra:

[1] ARNOLD, B.C., BALAKRISHNAN, N., and NAGARAJA,H.N.: A First Course in Order Statistics, John Wiley and Sons, New York, 1992

[2] GILCHRIST, W.G.: *Statistical modelling with quantile functions*, Chapman & Hall, 2000

[3] SIPKOVÁ, Ľ. – SODOMOVÁ, E.: Modelovanie kvantilovými funkciami, Vydavateľstvo Ekonóm, Bratislava, (v tlači pre rok 2007)

[4] SIPKOVÁ, Ľ.: Kvantilové modelovanie rozdelenia príjmov domácností: dizertačná práca, školiteľ Eva Sodomová, KŠ FHI, Ekonomická univerzita Bratislava, 2004

Kontakt : Ing. Ľubica Sipková, PhD.
Katedra Štatistiky, FHI, Ekonomická univerzita v Bratislave
tel: 02/67295717, e-mail: lubica.sipkova@euba.sk

Prispevok je súčasťou riešenia projektu VEGA č. 1/4586/07 pod názvom „Modelovanie sociálnej situácie obyvateľstva a domácností v SR a jej regionálne a medzinárodné porovnania“.

Odhad cenovej kategórie ojazdených motorových vozidiel pomocou rozhodovacích stromov v SASe

Stankovičová Iveta, Žambochová Marta

Abstract: Decision trees are one of the most popular methods of predictive modeling for data mining purposes because they provide interpretable rules and logic statements that enable more intelligent decision making. A decision tree partitions data into smaller segments called terminal nodes or leaves that are homogeneous with respect to a target variable. Partitions are defined in terms of other variables called input variables, thereby defining a predictive relationship between the inputs and the target. This partitioning continues until the subsets cannot be partitioned any further using user-defined stopping criteria. Decision trees are just one of the advanced analysis models included in SAS Enterprise Miner.

Key words: *Decision trees, classification and regression trees, nodes, leafs, rules, SAS Enterprise Miner, CHAID and CART algorithms.*

1. Úvod

Rozhodovacie stromy sú vhodným a obľúbeným nástrojom na klasifikáciu a predikciu. Atraktívnosť týchto metód je vo veľkej miere založená na tom, že rozhodovacie stromy predstavujú pravidlá. Pravidlá sa dajú vyjadriť v anglickom alebo slovenskom jazyku a tak im ľudia ľahko rozumejú. Môžu sa tiež vyjadriť v databázovom jazyku ako je SQL, ktorý umožňuje vyhľadávať záznamy príslušnej kategórie v databáze.

Termín *„rozhodovací strom“* je však zavádzajúci. Autori stromového riešenia v 60-tych rokoch nazvali pravidlá delenia (*splitting rules*) ako rozhodovacie pravidlá (*decision rules*) a táto terminológia pretrvala a rozšírila sa dodnes. Je to však na úkor použitia myšlienok a terminológie z teórie rozhodovania (*decision theory*), ktorá sa neobmedzuje len na analýzu dát, ale jej obsah je širší. V článku sa budeme zaoberať len rozhodovacími stromami z oblasti analýzy dát, a preto pojem „*rozhodovací strom*“ by bolo lepšie nahradiť pojmom „*strom delenia dát*“, čo je výstižnejšie. V praxi sa však používa pojem „*rozhodovací strom*“.

2. Algorimy vytvárania rozhodovacích stromov

Rozhodovací strom je štruktúra, ktorá rekurzívne rozdeľuje skúmané údaje podľa určitých rozhodovacích kritérií. Koreň stromu (*root*) reprezentuje celý dátový súbor. Vnútorné uzly stromu reprezentujú podmnožiny tohto skúmaného súboru. V listoch stromu (*leafs*), t.j. v konečných uzloch stromu, ktoré sa už ďalej nedelia, môžeme zistiť hodnoty vysvetľovanej premennej. Jednotlivé algoritmy vytvárania rozhodovacích stromov sa líšia v závislosti od nasledovných charakteristík:

1. Typ vysvetľovanej, čiže modelovanej premennej (*angl. target*), od ktorej závisí opis obsahu jednotlivých listov stromu. Stromy v zásade delíme na:
    - *Klasifikačné stromy:* Ak závislá (modelovaná) premenná je kategoriálneho typu, tak v každom liste stromu môžeme zistiť bodový odhad pravdepodobnosti priradenia do modelovanej kategórie.

- *Regresné stromy:* Ak závislá (modelovaná) premenná je spojitá (intervalová), tak v každom liste môžeme zistiť priradenie číselnej konštanty (aritmetického priemeru a štandardnej odchýlky), čiže odhad hodnoty závislej premennej.

2. Typ vysvetľujúcich premenných (vstupov, prediktorov, *angl. indicators, inputs, regressors, independent variables*):
   - nominálne,
   - ordinálne,
   - spojité (intervalové).
3. Pravidlo delenia (*splitting rule*).
4. Kritérium pre zastavenie delenia (*stopping rule*).
5. Typ podmienok vetvenia:
   - multivariantné (testuje sa niekoľko prediktorov),
   - univariantné (v danom kroku sa testuje len jeden z prediktorov).
6. Spôsob vetvenia uzlov stromu:
   - binárne (každý vnútorný uzol, sa delí na dva následné uzly, čiže na dve vetvy),
   - *k*-árne (vnútorné uzly sa delia na viac ako dve vetvy).

Pre vytváranie rozhodovacích stromov bolo vyvinutých viacero algoritmov. Historicky najčastejšie používané algoritmy v praxi, ktoré sú, resp. boli súčasťou aj softvérových programov, sú: AID, SEARCH, CHAID, CART, ID3, C4.5, C5, QUEST a OCI. Autormi týchto algoritmov je jeden človek alebo skupinka ľudí, ktorí sa snažili vylepšiť predtým používaný algoritmus delenia stromu.

3. Algorimus CHAID

Ako prvý bol vynájdený a do formy programu zaznamenaný algoritmus AID (*Automatic Interaction Detection*) v roku 1963. Jeho autormi sú James Morgan a John Sonquist. Tretia verzia programu tohto algoritmu sa volá SEARCH a používala sa v období rokov 1971-1996. AID hľadá binárne delenia na základe nominálnych a ordinálnych vstupov (prediktorov) tak, aby sa čo najviac zredukovala suma štvorcov odchýlok od priemeru intervalovej modelovanej (*target*) premennej. AID sa pokúša ako prvé rozdeliť uzly, v ktorých je najväčšia suma štvorcov odchýlok (*sum-of-square errors*). Program sa zastaví, keď redukcia sumy štvorcov odchýlok je nižšia ako nastavená konštanta vynásobená celkovou sumou štvorcov odchýlok. Nastavená konštanta v systéme SAS je 0,006. Delenie môže byť zastavené aj tak, že užívateľ určí počet prvkov v uzle.

Algoritmus CHAID (*Chi-Squared Automatic Interaction Detection*) predstavuje zlepšenie algoritmu AID pre nominálnu vysvetľovanú (*target*) premennú. Autorom je Gordan Kass (1980) z Južnej Afriky. CHAID rekurzívne delí údaje s nominálnou modelovanou premennou na základe nominálnych a ordinálnych prediktorov. Výsledkom sú nebinárne stromy. Metóda využíva χ^2- test na delenie dát a na určenie počtu vetiev stromu. Z dôvodu obavy pred časovou náročnosťou pôvodného algoritmu AID, autor hľadá už len suboptimálne delenie namiesto prehľadávania všetkých možných delení a hľadania optimálneho delenia.

Algoritmus delenia v CHAID je nasledovný: V rámci jedného listu sa vytvorí kontingenčná tabuľka pre *m* hodnôt prediktora (*m* kategórií) a *k* kategórií závislej premennej, čiže rozmeru (*m* x *k*). Ďalej sa nájde dvojica kategórií prediktora, pre ktoré má subtabuľka rozmerov (*2* x *k*) najmenej významnú hodnotu χ^2- testu. Tieto dve kategórie sa zlúčia. Týmto nám vzniká nová kontingenčná tabuľka o rozmeroch [(*m-1*) x *k*]. Proces zlučovania sa opakuje až kým neklesne významnosť χ^2- testu pod dopredu zadanú hodnotu. Týmto je ukončený proces delenia jedného rodičovského uzla na niekoľko uzlov - potomkov. Ďalej sa pokračuje podobne pre každý listový uzol až do doby nevýznamného výsledku χ^2- testu.

CHAID používa Bonferroniho úpravu pre počet kategórií prediktorov, čím sa zníži skreslenie voči vstupom, ktoré majú veľa hodnôt (kategórií). S chýbajúcimi hodnotami (*missing values*) pracuje ako so špeciálnymi dodatočnými hodnotami.

Treba tiež poznamenať, že CHAID algoritmus nevyhodnocuje najviac významné delenie. Začiatkom 90-tych rokov minulého storočia prišiel Barry de Ville s algoritmom „vyčerpávajúceho" (*exhaustive*) CHAID stromu, ktorý používa najviac významné delenie. Tento algoritmus vytvára delenia s väčším počtom vetiev ako pôvodný CHAID algoritmus.

4. Klasifikačné a regresné stromy - algorimus CART

V roku 1984 bola publikovaná kniha a komerčný softvér od autorov **B**reiman Leo, **F**riedman Jerome, **O**lshen Richard a **S**tone Charles, na ktorom začali pracovať už začiatkom 70-tych rokov (*Pozn.:* Skratka algoritmu je tiež BFOS od iniciálok priezvisk autorov.). Ich program pre algoritmus CART (*Classification and Regression Trees*) tvorí binárne stromy na základe nominálnych a intervalových vstupov pre nominálnu, ordinálnu alebo intervalovú modelovanú premennú.

V každom kroku CART algoritmus robí vyčerpávajúce hľadanie najlepšieho delenia na základe všetkých vstupov. Meradlom pre najlepšie delenie je zvýšenie čistoty dát. To znamená, že jedno delenie je lepšie ako druhé, ak získame dva homogénnejšie súbory dát ako by sme získali pomocou iného delenia. Meradlo delenia je rôzne pre klasifikačné a pre regresné stromy.

V regresných stromoch, ktoré modelujú intervalovú premennú, ako meradlo delenia sa používa zníženie variability dát na základe redukcie chyby štvorcov alebo redukcie priemernej absolútnej odchýlky od mediánu. V klasifikačných stromoch, ktoré modelujú nominálnu premennú, sa ako meradlo delenia používa redukcia Giniho indexu alebo tzv. „*twoing*" pravidla. V klasifikačných stromoch, ktoré modelujú ordinálnu premennú, sa ako meradlo delenia používa redukcia tzv. poradového „*twoing*" pravidla („*ordering twoing*").

5. SAS algoritmy pre rozhodovacie stromy

V systéme SAS sú algoritmy vytvárania rozhodovacích stromov zabudované do nástroja SAS® Enterprise Miner (ďalej len SAS EM). SAS EM predstavuje vizuálne programovacie prostredie pre prediktívne modelovanie, v ktorom sa vytvárajú procesné diagramy zložené z uzlov (*nodes*). Uzol rozhodovacieho stromu (*Decision Tree Node*) má v sebe zabudované a ponúka tie najlepšie myšlienky z rekurzívneho delenia na základe jednej premennej, ktoré boli historicky vyvinuté a zakomponované do vyššie spomínaných algoritmov rozhodovacích stromov. SAS implementácia rozhodovacích stromov hľadá najlepšie delenia dát na základe nominálnych, ordinálnych a intervalových vstupov. Užívateľ si vyberá deliace kritériá a iné voľby, ktoré determinujú metódu konštrukcie stromu. Pomocou nastavenia volieb môžeme dosiahnuť, aby rozhodovací strom mal vlastnosti populárneho CHAID algoritmu alebo tiež algoritmu BFOS (resp. CART).

Ako modelovaná premenná môže byť v SAS EM použitá nominálna, poradová alebo intervalová premenná. Podľa typu modelovanej premennej sa bude automaticky tvoriť buď klasifikačný alebo regresný strom. V SAS EM sa automaticky zmenia nastavené (*default*) hodnoty (*Obrázok 1*). Napríklad ako nastavená (*default*) hodnota pre kritérium delenia (*Splitting Rule Criterion*) je pre nominálnu modelovanú premennú *ProbChisq*, pre intervalovú *ProbF* a pre ordinálnu *Entropy*.

Užívateľ určí maximálny počet vetiev pri delení, čiže či chce binárny strom, košatý strom, alebo niečo medzi tým. Delenia sú hodnotené buď na základe zníženia nečistoty dát (Použité sú nasledovné miery: chyba štvorcov (*Variance*), Giniho index (*Gini*) alebo entropia (*Entropy*)), alebo na základe štatistickej významnosti testu (Použité sú *p*-hodnoty pre nasledovné dva štatistické testy: χ^2- test (*ProbChisq*) a *F*-test (*ProbF*)). Pri testoch

významnosti sa povoľuje aj Bonferroniho úprava (ako to robí CHAID), aby sa čelilo skresleniu voči vstupom s veľkým počtom kategórií. V uzle rozhodovacieho stromu je možné urobiť aj iné zmeny v nastaveniach (napr. nastaviť hĺbku stromu a minimálnu početnosť listov, práca s chýbajúcimi hodnotami (*Missing values*) a s počtom vstupov a pod.).

SAS algoritmus v uzle rozhodovacieho stromu je iný ako originálny CHAID algoritmus od G. Kassa. Približný CHAID algoritmus v SAS EM môžeme dosiahnuť nasledovným postupom:

- Intervalové prediktory musíme najskôr zdiskretizovať do niekoľkých intervalov.
- Pre nominálnu modelovanú premennú musíme nastaviť *Splitting Rule Criterion* na *ProbChisq* a pre intervalovú modelovanú premennú zase na *ProbF*. Musíme tiež nastaviť hladinu významnosti pre tieto testy z hodnoty 0.2 (*default*) na 0.05
- Musíme zrušiť automatické orezávanie stromu (*pruning*) tak, že voľbu *Subtree Method* nastavíme z *Assessment* na *Largest*.
- Nastavíme maximálny počet vetiev (*Maximum Branch*) stromu na maximálny počet hodnôt kategórií prediktorov. (Default je 2 a tak sa tvoria len binárne stromy, pričom CHAID má vytvárať nebinárne stromy.)
- Necháme voľbu pre počet náhradných pravidiel (*Number of Surrogate rules*) na čísle 0 (default hodnota je 0).
- Aby sme zabezpečili heuristické hľadanie delenia stromu, tak nastavíme voľbu *Exhaustive* na číslo 0 (default hodnota je 5000).
- Nastavíme voľbu *Bonferroni Adjustment* na *Y* (=Yes) a *Time of Kass Adjustment* na *After* (potom).

CHAID algoritmus sa nedá aplikovať na ordinálnu modelovanú premennú v uzle rozhodovacieho stromu implementovaného v SAS EM.

Algoritmus CART (resp. BFOS metóda) sa tiež dá aplikovať pomocou uzla rozhodovacieho stromu v SAS EM. Vyžaduje nominálne a intervalové prediktory. Ordinálne prediktory sú považované za intervalové. Výsledok je podobný BFOS metodológii avšak bez použitia lineárnej kombinácie delení, „*twoing*“ a „*ordered twoing*“ pravidla, pretože SAS EM ich neponúka. SAS EM neponúka ani krížové overenie (*cross-validation*) pre hodnotenie kvality modelu stromu, pretože sa predpokladá, že v data miningu ide o veľké súbory dát. Na zhodnotenie kvality modelu sa využíva delenie súboru dát na trénovaciu, validačnú, prípadne aj testovaciu množinu dát prostredníctvom uzla *Data Partition*, ktorý sa umiestni pred uzol modelu rozhodovacieho stromu v procesnom diagrame SAS EM.

Pre dosiahnutie CART algoritmu v SAS EM treba nastaviť nasledovné voľby:

- Pre nominálnu modelovanú premennú musíme nastaviť *Splitting Rule Criterion* na *Gini* a pre intervalovú modelovanú premennú zase na *Variance*.
- Ponecháme maximálny počet vetiev (*Maximum Branch*) stromu na čísle 2.
- Zmeníme voľbu *Missing Values* z *Use in search* na *Largest branch*, čím sa pozorovania z chýbajúcimi hodnotami priraďujú do najpočetnejších vetiev stromu.
- Nastavíme voľbu pre počet náhradných pravidiel (*Number of Surrogate rules*) na číslo 5 (default hodnota je 0).
- Ponecháme voľbu pre automatické orezávanie stromu (*pruning*) na *Assessment* (časť *Subtree*, položka *Method*), ale zmeníme položku *Assessment Measure* z voľby *Decision* na *Average Square Error*.
- Nastavíme hodnotu pre voľbu *Node sample* (časť *Split Search*) na číslo väčšie ako je *n* (= všetky pozorovania v tréningovom súbore dát). Default hodnota je 5000.

Ani CART algoritmus sa nedá aplikovať na ordinálnu modelovanú premennú v uzle rozhodovacieho stromu implementovaného v SAS EM.

Property	Value
Node ID	Tree
Imported Data	
Exported Data	
Variables	
Interactive	
Splitting Rule	
Criterion	Default
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	6
Minimum Categorical Size	5
Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	
Split Search	
Exhaustive	5000
Node Sample	5000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25
P-Value Adjustment	
Bonferroni Adjustment	Yes
Time of Kass Adjustment	Before
Inputs	No
Number of Inputs	1
Split Adjustment	Yes
Output Variables	

Obrázok 1: *Okno nastavení (default) pre uzol rozhodovacieho stromu v SAS Enterprise Miner 5.2 (časť)*

6. Výsledky modelov rozhodovacích stromov pre cenové kategórie ojazdených áut v SAS Enterprise Miner 5.2

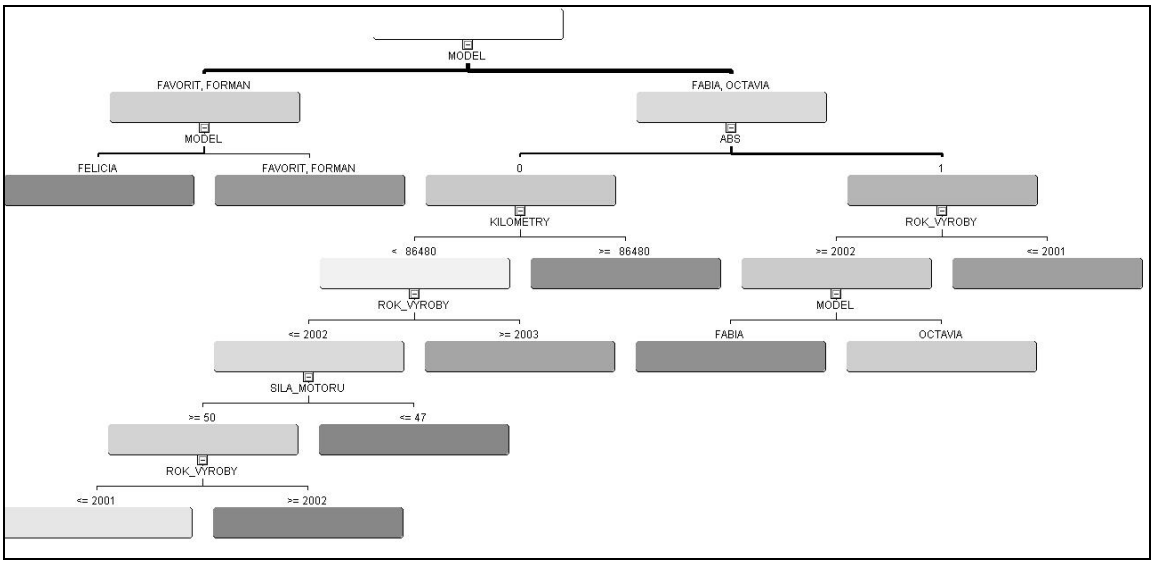
Na analýzu sme použili vzorku údajov o ojazdených motorových vozidlách značky Škoda vo vybratých autobazároch v ČR za máj 2007. Súbor obsahoval 498 pozorovaní a 21 premenných. Premenné obsahovali základné informácie potrebné pre kúpu a predaj auta, ako napr. model auta, typ karosérie, druh paliva, sila motora, rok výroby, najazdené km, výbava (airbag, ABS, ASR (t.j. typ protišmykového systému), elektrické otváranie okien, klimatizácia, atď.) a cena auta. Cena bola následne kategorizovaná do ordinálnej premennej so štyrmi cenovými kategóriami: 1. lacné autá (do 50 tis. Kč), 2. stredne drahé autá (50-150 tis. Kč), 3. drahšie autá (150-300 tis. Kč) a 4. drahé autá (nad 300 tis. Kč).

V prostredí SAS Enterprise Miner 5.2 sme sa pokúsili vytvoriť modely rozhodovacích stromov pre cenu ojazdených áut Škoda na základe ostatných vstupných informácií rôznymi algoritmami. Tvorili sme regresné stromy, kedy sme ako modelovanú premennú použili priamo cenu áut (CENA) a tvorili sme aj klasifikačné stromy, kedy sme ako modelovanú premennú použili premennú obsahujúcu nami vytvorené cenové kategórie áut (KATEGORIE). Na posúdenie kvality modelov sme použili rozdelenie vzorky dát na trénovaciu a validačnú časť v pomere 50:50 percent.

V prípade regresného stromu vytvoreného algoritmom CART (resp. metódou BFOS) sa ako významné premenné pre delenie stromu ukázali nasledujúce prediktory (zoradené podľa významnosti): sila motoru, ABS, elektrické okná, rok výroby, model, najazdené km, typ karosérie, ASR, typ paliva. Výsledný strom mal až 18 konečných listov, pričom niektoré

obsahovali len nízky počet pozorovaní (4 až 7). Kvalita modelu nebola uspokojivá, lebo boli rozdiely vo výsledkoch na trénovacej a validačnej vzorke dát.

V prípade klasifikačného stromu algoritmom CHAID aplikovaného na premennú KATEGORIE sme dosiahli lepšie výsledky. Ako významné prediktory sa ukázali (zoradené podľa významnosti): model, sila motoru, rok výroby, klimatizácia, najazdené km a ASR. Strom nebol binárny a vytvoril tiež pomerne veľa listov (24) a ich početnosť bola preto nízka. Percento chybnej klasifikácie (*Misclassifacation Rate*) bolo pomerne uspokojivé: pre trénovaciu vzorku dát 9,3% a pre validačnú 11,9%. Vytvorili sme preto binárny rozhodovací strom pomocou algoritmu SAS. Získali sme rozhodovací strom s 10-timi listami, v ktorom ako najvýznamnejšia premenná sa ukázal model auta Škoda. Do jednej vetvy sa zaradili lacnejšie modely Favorit, Forman a Felícia. Do druhej vetvy sa zaradili modely Fabia a Octavia, ktoré patria medzi drahšie modely značky Škoda. Druhá vetva sa potom ešte podrobnejšie delí na základe ďalších premenných ako sú: ABS, km, rok výroby a sila motoru (*Obrázok 2*). Percento chybnej klasifikácie (*Misclassifacation Rate*) bolo pomerne uspokojivé: pre trénovaciu vzorku dát 9,3% a pre validačnú 13,9%. Dosiahnutý kumulatívny lift (*Cumulative Lift*) pre všetky tieto modely bol však pomerne nízky, len okolo hodnoty 2,3.



Obrázok 2: *Binárny rozhodovací strom pre KATEGORIE pomocou SAS EM algoritmu*

Pokúsili sme sa zlepšiť kvalitu modelu rozhodovacieho stromu. Zmenili sme kategorizáciu ceny áut nasledovným spôsobom: 1. lacné autá (do 59000 Kč), 2. stredne drahé autá (59000-150100 Kč), 3. drahšie autá (150100-300250 Kč) a 4. drahé autá (nad 300250 Kč). Využili sme k tomu nástroje SAS EM, konkrétne uzol *Transform Variables*, ktorý zoptimalizoval kategorizáciu ceny pri delení do 4 intervalov. Získali sme iné hranice cenových intervalov, ako sme si pôvodne určili.

Pre takto vytvorenú modelovanú nominálnu premennú OPT_CENA_KAT sme vytvorili opäť niekoľko modelov rozhodovacích stromov. Použili sme algoritmus CART, CHAID a aj nastavený (*default*) SAS algoritmus. Získali sme rozhodovacie stromy s kumulatívnymi liftovými hodnotami (*Cumulative Lift*) vyššími ako 8. Ako najlepší z hľadiska interpretácie sa javí binárny strom získaný na základe SAS algoritmu, ktorý má len 5 listov (*Tabuľka 1*). Ako významné prediktory sa prejavili len štyri premenné: ABS, model, sila motoru, ASR. Na základe výsledných rozhodovacích pravidiel je zrejmé, že modely áut Favorit a Forman patria do kategórie lacných ojazdených áut (odhadovaná pravdepodobnosť je až 94,7%, pozri NODE 2 v *Tabuľka 1*). Medzi najdrahšie ojazdené automobily Škoda patria modely Fabia a Octavia s ASR, so silou motora 74 koní a ktoré sú vybavené ABS (odhadovaná pravdepodobnosť je 85,7%, pozri NODE 13 v *Tabuľka 1*).

Tabuľka 1: *Rozhodovacie pravidlá (English Rules) pre výsledný rozhodovací strom pre OPT_CENA_KAT pomocou SAS EM algoritmu*

IF MODEL IS ONE OF: FAVORIT FORMAN THEN NODE : 2 N : 38 05:4-59000: 94.7% 06:59000: 5.3% 07:150100: 0.0% 08:300250: 0.0%	IF ABS EQUALS 0 AND MODEL IS ONE OF: FABIA OCTAVIA FELICIA THEN NODE : 4 N : 115 05:4-59000: 1.7% 06:59000: 84.3% 07:150100: 13.0% 08:300250: 0.9%	IF SILA_MOTORU IS ONE OF: 55 66 81 50 47 AND ABS EQUALS 1 AND MODEL IS ONE OF: FABIA OCTAVIA FELICIA THEN NODE : 9 N : 71 05:4-59000: 0.0% 06:59000: 0.0% 07:150100: 88.7% 08:300250: 11.3%
IF ASR EQUALS 0 AND SILA_MOTORU EQUALS 74 AND ABS EQUALS 1 AND MODEL IS ONE OF: FABIA OCTAVIA FELICIA THEN NODE : 12 N : 15 05:4-59000: 0.0% 06:59000: 33.3% 07:150100: 60.0% 08:300250: 6.7%	IF ASR EQUALS 1 AND SILA_MOTORU EQUALS 74 AND ABS EQUALS 1 AND MODEL IS ONE OF: FABIA OCTAVIA FELICIA THEN NODE : 13 N : 7 05:4-59000: 0.0% 06:59000: 0.0% 07:150100: 14.3% 08:300250: 85.7%	

7. Záver

Cieľom príspevku bolo ukázať, že rôzne algoritmy na vytváranie rozhodovacích stromov vedú aj k rôznym výsledkom. Výsledok závisí aj od charakteru modelovanej premennej. V mnohých prípadoch je potrebné spojitú (intervalovú) premennú vhodne kategorizovať, aby bol výsledok modelovania prijateľný z hľadiska kvality modelu a aj interpretovateľnosti tohto modelu.

V prostredí SAS Enterprise Miner sú implementované tie najlepšie myšlienky z procesu vývoja algoritmov pre tvorbu rozhodovacích stromov. V tomto prostredí sa dajú dosiahnuť aj klasické algoritmy pre delenie rozhodovacích stromov, len je potrebné zmeniť niektoré nastavenia.

8. Literatúra

PADRAIC G. NEVILLE: Decision Trees for Predictive Modeling. SAS Institute Inc., 4 August 1999. Dostupné na internete: http://bus.utk.edu/stat/DataMining/Decision%20Trees%20for%20Predictive%20Modeling%20(Neville).pdf

SAS OnLine Doc 9.1.3. Dostupné na internete: http://support.sas.com/onlinedoc/913/docMainpage.jsp

STANKOVIČOVÁ, IVETA: Rozhodovacie stromy v marketingových analýzach. In: Nová ekonomika. Roč. 4, č. 1 (2005), s. 105-111

ŽAMBOCHOVÁ, MARTA: Odhad cen ojetých vozů pomocí rozhodovacích stromů. In: Sborník příspěvkú z Mezinárodní statisticko-ekonomické dny na VŠE v Praze, září 2007

Adresy autoriek:

Ing. Iveta Stankovičová, PhD.
Fakulta managementu UK v Bratislave
Katedra informačných systémov
Odbojárov 10
820 05 Bratislava, SR
iveta.stankovicova@fm.uniba.sk

RNDr. Marta Žambochová
Univerzita J.E. Purkyně, Fakulta sociálně ekonomická, Katedra matematiky a statistiky
Moskevská 54
400 96 Ústí nad Labem, ČR
zambochova@fse.ujep.cz

Využitie štatistického systému Stagraphics pre modelovanie poistných škôd

Alena Tartaľová

Úvod

Poisťovňa potrebuje na riešenie zásadných otázok svojej činnosti poznať predovšetkým zákony rozdelenia pravdepodobnosti počtu poistných plnení a výšky individuálnych poistných plnení pri rôznych typoch poistenia. Znalosť rozdelenia pravdepodobnosti výšky škôd je nutnou podmienkou riešenia všetkých aktuárskych problémov, napr. modelu kolektívneho rizika, zaistenia, simulačných úloh a pod. Pre praktickú aplikáciu pravdepodobnostného rozdelenia je potrebné odhadnúť parametre rozdelenia a otestovať jeho zhodu s empirickým rozdelením.

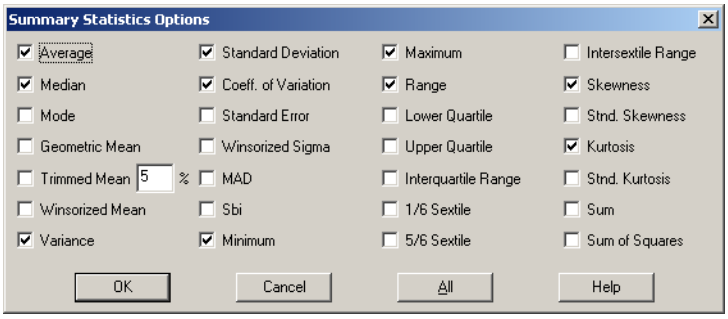
Modelovanie výšky škôd

Modelom výšky škôd pri určitom type poistenia rozumieme pravdepodobnostné rozdelenie individuálnych poistných plnení, ktorého typ určíme na základe známych údajov z poisťovne metódami štatistickej indukcie.

Všeobecný postup pri výbere vhodného rozdelenia výšky poistných plnení môžeme zhrnúť do troch krokov:

1. Navrhnúť predpokladaný typ rozdelenia na základe grafických metód
2. Odhadnúť parametre vybraného rozdelenia
3. Overiť vhodnosť vybraného rozdelenia pomocou testov dobrej zhody na základe výberových údajov

V praktickej analýze použijeme reálne údaje o výške 91 škôd pri havarijnom poistení motorových vozidiel. Najskôr potrebujeme mať konkrétnu predstavu o empirickom rozdelení výberových údajov. Získame ju na základe histogramu a výberových charakteristík. Program STATGRAPHICS Centurion XV. v procedúre *Analyze – One Variable Analysis* ponúka nasledovné možnosti pre výberové charakteristiky (pozri Obr.1), výstupom procedúry je Tab.1.



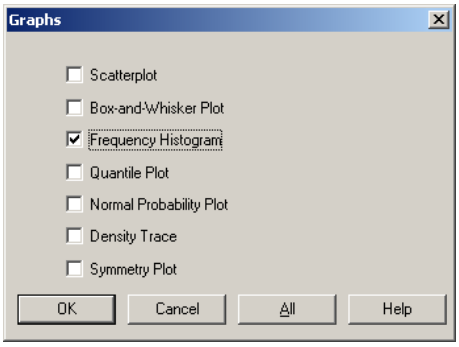
Obr.1: ***Možnosti pre výberové charakteristiky***

Summary Statistics for x

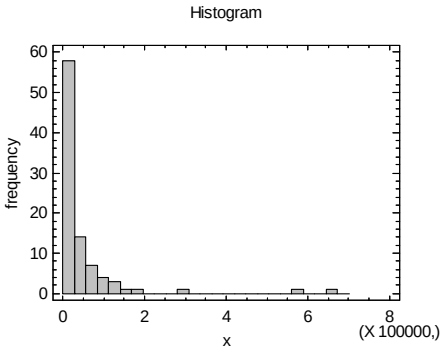
Count	91
Average	47111,2
Median	18900,0
Variance	9 417 548 306
Standard deviation	97044,1
Coeff. of variation	205,989%
Minimum	501,0
Maximum	648748,
Range	648247,
Skewness	4,78349
Kurtosis	25,6725

Tab.1: ***Vypočítané výberové charakteristiky***

Grafické nástroje, ktoré ponúka tento program sú:



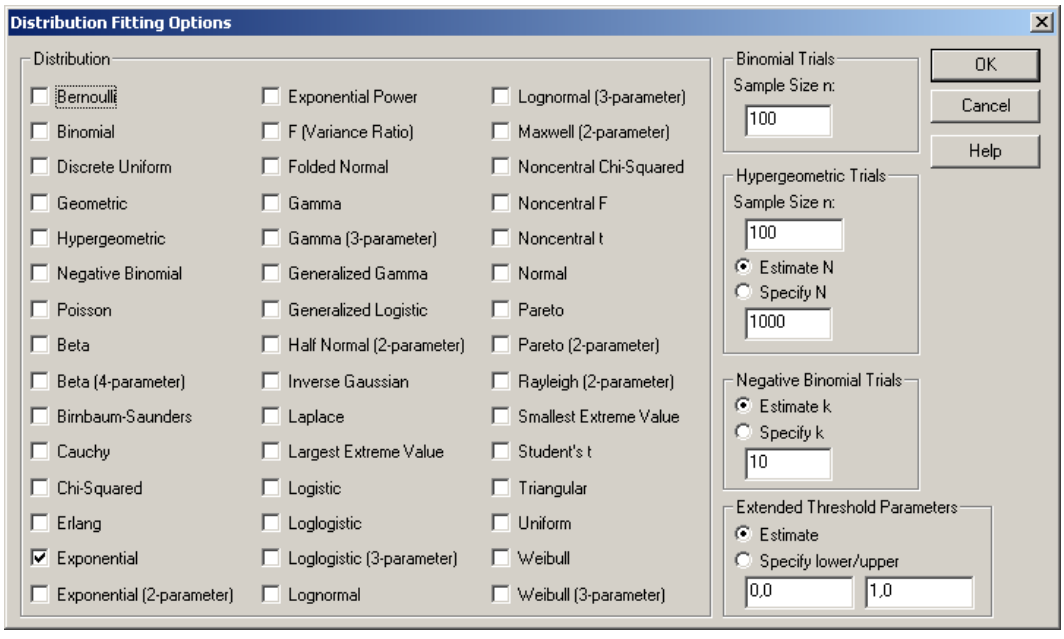
Obr.2: ***Grafické nástroje***



Obr.3: ***Histogram rozdelenia poistných škôd***

Základné informácie o rozdelení výšky poistných plnení môžeme na základe histogramu a výberových charakteristík zhrnúť takto: Väčšina poistných plnení má nízke hodnoty, ktoré majú veľkú pravdepodobnosť vzniku. Vyskytujú sa však aj extrémne hodnoty, ktoré majú relatívne malú pravdepodobnosť. To spôsobuje veľkú variabilitu poistných plnení - variačný koeficient výberového súboru má hodnotu až 205,989%. Rozdelenie je pravostranne (pozitívne) zošikmené, čo potvrdzuje hodnota koeficientu šikmosti (Skewness=4,78349). Vhodným modelom výšky poistných plnení môže byť niektoré pravostranne zošikmené rozdelenie, tzv. rozdelenie s tučným chvostom, napr. lognormálne, weibullovo, gamma, paretovo a pod.

Pri modelovaní výšky škôd je veľmi výhodné využiť procedúru *Distribution fitting* z ponuky *Analyze-Variable Data* programu STATGRAPHICS Centurion XV. Táto procedúra umožňuje nájsť pre empirické údaje vhodný model teoretického rozdelenia. Štatistický analytický systém STATGRAPHICS Centurion XV. má vo svojej ponuke 45 rozdelení, nás pritom zaujímajú spojité, pravostranne zošikmené rozdelenia:



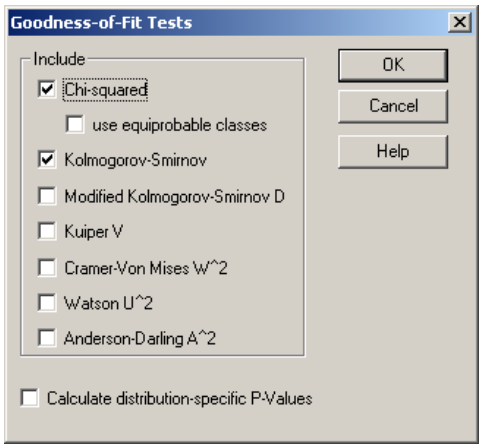
Obr. 4.:*Možnosti výberu teoretického rozdelenia*

Vhodnosť zvoleného teoretického rozdelenia overíme pomocou testov dobrej zhody.

Testy dobrej zhody

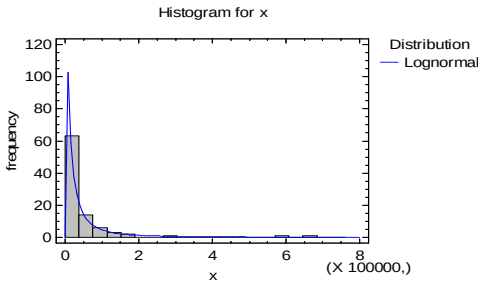
V ponuke procedúry *Goodness-of-Fit Tests* pomocou 7 testov dobrej zhody môžeme overiť, či zvolené pravdepodobnostné rozdelenie dobre modeluje histogram empirických údajov. V porovnaní so staršou verziou Statgraphics Plus, v novej verzii STATGRAPHICS Centurion XV. Pribudli k χ^2-testu a Kolmogorovovmu-Smirnovovmu testu aj nasledujúce testy dobrej zhody:

- Kuiper V
- Cramer-von Mises W^2
- Watson U^2
- Anderson-Darling A^2

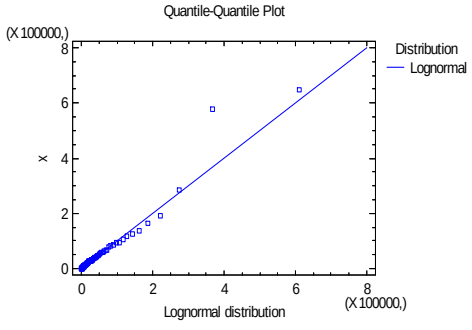


Obr. 5:*Možnosti výberu testov dobrej zhody*

Postup testovania χ^2-testu a Kolmogorovovho-Smirnovovho testu je podrobne popísaný v [5]. Naše porovnanie uskutočníme na základe týchto dvoch testov. Kľúčom k určovaniu bude p-hodnota. Výsledky χ^2-testu a Kolmogorovovho-Smirnovovho testu potvrdili zhodu pre *gama*, *lognormálne* a *weibullovo* rozdelenie. Vo všetkých prípadoch bola p-hodnota vyššia ako zvolená hladina významnosti 0,05. Najvhodnejším modelom je lognormálne rozdelenie s odhadnutými maximálne vierohodnými parametrami mean = 50171,2 a standard deviation = 139352,0. Pre toto rozdelenie dávajú obidva testy najvyššiu p-hodnotu, ktorá sa pri χ^2-teste rovná 0,94983 a pri K-S teste 0,993747. Výraznú zhodu s lognormálnym rozdelením potvrdzuje grafické porovnanie histogramu, na Obr.6 a kvantilový graf na Obr.7.



Obr. 6:*Histogram a lognormálne rozdelenie*

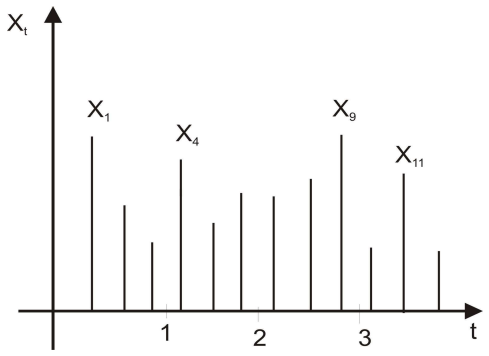


Obr. 7:*Kvantilový graf (QQ-plot)*

Ako sme už spomínali, vyskytujú sa tu aj extrémne vysoké poistné plnenia. Lognormálne rozdelenie nemodeluje dobre najväčšie hodnoty na pravom chvoste rozdelenia. Práve tieto hodnoty sú pre poisťovňu obzvlášť dôležité, pretože majú závažný finančný dopad. Teória extrémnych hodnôt je veľmi dobre spracovaná, napr. v monografiách [1] a [3]. Uvažujme postupnosť nezávislých, identicky rozdelených náhodných veličín, $X_1, X_2, \ldots, X_n$,ktoré predstavujú individuálne poistné plnenia. Pre registráciu extrémnych hodnôt sa najčastejšie využívajú nasledujúce metódy:

1. Metóda blokového maxima

V tejto metóde sa za extrémne hodnoty považujú maximá, ktoré sa nadobúdajú v za sebou nasledujúcich obdobiach rovnakej dĺžky, napr. v mesiacoch alebo rokoch. Potom $M_n = \max\{X_1, \ldots, X_n\}$ je blokové maximum (napr. ročné, mesačné a pod.). Táto metóda je založená na Fisher-Tippetovej vete, ktorá hovorí, že limitným rozdelením pre maximum náhodných veličín je Fréchetovo, Weibullovo alebo Gumbelovo rozdelenie.



Obr. 8: ***Metóda blokového maxima: X_1, X_4, X_9, X_{11} sú blokové maximá***

Veta 1 (Fisher-Tippet):

Nech $\{X_n\}_n$ je postupnosť nezávislých, identicky rozdelených náhodných veličín. Ak existujú normujúce konštanty $a_n \in R$, $b_n > 0$ a nedegenerovaná distribučná funkcia *H* taka, že platí:

$$b_n^{-1}(M_n - a_n) \xrightarrow{d} X, \text{ pre } n \to \infty.$$

Potom H prislúcha jednému z troch typov distribučných funkcií:

1. Fréchet: $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x > 0$, $\alpha > 0$ (inak $\Phi_\alpha(x) = 0$).
2. Weibull: $\Psi_\alpha(x) = \exp\{-(-x)^{\alpha}\}$, $x \le 0$, $\alpha > 0$ (inak $\Psi_\alpha(x) = 1$).
3. Gumbel: $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in R$.

Táto metóda sa využíva na analýzu sezónnych dát, napríklad v hydrológii, kde sa modelujú ročné teplotné maximá, maximálny úhrn zrážok, maximálna hladina rieky a pod. (pozri [1]). Nevýhodou tejto metódy je, že vyžaduje veľmi veľkú údajovú základňu, preto sa častejšie používa nasledujúca metóda.

2. Metóda excedentov ponad prah

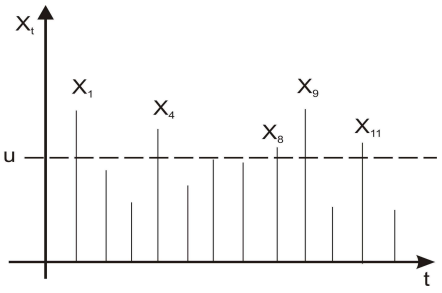
Modernejší prístup modelovania extrémnych hodnôt upriamuje svoju pozornosť nielen na najvyššie hodnoty ako v predchádzajúcej metóde, ale na všetky hodnoty vyššie ako vysoký vopred daný prah *u*. Táto metóda sa nazýva metóda excedentov ponad prah, čo je voľný preklad z angl. Peaks Over Treshold (POT). Táto metóda je založená na nasledujúcej vete:

Veta 2. (Pickands, Balkema and de Haan)

$F_u(x) = P(X - u \le x \mid X > u) = \dfrac{F(x+u) - F(u)}{\overline{F}(u)}$, $x \ge 0$ je distribučná funkcia excedentov ponad prah u práve vtedy, ak vieme nájsť kladnú, merateľnú funkciu $\beta = \beta(u)$ pre každé $\xi > 0$ tak, že platí:

$$\lim_{x \uparrow x_F} \sup_{0 \le x \le x - x_F} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \text{ ,}$$

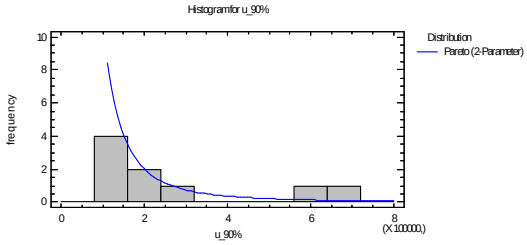
Kde $G_{\xi,\beta}$ je všeobecné dvojparametrové Paretovo rozdelenie.



Obr.9: ***Metóda excedentov ponad prah: X_1, X_4, X_6, X_9 ,X_{11} sú excedenty ponad prah u***

Problémom v praktických aplikáciách ostáva, ako zvoliť prah u. Nemal by sa voliť príliš veľký, pretože by ho mohlo prekročiť iba zopár pozorovaní, ale ani príliš malý, kedy by mohol skresliť odhad Paretovho rozdelenia. Podľa [1] sa za *u* volí hodnota medzi 90- až 95-percentilom.

Údaje, ktoré máme k dispozícii síce nepostačujú na solídnu analýzu údajov, ale urobíme si aspoň názornú ukážku. Počet excedentov, ktoré odpovedajú 90-percentilu ($x_{0,9}$=104 030), je iba 9 hodnôt. V programovom systéme STATGRAPHICS Centurion nie je procedúra, ktorá by usekla údaje, preto to musíme urobiť s využitím Excelu. Najvhodnejším modelom je dvoj parametrové Paretovo rozdelenie s parametrami: shape = 1,4576 a lower threshold = 104030, pre ktoré je p-hodnota K-S testu rovná 0,994418 , výraznú zhodu je vidieť aj na obrázku dole.



Obr.10*: Grafické porovnanie zhody s Paretovým rozdelením*

Paretovo rozdelenie sa používa na modelovanie výšky škôd v prípade, keď tieto škody dosahujú obzvlášť vysoké hodnoty. Má značné využitie aj pri zaistení [3].

Tento článok bol vypracovaný v rámci projektu VEGA 1/3815/06.

LITERATÚRA

[1] BEIRLANT, J. - GOEGEBEUR, Y. - SEGERS, J. - TEUGELS, J.: *Statistics of Extremes: Theory and Applications*, Wiley, New York, 2004.

[2] CURRIE, I.D.: *Loss Distributions*. London and Edingburgh: Institute of Actuaries and Faculty of Actuaries, 1993

[3] EMBRECHTS, P. - KLÜPPELBERG, C. - MIKOSCH, T.: *Modelling extremal events for insurance and finance*, Springer, Berlin, 1997.

[4] HOGG, R.V – KLUGMAN, S.A.: *Loss distributions*, John Wiley & Sons, New York, 1984

[5] PACÁKOVÁ, V.: *Aplikovaná poistná štatistika*, Edícia Ekonómia, Bratislava, 2004

[6] PACÁKOVÁ, V.: *Modelovanie a simulácie poistných rizík*, Edícia: Habilitačné a inauguračné prednášky, Bratislava, 2006

[7] TARTAĽOVÁ, A.: *Modelovanie extrémnych poistných udalostí*, diplomová práca na UPJŠ, Košice, 2006

Mgr. Alena Tartaľová

Katedra aplikovanej matematiky a hospodárskej informatiky

Ekonomická fakulta Technickej univerzity v Košiciach

Nemcovej 32

040 01 Košice

alena.tartalova@tuke.sk

Simulácia vývoja úrokových sadzieb pomocou programov SPSS a MS® Excel

Vladimír Úradníček, Rudolf Gavliak

Abstract: The paper deals with the simulations of overnight offered interest rate using the software packages SPSS 13.0 a MS® Excel. The simulations are utilizing the CIR model. The goal is to verify the justification of hypothetical expectations of interest rate decline. The expected decline from 5,75 % to current 5 % level in Euro area is expected, because of proposed interest rate convergence, which is one of the criteria for Euro adoption in Slovakia.

Keywords: *short rate models, Itô process, one-factor equilibrium models, the CIR model.*

1. Úvod

Príspevok sa zaoberá simuláciou vývoja úrokovej sadzby pre jednodňové refinančné obchody pomocou programov SPSS 13.0 a MS®Excel. Cieľom simulácie je okrem iného overiť, či je v súčasnosti reálne očakávať zníženie tejto sadzby z hodnoty 5,75 % na 5 % ako splnenie jednej z požiadaviek spojenej so zavedením Eura na území Slovenska.
Medzi tradičné prístupy modelovania vývoja krátkodobej úrokovej sadzby patria jedno- a viacfaktorové modely. Jednofaktorové modely vývoja krátkodobej úrokovej sadzby sa zvyknú klasifikovať do dvoch skupín – jednofaktorové rovnovážne modely a jednofaktorové bezarbitrážne modely. Do skupiny jednofaktorových rovnovážnych modelov patrí aj Cox-Ingersoll-Rossov model (CIR model), ktorého diskrétna verzia bola použitá na simuláciu vývoja danej úrokovej sadzby. Pri konštrukcii variantnej simulácie boli využité postupy kalibrácie jednotlivých parametrov uvedené v [1].

2. Teoretické východiská

Ako sme uviedli v našom príspevku [5], môžeme pri úrokových sadzbách v príslušnom časovom horizonte identifikovať tendenciu ich návratu k tzv. dlhodobým rovnovážnym sadzbám. Základné parametre nami aplikovaného modelu, ktoré to zabezpečujú, tvoria: parameter pre dlhodobú rovnováhu (b) a parameter pre dynamickosť približovania sa úrokových sadzieb k dlhodobej rovnováhe (a). Vo všeobecnosti patria uvedené procesy (nazývané tiež ako reverzné procesy – *mean reverting*) do všeobecnej kategórie Itôovho procesu a teda obsahujú aj Wienerov proces (bližšie pozri [12]).
Nech $W_t(\omega)$ je Brownov pohyb na pravdepodobnostnom priestore (Ω, F, P), kde F je σ-algebra merateľných množín Ω a P je pravdepodobnostná miera na Ω. Symbolom $\Upsilon = \Upsilon(S,T)$ označme triedu funkcií $f(t,\omega):<0,\infty)\times\Omega\to\mathbf{R}$, takých, že

(i) funkcia $(t,\omega)\to f(t,\omega)$ je $B \times F$-merateľná, kde B označuje borelovské množiny na $<0,\infty)$;

(ii) stochastický proces $f(t,\omega)$ je F_t^W - adaptovaný;

(iii) $E\left[\int_S^T f^2(t,\omega)\,dt\right]<\infty.$

Pre $f(t,\omega) \in \Upsilon$ bude definovaný Itôov integrál $\int_S^T f(t,\omega)\,dW_t(\omega)$. (i) a (iii) sú technické podmienky, podmienka (ii) precizuje vlastnosť, že proces $f(t,\omega)$ informačne nepredbieha Brownov pohyb.

Jednodimenzionálny Itôov proces je stochastický proces tvaru

$$X_t(\omega) = X_0 + \int_0^t u(s,\omega)\,ds + \int_0^t v(s,\omega)\,dW_s(\omega), \tag{1}$$

kde $u, v \in \Upsilon$ a $P\left[\int_0^t |u(s,\omega)|\,ds < \infty \;\forall t \geq 0\right] = 1.$

Vzťah (1) môžeme zapísať v diferenciálnom tvare

$$dX_t(\omega) = u(t,\omega)\,dt + v(t,\omega)\,dW_t(\omega). \tag{2}$$

Funkcia *u* sa zvykne nazývať „drift“ a funkcia *v* „volatilita“. [8]

Predpokladajme, že *short-rate r* vyhovuje stochastickej diferenciálnej rovnici (2). Jednofaktorové modely vychádzajú z predpokladu, že časová štruktúra úrokovej sadzby je určená len jedným faktorom, ktorým je krátkodobá úroková sadzba r_t, pričom proces r_t krátkodobej úrokovej sadzby sa v rizikovo neutrálnom svete riadi stochastickou diferenciálnou rovnicou (2).

V nami aplikovanom modeli predpokladáme, že r_t je Markovovský proces (nezávisí od minulosti, len od súčasného stavu) s deterministickou zložkou (driftom) $\mu(r_t,t)$ a volatilitou $\sigma(r_t,t)$. Vzťah (2) môžeme prepísať ako

$$dr_t = \mu(r_t,t)\,dt + \sigma(r_t,t)\,dW_t, \tag{3}$$

kde

$\mu(x,t)$ a $\sigma(x,t)$ sú deterministické funkcie.

Potom vývoj úrokovej sadzby r_t môžeme modelovať podľa všeobecného vzťahu:

$$r_t = r_{t-1} + dr_t. \tag{4}$$

Simuláciu vývoja úrokovej sadzby pre jednodňové refinančné obchody pomocou programov SPSS a MS®Excel sme realizovali prostredníctvom jednofaktorového rovnovážneho Cox-Ingersoll-Rossovho modelu (CIR model z roku 1985). Podľa tohto modelu sa dr_t vo vzťahu (4) definuje ako

$$dr_t = a(b - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t, \tag{5}$$

kde *a* predstavuje dynamickosť (rýchlosť), s akou je úroková sadzba *r* priťahovaná k dlhodobej (rovnovážnej) úrokovej sadzbe *b* a σ predstavuje volatilitu.

Prvým predpokladom simulácie je vhodná kalibrácia zvoleného modelu. V prípade *CIR-modelu* je vhodné najprv transformáciou redukovať pôvodné parametre napríklad nasledovne:

$$\rho = \frac{2ab}{\sigma^2},\ \xi = \frac{a+\lambda+\eta}{2\eta},\ \beta = e^{-\eta}, \tag{6}$$

kde $\eta = \sqrt{(a+\lambda)^2 + 2\sigma^2}$ a λ je trhová hodnota rizika. Bližšie pozri napr. [1].

Spätné odtransformovanie jednotlivých parametrov zo vzťahu (6) má tvar:

$$a = \eta(2\xi - 1) - \lambda,\ \sigma = \eta\sqrt{2\xi(1-\xi)},\ b = \frac{\rho\sigma^2}{2a}\ \text{, kde } \eta = -\ln\beta. \tag{7}$$

Pre kalibráciu jednofaktorového modelu môžeme použiť tzv. *stratovú funkciu* definovanú ako:

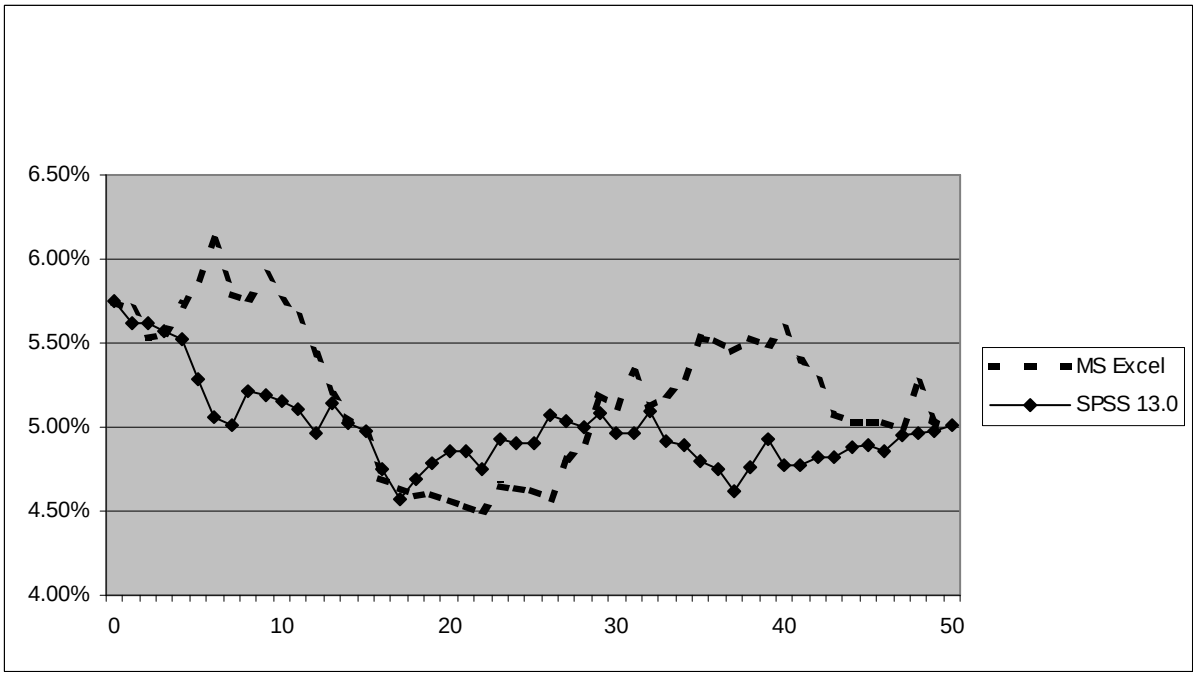
$$U(\beta,\xi,\rho)=\frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\left(R_j^i-\bar{R}_j^i\right)^2\tau_j^2\right], \tag{8}$$

kde $\tau = T - t \in <0,T>$ je čas do exspirácie a $\left\{R_j^i, j=1,2,...m\right\}$ sú reálne a $\left\{\bar{R}_j^i, j=1,2,...m\right\}$ sú kalkulované (teoretické) výnosové krivky i = 1, 2, ..., n zo vzťahu: $A_j e^{-B_j R_o^i} = e^{-\bar{R}_j^i \tau_j}$, kde $A_j = A(\tau_j)$ a $B_j = B(\tau_j)$, kde τ_j, j =1, 2, ..., m sú vzostupne zoradené časy do splatnosti jednotlivých dlhopisov, určujúcich tvar výnosovej krivky a R_0^i je krátkodobá, tzv. *over-night* úroková sadzba.

Pri minimalizácii stratovej funkcii (8) zároveň môžeme nájsť aj vektor (β,ξ,ρ) optimálnych hodnôt odhadov transformovaných parametrov. V záverečnom kroku použijeme vzťahy (7) na ich odtransformovanie.

3. Variantné výsledky simulácie pomocou programov SPSS a MS®Excel

Graf 1. Variantná simulácia vývoja úrokovej sadzby



Zdroj: Vlastné spracovanie

4. Záver

Ako vyplýva z výsledkov variantnej simulácie vývoja úrokovej sadzby pre jednodňové refinančné obchody pomocou programov SPSS 13.0 a MS®Excel, je reálne, že v relatívne krátkom časovom horizonte je objektívne vytvorený priestor pre zníženie tejto úrokovej sadzby na úroveň 5 %, čo je v súlade s očakávaním splnenia požiadavky ECB pre jej korekciu na požadovanú úroveň pri zavedení Eura na Slovensku.

5. Literatúra

[1] ANTOL, M. 2006. *Analýza jedno a dvojfaktorových modelov derivátov úrokových mier*. *Diplomová práca*. Bratislava : FMFI, 2006. 38 s.

[2] BRIGO, D. – MERCURIO, F. 2001. *On deterministic-shift extensions of short-rate models.* www.damianobrigo.it/detshiftrep.pdf. [dostupné na internete 03.09.2007], 25 s.

[3] BOHDALOVÁ, M. – STANKOVIČOVÁ, I. 2007. *The using of the PCA method for measuring risk of the financial portfolios*. In: Forum Statisticum Slovacum 2/2007, s. 20 – 25. Bratislava : SŠDS, 2007. ISSN 1336-7420.

[4] BOHDALOVÁ, M. – STANKOVIČOVÁ, I. 2006. *Using the PCA in the Analyse of the risk Factors of the investment Portfolio*. In. Forum Statisticum Slovacum 3/2006, s. 41 – 52. Bratislava : SŠDS, 2006. ISSN 1336-7420.

[5] GAVLIAK, R. – ÚRADNÍČEK, V. 2007. *Vybrané jednofaktorové modely krátkodobej úrokovej sadzby*. In: Forum Statisticum Slovacum 4/2007, s. 34 – 38. Bratislava : SŠDS, 2007. ISSN 1336-7420.

[6] HULL, J. 2004. *Options, futures et autres actifs dérivés*. Paris : Pearson Education France. 5e édition. 2004. 806 s. ISBN 2-7440-7018-1.

[7] CHAJDIAK, J. – LUHA, J. 1997. *Realizácia štatistických testov v prostredí SPSS a SAS*. In: Zborník príspevkov EKOMSTAT´97, Trenčianske Teplice. Bratislava : SŠDS, 1997.

[8] MELICHERČÍK, I. – OLŠAROVÁ, L. – ÚRADNÍČEK, V. 2005. *Kapitoly z finančnej matematiky*. Bratislava : Epos, 2005. 242 s. ISBN 80-8057-651-3.

[9] STANKOVIČOVÁ, I. – VOJTKOVÁ, M. 2007. *Vybrané štatistické metódy s aplikáciami*. Bratislava : Iura Edition, 2007. 261 s. ISBN 978-80-8078-152-1.

[10] STANKOVIČOVÁ, I. 2002. *Štatistické metódy a hĺbková analýza dát (Data Mining)*. In: Prastan 2002. Bratislava : SŠDS, 2002, s. 71 – 85.

[11] VOJTKOVÁ, M. 2005. *Použitie kanonickej korelačnej analýzy pri riešení závislosti ukazovateľov veľkosti potravinárskych priemyselných podnikov SR*. In: Forum Statisticum Slovacum 2/2005, s. 147 – 152. Bratislava : SŠDS, 2005.

[12] ZMEŠKAL, Z. A KOL. 2004. *Finanční modely*. Praha : Ekopress, s.r.o., 2004. 236 s. ISBN 80-86119-87-4.

[13] ŽAMBOCHOVÁ, M. 2006. *Shlukování ve velkých souborech dat.* In: Forum Statisticum Slovacum 5/2006, s. 168 – 174. Bratislava : SŠDS, 2006. ISSN 1336-7420.

Príspevok vznikol v rámci riešenia vedeckého projektu č. 1/4634/07 financovaného grantovou agentúrou VEGA – Variantné metódy predikovania finančného vývoja malých a stredných podnikov po zavedení spoločnej európskej meny v Slovenskej republike a *vedeckého projektu č. 1/2594/05 financovaného grantovou agentúrou VEGA – Analýza vybraných otázok finančného a bankového trhu po vstupe SR do EÚ.*

Adresa autorov:

Vladimír Úradníček, Ing., Ph.D. – Rudolf Gavliak, Ing.
Oddelenie štatistiky a ekonomickej analytiky
Katedra kvantitatívnych metód a informatiky
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
vladimir.uradnicek@umb.sk, rudolf.gavliak@umb.sk .

ŠTATISTICKÁ ANALÝZA VÝVOJA VYŠETRENÍ NA KARDIOVASKULÁRNE OCHORENIA V ROKOCH 1997 - 2006 VO VYBRANOM REGIÓNE SLOVENSKA

Marianna Vavrová

Abstract: Cardiovascular diseases represent arterial reasons of death in Slovakia, Slovak republic belong to countries with the higher cardiovascular mortality in Europe. Regarding statistical facts of WHO Slovak republic belong to countries with the higher mortality by cardiovascular diseases.

Key words: *arterial reasons of death, cardiovascular mortality, higher mortality, cardiovascular diseases*

1. Úvod

Podľa štatistických prieskumov sa na Slovensku každý z nás najviac obáva onkologických ochorení. Avšak na tieto ochorenia zomiera približne pätina Slovákov. Žiaľ, podľa spomínaných prieskumov, na srdcovocievne ochorenia - odborne nazývané kardiovaskulárne ochorenia (KVO) - zomiera viac ako polovica populácie [3].

Hlavnými rizikovými faktormi podmieňujúcimi vznik týchto ochorení sú sedavý spôsob života a málo pohybu, fajčenie, pitie alkoholu, vysoká hmotnosť, vysoká hladina cholesterolu, vysoká hladina cukru v krvi, vysoký krvný tlak, dedičnosť, stres, a ďalšie. Najrizikovejšími faktormi však sú vysoký cholesterol (hyperlipoproteinémia), fajčenie (nikotinizmus) a vysoký krvný tlak (hypertenzia) [3].

2. Základné informácie

Analýza sa týkala denných záznamov pacientov zistených v priebehu rokov 1997 - 2006 v konkrétnom neštátnom zdravotníckom zariadení vybraného regiónu Slovenska založených na anamnézach, objektívnych vyšetreniach a laboratórnych nálezoch. Databáza údajov je vytvorená v operačnom systéme MS DOS.

Anamnéza ako vstupný rozhovor pomáha vytvoriť pozitívnu atmosféru medzi pacientom a lekárom, ktorý môže lepšie hodnotiť osobnosť pacienta, jeho prístup k ochoreniu, vzťah k liečbe a pod. [4].

Keďže počet pacientov, ktorí netrpeli kardiovaskulárnymi ochoreniami bolo zo štatistického hľadiska zanedbateľne málo, boli z analýzy eliminovaní.

Cieľom je poukázať na stúpajúci trend týkajúci sa vyšetrení pacientov na kardiovaskulárne ochorenia, odhadnúť jeho priebeh a zistiť ako disciplinovane a prečo sa pacienti počas sledovaného obdobia správali.

Výsledky všetkých štatistických analýz boli podrobené testu štatistickej významnosti. Publikované poznatky sú štatisticky signifikantné.

3. Výsledky analýzy

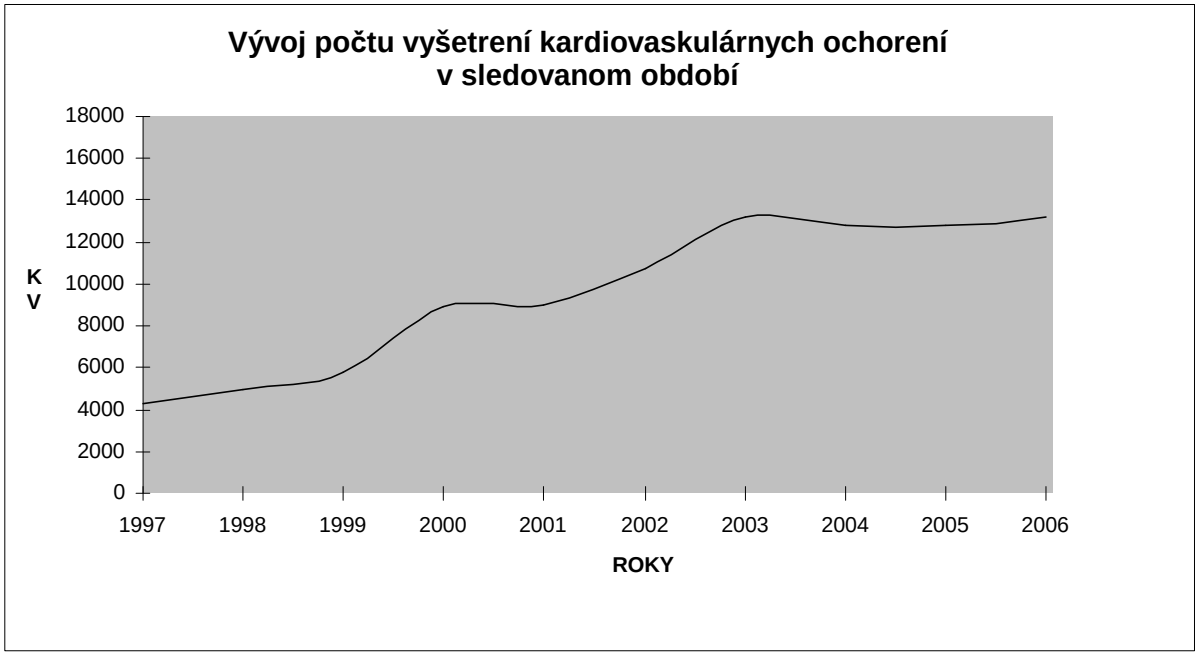
3. 1. Vývoj počtu vyšetrení kardiovaskulárnych ochorení (KVO)

Na základe poznatkov o zistenej údajovej základni sme sa rozhodli priebeh počtu vyšetrení na KVO v posledných troch sledovaných rokoch (2004 až 2006) vyrovnať priamkou, keďže v porovnaní s inými funkciami sme práve priamkou dosiahli najnižšiu chybu odhadu (priemerný mesačný počet je približne 600 vyšetrení).

V celom sledovanom období dochádza k viditeľným výkyvom v počte vyšetrení na KVO, čo bolo spôsobené v dôsledku organizačných zmien v danom regióne (roky 2000 a 2003), zavedenia poplatkov a doplatkov za lieky a v poslednom sledovanom roku 2006 zrušenie poplatkov. Tieto javy nastali v sledovanom období tak, ako to vyplýva z dostupnej dokumentácie konkrétneho lekára, či vývoja zmien v slovenskom zdravotníctve. Išlo o javy nepredvídateľné, avšak z hľadiska štatistického nie zanedbateľné.

Z hľadiska poklesu počtu vyšetrení na KVO sme uvažovali tiež i nad faktom, že pacienti sa od lekára odhlásili, prípadne zomreli, avšak tento úbytok bol priebežne nahrádzaný prijímaním nových pacientov; lekárovi sa tak v sledovanom období skôr zvyšoval počet pacientov. Pokles počtu vyšetrení na KVO teda nemohol byť spôsobený znížením počtu pacientov. Pre názornosť vývoja skutočných hodnôt daného javu uvádzame graf č. 1 Vývoj počtu vyšetrení kardiovaskulárnych ochorení v sledovanom období.

Graf č. 1



Ako vidno z grafu č. 1 trend počtu vyšetrení KVO má stúpajúci charakter i keď v skutočných hodnotách sú jasné výkyvy. Tieto sú spôsobené už spomínanými skutočnosťami (organizačné zmeny v danom regióne (roky 2000 a 2003), zavedenia poplatkov a doplatkov za lieky a v poslednom sledovanom roku zrušenie poplatkov), ktoré nemožno považovať za pravidelne sa opakujúci jav a nie je možné ich samostatne vopred predvídať.

Samotný stúpajúci trend v počte vyšetrení KVO je spôsobený aj faktom, že od roku 1997 do roku 2006 lekárovi stúpol samotný počet pacientov s KVO približne o 90%, teda skoro dvojnásobne. Na rok 2007 sme sa pokúsili odhadnúť prognózu počtu vyšetrení pacientov s KVO na základe vzťahu:

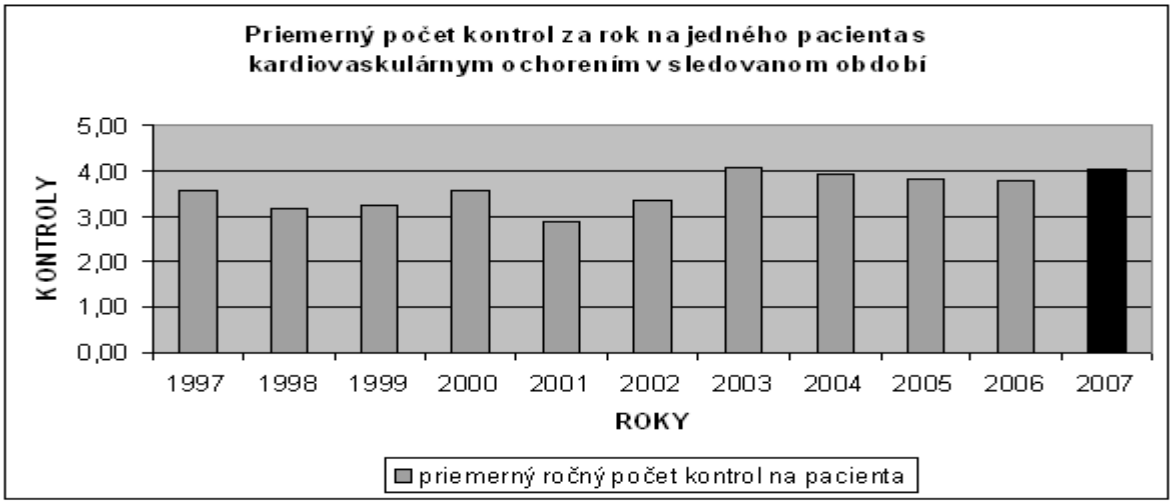
počet vyšetrení KVO 2007 = 13 200 + 0,106*t,

pričom analýzou rozptylu sa potvrdilo na akejkoľvek hladine významnosti (Significance F = 0,005378E-12), že modelom je vysvetlená podstatná časť celkovej variability. Štatisticky významné boli i oba regresné koeficienty. Testom bola tiež zamietnutá autokorelácia rezíduí skúmaného časového radu. Preto sme nemali dôvod zamietnuť predpoklad, že v roku 2007 by sa mohlo očakávať približne 13 839 vyšetrení pacientov s KVO.

3. 2. Priemerný ročný počet kontrol u lekára pripadajúceho na jedného pacienta s KV

V ďalšom kroku sme sa snažili zistiť ako často počas každého sledovaného roku absolvovali pacienti s KVO pravidelné kontroly. Priemerný počet kontrol, ktoré absolvovali pacienti počas roka nie je totožný s javom priemerný počet vyšetrení pacientov s KVO. Každá kontrola sa však považuje za vyšetrenie pacienta s KVO, ale do tohto pojmu nespadajú pacienti, ktorí s KVO prišli na prvé vyšetrenie, prípadne mimo harmonogramu kontrol „keď sa cítili zle“. Pri hlbšom zamyslení sa, tento rozdiel pochopí každý z nás. Každé vyšetrenie, ktoré bolo zaznamenané počas roka sa v tejto časti skúma z hľadiska, či išlo o opätovné vyšetrenie toho istého pacienta, ktorý už bol najmenej raz v sledovanom roku vyšetrený, avšak plnil si svoj harmonogram kontrol odporúčaných ošetrujúcim lekárom (frekvencia návštev pacienta u lekára počas roka). Tento jav je znázornený v grafe č. 2 Priemerný počet kontrol za rok na jedného pacienta s kardiovaskulárnym ochorením v sledovanom období.

Graf č. 2



Ako vidno z grafu č. 2, spomínané príčiny (organizačné zmeny v danom regióne (roky 2000 a 2003), zavedenia poplatkov a doplatkov za lieky a v poslednom sledovanom roku zrušenie poplatkov) sa odzrkadlili i v disciplinovanosti, resp. nedisciplinovanosti samotných pacientov. Kým v rokoch 2000 a 2003 sa pacienti správali viac disciplinovane a dodržiavali harmonogram kontrol u lekára (najvýraznejšie je tento jav vidno v roku 2003), rok 2001 je jasným príkladom nedisciplinovanosti samotných pacientov, keď najviac zanedbávali

kontroly u lekára a stačili im priemerne ani nie 3 ročne, čo je zároveň najmenej v celom sledovanom období. V roku 2002 priemerný ročný počet kontrol opäť stúpol a v roku 2003 sa pacienti správali najdisciplinovanejšie, kedy sa počet kontrol pacientov pohybuje v priemere viac ako 4 ročne. V tomto období však boli zavedené poplatky a doplatky za lieky a počet kontrol u lekára začal opäť klesať a zastavuje sa až v roku 2006, kedy začiatkom štvrtého štvrťroka boli zrušené poplatky u lekárov.

Frekventovanosť vyšetrení jedného pacienta v priebehu roka štatisticky významne závisela od organizačných zmien v danom regióne (roky 2000 a 2003), zavedenia poplatkov a doplatkov za lieky a v poslednom sledovanom roku zrušenie poplatkov (viacnásobný korelačný koeficient = 0,97 pri significance level = 0,001109E-9).

Keďže, ako bolo spomínané, databáza údajov je pod operačným systémom MS DOS, boli sme nútení osloviť odborníka v oblasti programovania, vďaka ktorému sa nám spoločne podaril pokus o prognózu tohto javu pre rok 2007 (graf č. 2 čierny stĺpec), očakáva sa približne rovnaký nárast ako v roku 2003, čo - ako sa domnievame - môže byť spôsobené i už spomínanou skutočnosťou zrušenia poplatkov u lekára a tým i disciplinovanejšie správanie sa pacientov. Hlbšia analýza tohto javu žiaľ, nebola možná, keďže pri každom ďalšom pokuse o spracovanie by mohlo dôjsť k strate celých záznamov o pacientoch a lekár by tak riskoval stratu dôležitých informácií.

4. Záver

Cieľom tohto príspevku bolo poukázať na využiteľnosť poznatkov z databáz v oblasti zdravotníctva a na základe anamnéz pacientov zistiť, resp. potvrdiť stúpajúcu tendenciu vyšetrení týkajúcich sa kardiovaskulárnych ochorení (KVO) zistených v sledovanom období konkrétnym lekárom. Tento cieľ sa nám podarilo splniť (predpoklad, že v roku 2007 by sa mohlo očakávať približne 13 839 vyšetrení pacientov s KVO) a aspoň čiastočne odhaliť príčiny správania sa samotných pacientov, ktorých sa tieto ochorenia týkajú, teda to, že priemerná frekventovanosť vyšetrení jedného pacienta v priebehu roka štatisticky významne závisela od organizačných zmien v danom regióne (roky 2000 a 2003), zavedenia poplatkov a doplatkov za lieky a v poslednom sledovanom roku zrušenie poplatkov (viacnásobný korelačný koeficient = 0,97 pri significance level = 0,001109E-9).

Treba si uvedomiť, že každá analýza, ktorá sa v danej publikácii zdá triviálna bola získaná zo súvislej výpovede pacienta (anamnézy), čím nešlo o maticu údajov, ktorá obsahovala údaje zoradené do stĺpcov a riadkov, teda neplatí, že pacient = riadok a premenná = stĺpec! V žiadnom prípade nemožno hovoriť o bežne spracúvanej matici údajov v štatistike.

Pri spracúvaní údajovej bázy v tejto publikácii sa nám potvrdil fakt, že i keď štatistik má akokoľvek ušľachtilý cieľ, žiaľ, úroveň spracúvaných údajov, resp. technické vybavenie (zastaralý hardware) a programové vybavenie (software fungujúci pod operačným systémom MS DOS) daného pracoviska nebýva na takej úrovni, aby zanalyzoval konkrétny jav tak, aby na druhej strane nespôsobil samotnému poskytovateľovi údajov viac škody svojou prácou, ako úžitku.

5. Literatúra

[1] Bakytová, H. - Bodjanová, S. - Rublíková, E.: Viacrozmerná analýza, ES VŠE v Bratislave, Bratislava, 1988.

[2] Danzig, V. - Šimek, S. - Šimková, R. a kol.: ISCHEMICKÁ CHOROBA SRDEČNÍ U DIABETIKŮ, MAXDORF, Praha, 2006, ISBN 80-7345-079-8.

[3] Kompendium medicíny 3. ročník 2007, Review z odborných kongresov, sympózií a prieskumov.

[4] Lukl, J.: Klinická kardiologie stručně, Univerzita Palackého v Olomouci, Olomouc, 2004, ISBN 80-244-0876-7.

[5] Opatrenenie MZ SR č. 07045/2003 OAP z 31. 12. 2003, ktorým sa ustanovuje rozsah regulácie cien v oblasti zdravotníctva v znení neskorších predpisov

[6] Špinar, J. - Vítovec, J. a kol.: ISCHEMICKÁ CHOROBA SRDEČNÍ, Grada Publishing, Praha, 2003, ISBN 80-247-0500-1.

[7] Štejfa, M. a spolupracovníci: KARDIOLOGIE, Grada Publishing, Praha, 1995, ISBN 80-7169-110-0.

Kontakt:

Ing. Marianna Vavrová
INTERMEDI CENTRUM, s. r. o.
A. Hlinku 27
920 01 Hlohovec
e-mail: vm0011@gmail.com

ŠTATISTICKÁ ANALÝZA VÝVOJA MEDIKAMENTÓZNEJ LIEČBY KARDIOVASKULÁRNYCH OCHORENÍ V SLEDOVANOM OBDOBÍ VO VYBRANOM REGIÓNE SLOVENSKA

Marianna Vavrová

Abstract: During treatment of cardiovascular diseases is important patients awareness of their disease and keeping non pharmacological measures - no smoking, reduction of weight, physical activity, regime (dietary, ...). Priority in treatment of cardiovascular diseases has pharmacological therapy we are spending on this therapy big value of funds.

Key words: *cardiovascular diseases, patients, non pharmacological measures, pharmacological therapy.*

1. Úvod

Nárast kardiovaskulárnych ochorení vo všeobecnosti stúpa i keď vo vyšších vekových skupinách má všeobecne klesajúcu tendenciu, u populácie v strednom veku toto tvrdenie neplatí. V každom prípade nastupuje farmakologická liečba týchto ochorení, ktorá sa vždy kombinuje s nefarmakologickými opatreniami, ako sú zákaz fajčenia, redukcia hmotnosti, dostatok pohybu, životospráva, či prísna diéta, ... Odborníci, zaoberajúci sa skúmaním vynakladaných prostriedkov na zdravotníctvo, sa vo všeobecnosti zhodujú, že náklady na liečbu v tejto oblasti majú stúpajúcu tendenciu [4].

Cieľom tohto článku je poukázať na konkrétne lieky predpisované lekárom na recept na liečbu kardiovaskulárnych ochorení (KVO) vyskytujúcich sa v konkrétnom regióne Slovenska v sledovanom období rokov 2001 až 2006 a to vzhľadom na účinnú látku i na počet balení predpisovaných liekov na recept, ako i na vývoj cien týchto liekov. Článok vznikol hlavne na základe monitoringu jedným konkrétnym lekárom.

Analýza bola vykonaná na základe spracovania databázy údajov získaných z denných lekárskych záznamov o každom jednom pacientovi, ktorému boli predpísané lieky na recept. Databáza je vytváraná v operačnom systéme MS DOS. Citlivosť software, ktorým sú jednotlivé záznamy v databáze vytárané nedovoľuje vyberať niektoré vytypované údaje na ďalšie spracovanie treťou osobou bez toho, aby došlo k úplnej strate informácií o pacientoch, čím by lekár prišiel o kompletnú dokumentáciu daného pacienta, či skupiny skúmaných pacientov.

2. Sumár 10-tich najpredpisovanejších liekov na recept v každom sledovanom roku

Tabuľka č. 1 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2001 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
9	1	RHEFLUIN	552	54,80	51,00	3,80
6	2	ANOPYRIN	492	37,20	37,20	0,00
3	3	VASOCARDIN	433	126,70	126,70	0,00
2	4	DIACORDIN	372	123,00	118,50	4,50
1	5	SORBIMON	340	89,20	85,00	4,20
8	6	CORVATON	327	58,10	57,00	1,10
4	7	TRITACE	304	201,50	150,00	51,50
4	8	TENSIOMIN	274	122,70	115,50	7,20
1	9	NITROMAC R.	218	69,00	56,30	13,30
10	10	DOPEGYT	198	138,50	138,50	0,00

Tabuľka č. 2 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2002 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
7	1	SIMVOR	439	1326,20	866,00	460,2
6	2	ANOPYRIN	437	38,40	38,40	0,00
4	3	MONOPRIL	367	469,10	373,30	95,80
3	4	VASCARDIN	344	130,60	126,70	3,9
9	5	RHEFLUIN	333	56,50	51,00	5,50
4	6	TRITACE	308	207,80	150,00	57,80
8	7	CORVATON	268	59,90	57,00	2,90
7	8	LIPANTHYL	205	673,00	673,00	0,00
2	9	LACIPIL	184	465,90	294,00	171,90
3	10	LOKREN	154	245,70	182,00	63,70

Tabuľka č. 3 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2003 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
6	1	ANOPYRIN	686	45,00	45,00	0,00
7	2	SIMVOR	477	380,00	352,00	28,00
2	3	AMLOPIN	460	160,00	138,00	22,00
4	4	MONOPRIL	430	397,40	373,30	24,10
3	5	VASOCARDIN	398	139,00	136,70	2,30
2	6	APRESA	298	287,40	138,00	149,40
8	7	CORVATON	286	63,80	43,00	20,80
3	8	CONCOR	269	156,90	85,50	71,40
3	9	LOKREN	246	237,10	193,20	43,90
4	10	RENPRES	222	312,90	300,00	12,90

Tabuľka č. 4 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2004 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
2	1	AMLOPIN	1480	160,00	138,00	22,00
6	2	ANOPYRIN	1090	45,00	45,00	0,00
3	3	CORONAL	664	99,00	85,50	13,50
4	4	MONOPRIL	644	397,40	373,30	24,10
7	5	SIMVOR	624	380,00	352,00	28,00
5	6	LAKEA	599	201,00	186,00	15,00
7	7	SUPRELIP	516	349,00	270,00	79,00
8	8	CORVATON	465	63,80	43,00	20,80
3	9	LOKREN	423	237,10	193,20	43,90
9	10	RHEFLUIN	368	71,30	54,00	17,30

Tabuľka č. 5 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2005 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
2	1	AMLOPIN	820	236,40	236,40	0,00
5	2	LAKEA	651	177,30	168,00	9,30
6	3	ANOPYRIN	621	48,70	48,70	0,00
7	4	SIMVOR	527	220,00	200,00	20,00
3	5	CORONAL	457	67,50	54,00	13,50
7	6	TULIP	355	364,00	345,00	19,00
8	7	CORVATON	341	62,80	42,00	20,80
4	8	PIRAMIL	248	125,00	120,00	5,00
4	9	MONOPRIL	207	270,00	194,10	75,90
3	10	LOKREN	191	233,00	182,00	51,00

Tabuľka č. 6 Sumár najčastejšie predpisovaných liekov na recept daným lekárom za rok 2006 vo vybranom regióne Slovenska

Kód účinnej látky	Poradové číslo lieku	Názov lieku	Množstvo balení predpísaného lieku za rok	Cena lieku za balenie (v Sk)	Z toho (za balenie):	
					Hradené poisťovňou (v Sk)	Hradené pacientom (v Sk)
7	1	SIMVOR	1064	178,10	160,00	18,10
6	2	ANOPYRIN	796	50,80	40,00	10,80
2	3	AMLOPIN	510	102,10	99,00	3,10
8	4	CORVATON	550	54,30	36,00	18,30
4	5	MONOZIDE	401	323,50	268,80	54,70
5	6	LAKEA	382	153,10	147,00	6,10
4	7	PIRAMIL	343	97,10	90,00	7,10
7	8	TULIP	320	295,00	252,00	43,00
4	9	IRUZID	308	334,30	306,00	28,30
3	10	CARVEDILOL	263	223,30	145,60	78,70

Uvedené tabuľky (č. 1 - 6) zachytávajú desať najpredpisovanejších liekov lekárom na predpis v uvedenom neštátnom zdravotnom zariadení vo vybranom regióne Slovenska.
V každom sledovanom roku sú uvedené lieky usporiadané podľa najpredpisovanejšieho priemerného ročného počtu balení predpísaných lekárom na recept. Pri každom lieku je uvedený i kód účinnej látky. Podľa tejto skupiny sa dá určiť, aké kombinácie predpísaných liekov môžu byť predpisované lekárom.

V tab. č. 1 môže byť **podľa účinnej látky** predpísaný 1 liek až kombinácia 8 uvedených liekov. V tab. č. 2 môže byť situácia obdobná, v tab. č. 3 môže byť predpísaný iba jeden, ale i 7 liekov. V tab. č. 4 sa situácia opakuje ako v tab. č. 3.. V tab. č. 5 môže byť predpísaných od 1 až po kombináciu 6 liekov a rovnako tak i v tabuľke č. 6.. Pre obyčajného pacienta s KVO to znamená, že pravidelne za medikamentóznu liečbu nemusí doplácať vôbec, alebo v niektorých prípadoch i niekoľko sto korún.

V ďalšom kroku sme na zistenie asociácie medzi typom predpisovaného lieku (charakterizujúceho typ ochorenia KVO) a účinnou látkou pomocou výpočtu štvorcovej kontingencie a použili sme chi-square test (Chi-square = 3098.8) z ktorého vyplýva, že na akejkoľvek štatisticky významnej hladine (α = 2,5%) môžme povedať, že medzi typom predpisovaného lieku (charakterizujúceho typ ochorenia KVO) a účinnou látkou (skupinou/triedou látok) existuje významná asociácia.

Tab. č. 7
Summary Statistics for Contingency Tables

Statistic	Value	Significance
Contingency Coeff.	0.81279	
Cramer's V	0.87546	
Conditional Gamma	0.88023	
Pearson R	0.83876	
Kendall's Tau B	0.79817	0.00000368
Kendall's Tau C	0.83964	

Z vyššie uvedených výsledkov štatistickej analýzy (tab. č. 7) vyplýva, že medzi typom predpisovaného lieku (charakterizujúceho typ ochorenia KVO) a účinnou látkou (skupinou/triedou látok) existuje viac ako stredne silná závislosť, keďže kontingenčný koeficient je rovný 0,81 a ostatné miery asociácie sú tiež vysoko štatisticky významné. Podľa tohto zistenia sa potvrdilo, že lekár predpisuje lieky na KVO podľa účinnej látky.

Podľa toho, akú účinnú látku liek obsahuje, podľa toho sú lieky zaradené do skupín. Zo všetkých možných nás zaujímali iba tie, ktoré sa týkajú kardiovaskulárnych ochorení a boli uvedené v predchádzajúcich tabuľkách.

Uvažovaných bolo 10 skupín účinných látok [3]:
1. NITRÁTY, 2. CA BLOKÁTORY, 3. β - BLOKÁTORY, 4. AC INHIBÍTORY, 5. AT_1 BLOKÁTORY, 6. ANTITROMBOTIKÁ, 7. HYPOLIPIDEMIKÁ, 8. KARDIOPROTEKTÍVNE LÁTKY, 9. DIURETIKÁ, 10. CENTRÁLNE HYPERTENZÍVA (DOPEGYT).

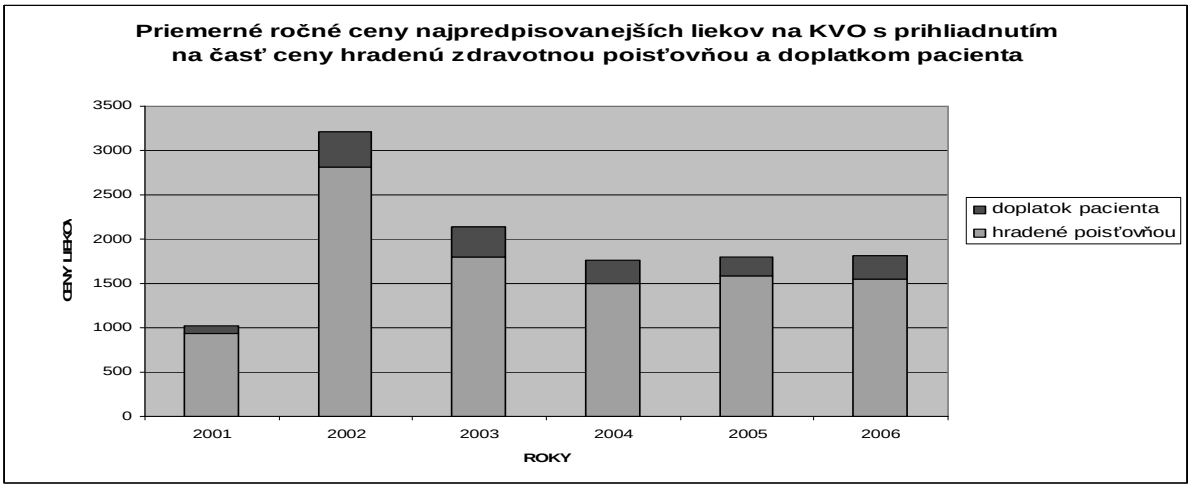
Odhliadnuc od tabuliek zistených najpredpisovanejších liekov lekárom na predpis, najčastejšie tento lekár predpisoval lieky (neberúc do úvahy ročné počty 10 najpredpisovanejších balení liekov) podľa nasledovných účinných látok v posledných dvoch sledovaných rokoch takto:

1. HYPOLIPIDEMIKÁ až 80%,
2. CA BLOKÁTORY až 60%,
3. AC INHIBÍTORY až 60%,
4. ANTITROMBOTIKÁ až 60%,
5. KARDIOPROTEKTÍVNE LÁTKY až 40%,
6. AT_1 BLOKÁTORY až 40%,
7. CENTRÁLNE ANTIHYPERTENZÍVA až 10%.

Táto skutočnosť nás zaujímala z dôvodu možností poskytnutého výberu liekov spadajúceho do konkrétnej kategórie účinnej látky a rozhodovania lekára, ktorý liek z tej - ktorej skupiny účinnej látky môže predpísať zoberúc do úvahy možnú kombináciu s inými liekmi, stúpajúcu tendenciu cien liekov, ...

Vrátiac sa k výsledkom v tabuľkách č.1 až 6, ako vidno ceny liekov mali na začiatku sledovaného obdobia stúpajúcu tendenciu, hoci častokrát bez vplyvu na zvýšenie doplatku pacienta zaň. Najmenej priemerne ročne pacienti s kardiovaskulárnymi ochoreniami doplácali za lieky v roku 2001 (85,60 Sk), najviac už hneď o rok 2002 (405,60 Sk), v roku 2003 doplatky pacientov klesli na priemernú hodnotu 332,40 Sk a odvtedy sa dá skonštatovať, že zvyšné tri roky majú približne rovnakú priemernú hodnotu doplatkov pacientov za desať najpredpisovanejších liekov lekárom na recept. Na porovnanie výšky úhrady cien liekov zdravotnou poisťovňou a doplatkom pacienta uvádzame graf č. 1.

Graf č. 1



Od staršieho k novšiemu sledovanému roku možno badať i zmeny v predpisovaných liekoch i keď rovnakej účinnej látky, čo môže znamenať - nahradenie doterajšieho lieku

novým, prípadne cenovú úpravu liekov smerom hore, čo možno interpretovať z dvoch hľadísk: **z hľadiska pacienta**, ktorý odmieta za rovnaký predpisovaný liek (zmenou jeho ceny) doplácať, keďže doteraz tak nerobil, čo vedie lekára nahrádzať pôvodne predpisované lieky za nové; **z hľadiska lekára,** zvyšovanie liekov prináša pre lekárov problémy, ako predpísať pacientovi dané lieky tak, ako si to jeho diagnóza vyžaduje a zároveň brať ohľad na finančnú situáciu pacienta a v neposlednom rade mať na zreteli vlastnú preskripciu, ktorej hranicu by žiaden lekár nemal prekročiť.

Zo štatistického hľadiska existuje medzi **množstvom predpisovaného lieku** (resp. ich kombináciou) a **cenou lieku** (resp. úhradou časti ceny lieku poisťovňou a doplatkom za liek pacientom) štatisticky významná až silná korelácia (korelačný koeficient = 0,98627 pri significance level = 0,284901E-9).

3. Záver

Tento článok splnil cieľ zistiť 10 konkrétnych najpredpisovanejších liekov lekárom na recept na liečbu kardiovaskulárnych ochorení (KVO) vyskytujúcich sa v konkrétnom regióne Slovenska v sledovanom období rokov 2001 až 2006, vzhľadom na ročný počet balení, ich účinnú látku, a tiež i na cenové relácie liekov. V analýze boli zohľadnené i zmeny cien liekov počas každého sledovaného roku. Zo štatistického hľadiska existuje medzi **množstvom predpisovaného lieku** (resp. ich kombináciou) **a cenou lieku** (resp. úhradou časti ceny lieku poisťovňou a doplatkom za liek pacientom) štatisticky významná až silná korelácia (korelačný koeficient = 0,98627 pri significance level = 0,284901E-9).

Článok je zameraný hlavne na monitoring jedným konkrétnym lekárom. Analýza bola vykonaná na základe spracovania databázy údajov získaných z denných lekárskych záznamov o každom jednom pacientovi, ktorému boli predpísané lieky na recept. Uvedená databáza údajov je spracovávaná pod operačným systémom MS DOS. Citlivosť software, ktorým sú jednotlivé záznamy v databáze vytárané nedovoľuje vyberať niektoré vytypované údaje na ďalšie spracovanie treťou osobou bez toho, aby došlo k úplnej strate informácií o pacientoch, čím by lekár prišiel o kompletnú dokumentáciu daného pacienta, či skupiny skúmaných pacientov.

Citlivosť software, ktorým sú jednotlivé záznamy v databáze vytárané nedovoľuje vyberať niektoré vytypované údaje na ďalšie spracovanie treťou osobou bez toho, aby došlo k úplnej strate informácií o pacientoch, čím by lekár prišiel o kompletnú dokumentáciu daného pacienta, či skupiny skúmaných pacientov.

Treba si uvedomiť, že každá analýza, ktorá sa v danej publikácii zdá triviálna bola získaná zo súvislej výpovede pacienta (anamnézy), čím nešlo o maticu údajov, ktorá obsahovala údaje zoradené do stĺpcov a riadkov, **teda neplatí**, že pacient = riadok a premenná = stĺpec! V žiadnom prípade nemožno hovoriť o bežne spracúvanej matici údajov v štatistike, na akú je zvyknutá väčšina súčasných spracovávateľov štatistických údajov.

4. Literatúra

[1] Bakytová, H. - Bodjanová, S. - Rublíková, E.: Viacrozmerná analýza, ES VŠE v Bratislave, Bratislava, 1988.

[2] Danzig, V. - Šimek, S. - Šimková, R. a kol.: ISCHEMICKÁ CHOROBA SRDEČNÍ U DIABETIKŮ, MAXDORF, Praha, 2006, ISBN 80-7345-079-8.

[3] Jonáš, P. a kolektív: ARTÉRIOVÁ HYPERTENZIA v praxi - diagnostika a liečba, MAXDORF, Praha, 2001, ISBN 90-85912-47-3.

[4] Kompendium medicíny 3. ročník 2007, Review z odborných kongresov, sympózií a prieskumov.

[5] Lukl, J.: Klinická kardiologie stručně, Univerzita Palackého v Olomouci, Olomouc, 2004, ISBN 80-244-0876-7.

[6] Opatrenenie MZ SR č. 07045/2003 OAP z 31. 12. 2003, ktorým sa ustanovuje rozsah regulácie cien v oblasti zdravotníctva v znení neskorších predpisov.

[7] Špinar, J. - Vítovec, J. a kol.: ISCHEMICKÁ CHOROBA SRDEČNÍ, Grada Publishing, Praha, 2003, ISBN 80-247-0500-1.

[8] Štejfa, M. a spolupracovníci: KARDIOLOGIE, Grada Publishing, Praha, 1995, ISBN 80-7169-110-0.

Kontakt:

Ing. Marianna Vavrová
INTERMEDI CENTRUM, s. r. o.
A. Hlinku 27
920 01 Hlohovec
e-mail: vm0011@gmail.com

Estimation of absolute and relative frequency for the sampling wi-thout replacement

Marek Luboš, Vrabec Michal

Abstract: The one of the problem, when we apply the sampling without replacement, is the estimation of relative frequency in the parent population of size N on the basis of a random sample of size n. We estimate the parameter π by two ways: as the point estimation or as the interval estimation. We describe the building of confidence interval in our paper. There are some possibilities how to build the bounds of this interval. We can use normal approximation, binomial approximation or we can calculate the exact bounds in MS Excel. Results for these three methods are compare at the end of our paper.

Key words: *Alternative distribution, hypergeometric distribution, relative frequency, confidence interval, approximations, MS Excel*

1. Confidence interval for relative frequency (sampling without replacement)

We often need to build a point estimate or confidence interval for the relative frequency for the sampling without replacement. We can use the similar way as for the sampling with replacement: the proportion p is used as an estimate of the unknown parameter value π. Therefore, the unknown absolute frequency in the parent population M we estimate by the pN, where N is the size of parent population. We can use the knowledge of random value distribution m (absolute frequency in sample). The probability distribution of m is the hypergeometric distribution.

The probability function for random value m has the form

$$P(m=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x=\max\left[0, n-\left(N-M\right)\right], \ldots, \min\left[n, M\right]. \tag{1}$$

The accurate bounds of $(1-\alpha)$% confidence interval for the relative frequency value π we obtain by solving the following equation (for given m and α_1; m is the number of "successes" in n trials)

$$\sum_{x=m}^{n}\frac{\binom{Np'}{x}\binom{N\left(1-p'\right)}{n-x}}{\binom{N}{n}}=\alpha_1 \quad \text{and} \quad \sum_{x=0}^{m}\frac{\binom{Np''}{x}\binom{N\left(1-p''\right)}{n-x}}{\binom{N}{n}}=\alpha_1 . \tag{2}$$

For $0\le\alpha_1\le 1$, $0\le\alpha_2\le 1$ a $\alpha_1+\alpha_2=\alpha$ the next equations are valid

$$P(p' \le \pi \le p'') = 1 - (\alpha_1 + \alpha_2). \tag{3}$$

$$P(m' \le M \le m'') = 1 - (\alpha_1 + \alpha_2) \tag{4}$$

where

$$m' = Np' \quad \text{a} \quad m'' = Np''. \tag{5}$$

The solving of these equations is difficult. From this reason we cannot find complete tables for p' a p'' in literature. The complete tables would be very large – see ČERMÁK – VRABEC (2003). We can use partly the publication CHUNG – DELURY (1950) or we can compute some approximation. Two most frequently approximations we describe in the next text.

Normal approximation

The first is the approximation by normal distribution. The statistic (6) has asymptotic normal standardized distribution $N(0,1)$ for sampling with replacement,

$$\frac{p - \pi}{\sqrt{p(1-p)}} \sqrt{n-1}, \tag{6}$$

and statistic (7) has asymptotic normal standardized distribution $N(0,1)$ for sampling without replacement.

$$\frac{p - \pi}{\sqrt{\frac{p(1-p)}{n-1} \frac{N-n}{N}}}. \tag{7}$$

For the large n is valid the next formula with confidence approximately $1 - \alpha$

$$p - \frac{1}{2n} - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n-1} \frac{N-n}{N}} < \pi < p + \frac{1}{2n} + u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n-1} \frac{N-n}{N}}. \tag{8}$$

We used the correction term, equal to $1/(2n)$, with respect to discontinuity. The normal approximation gives the very close interval without this correction. The question is, what n is sufficient large.

Binomial approximation

The hypergeometric distribution tends to binomial distribution for $N \to \infty$ (and for fixed n). From this reason, we can determine for sufficiently large N (let's say for $N > 1\,000$) the approximate lower and upper bounds for confidence interval with using tables of values p' and p'' for the sampling with replacement. We use the following procedure: When the sampling quotient f is small, say $f < 0.05$, we use directly tabled values without correction. When the sampling quotient f is sufficient large, say $f > 0.05$, we use the correction – see ARKIN (1963).) The final form of confidence interval is

$$p - \frac{1}{2n} - \left(p - \frac{1}{2n} - p' \right) \sqrt{\frac{N-n}{N}} < \pi < p + \frac{1}{2n} + \left(p'' - p - \frac{1}{2n} \right) \sqrt{\frac{N-n}{N}}. \tag{9}$$

The quality of this approximation was tried by BURSTEIN (1975). He was be satisfied that the usual confidence coefficients (0.8 till 0.9999) give such values, that are very close to the real confidence bounds.

Computing confidence bounds in MS Excel

The other way, how to obtain the confidence bounds, is computing them in MS Excel. We can compute the exact results by solving the equations (2) in Solution procedure.

Practical calculation

We compute the confidence interval for π parameter by the all of described ways and we compare the obtained results in two examples.

Example 1: We work with population with size $N = 10\,000$. We use the sampling without replacement, the sample size is $n = 1\,00$, $m = 50$. We compute 95% confidence interval for parameter π.
We found the exact bounds for p' and p'' as $p' = 0.400$ and $p'' = 0.601$ (CHUNG and DE LURY), the approximate bounds are

a) Normal approximation – formula (8)

$$p' = 0.5 - 0.005 - 1.96 \cdot 0.04767 = 0.402$$

$$p'' = 0.5 + 0.005 + 1.96 \cdot 0.04767 = 0.598.$$

b) Binomial approximation – formula (9)

$$p' = 0.5 - 0.005 - (0.5 - 0.005 - 0.399)\ \sqrt{0,9}\ = 0.403,$$

$$p'' = 0.5 + 0.005 + (0.601 - 0.5 - 0.005)\ \sqrt{0,9}\ = 0.602.$$

c) MS Excel computing - formula (2)

$$p' = 0.403$$

$$p'' = 0.597.$$

Example 2: We work with population with size $N = 500$. We use the sampling without replacement, the sample size is $n = 100$, $m = 9$. We compute 95% confidence interval for parameter π.
We found the exact bounds for p' and p'' as $p' = 0.045$ and $p'' = 0.157$, the approximate bounds are

a) Normal approximation – formula (8)

$$p' = 0.090 - 0.005 - 1.96 \cdot 0.025726 = 0.035,$$

$$p'' = 0.090 + 0.005 + 1.96 \cdot 0.025726 = 0.145.$$

b) Binomial approximation – formula (9)

$$p' = 0.090 - 0.005 - (0.090 - 0.005 - 0.042)\ \sqrt{0,8}\ = 0.047,$$

$$p'' = 0.090 + 0.005 + (0.164 - 0.090 - 0.005)\ \sqrt{0,8}\ = 0.157.$$

c) MS Excel computing - formula (2)

$$p' = 0.048$$

$$p'' = 0.154.$$

3. Conclusions

When determining the confidence interval for parameter value π of an alternative distribution, approximation by normal distribution can be successfully applied (with possible correction for discontinuity) if the actual value of π lies between 0.1 and 0.9. However, if $\pi < 0.1$ or $\pi > 0.9$, tabled values must be utilised. In both instances, MS Excel Solution can be used.

4. References

COCHRAN, W.G.: Sampling Techniques, 3rd ed. New York, Wiley 1977, Chap. 2.
HÁJEK, J.: Theory of probabilistic sampling with applications to sampling surveys (in Czech). Prague, ČSAV 1960
CHUNG, J.H., DELURY, D.B.: Confidence Limits for the Hypergeometric Distribution. Toronto, University of Toronto Press, 1950.
JANKO, J.: Elements of statistical induction (in Czech). Prague, State Statistical Office 1937
JANKO, J.: Statistical tables (in Czech). Prague, ČSAV 1958
BUONACORSI, J.P.: A note on confidence intervals for proportions in finite populations. American Statistician, 41, 1987, 3, 215 – 218.
BURSTEIN, H.: Finite population correction for binomial confidence limits. JASA, 70, 1975, 349, 67 – 69.
QUESENBERRY, C.P., HURST, D.C.: Large sample simultaneous confidence intervals for multinomial proportions. Technometrics, 6, 1964, 1, 191 – 195.

Address of authors:

Luboš Marek, Doc., RNDr., CSc.
VŠE Praha
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4
130 67 Praha 3
marek@vse.cz

Michal Vrabec, Mgr., CSc.
VŠE Praha
Vysoká škola ekonomická v Praze
nám. W. Churchilla 4
130 67 Praha 3
vrabec@vse.cz

Otevřená statistika 2007

Jiří Žváček

Abstract: Open statistics contains short description of statistical opensource software, statistical wikipages, online statistical programs and short overwiew of situation in construction os statistical educational pages and textbooks.

Key words: statistical opensource software, statistical wikipages, online statistical programs statistical educational pages, statistical textbooks

1. Úvod

V tomto článku se zabýváme statistickým softwarem a znalostmi, které jsou na internetu zdarma a online.V češtině i slovenštině, které trpí germánskou přesností, máme trochu problém s terminologií. Anglické slovo **free** znamená sice primárně **svobodný**, ale zároveň i **zdarma**. Navíc sem patří i termín **otevřený**, který může znamenat i svobodu modifikovat zdrojový kód či data. Takže vzhledem k tomu, že zároveň je požadavkem přístupnost, používáme termín otevřená statistika.

V dnešní době mediální komercializace je překvapivá dynamika této oblasti, která bourá mnohé klasické představy.

Budeme se stručně zabývat čtyřmi oblastmi:

- opensource software, na jehož vývoji se lze podílet,
- wiki stránky, které umožňují kooperativně měnit obsah,
- online výpočty, které umožňují přístup ke statistickému softwaru z internetu,
- individuální tvorba, které umožňují tvorbu dynamických a interaktivních stránek s využitím svobodných statistických komponent (včetně učebnic).

2. Opensource software

Konkrétní svoboda softwaru má mnoho forem, od omezení na typ uživatele (studenti, osoby, ...), až po omezení na typ použití (viz třeba **Categories of Free and Non-Free Software** (www.gnu.org/philosophy/categories.html)), základem však je, že je zdarma a lze jej používat v rámci určité licence.

Nejdynamičtější oblastí uživatelského softwaru je tzv. **opensource** (cs.wikipedia.org/wiki/Open_source_software) software (viz **Why Open Source** (www.dwheeler.com/oss_fs_why.html)). V OS software může potenciálně každý zasahovat do zdrojového kódu aplikace, převzít jej a modifikovat.

Pokud je produkt užitečný či zajímavý, obvykle se kolem skupiny nadšenců vytvoří okruh přispívajících a produkt se velmi rychle vyvíjí. Příkladů je mnoho a často se jedná o rozsáhlé

projekty, které zásadně ovlivňují vývoj webu i softwaru (Linux, Firefox, Open Office, PHP, Apache, Python, Java atd). Okruh OS softwaru se stále rozšiřuje, protože kooperace při vývoji a požadavky a zkušenosti uživatelů jsou při rostoucí komplexnosti produktů velkou výhodou (v poslední době přibyla např. Java). Výhody si uvědomil i všemocný Microsoft a přišel se svou verzí **Shared Source licence** (www.microsoft.com/resources/sharedsource/default.mspx) a obdobně **Google and Open Source** (code.google.com/).

Přehled opensource projektů je na **SourceForge.net** (sourceforge.net/index.php), která eviduje více než 132 000 projektů a poskytuje základní služby vývojářům.

2.1. Statistické OS projekty

Řádově 900 opensource projektů souvisí se statistickými výpočty. Velký přehled statistických OS nalezneme na Wikipedii (en.wikipedia.org/wiki/List_of_open_source_software_packages#Statistics), kde je seznam a další odkazy (včetně alternativních přehledů).

Mezi nejvýznamnější jednotlivé statistické OS patří

The R Project (www.r-project.org/)
: Jazyk R je opensource klon komerčního paketu S+ s velmi širokou škálou navazujících produktů a dynamickým vývojem. Nové verze by měly vycházet vždy 1.4. a 1.9., lze se zúčastnit vývoje na beta verzích a objednat si novinky. Současná verze je 2.5.0, vše podstatné je na specializované stránce **rwiki** (wiki.r-project.org/rwiki/doku.php). R má wikibook **Statistical Analysis using R** (en.wikibooks.org/wiki/Statistical_Analysis:_an_Introduction_using_R)

OCTAVE (octave.sourceforge.net/)
: je opensource varianta paketu MATLAB. Viz **wiki/GNU_Octave** (en.wikipedia.org/wiki/GNU_Octave) Časté inovace. V březnu 2007 reorganizován do paketového systému. Vznikla česká podpůrná stránka **Octave** (www.octave.cz/). MATLAB klonů je více, např. **SciLab** (www.scilab.com/) (popis viz **Scilab-Wikipedia** (en.wikipedia.org/wiki/Scilab)) .

Gnumeric (www.gnome.org/projects/gnumeric/)
: je spredsheet s mnoha funkcemi. Viz též **wiki** (en.wikipedia.org/wiki/Gnumeric). Pracuje i s makry Excelu, obsahuje navíc 154 dalších funkcí, včetně pokročilé statistické analýzy, mnoha generátorů náhodných čísel a procedur pro řešení lineárních a nelineárních úloh.

wxMaxima (wxmaxima.sourceforge.net/wiki/index.php/Main_Page)
: je grafické rozhraní k OS matematickému systému **Maxima** (cs.wikipedia.org/wiki/Maxima), vhodný ke psaní matematických textů (lze využít třeba symbolickou matematiku).

matematické systémy
: jsou poměrně často zdarma, viz přehled na **wiki** (en.wikipedia.org/wiki/List_of_computer_algebra_systems)

Z mnoha dalších jsou "živé" například **Gretl** (en.wikipedia.org/wiki/Gretl) pro časové řady a ekonometrii (s možným výstupem do TeXu), **OpenEpi** (www.openepi.com/Menu/OpenEpiMenu.htm) pro epidemiologickou statistiku, **Tanagra** (chirouble.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html) pro dataminig, **PAST** (folk.uio.no/ohammer/past/) pro statistiku v paleontologii atd.

2.2. Internetové distribuované výpočty

Internetové distribuované výpočty jsou krásným příkladem síly opensource projektů. Volnou výpočetní kapacitu počítače lze využít k tomu, aby se zapojil do výpočtů organizovaných pomocí internetu. Práce se pak může zúčastnit každý na síti, Ty nejúspěšnější projekty tak disponují nejvýkonnějším superpočítačem planety, protože se jich účastní miliony počítačů. Zúčastnit se může každý, případně i s vlastním projektem.

Dnes je dokonce možno si na vlastní server nainstalovat vlastní projekt **BOINC** (boinc.berkeley.edu/), což je opensource platforma pro realizaci komunitních výpočetních projektů.

Pro statistiku by takovýto projekt byl jistě zajímavý třeba v oblasti dataminingu (třeba metoda GUHA je dosti výpočetně náročná).

3. Wikii

Úspěch internetové encyklopedie **Wikipedia** (www.wikipedia.com/) velmi zpopularizoval systémy pro kolektivní správu obsahu, inicioval vytvoření nového mezinárodního slova wiki a celé kategorie softwaru, který wiki stránky podporuje.

V podstatě se jedná o otevřený publikační systém. Vzniká jich velmi mnoho a existuje mnoho variant. Jen na samotné Wikipedii vznikly celé celé kategorie specializovaných stránek jako jsou **Wikibooks** (wikibooks.org/) - knihy, **Wiktionery** (wiktionery.org/) - slovníky, **Wikinews** (en.wikinews.org/wiki/Main_Page) - novinky, **Wikiversity** (en.wikipedia.org/wiki/Wikiversity) - výukové materiály, a jistě budou rychle přibývat.

Stejně tak rychle přibývají další varianty s odlišným softwarem a potenciální uživatel má obrovský problém který systém si vybrat (hodně pomůže **Wikimatrix** (www.wikimatrix.org/)).

Velmi vhodné jsou tyto stránky zejména tam, kde je potřeba zpracovat velmi rozsáhlou oblast nebo tam, kde se informace v čase mění a nepotřebujeme zastaralé informace, ale aktuální stav.

3.1. Statistická Wikipedia

V rámci samotné Wikipedie existuje významná kategorie **Statistics** (en.wikipedia.org/wiki/Statistics). Tvoří vrcholek statistických stránek na Wikipedii, které již dnes poskytují základní informace a postupně získávají na rozsahu a kvalitě.

Kromě samotných hesel, které tvoří jakýsi **statistický slovník** jsou zde i velmi užitečné specializované stránky.

Výborné jsou třeba stránky **List** (en.wikipedia.org/wiki/Statistical_package) či zejména **Comparison** (en.wikipedia.org/wiki/Comparison_of_statistical_packages) , seznam a srovnání statistických paketů, a navíc každý zajímavý softwarový produkt již má svou vlastní wiki stránku s aktuálními informacemi.

V počátečním stádiu jsou rozsáhlejší projekty, např. wikiknihy **Handbook of Descriptive Statistics** (en.wikibooks.org/wiki/Handbook_of_Descriptive_Statistics), **Probability** (en.wikibooks.org/wiki/Probability) , **Statistics** (en.wikibooks.org/wiki/Statistics), Wikiversity: **Fundamentals of Probability, Statistics, Experiments and Data** (en.wikiversity.org/wiki/Fundamentals_of_Probability%2C_Statistics%2C_Experiments_and_Data)

Nalezl jsem už i první použití při výuce:**StatWiki** (statwiki.wiwi.hu-berlin.de/index.php/Hauptseite), statistický slovník pro podporu výuky (**přehled hesel** (statwiki.wiwi.hu-berlin.de/index.php/StatWiki_Index)).

Česká wiki zatím zaostává, ale to je i věc českých statistiků.

4. Online software

Online software nebo přesněji **webová aplikace**, je software, který ovládáme interaktivně prostřednictvím internetu (program a případně i data jsou mimo počítač uživatele).

Dnes už se nejedná o pouhé rozhraní k počítačovému programu, ale o komplexní služby, které umožňují přenést mnohé činnosti a data na internet a případně je sdílet s dalšími uživateli. Je to mnohem ekonomičtější než klasický model distribuce softwaru. V neposlední řadě také reakce na přechod k mobilním zařízením a současné práci na více počítačích.

Tímto směrem vrhly i velké softwarové firmy, jmenujme například Google, Microsoft, AOL atd. a komplexní systémy tohoto druhu rychle přibývají.

4.1. Statistika online

Prakticky vše ze statistiky lze dnes spočítat online. Klasický přehled má Pezullo na stránkách **Interactive Statistical Calculation Pages** (statpages.org/).

Z jednotlivých produktů je zejména třeba jmenovat

R Online (user.cs.tu-berlin.de/~ulfi/cgi-bin/r-online/r-online.cgi)
: Funkční verze interaktivního prostředí pro jazyk R **R commander** (en.wikipedia.org/wiki/R_Commander) (je opensource a každý si jej může instalovat na vlastní server).

SOCR (www.socr.ucla.edu/)

je zkratka Statistics Online Computational Resource, statistický online výpočetní zdroj. Obsahuje mnoho ilustrací, výpočtů grafů atd. (viz **SOCR brožura** (www.socr.ucla.edu/docs/SOCR_Brochure_Integrated2Page_2006.pdf)).

SISA (www.quantitativeskills.com/sisa/)
Simple Interactive Statistical Analysis, mnoho interaktivních výpočtů

Dr. Arsham's Statistics Site (home.ubalt.edu/ntsbarsh/Business-stat/opre504.htm)
konkrétně **A Collection of JavaScript E-labs Learning Objects** (home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/scientificCal.htm) obsahuje mnoho statistických interaktivních výpočtů v Javascriptu

DanielSoper.com (www.danielsoper.com/statcalc/default.aspx)
43 online statistických kalkulaček

Statistics to Use (www.physics.csbsju.edu/stats/)
mnoho interaktivních výpočtů

Existují celé servery, které se věnují online statistickému softwaru jako třeba

XURU (www.xuru.org/)
který obsahuje mnoho jednoduchých výpočtů,

Wessa (www.wessa.net/stat.wasp)
další obsahuje mnoho hotových interaktivních statistických výpočtů, které jsou napsány v jazyce R. Je zde možno publikovat i vlastní algoritmy (a oni se postarají o úpravy při změnách verzí R).

Create a Graph (nces.ed.gov/nceskids/createagraph/)
interaktivní kreslení statistických grafů (populární výklad pro děti, obsahuje vše potřebné)

fyzikální generátor náhodných čísel online (www.randomnumbers.info/)
měl by být kvalitnější než numerické generátory

4.2. Interaktivní online učebnice

Variantou interaktivních online výpočtů jsou statistické interaktivní učebnice. Počáteční nadějný rozlet se zastavil a mnoho nového se neděje. Stále dominují **Hyperstat** (davidmlane.com/hyperstat/), **MD*stat*** (www.quantlet.com/mdstat/mmstat.html) a jeho česká verze (www.quantlet.com/mdstat/scripts/mmcze/java/start.html) a **Seeing Statistics** (www.duxbury.com/authors/mcclellandg/seeing.html). Na počátcích je experimentální interstat (www.stahroun.me.cz/interstat/index.htm).

Pozornost se totiž upřela jinam. Budoucnost výuky budou komplexní publikační systémy jako je například Moodle (moodle.cz), což je opensource projekt pro konstrukci komplexních výukových kurzů. Zde je možno pracovat obdobným způsobem jako ve wiki, speciální znalosti z programování nejsou nutné. Navíc je možno v rámci diskusních fór ovlivnit další vývoj

5. Další online produkty

Nejedná se samozřejmě pouze o výpočty. Online lze dnes absolvovat statistické kurzy (zejména **statistics.com** (www.statistics.com/) - není to sice zadarmo, ale lze takto získat i

univerzitní kredity, lze získat libovolnou literaturu (viz třeba **suveco.cz** (www.suweco.cz/), ale za peníze lze získat online i cokoliv od **Springera** (www.springer.com/), který už svou činnost doplnil na Academic Journals, Books and Online Media a možnosti narůstají.

6. Individuální tvorba

Pomalu přibývají i statistikové, kteří nejenom že mají vlastní stránku, ale začínají i tušit jaké možnosti publikování na webu umožňuje a začínají ji používat k podpoře výuky. Zatím to je u nás v embryonální podobě, ale vzhledem k tomu, že zdarma jsou dnes i velmi silné vývojové prostředky pro publikování na internetu (třeba Silverlight), jistě se brzy dočkáme modernějších přístupů

Již pro trochu kvalifikované se nabízí spousty statistických komponent, kterými lze doplnit vlastní stránky. Několik příkladů:

CAPDM (www.capdm.com/demos/software/html/capdm/)
: interaktivní ilustrace, applety lze použít ve vlastních stránkách.

Mathematica
: výrazně inovovala svůj prohlížeč matematických dokumentů **Mathematica Player** (www.wolfram.com/products/player/download.cgi), který umožňuje prohlížet zdarma tisíce (zatím) interaktivních a dynamických matematických dokumentů. Zahájila velkou ofenzívu a na stránce **WOLFRAM DEMONSTRATIONS PROJECT** (demonstrations.wolfram.com/) se kumuluje velké množství interaktivních ilustrací a výpočtů z mnoha oborů. 20.10.2007 zde bylo 88 ilustrací ze **statistiky** (demonstrations.wolfram.com/topic.html?topic=Statistics&limit=100) a 41 z **pravděpodobnosti** (demonstrations.wolfram.com/topic.html?topic=Probability&limit=50). (a poměrně rychle přibývají další)!

Learning by Simulations (www.vias.org/simulations/simu_stat.html)
: mnoho statistických experimentů (možno stáhnout zdarma)

6. ZÁVĚR

Ekonomy možná překvapí, že koncept svobodného softwaru funguje. Miliony kvalifikovaných lidí (na Wikipedii je registrováno skoro 6 milionů autorů) věnují část své kapacity zdarma ostatním. Na druhé straně udělat něco, co bude možná zajímat další miliony lidí je lákavé i když to ocení jen pár informovaných kolegů a systémy hodnocení zděděné z dávné minulosti to ignorují.

Důsledkem současného vývoje softwaru ovšem je, že česká statistika (slovenská, ...) bude taková, jakou si ji domorodci sami udělají. Web2 totiž předpokládá aktivitu uživatelů.

Problém vidím v tom, že současní specializovaní statistikové nemají potřebné znalosti a nadšení pro práci s internetem.

Naději vidím v mladé generaci. První známky již vidíme ve wikiaktivitě studentů MFF **wiki.matfyz.cz/** (wiki.matfyz.cz/).

Interaktivní aktualizovaná (odkazy se občas mění) verze tohoto článku je na **www.stahroun.me.cz/eseje/2007otevrena/index.htm** .

KONTAKT

Doc.Ing.Jiří Žváček, CSc.
V úvalu 84, nem.Motol, LDN, 7.st.
Praha 5
15000
jzvacek@seznam.cz

PREHLIADKA PRÁC MLADÝCH ŠTATISTIKOV A DEMOGRAFOV

Demografické osobitosti urbánnych a rurálnych štruktúr Slovenska

Katarína ČUPEĽOVÁ

Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra humánnej geografie a demogeografie, 1. ročník magisterského štúdia, katka_cupelova@centrum.sk

Abstract: This work analyses demographic structures of urban and rural population according to biological and cultural characteristics. It entertains its historical developement and present extension in space of Slovakia. Analysis is elaborated in level of two subpopulation – urban and rural and in level of municipalities, that are classed into categories according to number of inhabitants. Synthesis of acquired knowledges evaluates demographic behaviour of urban and rural population and its dependence on number of inhabitants in municipality. Specially this work evaluates differences in demographic behaviour of population in town and country or in small municipalities and big cities.

Key words: demographic structures of population, urban and rural population, categories of municipalities according to number of inhabitants, demographic behaviour

Úvod

Mestské a vidiecke prostredie sa vyznačujú špecifikami, ktoré sa odrážajú v štruktúrnej skladbe a následne aj populačnom správaní ich obyvateľstva. Vidiecke prostredie s prevládajúcimi osobnými vzťahmi a silnou pozíciou tradícií a zvykov versus mestské prostredie anonymity v spoluprácii so súčasnými trendami modernej spoločnosti (individualizmus, sekularizácia), náchylné na zmenu priorít a spôsobu života.

Cieľom tejto práce je potvrdiť, resp. vyvrátiť autenticitu vidieka a mesta v demografickom správaní, pričom poukazujeme na špecifické črty vybraných demografických štruktúr populácií. Porovnávané sú štruktúry podľa biologických a kultúrnych znakov na úrovni podsúborov mestského a vidieckeho obyvateľstva a na úrovni veľkostných kategórií obcí.

Metodika

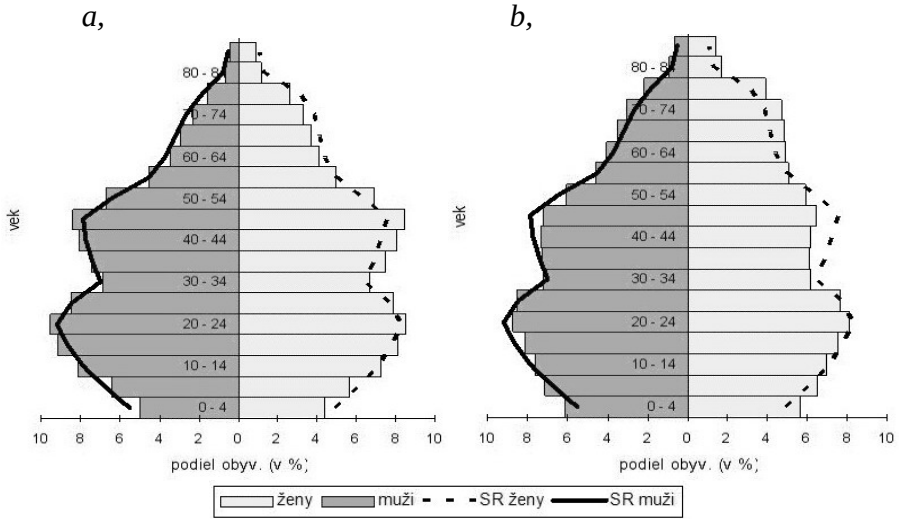
Základnými metodickými postupmi bolo preštudovanie dostupnej literatúry týkajúcej sa vidieckeho a mestského obyvateľstva, spracovanie dát zo Sčítania obyvateľov, domov a bytov 2001 a následné vyvodenie zákonitostí v usporiadaní štruktúr obyvateľstva v priestore.

Jednotlivé demografické štruktúry obyvateľstva boli na základe spracovaných dát prezentované graficky a štatisticky. Za mestské obyvateľstvo považujeme sumu obyvateľstva štatutárnych miest Slovenska a za vidiecke obyvateľstvo sumu obyvateľov vidieckych obcí.

Veková a pohlavná štruktúra

Pri sčítaní v roku 2001 bol na Slovensku podiel žien 51,44 %, index femininity dosahoval hodnotu 1059. Podiel žien bol vyšší od vekovej kategórie 45 – 49 rokov. Pri porovnaní vekovej štruktúry mestského a vidieckeho obyvateľstva môžeme vidieť, že v mestách je podiel žien vyšší už vo vekovej kategórii 30 – 34 rokov – 51,23 %. U vidieckeho obyvateľstva ženy prevládajú od vekovej kategórie 50 – 54 rokov, kde je ich podiel 50,29 %. Tieto rozdiely v zložení pohlaví sú spôsobené najmä migráciou žien v produktívnom veku do miest, čo dosvedčujú aj nízke hodnoty indexu femininity vidieckeho obyvateľstva vo vekových kategóriách, kedy je toto obyvateľstvo ešte ekonomicky aktívne. V kategóriách 30 – 44 rokov klesá pod hodnotu 900.

Graf č.1: Veková štruktúra mestského (*a*) a vidieckeho (*b*) obyvateľstva Slovenska (26. 5. 2001)



Zdroj: Autor, spracované podľa Sčítanie, obyvateľov domov a bytov 2001

Slovenská populácia sa vyznačuje starnutím, no to má iný priebeh v mestách a na vidieku. Podľa podielov jednotlivých vekových kategórií, vybraných ukazovateľov starnutia v jednotlivých veľkostných kategóriách obcí a vekových pyramíd (Graf č.1) mestského a vidieckeho obyvateľstva môžeme konštatovať, že intenzívnejšie starne vidiecke obyvateľstvo. Mestské obyvateľstvo sa vyznačuje vyšším podielom obyvateľstva v produktívnom veku – 65,4%. V detskej vekovej kategórii a v starších vekových kategóriách má vyššie podiely vidiecke obyvateľstvo.Takéto rozloženie obyvateľstva je výsledkom dlhodobých migračných trendov vidieckeho obyvateľstva do miest, kde najmä obyvateľstvo v produktívnom veku hľadá uplatnenie takmer vo všetkých tam sústredených ekonomických aktivitách. Naopak vidiecke sídla, najmä kvalitné bývanie a životné prostredie, môžu byť v budúcnosti atraktívne pre obyvateľstvo v penzijnom veku, čím sa podiel staršieho obyvateľstva môže na vidieku ešte zvýšiť.

Všeobecne konštatujeme, že čím je sídlo väčšie, tým mladšie je jeho obyvateľstvo. Podiel detskej zložky stúpa do kategórie sídel do 5000 obyvateľov (20,9 %), potom klesá. Produktívne vekové kategórie sa sústredia do väčších sídel (nad 5000 obyvateľov), kde majú vyšší podiel ako celoslovenský priemer. Staršie vekové kategórie sú sústredené do najmenších obcí. V týchto obciach sa výrazne prejavuje starnutie zhora aj zdola. V najväčších mestách – Košice a Bratislava – je najnižší podiel detskej zložky (15,9 %), čo indikuje starnutie zdola [3].

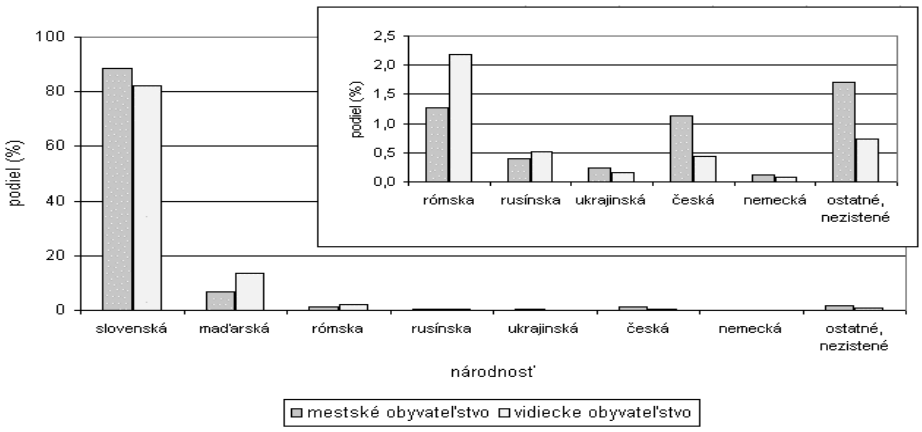
Národnostná štruktúra

V mestách aj na vidieku má dominantné postavenie slovenské obyvateľstvo (viac ako 80%). Jeho podiel na vidieku je znižovaný vysokým podielom obyvateľstva iných národností, ktoré sa priestorovo viažu práve na vidiecke prostredie (najnižšie podiely malo obyvateľstvo obcí s 5000 – 9999 a 200- 499 obyvateľmi, pričom ide o obce s vysokou koncentráciou maďarského obyvateľstva a v druhom prípade sa k nemu pridávajú aj obyvatelia rómskej, rusínskej a ukrajinskej národnosti).

U vidieckeho obyvateľstva dosahuje maďarské obyvateľstvo viac ako dvojnásobný podiel ako u mestského obyvateľstva. Tento jav je spôsobený najmä koncentráciou maďarského obyvateľstva do poľnohospodársky najrozvinutejších oblastí Slovenska. Najvyšší podiel dosahuje

v obciach s 5000 – 9999 obyvateľmi – 16,2 %. Podobne sa na vidiecke prostredie viaže aj obyvateľstvo rómskej a rusínskej národnosti. Menej početné národnosti (česká, ukrajinská, nemecká a iné) sú situované do väčších miest, kde intenzívnejšie dochádza k ich asimilácii najmä so slovenským obyvateľstvom.

Graf č. 2: Národnostná štruktúra mestského a vidieckeho obyvateľstva Slovenska (k 26. 5. 2001)



Zdroj: Autor, spracované podľa Sčítanie obyvateľov, domov a bytov, 2001

Religiózna štruktúra

Výsledky sčítania dokumentujú vyššiu mieru religiozity vidieka (93%) ako miest (77%), pričom s rastom veľkosti obce klesá podiel veriacich, v našich dvoch najväčších mestách klesol pod hodnotu 70 %, zatiaľ čo obyvateľstvo bez vyznania tvorí až 25,8 %. Najvyšší podiel veriacich je v obciach s 200 – 1000 obyvateľmi, takmer 94 %.

Pomerné zastúpenie obyvateľov rímskokatolíckeho vyznania na vidieku a v mestách sa výrazne nelíši, rímskokatolíci tvoria profilovú populáciu Slovenska. Najnižšie hodnoty dosahujú podiely v najmenších veľkostných kategóriách, v obciach s menej ako 200 obyvateľmi je ich podiel len 49,9 %. V týchto obciach je to spôsobené najmä vysokým podielom obyvateľstva gréckokatolíckeho a evanjelického vierovyznania.

Evanjelické augsburské vierovyznanie je podobne ako rímskokatolícke rovnomerne rozložené v mestách aj na vidieku, s miernou prevahou u vidieckeho obyvateľstva (0,52 percentuálneho bodu). Gréckokatolícke a pravoslávne vierovyznanie sa často hodnotia spolu pre ich spoločnú minulosť, priestorovo sú viazané na rusínske a ukrajinské obyvateľstvo východného Slovenska, pričom gréckokatolícke obyvateľstvo ja viac sústredené na vidieku (v obciach pod 200 obyvateľov až 21,3 % z veriaceho obyvateľstva) a pravoslávne v mestách. Reformované kresťanstvo je viazané na obyvateľstvo maďarskej národnosti, a tým aj na vidiecke priestory najmä južného Slovenska.

Ostatné vierovyznania vykazujú podiely menej ako 1 % a sú viazané v prevažnej miere na mestské prostredie, kde je najmä v dôsledku migrácií a vysokej anonymity religiózna štruktúra obyvateľstva pestrejšia.

Tabuľka č. 1: Religiozita a podiel najpočetnejších vierovyznaní v jednotlivých veľkostných kategóriách obcí Slovenska (k 26. 5. 2001)

Veľkostné kategórie obcí	Zastúpenie kategórií (v %)			Podiel z veriacich (v %)			
	veriaci	bez vyznania	neudané	RK	GK	EAV	RK
0 - 199	93,6	4,6	1,8	49,9	21,3	16,2	4,1
200 - 499	93,9	4,4	1,7	66,9	12,8	12,6	4,4
500 - 999	93,8	4,5	1,7	76,2	6,7	10,3	4,2
1000 - 1999	93,1	5,1	1,8	85,2	3,9	6,8	3,0
2000 - 4999	91,1	6,7	2,1	88,5	2,2	6,0	2,1
5000 - 9999	86,3	11,0	2,7	82,9	3,7	7,3	3,7
10000 - 19999	79,2	17,2	3,6	82,3	4,2	9,6	1,0
20000 - 49999	79,9	16,7	3,5	80,7	5,8	8,1	2,2
50000 - 99999	74,5	21,1	4,4	86,1	2,6	8,9	0,2
100000 a viac	69,4	25,8	4,8	82,5	4,5	7,4	1,8
SPOLU	**84,0**	**13,0**	**3,0**	**82,0**	**4,9**	**8,2**	**2,4**

Poznámka: RK – rímskokatolícke
GK – gréckokatolícke
EAV – evanjelické augsburského vyznania
RK – reformované kresťanské

Zdroj: Sčítanie obyvateľov, domov a bytov 2001. In: Mládek, J. a kol. (2006)

Záver

Štatistické vyhodnotenie dát zo Sčítania obyvateľov, domov a bytov 2001 potvrdilo v úvode vyslovenú hypotézu, podľa ktorej sa mestské a vidiecke štruktúry vyznačujú špecifickou štruktúrou obyvateľstva a tá je závislá aj na populačnej veľkosti obce. Jednotlivé demografické štruktúry majú značný vplyv na jeho populačnú dynamiku, preto by mohla byť táto práca základom pre skúmanie odlišného demografického správania obyvateľstva na vidieku a v mestách a aj v jednotlivých veľkostných kategóriách obcí Slovenska.

Použitá literatúra

[1] BEDNÁRIK, R. (1995): Sociologické aspekty religiozity na Slovensku. In: Fenomén národnosti (etnicity) a náboženstva v demografii strednej Európy. Zborník príspevkov 5. Demografickej konferencie. Slovenská štatistická a demografická spoločnosť, Bratislava, s. 4 - 9, ISBN 80-967343-2-6.

[2] MLÁDEK, J. (1995): Národnostná štruktúra obyvateľstva Slovenska a procesy jeho starnutia. In: Fenomén národnosti (etnicity) a náboženstva v demografii strednej Európy. Zborník príspevkov 5. Demografickej konferencie. Slovenská štatistická a demografická spoločnosť, Bratislava, s. 98 – 104, ISBN 80-967343-2-6

[3] MLÁDEK, J. a kol. (2006): Demografická analýza Slovenska. Vydavateľstvo UK, Bratislava, s. 222, ISBN 80-223-2191-5.

[4] Sčítanie obyvateľov, domov a bytov 2001. Definitívne výsledky za SR, NUTS2, kraje, okresy a obce. Štatistický úrad Slovenskej republiky, Bratislava, 2003.

Kritéria rozhodovania investorov pri nákupe podielových fondov

Ivan Gurník

Abstract: The aim of this work is to reveal and measure the most significant parameters of the choice process by mutual funds. This is a reaction on the debate announced in last weeks, which revealed a poor financial awareness by Slovaks. Therefore is significant to know how Slovak investors approach to selection between mutual funds offered on the Slovakian market. Results of this work can help asset managements by arrangement of marketing strategy and so could be also helpful for clients by determining which criterion they should avoid to stay a rational investor. I focused on the funds operating on the money, bond and capital assets markets processed into panel data.

Keywords: *Mutual funds, selection criterion, investment strategy, asset management, money market, bond market, capital asset market, panel data*

1. Úvod

Mojím cieľom je odhad, verifikácia a aplikácia ekonometrického regresného modelu, ktorý bude kvantifikovať kritéria, ktoré uprednostňujú klienti pri nákupe podielových fondov. Z palety produktov som si vybral peňažné fondy, dlhopisové a akciové fondy, ktoré reprezentujú najviac dopytované fondy. Poznanie preferencií a spôsobu výberu podielových fondov môže byť užitočné pre správcov fondov, ktorý by na základe daných kritérií mohli upraviť nielen svoju marketingovú kampaň ale aj samotnú správu podielových fondov s ohľadom na klienta, pretože je možné, že nielen výnos ale aj ostatné vlastnosti fondu, ako mena, hodnota aktív či poplatky môžu zohrávať významnú rolu pri výbere. Keďže mi nie je známy univerzálny vzorec na investovanie do podielových fondov(rozhodovanie podľa fundamentálnej a technickej analýzy), môže byť táto práca prínosom aj pre samostatných klientov, aby bolo možné rozpoznať chyby pri výbere a tak optimalizovať svoje peňažné zdroje s ohľadom na výnosnosť i riziko.

Preto ako závislá premenná bude slúžiť absolútna hodnota z čistých ročných predajov a regresormi budú údaje o podielových fondoch zozbieraných Slovenskou asociáciou správcovských spoločností, ktoré bývajú následne zverejňované vo väčšine periodík, ako to správcom káže Zákon o kolektívnom investovaní. Ako zdrojové dáta mi poslúžili týždenné údaje o otvorených podielových fondoch z dňa 23.októbra 2007. Do modelu som pridal viaceré umelé premenné, a to pre distribúciu buď bankou, alebo finančným poradenstvom, denominačnú menu fondu: SKK, EUR alebo USD.

2. Ekonometrický model

Zásadným krokom pri tvorbe modelov bolo vytvorenie semi-logaritmického modelu (logaritmus z absolútnych ročných čistých predajov), čo podstatne zvýšilo indexy determinácie, a tým percentuálnu hodnotu vysvetlenia.

2.1. Peňažné fondy

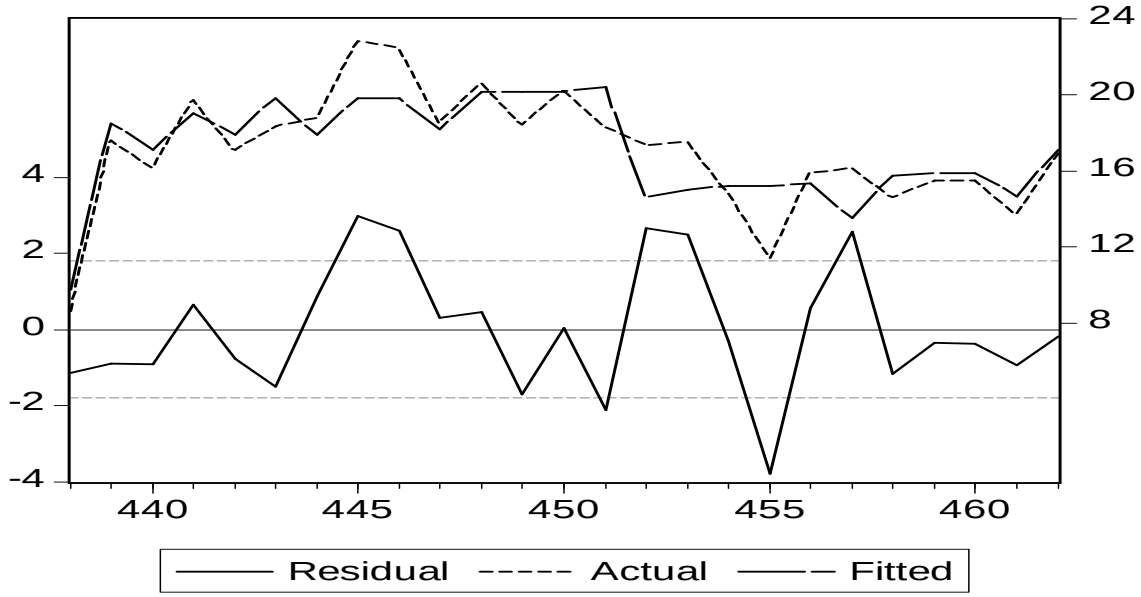
Peňažné fondy sú najviac predávané podielové fondy na Slovensku, pretože predstavujú výnosnejšiu alternatívu k termínovaným vkladom a Slováci sa radia skôr ku konzervatívnym klientom. Preto z dostupných zdrojov dát vyšiel ako najvhodnejší nasledujúci model:

$$\log(|\widehat{čp}_t|) = 17,36 + 5,99.skk_t + 1,94.ba_t + 1,94.\log(|1t_t|) \quad t = 1,2,...25 \qquad (1)$$

t- stat. (13,32) (6,49) (2,64) (3,78)

kde
Čp - hodnota čistých ročných predajov,
Skk – fond je denominovaný v SKK(umelá premenná),
1t – hodnota týždennej výnosnosti fondu v denominovanej mene,
Ba – fond je distribuovaný bankou(umelá premenná).

Graf č.1: Aktuálne a odhadnuté hodnoty a rezíduá pre peňažné fondy



Zdroj: vlastné výpočty v EViews 4.1

Správnosť výsledného modelu potvrdil Ramseyho RESET test (F-štatistika 0,1677 a výsledná p-hodnota 0.68). Model spĺňa aj požiadavky štatistickej verifikácie, kedy je model významný ako celok, všetky parametre sú významné na všetkých hladinách a AIC má hodnotu 4,15, ktorá bola najnižšia z testovaných modelov pri splnení ostatných kritérií. Model vysvetľuje variabilitu závislej premennej zvolenou regresnou funkciou na 71%. Kleinov test ani test Farrara-Glaubera nepotvrdili prítomnosť multikolinearity. Taktiež Whiteov test heteroskedasticity nepotvrdil jej prítomnosť. Nakoľko model bol rátaný z prierezových dát, nevyskytla sa v ňom autokorelácia, čo bolo aj potvrdené Breusch-Godfreyovým LM testom.

Z modelu vyplýva, že peňažné fondy sú prevažne distribuované bankami ako alternatíva k termínovaným vkladom, čo je potvrdené aj kladným znamienkom pred hodnotou týždennej výkonnosti, čo je pravdepodobne spôsobené vyššími výnosmi peňažných fondov v porovnaní s termínovanými vkladmi pri takmer rovnakej rizikovosti. Daná súvislosť by mohla potvrdzovať výber fondu klientom v samotnej banke v čase poskytnutia informácie od klientskeho pracovníka. Významný parameter o denominácii fondu v SKK je pozitívnym zistením o tom, že investori začali brať do úvahy aj posilňovanie SKK, čo znižovalo výnosnosť fondov denominovaných v iných menách a môže nasvedčovať o zvyšujúcej sa finančnej gramotnosti vtom, že ľudia aktívnejšie začali vnímať súvislosti medzi denominovanou menou a výnosnosťou fondu.

Rozhodovania sa na základe krátkodobej výnosnosti (týždennej) môže byť náznak racionálneho rozhodovania, pretože tituly v peňažných fondoch sa obmieňajú pomerne často z dôvodu držby krátkodobých cenných papierov (pokladničné poukážky).

2. 2. Dlhopisové fondy

Model najlepšie vysvetľujúci výberové kritériá má nasledovné hodnoty parametrov:

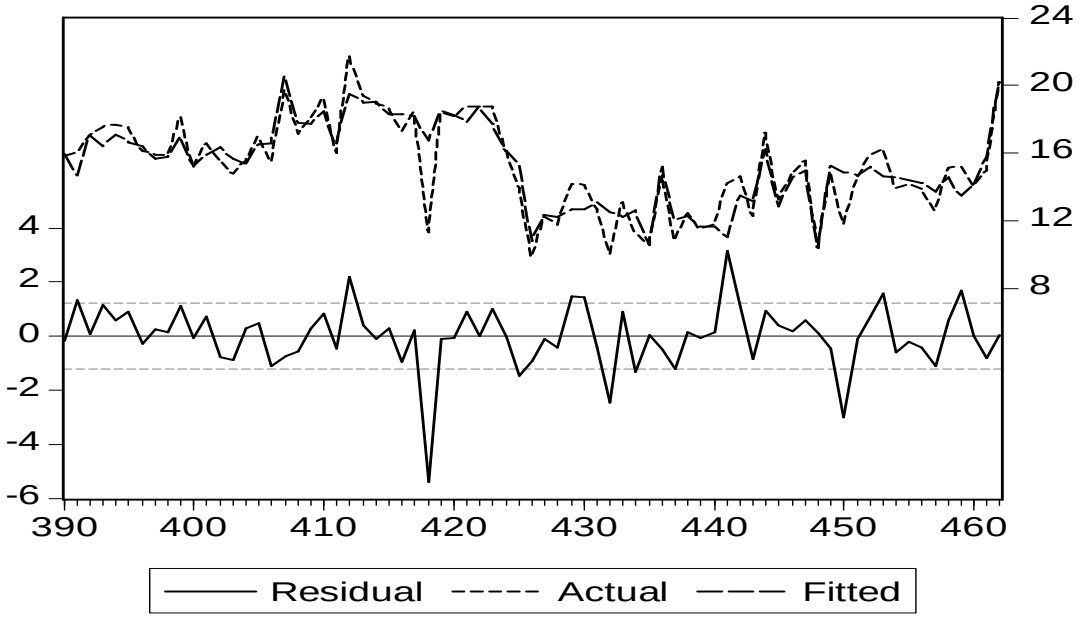
$$\log(|\widehat{čp}_t|) = 0,85.\log(has_t) - 0,28.pv_t + 0,00004.mi_t - 0,05.1r_t,\ t = 1,2,...73 \qquad (2)$$

t- stat. (15,67) (-2,29) (2,18) (-2,12)

kde

Čp – hodnota absolútnych čistých ročných predajov,
Has – hodnota aktív fondu na Slovensku,
Pv – poplatok pri vydaní,
Mi – minimálna investícia,
1r – výnosnosť fondu za posledných 12 mesiacov p.a. v denominačnej mene.

Graf č.2 :Aktuálne hodnoty, odhadnuté hodnoty a rezíduá



Zdroj: vlastné výpočty v EViews 4.1

RESET test potvrdil správnosť modelu a je významný ak celok. Jednotlivé parametre sú významné okrem konštanty (p-hodnota = 0,13), preto som ju nezahrnul ani do výsledného modelu. AIC dosiahlo hodnotu 3,3. Model vysvetľuje variabilitu závislej premennej na 80%, čo je najvyššia hodnota vysvetľovania z troch vytváraných modeloch. Na základe Kleinovho testu či testu Farrara-Glaubera sa nepotvrdila prítomnosť multikolinearity. Whiteov test nezamieta nulovú hypotézu o homoskedasticitu. Problém autokorelácie nebol taktiež diagnostikovaný.

Pri dlhopisových fondoch čiastočne prekáža záporné znamienko pri ročnej výnosnosti, no môže to byť spôsobené tým, že dlhové cenné papiere dlhodobé, ako sú dlhopisy, musia určitým spôsobom reagovať na zmenu úrokových sadzieb, a tým je zmena hodnoty samotného cenného papiera v krátkom období, až do obdobia kým sa vyššia úroková miera neprejaví aj rastom kupónu. Preto pri poklese ročnej výnosnosti fondu môžu investori kupovať podielové jednotky v domnienke neskoršieho rastu výnosnosti, čo sa prejaví na raste predajnosti. Pochopiteľné je aj záporné znamienko pred parametrom poplatok pri vydaní, kedy sa rast poplatkov pri vstupe do fondu prejaví nižším záujmom klientov o investíciu do

tohto typu fondov alebo do daného konkrétneho fondu. Výška vstupných poplatkov môže výrazne znížiť celkovú výnosnosť za obdobie držby podielových listov.

Predaj dlhopisových fondov je závislý od viacerých parametrov, čo môže súvisieť aj s časom, na ktorý investori viažu svoje peňažné prostriedky do fondov. Rozhodovanie určite ovplyvňuje viac faktorov. Prínosom je aj významnosť parametrov ako minimálna investícia, či hodnota aktív fondu na Slovensku, ktoré môžu súvisieť so snahou správcovských spoločností o zníženie nákladovosti správy z dôvodu väčšieho počtu investorov s nižšími vkladmi. No stále budú tieto faktory skôr vecou marketingového prístupu z dôvodu existencie fondov s minimálnou investíciou 500 Sk, no pri mediánovej hodnote 10 000 Sk. Potreba výšky minimálnej investície môže súvisieť aj s charakterom cenných papierov ako dlhodobejšej investície s väčším objemom viazaných prostriedkov v porovnaní s cennými papiermi na peňažnom trhu. Hodnota aktív na Slovensku môže naznačovať orientáciu investorov nasledovať ostatných majiteľov podielových listov, kedy sa objem vložených prostriedkov prejaví aj zmenou u celkovej ročnej predajnosti nižšou intenzitou ($has_t^{0,85}$).

2. 3. Akciové fondy

Regresný model pre akciové fondy má nasledujúci tvar:

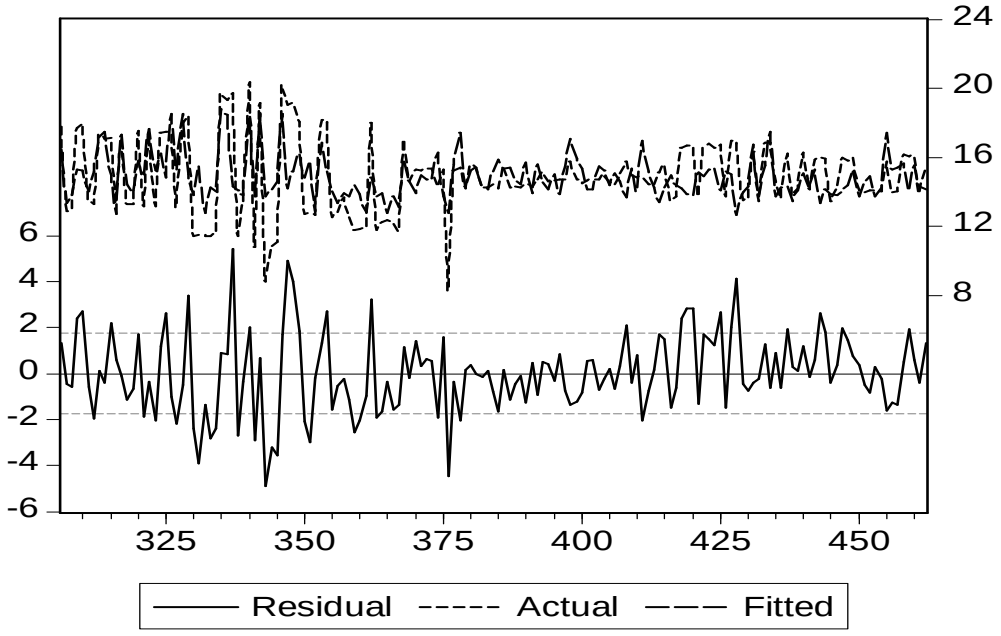
$$\log(|\widehat{čp}_t|) = 13,7 + 0,003.(p_t)^2 + 3,67.skk_t + 0,03.3r_t - 1,45.pv_t + 1,09.d_t \quad t=1,2...157 \quad (3)$$

t- stat. (51,46) (5,05) (8,79) (3,98) (-3,68) (3,84)

kde

Čp – hodnota čistých ročných predajov,
P – hodnota aktív fondu na Slovensku/ hodnota aktív fondu,
Skk – fond je denominovaný v Sk,
3r – trojročná výnosnosť fondu v denominovanej mene p.a.,
Pv – poplatok pri vrátení ,
D – umelá premenná za podmienky, že čisté predaje sú väčšie ako 0(iba predaj).

Graf č.3: Aktuálna hodnota, odhadnutá hodnota a rezíduá



Zdroj: vlastné výpočty v EViews 4.1

Aj pri tomto modely Ramseyho test potvrdil správnu špecifikáciu. Model je významný ako celok, a tiež jednotlivé parametre sú významné na všetkých hladinách významnosti pri hodnote AIC 4,01. Problém tohto modelu je vtom, že vysvetľuje iba 36% variability objemu predaja, no neobsahuje problém multikolinearity, heteroskedasticity ani autokorelácie.

Významnosť záporného znamienka pri výške poplatkov pri vrátení svedčí o tom, že investori negatívne reagujú na rast poplatkov. Ale napriek tomu priaznivo reagujú na portfólia denominované v Sk kvôli posilňovaniu meny a uprednostňujú portfóliá, ktoré rastú v priebehu posledných troch rokov, čo spĺňa predpoklad, že akciové portfólia sú skôr určené na dlhodobejšie investovanie . Pomer medzi hodnotou aktív fondu na Slovensku a celkovou hodnotou aktív fondov môže nasvedčovať o spôsobe investovania, kedy jednoducho investori vkladajú svoje prostriedky podľa príkladu ostatných a investujú do najviac predávaných fondov.

Už spomínaná problematika nízkej vysvetlenej variability závislej premennej pri akciových fondoch môže byť zapríčinená aj formou cenných papierov- akcií, kedy na ich trhovú hodnotu vplýva množstvo faktorov od makroekonomických až po nezávislé subjektívne hodnotenie investorom. Portfóliá navyše obsahujú viacero akciových titulov, ktoré sa časom obmieňajú, a to môže znižovať orientáciu potenciálnych investorov, a tí sa preto najviac rozhodujú podľa trojročnej výnosnosti, čo nie je najšťastnejším výberovým kritériom, pretože súčasná hodnota hodnoty podielovej jednotky nie je garanciou budúcej výnosnosti a minulé výnosy môžu byť nasledované dlhodobejším poklesom, ako sa mnoho klientov mohlo presvedčiť na prelome tisícročí. No zaradenie SKK a poplatku pri vrátení do rozhodovania je z môjho pohľadu dobrou voľbou pri výbere podielového fondu.

3. Záver

V príspevku sme čiastočne potvrdili racionálny výber pri investovaní do podielových fondov, najmä pri zohľadňovaní meny v ktorej sa investuje. Orientácia investorov podľa historickej výnosnosti nemusí byť najvhodnejším krokom, no s ohľadom na dostupnosť týchto informácií určených pre investorov vlastniacich už podielové listy jednotlivých fondov, je daný výber ako orientačné kritérium pochopiteľný. Pôvodná domnienka, že väčšina podielových fondov je distribuovaná prevažne bankami, sa potvrdila iba pri peňažných fondoch. Keďže modely úplne nevysvetlili objem predajov podielových fondov, je zrejmá prítomnosť aj iných faktorov, ako môžu byť nálada investorov či postoj k riziku, ktoré je už zložitejšie kvantifikovať.

4. Použitá literatúra

GREEN, W. H. 1997. Econometric Analyses. Londýn: Prentice – Hall, 1997. 1076 s. ISBN 0-13-7246659-5.

HATRÁK, M. 2007. EKONOMETRIA.. BRATISLAVA: IURA, 2007. ISBN 978-80-8078-150-7.

HUŠEK, R.: Ekonometrická analýza. Praha: Ekopress, 1999. ISBN 80-86119-19-X.

CHOVANCOVÁ, B. a kol.: Finančný trh. Nástroje, transakcie, inštitúcie. Bratislava: EUROUNION, 1999. ISBN 80-88984-03-3.

Adresa autora
Ivan Gurník
Ekonomická Fakulta
Univerzita Mateja Bela Banská Bystrica
Tajovského 10
975 90 Banská Bystrica
i.gurnik@centrum.sk

Sledovanie časového a priestorového šírenia prvých ľudí druhu Homo sapiens sapiens pomocou najnovších vedeckých výskumov analýzy DNA

Lucia Hoffmanová

Abstract: Where do we come from? How did we get to where we live today? This questions could be answered by new genetical research methods. The scientists are now working with genetical codes in DNA of the people on the world. And they can finally give us the real answer of ancient migrations of the first modern people to the whole world.

Kľúčové slová: DNA, Homo sapiens sapiens, chromozóm Y, mutácia, mutačný znak

Z hľadiska skúmania šírenie súčasného človeka z Afriky na ostatné kontinenty sveta sa využívajú v súčasnej dobe nové poznatky z oblasti biológie. Odvetvie biológie, a najmä genetika, sa teda stáva veľmi významné pre výskum evolúcie človeka. Genetika sa zaoberá premenlivosťou živých sústav a sleduje variabilitu, rozdielnosť a prenos druhových a dedičných znakov medzi rodičmi a ich potomkami.

Ľudská morfológia je určite závislá na genetickej zložke, ale je zrejmé, že jej variabilitu kontrolujú desiatky a možno stovky samostatných génov. Zmeny v genetickej výbave boli pre štúdium variability znakov u ľudí zásadné, pretože práve tieto zmeny sú zdrojom evolúcie. A evolúcia je vo svojej najjednoduchšej podobe zmena genetickej výbavy druhov, ktorá prebieha v čase. Ak chceme určiť mieru príbuznosti, musíme vedieť niečo o ich génoch. Pokiaľ majú zhodné gény, patria k jednému druhu (gén je úsek DNA).

DNA sa nachádza v chromozómoch, ktorých má človek 23 párov (posledný pár určuje pohlavie človeka). Tieto chromozómy sa nachádzajú v jadre bunky. Primárnym zdrojom genetickej variability je DNA. Rozhodujúcou silou sú mutácie, bez ktorých by variabilita neexistovali. Mutácie sú náhodné zmeny v sekvencii DNA, vznikajú chybným kopírovaním počas bunkového delenia. Každý človek je nositeľom okolo 30 úplne nových mutácií, ktoré ho odlišujú od jeho rodičov.

Jeden kúsok DNA predstavuje neoceniteľný nástroj pre získavanie podrobností z histórie človeka. Je to časť DNA, ktorá je predávaná iba z otca na syna. Preto teda definuje výhradne samčiu líniu. Hovorí sa mu chromozóm Y a je to pohlavný chromozóm v 23. páre, ktorý sa nachádza v jadre bunky.

U samcov cicavcov dochádza k spojeniu nerovnakých chromozómov – jedného X a jedného Y. U samíc je chromozóm X prítomný vo dvoch kópiách tak ako ostatné chromozómy, čo umožňuje normálnu rekombináciu. U samcov chromozóm Y zodpovedá chromozómu X iba v krátkych úsekoch na oboch koncoch čo slúži k tomu, aby oba pohlavné chromozómy boli behom bunečného delenia spárované. Zvyšok chromozómu Y, ktorý je úplne odlišný od X, je teda neustále predávaný z generácie na generáciu nezmenený. Chromozóm Y je veľký a preto sa v ňom nachádza mnoho miest, na ktorých mohlo v minulosti dôjsť k mutáciám. A ak nedochádza k rekombinácii, sme schopní odvodiť poradie v akom sa mutácie na Y objavovali.

Väčšina genetických polymorfizmov nájdených u človeka sa vyskytuje výlučne u Afričanov, zatiaľ čo Európania, Aziati i pôvodní obyvatelia Ameriky sú nositelia iba malej časti mimoriadnej rozmanitosti. To znamená, že dlhšie časové obdobie prináša väčšiu zmenu.

Výhodou genetických metód je možnosť sledovať chod molekulárnych hodín. Keďže každý gén musí mať svojho predka, vedci tak môžu zostaviť ďaleko úplnejší obraz našej vzdialenej histórie.

Priestorové a časové šírenie prvých ľudí pomocou sledovania znakov na chromozóme Y

Väčšina ľudskej evolúcie sa pravdepodobne odohrala v Afrike. Práve genetické údaje v každom človeku nám poskytujú jasný obraz postupného presunu z Afriky do Eurázie a Ameriky. Sledovaním poradia mutácií, teda doby v ktorej k nim dochádzalo, a demografických detailov (ako zánik populácií či ich expanzia) môžeme zistiť podrobnosti o ľudkom sťahovaní. Sledovanie mutačných znakov na chromozóme Y je z môjho pohľadu výhodnejšie vďaka širšej zdrojovej základni.

Osídľovanie Afriky

Väčšina vedcov sa zhoduje, že anatomicky moderní ľudia sa prvýkrát objavili v Afrike asi pred 200 0000 rokmi. Najranejšie fosílie moderných ľudí boli nájdené vo Východnej Afrike na území Etiópie. Už pred 120 000 rokmi sa moderní ľudia dostali mimo územie Afriky. Ale zdá sa, že táto expanzia nepokračovala.

Genetické zdroje naznačujú, že ďalšia skupina odišla z Afriky pred 80 000 až 50 000 rokmi a že jej príslušníci boli predkami všetkých neafrických etník. Výskumy naznačujú, že sa dostali do Ázie cez Arabský polostrov. Vo vrchnom paleolite nastal jeden z masívnych nárastov populácie, čo mohlo viesť týchto ľudí k hľadaniu si nových lovných oblastí.

Prvý dôkaz pochádza od istého muža, ktorý žil asi pred 31 000 až 79 000 rokmi a mal pomerne významnú náhodnú mutáciu vo svojom chromozóme Y. Bol pomenovaný znakom M168. Je to akýsi prvý muž tzv. "Adam" a predok všetkých mužov neafrického pôvodu.

Do dneška sa dochovali dve časti africké mutační línie: M91 a M60. Výrazným africkým znakom je M91 (Mapa 1). Je to priama genetická línia vedúca k najstarším obdobiam moderného človeka, a teda k pôvodnému "Adamovi". Mnoho súčasných ľudí s týmto znakom hovorí pôvodnými jazykmi bušmenov či krovákov. M60 sa dnes nachádza najviac na východe strednej Afriky, M91 skôr na západe a juhu.

Osídľovanie Austrálie

Zaujímavé je, že na vetvách chromozómu Y nasledujú hneď po vetvení M168 ďalšie tri vetvenia. Dve z nich pritom rozdeľujú eurázijské vetvy na rôzne skupiny. Tá tretia, ktorá je definovaná znakom YAP alebo MI, sa nachádza hlavne v Afrike. Pri skúmaní dvoch neafrických vetiev sa zistilo, že jedna z týchto skupín je bežnejšia ako ostatné. Sú to ľudia rozšírení po celom svete a teda v Európe, Indii alebo aj v Južnej Amerike. Ľudia sú teda od seba tak vzdialení, a pritom ich spája určitý spoločný znak. Ľudia vzácnejšej línie sa v súčasnosti nachádzajú v Ázii, Austrálii a v Amerike.

Znak M130 (Obr. 1), nadväzujúci na znak M168, pravdepodobne sledoval pobrežnú líniu a v súčasnosti ho sledujeme v Ázii a Amerike. Podstatnú časť populácie tvorí v Austrálii (pôvodní obyvatelia). Objavy na dvoch dávnych náleziskách, artefakty z Malakananje a fosílie od jazera Mungo ukazujú, že do Austrálie prišli prví ľudia dnešného typu pred viac ako 50 000 až 60 000 rokmi. Je to neuveriteľné, pretože to znamená že hneď ako ľudia vykročili z Afriky, museli sa pomerne rýchlo dostať až do Austrálie. Podľa genetického datovania sa vtedy naši predkovia nevyskytovali nikde mimo Afriky. To znamená, že moderní ľudia použili cestu, ktorá umožnila veľmi rýchly presun. Pozostatky, ktoré sú mladšie než austrálske, nachádzame po celej predpokladanej pobrežnej cesty do Austrálie. Je zaujímavé, že čím bližšie sme k cieľu, tým staršie artefakty nachádzame.

Osídľovanie Ázie

Bezprostredne po M168 sa na trase do Eurázie objavil i znak M89. Tento znak geneticky spája populácie, ktoré žijú v severovýchodnej Afrike (Etiópia, Sudán) s populáciami v Levante (Stredomorie). 90 až 95% všetkých ľudí žijúcich mimo Afriky môže zaradiť svoj pôvod k tejto druhej migrácii cez stredný východ. Tento znak nachádzame od východného Francúzska až po Kóreu.

Keďže vtedajší ľudia boli lovcami, ich hlavným cieľom bolo nasledovanie zveri po stepiach Eurázie. Práve tu sa v období pred 40 000 rokmi v oblasti dnešného Iránu objavuje v línii M89 nový znak M9. Ľudia s týmto znakom sa postupne v priebehu ďalších 30 000 rokov rozšírili až na koniec pevniny. Nazývajú sa aj Eurázijský klan, ktorý zaľudnil väčšinu planéty. V súčasnosti väčšina ľudí, žijúca na severnej pologuli, je nositeľom znaku M9.

Pred asi 40 000 rokmi ľudia prenikli severným smerom do Strednej Ázie a prišli do trávnatých stepí severne od Himalájí. Zároveň putovali na sever cez juhovýchodnú Áziu a Čínu, aby potom dospeli do Japonska a na Sibír. Genetické vodítka naznačujú, že malé skupiny týchto ľudí nakoniec migrovali na americký kontinent zo severnej Ázie.

V Ázii bolo objavené nespočetné množstvo rôznych mutačných znakov. Takmer neprekonateľnou bariérou sa stal Pamírsky uzol, ktorý rozdelili eurázijských pútnikov na dve skupiny. Jedna putovala na sever od Hindukúša a druhá na juh do dnešného Pakistanu a na indický polostrov.

Ľudia smerujúci na juh mali na svojom chromozóme ďalšiu mutáciu so znakom M20. Najväčší výskyt ľudí s týmto znakom sledujeme v Indii (presnejšie v južnej Indii) a je to asi 50%. Týchto ľudí môžeme nazvať indický klan. Táto línia sa preto určite musela stretnúť s pobrežnou, ktorá sem prišla dávno

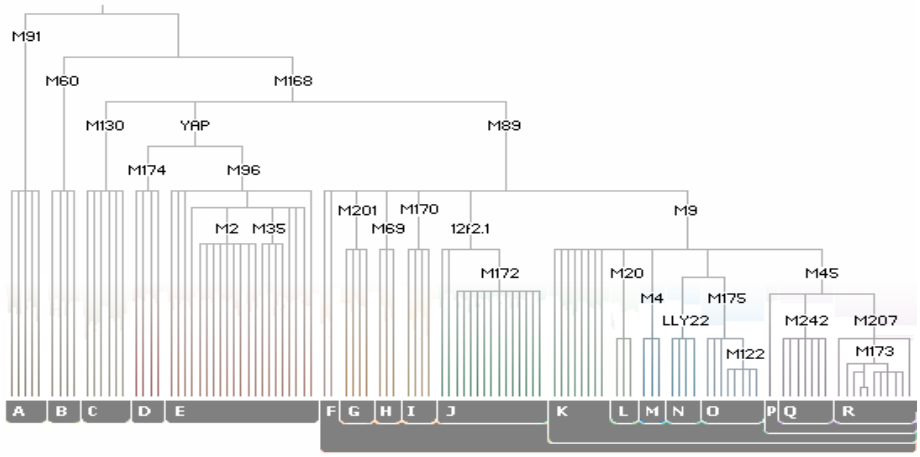
predtým. Ďalší znak sa objavil na severnej trase a bol nazvaný M45. Tento znak sa objavuje u ľudí na sever od Hindukúša a v strednej Ázii. Mutácia so znakom M45 sa objavila asi pred 35 000 rokmi. Potomkovia tohto klanu sa objavujú na Blízkom východe, východnej Ázii a s trochu väčšou početnosťou v Indii.
Súčasní obyvatelia Východnej Ázie sú nositeľmi znaku M175. Ich súčasní potomkovia sú rozšírení iba vo východnej Ázii, pričom sa vôbec nevyskytujú v západnej Ázii ani v Európe. V súčasnosti sa tento znak s najväčšou početnosťou vyskytuje v kórejskej populácii (okolo 30%) a v Číne tvorí 80 až 90%všetkých ľudí.

Osídľovanie Európy
Keď sa prví moderní ľudia objavili pred 50 000 až 40 000 rokmi v Európe, stretávali sa s neandertálcami, ktorí tam žili už tisícky rokov. Nikto presne nevie, ako tieto dve skupiny na seba pôsobili a čo spôsobilo vyhynutie neandertálcov pred asi 30 000 rokmi. Žiadne genetické svedectvá však nenaznačujú, že moderní ľudia majú nejakých neandertálskych predkov.
Pôvodné populácie na Blízkom východe pravdepodobne len málo ovplyvnili osídlenie Európy. I napriek tomu, že by to bolo najvhodnejšie miesto pre vstup do Európy. Jedným zo znakov, ktoré sa na chromozómoch Y u mužov Európanov nachádzajú je znak M173. Asi pred 35 000 rokmi línia nesúca tento znak predstavovala prvé veľké osídlenie Európy modernými ľuďmi. S vysokou frekvenciou sa nachádza po celej západnej Európe. Ďalšie významné línie chromozómu Y, ktoré sa v Európe nachádzajú, sú mladšie ako M173 . To znamená, že M173 je pravdepodobne znakom prvých moderných Európanov. Predchodcom M173 je M45, ktorý tvorí z Európanov podskupinu klanu strednej Ázie.
Veľkú početnosť v juhovýchodnej a centrálnej Európe má ďalší znak M170. Tento znak je tiež veľmi bežný u mužských obyvateľov Balkánu. Objavuje sa na línii so znakom M89. Uvažuje sa nad tým, že nositeľmi tohto znaku boli i Kelti (500 rokov p. n. l.). Keltskú expanziu z centrálnej Európy na západ mohlo spôsobiť rozšírenie tejto línie do mnohých dnešných lokalít.
Za prvých farmárov sú považovaní prví nositelia znaku M172. Táto línia vznikla asi pred 15 000 až 10 000 rokmi počas neolitickej revolúcie. Pravdepodobne viedli k vzniku usadených komunít a miest. Nachádzame ho na severe Afriky a juhu Európy a tiež na strednom východe.

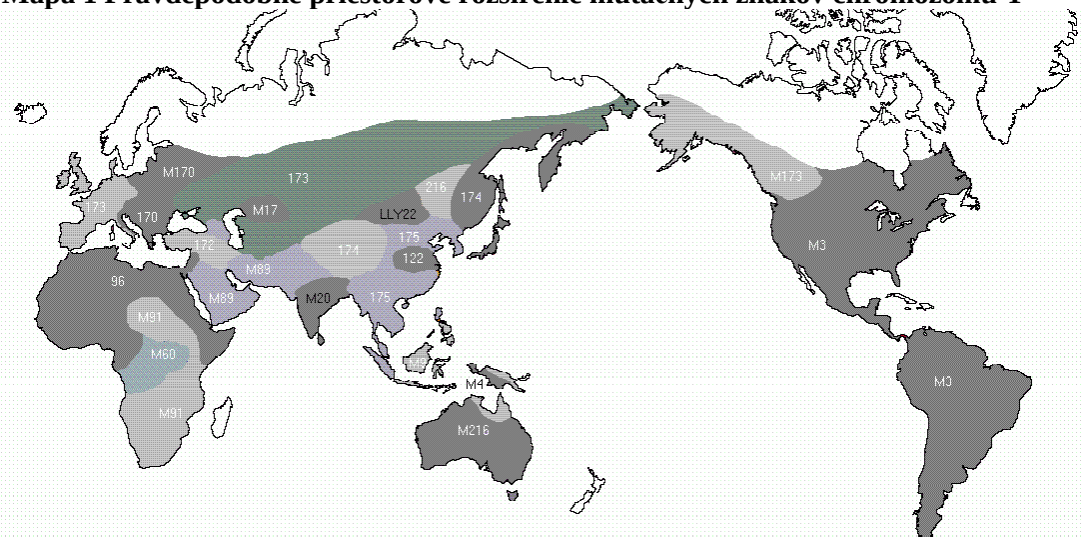
Osídľovanie Ameriky
O tom, kedy prišli ľudia na americký kontinent, sa diskutuje, ale genetika naznačuje, že to bolo asi pred 25 000 až 15 000 rokmi, kedy hladina morí bola nízka a Sibír bola súšou spojená s Aljaškou. Niektoré nálezy sa síce zdajú byť staršie ako 50 000 rokov, ale najlepšie datované dôkazy z Chile a USA nie sú staršie ako 19 000 rokov.
Z hľadiska osídlenia Ameriky je zaujímavé pochopiť smery putovania prvých ľudí. Na rozdiel od iných kontinentov tu totiž existovala iba jedna možná vstupná brána pre nových migrantov.
Podľa výskumov však nebola Amerika osídlená naraz. Osídľovanie prebiehalo v niekoľkých vlnách (presnejšie dvoch). Tá prvá viedla k osídleniu ako Severnej tak Južnej Ameriky, zatiaľ čo druhá vlna zanechala genetické stopy iba v Severnej Amerike.
Znak, ktorý je bežný v oboch Amerikách sa nazýva M3. Tento znak nesie viac ako 90% skúmaných obyvateľov Južnej a Strednej Ameriky, zatiaľ čo z obyvateľov Severnej Ameriky malo túto líniu asi iba 50%. Znak M3 však nebol v Ázii nájdený. Pri sledovaní predchodcu znaku M3 bola zistená líniová príbuznosť znaku M45. Toto bol znak zo strednej Ázie, z ktorého vznikol neskôr i znak M173 pre Európsku vetvu.
Nedávna analýza línie M45, ktorú uskutočnil Mark Seielstadem, určila ďalší znak M242, ktorý je potomkom M45. Tento vznikol asi pred 20 000 rokmi v strednej Ázii. Je rozšírený v celej Ázii, z južnej Indie cez Čínu až na Sibír a v Amerike. Jeho najvyššia početnosť bola zistená na Sibiri (sibírsky znak). Je bezprostredným predchodcom znaku M3. Najstarším genetickým znakom v Amerike je teda M242. Ameriky spája vysoká početnosť znakov M242 a M3, čo ich zaraďuje k príslušníkom sibírskeho klanu.
Druhú migračnú vlnu tvoril pravdepodobne pobrežný znak M130. V Južnej Amerike sa však znak M130 nevyskytuje. Genetické dáta nasvedčujú tomu, že sa dostal do Severnej Ameriky v priebehu posledných 10 000 rokov a pochádza z oblasti severnej Číny alebo juhovýchodnej Sibíri.

Za 40 000 rokov teda človek doputoval zo severnej Afriky až do Ameriky. Behom tejto púte človeku dobre poslúžila jeho vynaliezavosť a schopnosť adaptácie na život v rôznych podmienkach.

Obr. 1 Genetický strom postupnosti mutačných znakov chromozómu Y



Mapa 1 Pravdepodobné priestorové rozšírenie mutačných znakov chromozómu Y



Použitá literatúra:

Shreeve, J. (2005). Kudy kráčel člověk? National Geographic Česko, november, 36 – 47

Wells, S. (2005). Adam a jeho rod (Genetická odysea člověka). Praha (nakladateľstvá Dokořán a Argo)

National geographic society (1996 – 2005) a. Atlas of the human journey. [online].Dostupné na internete: <https://www3.nationalgeographic.com/genographic/atlas.htm>

Prírodovedecká fakulta Univerzity Komenského
2. ročník magisterského štúdia

Lucia Hoffmanová
Tupolevova 13
Bratislava
851 01
hoffman@chello.sk

Why seasonal adjustment?

Beáta Horváth (Hungarian Central Statistical Office)

Eötvös Loránd University, Budapest, 7th semester in second degree

Statistics represent nowadays a key tool for economic policy-making, business cycle analysis modelling, and forecasting. However, short-term statistics are often characterised by seasonal fluctuations and other calendar/trading-day effects, which can mask relevant short and long-term movements of the series, and hinder a clear understanding of economic phenomena, like trend, turning points and consistency between other indicators. The main aim of seasonal adjustment is to remove changes that are due to seasonal or calendar influences to produce a clearer picture of the underlying behaviour.

Brief Historical Background

The analysis of the components of time series has a long history going back to work in astronomy, meteorology, and economics in the 17th through 19th centuries, and to early seasonal analysis by Buys-Ballot (1847). The early work concentrated on first removing the spurious correlation between two variables.

The first overall seasonal adjustment methodology was created by Macauly (1930). This approach is nowadays commonly referred to as "Classical Decomposition" and laid the foundations of many modern-day approaches including the X11-ARIMA method.

Two major developments came during the early 1950s. The first one was the introduction of exponential smoothing techniques which simplified the tedious computations previously needed. The second development was the introduction of computers, which also provided an impetus to decomposition methods since calculations that previously took days could be performed in a few seconds.

The modelling of time series can be traced back at least to Yule (1927) who introduced autoregressive models and Slutsky(1937) who proposed moving average models. It was up to Wold (1938) to fit such moving average models to data and he also described the use of mixed ARMA models (1954). Box and Jenkins (1970) findings provided a set of criteria to determine the type of and order of an ARIMA model, which should be applied, to any time series.

By ARIMA (Auto Regressive Integrated Moving Average) model based approach, one starts by modelling the time series and derives the models for the components from this estimated model. In the structural model based approach, one starts directly with the estimation of the components (Engle (1978), Harvey and Todd (1983)). The most received structural models are BAYSEA, DECOMP and STAMP.

ARIMA modelling used non-seasonal and seasonal differencing in order to model non-stationary series. The first practical realization of this modelling took place in the Bank of England in the 1980s. Further developments were made at the Bank of Spain under the control of Augustin Maravall, and resulted in the TRAMO/SEATS program (Gómez and Maravall).

About seasonality

A seasonal effect represents intra-year fluctuations more or less stable year after year with respect to timing, direction and magnitude.

The large seasonal movements are masking smaller movements that could be to analyze series. Seasonal adjustment makes easier to see smaller movements in the series, to determine turning points and to compare different series. If we are able to remove the seasonal fluctuations, then masking effects would be gone and it would be have a better view of the behaviour of each series. For a time series to be analysis we need data that are comparable over time and across domain.

Month to month (or quarter to quarter) change of the seasonally adjusted data provides more meaningful comparison over short time frame. Seasonal adjustment contributes to make international comparability in economies and sectors, because for example when we are freezing in winter, Australians are burning on the beach.

Many of the important questions in economics involve understanding whether the economy or particular aspects of the economy are in growth or decline. A fundamental task in economics is predicting business cycles or understanding how far along the economy is in a particular business cycle.

Possible causes of seasonality include **natural factors** such as changes in weather including temperature, hours of daylight, agriculture and heating. **Administrative measures** for example the starting and ending dates of school years or tax deadlines cause seasonal change. **Social, cultural and religious traditions** affect seasonality such as the effect of Christmas or Easter, and last but not least **indirect seasonality** can be done due to the seasonality that affects other sectors such as toy industry is affected a long time before Christmas.

There are a variety of problems that can arise in time series data that will affect the quality of the seasonal adjustment like **outliers** which are extreme values that usually have identifiable causes, such as strikes, war, or extreme weather conditions which can distort the seasonal adjustment; **trend breaks** (also known as level shifts) where the trend component suddenly increases or decreases in value or **seasonal breaks** where there are changes in the seasonal pattern.

Decomposition of a Time Series

The basic goal of seasonal adjustment is to decompose a time series into a several components for the purpose of removing seasonality. The time series can be decomposed into four main unobserved components:

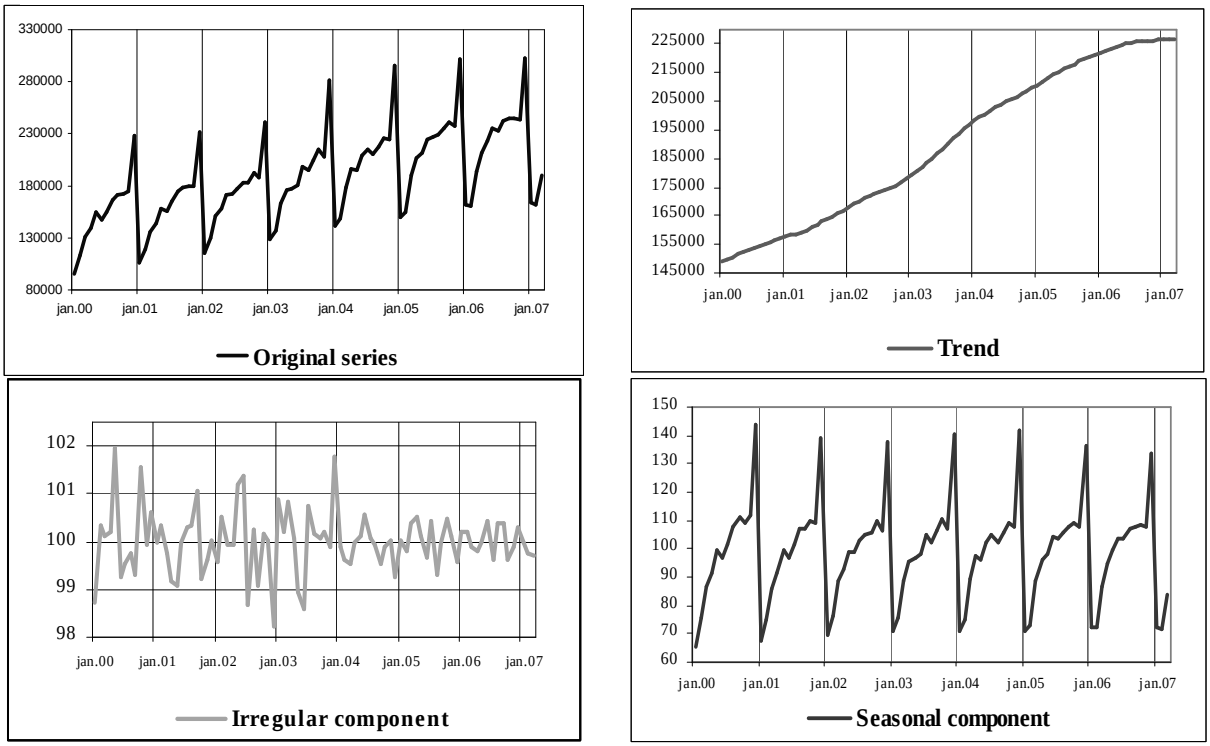
- **Trend (*T*)** indicates the long-term tendency, represents the structural variations of low frequency in a time series.
- **Cyclical component (*C*)** indicates the medium term fluctuation. The cyclical component is worth examining only in case of very long time series. In accordance with the general practice, the trend component is assumed to include also the cyclical component. Sometimes the trend and cyclical components together are called as trend-cycle.
- **Seasonal component (*S*)** is that part of the variations in a time series which represents intra-year fluctuations more or less stable year after year with respect to timing, direction and magnitude. It is also referred to as the seasonality of a time series. It reflects normal variations that recur every year to the same extent, e.g. weather fluctuations that are representative of the season, length of months, Christmas effect,

etc. It may also include calendar related systematic effects that are not regular in their annual timing and are caused by variations in the calendar from year to year.

– **Irregular component (*I*)** includes unpredictable effects, which are considered as random variables; it is assumed that the expected value of these factors is 0 (for an additive model) or 1 (for a multiplicative model). The irregular component of a time series is the residual time series after the trend, the cyclical and the seasonal components (including calendar effects) have been removed.

For an example[1] of the decomposition see Figure 1 graphed from January 2000 to March 2007.

Figure 1: Components of a Time Series



The previous mentioned components may be mutually linked in several ways. The most frequently specified models are the additive and the multiplicative model.

Additive model is used when the components are linked additively:

$$Y = T + C + S + I,$$

and multiplicative model is used when the components are linked through multiplication:

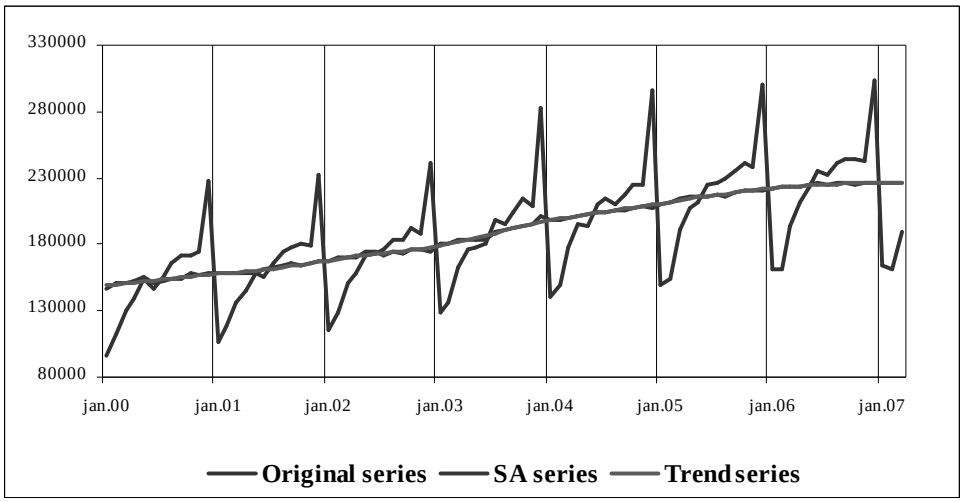
$$Y = T \cdot C \cdot S \cdot I.$$

The seasonally adjusted time series (SA) is calculated with the help of the above mentioned components. In this case the initial time series is adjusted for seasonal variations (including

[1] Source: Hungarian Retail Sale of Non Food Product

calendar effects, if present), that is either the value of the seasonal effect related to the given time period is deducted from the initial time series (SA=Y–S) or the initial time series values are divided by the seasonal component (SA=Y/S). Consequently the time series obtained includes the trend and random components. The resulting seasonally adjusted series would not have the large seasonal movements which can mask relevant information.

Figure 2: Types of a Time Series[2]



The desire for economic measure independent of seasonal variations led research to develop methods for adjusting economic data by estimating and removing the effect of seasonal changes from the original data. Seasonal adjustment is widely used in official statistics for enabling timely interpretation of time series data. The most commonly used methods in the European Statistical System are TRAMO/SEATS and X12-ARIMA.

Seasonal adjustment is a very complex procedure that does not provide one unique solution. To produce seasonal adjustment it is also need to take into account other calendar effects, like working or trading day effect. Since there is more than one solution it is helpful to have diagnostics to evaluate different possible adjustments.

Bibliography

Fischer, B. (1995). Decomposition of Time Series – Comparing Different Methods in Theory and Practice. Luxembourg.

Bauer, P., Földesi, E. (2005). Szezonális kiigazítás. Módszertani Füzetek. KSH (Seasonal Adjustment Methods. Methodological Paper. Hungarian Central Statistical Office. only in Hungarian)

[2] Source: Hungarian Retail Sale of Non Food Product graphed from January 2000 to March 2007.

SDMX (2006). SDMX Content-Oriented Guidelines: Metadata Common Vocabulary. Draft.

Eurostat (2007) Seasonal Adjustment: Methods and Practices. Handbook written by Bauer, P., Földesi, E., Horváth, B., Urr, B.

beata.horvath@ksh.hu

Hungarian Central Statistical Office

H-1024 Budapest

Keleti Károly u. 5-7

Vývoj akciových trhov krajín V4 a parciálna kointegrácia

Peter Hrubina

Abstract:
The aim of this paper is to reveal the issues regarding V4 countrie's stock markets in comparison with the stock markets of Western European Union countries. I will mainly focus on average revenues, volatility and partial cointegration as factor of long-term relationship between the 'new' and the 'old' European Union countries. Particularly, I will analyze stock markets of the following countries: Slovak Republic, Czech Republic, Poland, Hungary, Germany and France.

Key words: Stock market index, Revenues, Volatility, Cointegration, Cointegration vector

1. Úvod

Cieľom tohto príspevku je popis vývoja akciových trhov krajín V4 v porovnaní so „staršími" členskými krajinami EÚ, konkrétne Nemeckom a Francúzskom. Budem pritom, v prvej časti, zisťovať priemerný ročný výnos a riziko na týchto trhoch, dlhodobý vzťah týchto trhov na základe metodiky kointegrácie v druhej časti tejto práce. Štruktúru práce rozdelím na zistenia plynúce z overovania nasledovných troch hypotéz:

H_1: Riziko a výnos Francúzska a Nemecka je v priemere nižší ako je tomu v krajinách V4.
H_2: Vývoj akciových trhov krajín V4 je menej korelovaný ako v prípade starších členských krajín EÚ.
H_3: Vývoj akciových trhov Nemecka a Francúzska je kointegrovaný, pričom medzi Slovenskom a Českom neexistuje kointegračný vzťah[1].

Časové rady som zvolil na dennej frekvencii obchodovaných dní od 01. 04. 1999 do 01. 10. 2007. Pracovné prostredie je rozhranie ekonometrického softvéru Eviews 4.1 Student Edition a program MS Excel 2003.

Vzhľadom na obmedzený rozsah tejto práce, budem akciové trhy daných krajín aproximovať na akciové indexy. Vývoj príslušných akciových indexov budem prirovnávať k vývoju akciových trhov daných krajín. Pre každú krajinu boli zvolené nasledovné indexy: Slovenská republika – SAX, Česká republika – PX, Maďarsko – BUX, Poľská republika – WIG, Francúzsko - CAC 40 a Nemecko – DAX 30.

2. Vývoj analyzovaných akciových indexov

Analyzované indexy šiestich krajín sú indexy *blue chip* emisií, teda emisií, resp. aktív s relatívne nízkym rizikom a pomerne stabilným výnosom. Všetky sú kapitálovo vážené so základnou hodnotou stanovenou k začiatku kótovacej histórie daného indexu[2]. Konkrétne informácie o daných indexoch, teda o ich metodike výpočtu resp. zložení obchodovaných tituloch, môže čitateľ nájsť na hlavných stránkach národných búrz analyzovaných krajín. Potrebné je však povedať, že medzi indexami existujú veľké rozdiely z hľadiska trhovej kapitalizácie obchodovaných titulov či už samotný počet týchto titulov. Kým Slovenský SAX reprezetnuje 5 spoločností resp. Český PX 11 tak Poľský WIG odzrkadľuje vývoj 312 titulov. Maďarský BUX reprezentuje 12 titulov, Nemecký DAX s 30 a Francúzsky CAC so 40 titulmi[3].

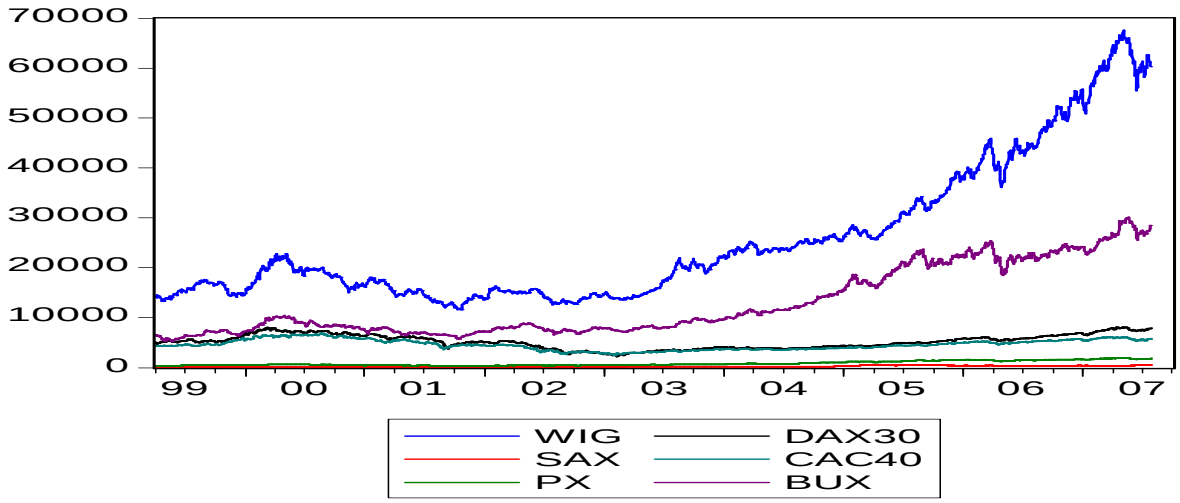
3. Hypotéza číslo 1

Predpokladá sa, že riziko a výnos Francúzska a Nemecka sú v priemere nižšie ako je tomu v prípade krajín V4.

---

[1] V prípade potvrdenia existencie kointegrácie si znázornime kointegračný vektor.
[2] Napríklad Slovenský SAX sa začal kótovať 14. 09. 1993 s bázickou hodnotou 100 b.b. Český proťajšok PX má základnú hodnotu 1000 b.b., ktorá sa viaže na 5. 4. 1994.
[3] Výber 40 titulov je založený na ohodnotení výkonnosti titulov a výberu najlepších 40 z množiny 100 titulov.

Denné hodnoty indexov od apríla 1999 po september 2007 sú nasledovné.



Graf č. 1: Historický vývoj hodnôt všetkých titulov.
Zdroj: Vlastné spracovanie.

Všetky indexy, okrem poľského WIG-u a maďarského BUX-u, vykazujú určitú stabilitu. Taktiež je zrejmý trend resp. rastúci priebeh denných hodnôt smerom k novším hodnotám. To do akej miery sú dané časové rady korelované resp. kointegrované zistím pri overovaní ostatných dvoch hypotéz. Denné hodnoty akciových indexov som upravil na stacionárny časový rad relatívnych zmien logaritmizáciou. O logaritmických výnosoch predpokladám, že sa riadia jednorozmerným normálnym rozdelením $N(\bar{u},\sigma^2)$[4]. Denné výnosy a variancie som upravil metódou Square-root time rule na ročné. Priemerné ročné výnosy a smerodajné odchýlky sledovaných titulov prezentuje tabuľka číslo 1.

	SAX	PX	BUX	WIG	CAC 40	DAX 30
E(r)year	17,19%	18,31%	16,99%	16,63%	3,58%	5,08%
S(r)year	21,01%	19,65%	22,57%	20,54%	21,83%	24,32%

Tabuľka č. 1: Výnos a riziko analyzovaných indexov.
Zdroj: Vlastné spracovanie.

Môžeme vidieť, že akciové trhy Slovenskej republiky, Českej republiky, Poľska a Maďarska dosahujú takmer identickú výkonnosť. Taktiež je zrejmé, že Francúzsko a Nemecko dosahujú najnižší výnos, pri relatívne najvyššom riziku. Na lepšiu orientáciu údajov z tabuľky číslo jeden prikladám priemerné hodnoty za skupiny krajín.

	V4	OC
E(r)year	16,70%	4,22%
S(r)year	20,59%	22,89%

Tabuľka č. 2: Výnos a riziko pre skupiny krajín.[5]
Zdroj: Vlastné spracovanie.

Pre overenie hypotézy som použil párové testy o zhode dvoch stredných hodnôt resp. disperzií, kde som definoval nasledujúce hypotézy napríklad: H_0: $\bar{u}_{V4}= \bar{u}_{OC}$ a H_1: $\bar{u}_{V4}> \bar{u}_{OC}$. Disperzie analogicky. Som si vedomý faktu, že pri párovom porovnávaní individuálna hladina významnosti nie je korektná, nakoľko celková hladina významnosti zistenej skutočnosti je aproximáciou pochybení všetkých párových testov. Na všetkých bežných hladinách významnosti môžem nulové hypotézy zamietnuť.

[4] Na základe testov normality(Jarque-Bera) sme zistili, že jednorozmerné normálne rozdelenie nie najpriliehavejšie popisuje dané časové rady. Do úvahy prichádza troj-parametrické Studentovo T-rozdelenie.
[5] V4 – krajiny Vyšegradskej štvorky, OC – „staré“ krajiny Európskej únie.

Na základe horeprezrentovaných údajov, **môžem potvrdiť**, že výnosy akciových trhov krajín V4 dosahujú v priemere vyššie hodnoty ako staré krajiny Európskej únie a riziko taktiež.

4. Hypotéza číslo 2
Predpokladal som, že vývoj akciových trhov krajín V4 je menej korelovaný ako v prípade starých krajín Európskej únie krajín.

Na analýzu danej hypotézy som použil korelačné matice. Skúsil som korelovať vývoj hodnôt indexov a taktiež vývoj výnosov. Korelačnú maticu výnosov analyzovaných indexov prezentuje tabuľka číslo 3.

	ESAX	EPX	EBUX	EWIG	ECAC40	EDAX30
ESAX	1.000000	0.027783	0.019767	0.038272	-0.035870	0.015079
EPX	0.027783	1.000000	0.035411	0.216401	0.013261	-0.017479
EBUX	0.019767	0.035411	1.000000	0.069289	-0.002644	-0.011120
EWIG	0.038272	0.216401	0.069289	1.000000	0.014367	0.036390
ECAC40	-0.035870	0.013261	-0.002644	0.014367	1.000000	0.071817
EDAX30	0.015079	-0.017479	-0.011120	0.036390	0.071817	1.000000

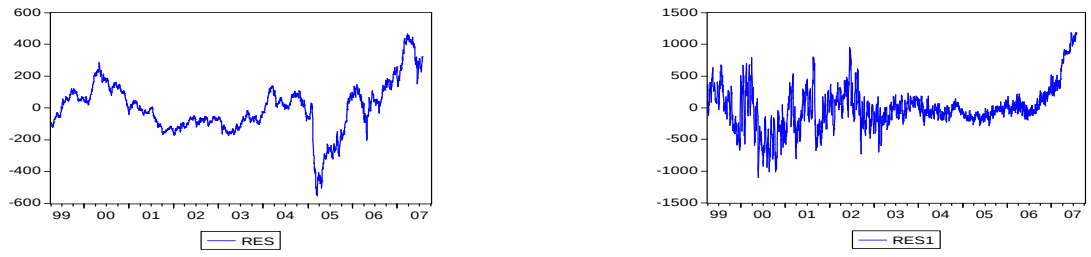
Tabuľka č. 3: Korelačná matica výnosov analyzovaných indexov.
Zdroj: Vlastné spracovanie.

Priemerná hodnota korelácie akciových trhov medzi Nemeckom a Francúzskom presahovala o 7 percentuálnych bodov koreláciu akciových trhov v rámci krajín V4.

Na základe tohto zistenia **môžem hypotézu,** o vyššej korelácií v rámci krajín V4 ako medzi starými krajinami Európskej únie **potvrdiť**.

5. Hypotéza číslo 3
Ďalej som predpokladal, že vývoj akciových trhov Nemecka a Francúzska je kointegrovaný, pričom medzi Slovenskom a Českom neexistuje kointegračný vzťah[6].

Tretia hypotéza predpokladá rovnaký vývoj hodnôt akciových indexov v Nemecku a Francúzsku v dlhodobom období a nepredpokladá takýto vzťah u krajín V4(partikulárne Slovenskej repuliky a Českej republiky). Danou hypotézou sa snažím zistiť, či historický vývoj starých krajín Európskej únie vykazuje určitú spoločnú trendovosť resp. integritu. Teda, že tieto „krajiny majú rovnaký vývoj v zhruba rovnakom čase.“[7] Zvolil som metodiku, založenú na testovaní stacionarity/nestacionarity reziduálnej veličiny regresného modelu najmenších štvorcov dvoch indexov a taktiež som sa pokúsil nájsť kointegračný vektor. Pričom je predpoklad, že medzi Slovenskom a Českom kointegračný vektor nebude existovať.

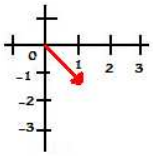


Graf č. 3: Reziduálna veličina regresného vzťahu metódy OLS PX a SAX(RES). Reziduálna veličina regresného vzťahu DAX 30 a CAC 40(RES1).
Zdroj: Vlastné spracovanie.

[6] V prípade potvrdenia kointegrácie znázorním kointegračný vektor.
[7] William H. Greene – Econometric analysis, 3rd edition, strana 852.

Pre úplnosť overenia som ešte aplikoval Johansenov kointegračný test. V prípade, že časové rady reziduí sú stacionárne, potom kointegračný vektor je [1; -β][8].

Na základe ADF testu vykonaného na časových radoch reziduí regresného vzťahu PX a SAX ($\varepsilon_t = PX_t - \beta.SAX_t$) môžeme prijať nulovú hypotézu o nestacionarite časového radu[9]. T.j. vývoj Slovenského akciového indexu a Českého akciového indexu nie je kointegrovaný. Teda neexistuje ani kointegračný vektor. Pokiaľ ide o časové rady reziduí regresného vzťahu DAX 30 a CAC 40 ($\varepsilon_t = DAX30_t - \beta.CAC40_t$) nemôžeme na základe hodnoty ADF štatistiky a kritických hodnôt alternatívnu hypotézu o stacionarite zamietnuť[10]. T.j. vo vzťahu akciových indexov Nemecka a Francúzska existuje kointegračný vzťah s vektorom o súradnici [1;- 1,31].



Obrázok č. 1: Kointegračný vektor medzi DAX30 a CAC40.
Zdroj: Vlastné spracovanie

Johansenov kointegračný test[11] jednoznačne potvrdil existenciu kointegračného vzťahu medzi Nemeckým a Francúzskym akciových trhom. A taktiež jednoznačne nezistil kointegrovanosť Slovenského a Českého akciového trhu.

Na základe zistených faktov **môžeme potvrdiť hypotézu číslo 3** o existencii kointegrácie medzi vývojvom akciových trhov Nemecka a Francúzska a neexistencii kointegračného vzťahu medzi vývojom akciových trhov Slovenska a Českej republiky.

6. Literatúra

Granger, C.W. – Newbold, P. 1974. Spurious Regression in Econometrics. In: Journal of Econometrics, č. 2, 1974, s. 111 – 120.
Green, W. h. 1997. Econometric Analyses. Londýn: Prentice – Hall, 1997. 1076 s. ISBN 0-13-7246659-5.
Kanderová, M. 2004. Metódy prognózovania sezónnosti v ekonomických časových radoch. In: ACTA FACULTATIS AERARI PUBLICI. 2004, č. 1. Banská Bystrica : Fakulta financií UMB, 2004, s. 53 – 60.
Verejné dostupné databázy historických dát sledovaných indexov, oficiálne stránky búrz, www.bcpb.sk, www.bcpp.cz, deutsche-boerse.com, www.money.pl, www.bse.hu, finance.yahoo.com, 2007

Adresa autora:
Peter Hrubina
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
peterhrubina@gmail.com

[8] β koeficient odhadnutý z modelu $\varepsilon_t = PX_t - \beta.SAX_t$ resp. $\varepsilon_t = DAX30_t - \beta.CAC40_t$
[9] Augmented Dickey-Fuller Test, Lagový operátor – L1 až L4. Testovacia štatistika ADF = -1,57. Na všetkých bežných hladinách významnosti bola testovacia štatistika ADF väčšia ako príslušné kritické hodnoty, čo nezamieta nulovú hypotézu o existencii jednotkového koreňa(unit root) a teda nestacionarite časového radu.
[10] Existencia jednotkového koreňa sa nepotvrdila.
[11] Green, W. h. 1997. Econometric Analyses, strana 873

Populačná explózia v rozvojových krajinách od roku 1950

Andrej Chromeček

Abstrakt:
The different demographic situations facing developed and developing countries today reflect the population trends of the 20th century, and especially the past 50 years. These trends not only shaped the current profile of these countries' populations, but also will influence their demographic futures.

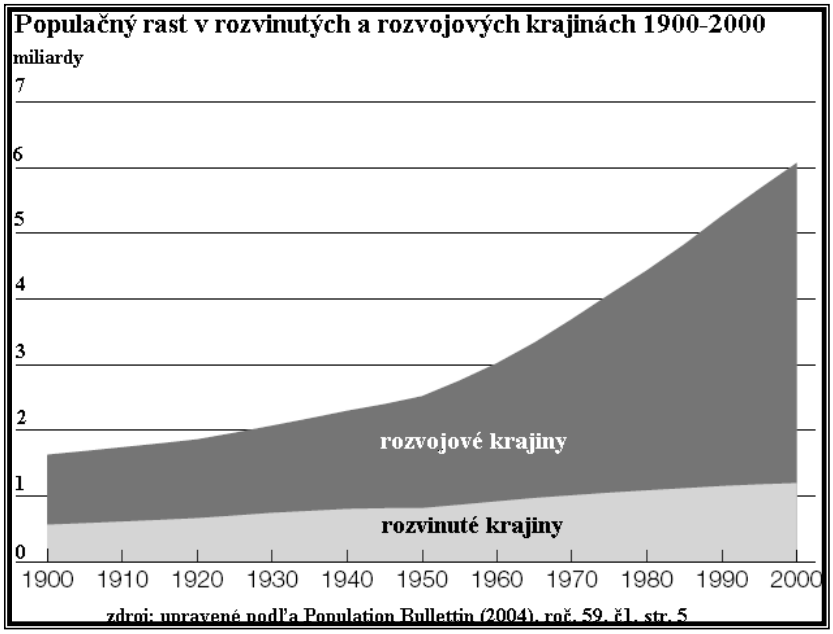
Úvod:

Druhá polovica 20. storočia nám priniesla niekoľko nových trendov v demografickom správaní. Na jeho konci už takmer všetky vyspelé krajiny prekonali demografický prechod k nízkym mieram pôrodnosti a úmrtnosti. Prirodzené prírastky, ktoré dosahujú sú nízke a v niektorých krajinách dochádza aj k poklesu počtu obyvateľov. Populácie v týchto štátoch tiež rýchlo starnú. Naopak rozvojovým krajinám priniesla druhá polovica minulého storočia rapídny populačný rast. Niektoré rozvojové krajiny postupne rôznym tempom prechádzajú jednotlivými štádiami demografickej revolúcie po vzore rozvinutých krajín, zatiaľ čo iné sledujú nové cesty premeny nemajúce obdobu v demografickej histórii rozvinutých krajín.

Klasifikácia:

Podľa klasifikácie Organizácie Spojených Národov medzi rozvojové krajiny patria všetky krajiny Afriky, Latinskej Ameriky a Karibiku, Ázie (okrem Japonska), a Oceánie (okrem Austrálie a Nového Zélandu)

Výsledky a diskusia:



Vzory pôrodnosti a úmrtnosti v rozvojových krajinách sú veľmi odlišné v porovnaní s rozvinutými krajinami za posledných 50 rokov. Nárast strednej dĺžky života po narodení, a s ňou spojený pokles dojčenskej úmrtnosti nastal vo väčšine rozvojových krajín až po roku 1950. V päťdesiatych a šesťdesiatych rokoch minulého storočia sa začali rozpadávať koloniálne impériá európskych štátov, a novovzniknuté krajiny sa museli potýkať s množstvom sociálnych, ekonomických a v neposlednom rade aj demografických problémov.

Vývoj pôrodnosti v rozvojových krajinách v druhej polovici 20 storočia bol v znamení poklesu. Tento pokles bol dokonca ešte rýchlejší ako v rozvinutých krajinách. Avšak

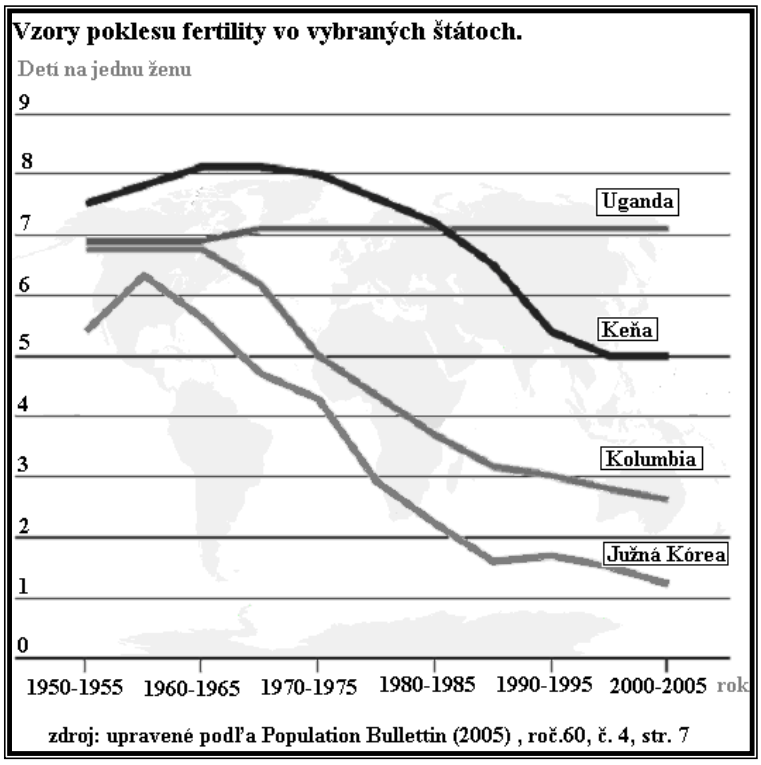
východiskové hodnoty v rokoch 1950-1955 boli oproti rozvinutým krajinám rádovo 2x vyššie. V súčasnosti je pôrodnosť v rozvojových krajinách tiež približne dvojnásobná v porovnaní rozvinutým svetom a jej hrubá miera dosahuje 22,4‰. Pokiaľ však spomedzi rozvojových štátov vylúčime Čínu ktorá nám veľkosťou svojej populácie veľmi zaťažuje výslednú hodnotu, priemerná hrubá miera pôrodnosti v týchto krajinách nám vyskočí až na 27‰. Päťdesiatka najchudobnejších krajín však vykazuje hrubú mieru pôrodnosti až na úrovni 36‰. Takéto veľmi vysoké čísla sú typické práve väčšinu krajín v Afrike, Ázii a Latinskej Amerike, ktoré ostávajú na nízkom hospodárskom stupni a kde stále prevláda poľnohospodárska výroba. Spoločenské normy preferujú mnohopočetné rodiny. Myšlienka ovplyvňovania plodnosti tu doteraz nie je príliš rozšírená, alebo akceptovaná. Preto miery pôrodnosti ostávajú stále vysoké a v niektorých oblastiach narastajú ako dôsledok zlepšenia zdravotnej starostlivosti o matky. Tieto vysoké miery pôrodnosti spolu s klesajúcou úmrtnosťou spôsobujú v rozvojových krajinách obrovské prírastky obyvateľstva ktoré nemajú obdobu v histórii dnešných rozvinutých krajín. V druhej polovici dvadsiateho storočia dosahovali niektoré rozvojové krajiny bežne rast vyšší ako 3%. Pri takomto raste sa veľkosť populácie zdvojnásobí za 23 rokov.

„Reprodukčná revolúcia“ v rozvinutých krajinách je jednou z najzaujímavejších udalostí druhej polovice dvadsiateho storočia. Rozvoj metód rodinného plánovania zahŕňajúci dostupnú, jednoduchú a účinnú antikoncepciu či zjednodušené možnosti sterilizácie dovolili ženám predchádzať nechceným tehotenstvám. Jednoduchý prístup k týmto metódam, ako aj socio-ekonomické zmeny motivovali páry v rozvinutých krajinách k obmedzovaniu veľkostí rodín. Oveľa zložitejšia situácia je však v rozvojových krajinách kde politické, ekonomické a kultúrne bariéry často limitujú prístup k metódam rodinného plánovania. Okolo 40% dievčat v rozvojových krajinách podľa OSN porodí prvé dieťa ešte pred dovŕšením dvadsiateho roku života. Napriek tomu od roku 1998 okolo 60% žien v reprodukčnom období v rozvojových krajinách využíva niektorú z metód plánovaného rodičovstva. Štátne politiky znižovania pôrodnosti boli často v rozpore s tradičnými a náboženskými zvyklosťami. Tieto vládne snahy boli často veľmi obtiažne uskutočniteľné, vyvolávali ideologické konflikty a hraničili s porušovaním ľudských práv.

Pozoruhodné zvýšenie strednej dĺžky života po narodení v rozvojových krajinách po roku 1950 bolo najmä odrazom zníženia dojčenskej a detskej úmrtnosti. Hlavným príspevkom k tomuto poklesu bol masívny celosvetový imunizačný program pre deti. V roku 1975 Svetová Zdravotnícka organizácia spustila rozsiahly Program imunizácie proti šiestim chorobám, ktoré dovtedy vzali život mnohým miliónom mladých ľudí. Boli to: tuberkulóza, hnačka, záškrt, týfus, čierny kašeľ a detská obrna. Do roku 1981 sa podarilo proti týmto chorobám zaočkovať 1/5 všetkých detí, v polovici deväťdesiatych rokov to boli už 4/5. Hnačka a ostatné nákazlivé choroby síce stále ostávajú hlavnou príčinou detskej úmrtnosti, ale epidémie týchto chorôb sú dnes už menej časté a menej smrtiace. Detská obrna vymizla takmer úplne. Deti majú dnes väčšiu šancu dožiť sa dospelosti ako kedykoľvek predtým. Na druhej strane epidémia HIV/AIDS predstavuje novú hrozbu pre zdravie detí. Vírus sa môže preniesť počas tehotenstva z matky na dieťa. Každý tretí novorodenec narodený matke s touto chorobou, je tiež infikovaný. Podľa odhadov OSN žije 90% detí infikovaných vírusom HIV v Afrike, ale čísla infikovaných detí v Indii a Juhovýchodnej Ázii sú tiež veľmi vysoké.

Celkový pokles mier úmrtnosti po roku 1950 mal za následok „populačnú explóziu“ vo väčšine rozvojových štátov. Napríklad v Mexiku vďaka zavedeniu modernej zdravotníckej starostlivosti dosiahli zníženie mier úmrtnosti oveľa rýchlejšie ako vo vyspelých krajinách. Miery pôrodnosti však zostali vysoké, vďaka čomu vzrástol prirodzený

prírastok neuveriteľne prudko. Mexická populácia narastala v šesťdesiatych a sedemdesiatych rokoch takmer o 3%ročne. Rozvojové krajiny ako celok dosahovali v šesťdesiatych a začiatkom sedemdesiatych rokov každoročné prírastky okolo 2%. Od roku 1950 do roku 2005 vzrástol počet obyvateľov v týchto krajinách z 1,7 na 5,2 miliardy. Populačné prírastky boli enormné, až pokým nezačala úhrnná plodnosť v týchto krajinách postupne klesať. Veľkosť a spôsob tohoto poklesu závisela od stupňa a úspešnosti ekonomického a sociálneho rozvoja, štátnej populačnej politiky, plánovaného rodičovstva a ďalších faktorov.



V rôznych rozvojových regiónoch nastával pokles úhrnnej plodnosti rozličnými spôsobmi a rozdielnym tempom, čo môžeme ilustrovať na príklade niekoľkých krajín. V Južnej Kórei dosahuje úhrnná plodnosť úroveň porovnateľnú s najnižšími hodnotami v rozvinutých krajinách sveta. Podobne bol dramatický pokles mier ÚP zaznamenaný aj v Číne, Kostarike, Srí Lanke, Tunisku, alebo Thajsku. Pokroky v šírení antikoncepcie, vzdelávaní a zdravotnej starostlivosti tu boli rýchlo nasledované aj poklesom úhrnnej plodnosti. Iné krajiny ako Keňa zaznamenali tiež významné zníženie mier úhrnnej plodnosti, ktoré sú však ich prípade aj napriek tomu stále relatívne vysoké. Krajinám ako Egypt, Bangladéš, India, Indonézia či Filipíny sa podarilo znížiť úhrnnú plodnosť na strednú úroveň, následne sa však pokles zastavil. Táto stagnácia môže byť dočasná ako v prípade Egypta, alebo dlhodobá ako v Keni. Súčasné analýzy dokumentujú, že k takáto stagnácia je pozorovaná v krajinách kde sa zastavil socio-ekonomický pokrok ktorý zahŕňa aj zvyšovanie príjmov obyvateľstva a úroveň vzdelávania. Nakoniec nám ostávajú krajiny ako Niger, Afganistan, Jemen, Nigéria či Uganda kde nebolo pozorované žiadne, alebo iba minimálne zníženie mier úhrnnej plodnosti za dlhé desaťročia. Takýto scenár je typický pre najchudobnejšie, prevažne poľnohospodársky zamerané krajiny s minimálnym rozšírením antikoncepcie ako aj nízkou úrovňou vzdelávania.

Krajiny s najväčšími prírastkami obyvateľstva dnes tvoria asi 8% svetovej populácie, avšak pre mnohé z nich sa predpokladá zdvojnásobenie, či strojnásobenie súčasnej populácie do 50tich rokov. V polovici tohto storočia budú potom tieto krajiny tvoriť už 20% obyvateľstva sveta. S výnimkou niekoľkých ropu exportujúcich krajín, ktoré sa tešia značnému ekonomickému rastu, je väčšina štátov s vysokými populačnými prírastkami uvedená na zozname najmenej rozvinutých krajín OSN. Krajiny uvedené na tomto zozname majú najnižšie príjmy na osobu, najnižšiu úroveň gramotnosti a v ich ekonomike dominuje poľnohospodárstvo na úkor priemyslu alebo výroby. Len máloktorá z týchto najchudobnejších krajín nemá vysoký populačný prírastok sprevádzaný zároveň aj vysokou úmrtnosťou alebo masívnou emigráciou obyvateľstva do miest za účelom hľadania

zamestnania. Väčšina krajín s najvyššími prírastkami sa nachádza v subsaharskej Afrike. Podarilo sa im znížiť úmrtnosť, ktorá však pri nezmenenej úrovni fertility spôsobuje roztváranie „demografických nožníc“ a tým pádom aj obrovské prírastky. Aj mimo Afrického kontinentu nájdeme príklady takýchto krajín ako sú Afganistan, Guatemala, alebo Haiti.

Budúca veľkosť svetovej populácie však závisí najmä od vývoja ÚP v krajinách s vysokou a stredne vysokou pôrodnosťou. Úmrtnosť má dnes malý vplyv na vývoj budúceho populačného prírastku, s výnimkou krajín s veľmi vysokou mortalitou spôsobenou vysokým rozšírením choroby AIDS či malárie. Úmrtia na chorobu AIDS dramaticky spomalili prírastok obyvateľstva v niektorých prevažne juhoafrických krajinách. Nikto však dnes nevie s istotou predpovedať ako táto choroba bude vplývať na budúci celosvetový populačný vývoj.

Záver:

Takzvaný rozvojový svet dnes tvorí neuveriteľne pestrá paleta štátov, ktoré sa odlišujú v neposlednom rade aj svojim demografickým správaním. Zatiaľ čo v niektorých je už demografická revolúcia ukončenou kapitolou v iných sa dá pozorovať veľmi pozvoľný nástup tohoto procesu. Väčšina krajín sa samozrejme nachádza niekde medzi týmito dvoma extrémami. To samozrejme ovplyvňuje hodnotu ich prirodzeného prírastku. V globálnom merítku je ale ešte stále dobre pozorovateľná široká priepasť medzi rozvojovými a rozvinutými krajinami.

Použité zdroje literatúry:

Kent, M. M, Haub, C. (2005): Global demographic divide. in: Population Bulletin, vol 60, no.4

Population Reference Bureau staff (2004): Transitions in World Population in: Population Bulletin, vol59, no.1

World Population Prospects The 2006 Revision

autor:

Bc. Andrej Chromeček
Katedra humannej geografie a demogeografie
Prirodovedecka fakulta UK
Mlynska dolina
842 15 Bratislava
email: KimiMouse@azet.sk

Analýza faktorov determinujúcich ceny bytov

Miroslav Kello

Abstract: The aim of this paper is to present the relation between the price of flats in the capital city of Slovakia and Czech Republic and chosen explanatory variables. We have chosen two capital cities – Prague and Bratislava, which are said to have the highest prices in both republics, because of functional market mechanism and enough accessible data. As the estimation method we used the ordinary least squares and weighted least squares method and all analysis procedures and tests were performed in EViews 4.1 Student Version and MS® Excel environment.

Key words: *real estate prices, elasticity estimation, regression analysis, least square method, heteroskedasticity, multicollinearity*

1. Úvod

Ekonomická prosperita, zvyšovanie počtu pracovných príležitostí, rast príjmov a ochoty míňať priniesli so sebou najrozsiahlejšiu stavebnú aktivitu od roku 1989. Investícia do kúpy bytu si však vyžaduje nemalé finančné prostriedky, keďže ceny bytov sa dnes pohybujú v astronomických výškach. Suma, ktorú dnes za byt zaplatíme, už dávno nezávisí iba od počtu izieb v byte, ale aj od mnohých ďalších premenných. Cieľom tohto príspevku je porovnať závislosť výšky cien bytov v závislosti od vybraných faktorov, a to v hlavnom meste Slovenskej ako i Českej republiky a porovnať zistené rozdiely.

2. Dáta a metodický postup

Prvým krokom pri špecifikácii ekonometrického modelu je určenie a špecifikácia všetkých premenných zahrnutých do modelu. V našom prípade je závislou premennou cena bytov. Za regresory použité v danom modeli sme sa rozhodli použiť: počet izieb, vzdialenosť od centra (pričom za centrum bola v Bratislave považovaná Obchodná ulica, v Prahe to bolo Václavské Námestie; vzdialenosti boli merané špecializovaným softvérom MS® AutoRoute a sú udané v km) a rozloha bytu v m^2. Umelými premennými sú: konštrukcia bytu (tehla alebo panel); to, či daný byt má parkovacie miesto, balkón alebo výťah a v neposlednom rade stav bytu (teda to, či je byt v pôvodnom stave alebo rekonštruovaný, resp. či ide o byt v novostavbe).

Ako druhý krok uvedieme stanovenie očakávaných znamienok odhadnutých parametrov modelu. V prípade vzdialenosti bytu od centra očakávame záporné znamienko, keďže s postupným vzďaľovaním predpokladáme pokles „príťažlivosti“ lokality, a teda aj pokles ceny bytu. V prípade ostatných parametrov je vhodné predpokladať kladný vplyv na cenu – je logické, že byt s väčšou rozlohou by mal stáť viac ako byt s menšou a pod. Samozrejme, do ceny bytov vstupuje aj mnoho iných faktorov, než tie, ktoré sme uviedli. Ide napríklad o to, aká kriminalita je v oblasti, možnosti nakupovania či výhľad – a všetky tieto premenné môžu ovplyvniť výslednú cenu bytu. Takže sa môže stať, že dva byty s rovnakou rozlohou i polohou budú mať rozdielne ceny práve v dôsledku týchto premenných.

Ďalším krokom bolo odhadnutie samotného modelu. Na kvantifikovanie parametrov ekonometrického modelu sme využili metódu viacnásobnej lineárnej regresie a v rámci nej techniku najmenších štvorcov, resp. techniku vážených najmenších štvorcov. Na ilustráciu týchto metód využijeme lineárny model s dvoma premennými.

Metóda najmenších štvorcov za účelom získania odhadov parametrov minimalizuje

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{1}$$

Naproti tomu, metóda vážených najmenších štvorcov minimalizuje váženú sumu štvorcov rezíduí

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \tag{2}$$

kde β_1 a β_2 sú odhady parametrov.

Za váhy w_i môžeme stanoviť napr.:

$$w_i = \frac{1}{\sigma_i^2} \tag{3}$$

Za prvotný odhadnutý model sme zvolili klasický lineárny regresný model:

$$\mathbf{Y} = \beta_1 + \beta_2 \mathbf{X}_2 + ... + \beta_k \mathbf{X}_k + \mathbf{u}. \tag{4}$$

kde $\mathbf{u}$ je vektor náhodných zložiek

β_j je j -ty regresný koeficient pre $j = 1, 2, ..., k$

Následne sme daný model modifikovali s ohľadom na dosiahnutie čo najlepších výsledkov v týchto kritériách: významnosť modelu ako celku (*F- štatistika, koeficient determinácie* R^2), významnosť jednotlivých koeficientov (*t - štatistiky*), Akaikeho informačné kritérium a Schwarzovo informačné kritérium (volia sa modely s čo najnižšími hodnotami týchto kritérií).

Po zvolení nami predpokladaného najlepšieho modelu sme pristúpili k testovaniu porušenia predpokladov klasického lineárneho regresného modelu, ako je multikolinearita a heteroskedasticita. Keďže problém autokorelácie sa pri prierezových dátach prakticky nevyskytuje, nevenovali sme tejto oblasti pozornosť.

Pod pojmom multikolinearita dnes rozumieme vysokú koreláciu medzi vysvetľujúcimi premennými. Zároveň sú závislé aj hodnoty parametrov, ktoré majú vysoké smerodajné odchýlky a nízke hodnoty t- štatistík, čo budí dojem nevýznamnosti. Multikolinearita zvyšuje rozptyly odhadov parametrov, čo má za následok zníženie presnosti odhadov v zmysle dlhších intervalov spoľahlivosti, zdanlivý rozpor medzi nevýznamnými výsledkami *t*-testov a významným výsledkom celkového *F*-testu, a nestabilné odhady regresných koeficientov (pri malej zmene špecifikácie vznikajú výrazné posuny vo významnosti parametrov).

Na testovanie prítomnosti multikolinearity sme využili nasledujúce postupy:
Ako prvé použijeme na testovanie Kleinovo kritérium. Princípom tohto testu je porovnávanie koeficientu viacnásobnej determinácie a párovými koeficientmi determinácie medzi jednotlivými vysvetľujúcimi premennými pre i ≠ j. Ak $R_y^2 < r_{x_i,x_j}^2$, potom je opodstatnený predpoklad existencie multikolinearity.

Keďže Kleinovo kritérium nie je exaktným štatistickým testom, použili sme preto aj test Farrara a Glaubera. Tento test je založený na testovaní nasledujúcej nulovej hypotézy:
H_0: D=1 (premenné sú ortogonálne) H_1: D<1(premenné sú závislé)
Testovacia štatistika :

$$FG = -(n-1-\frac{1}{6}(2K+5))\ln D \tag{5}$$

kde *n* je rozsah výberu, *K* je počet vysvetľujúcich premenných spolu s absolútnym členom (K = k+1), D je determinant korelačnej matice vysvetľujúcich premenných.

Testovacia štatistika má za predpokladu H_0 chí-kvadrát rozdelenie s 1/2K(K-1) stupňami voľnosti. Hypotézu H_0 zamietame na príslušnej hladine významnosti, ak hodnota vypočítanej testovacej štatistiky je väčšia ako tabelovaná hodnota chí-kvadrát rozdelenia.

Heteroskedasticita vzniká najmä v prípade prierezových dát. Vo svojej podstate znamená, že je porušený klasický predpoklad $D[\varepsilon_i]=\sigma^2$, t.j., že disperzia náhodných chýb nie je konštantná a konečná u všetkých pozorovaní. Na testovanie problému heteroskedasticity sme použili:
Whiteov test heteroskedasticity
Pri tomto teste sa snažíme o prijatie nulovej hypotézy :
H_0: Konštantný rozptyl rezíduí (túto hypotézu využívame i v nasledovných testoch)
Goldfeldov – Quandtov test
Pre účely tohto testu sme premenné usporiadali vzostupne podľa y*=log(cena)-rezíduá. Následne sme vynechali prostredných 20 pozorovaní (pre n > 30 sa odporúča vynechať približne ¼ až 1/8 dát). Zvyšok pozorovaní sme rozdelili na dve rovnaké podskupiny a na každej z nich previedli regresiu.
Testovacia štatistika :

$$F = \frac{\frac{RSS_2}{(n-g-2k)/2}}{\frac{RSS_1}{(n-g-2k)/2}} \sim F_{\frac{(n-g-2k)}{2},\frac{(n-g-2k)}{2}} \tag{6}$$

Breusch – Paganov test
Úspech Goldfeld – Quandtovho testu závisí nielen od počtu vynechaných prostredných pozorovaní, ale aj od identifikácie správnej premennej, podľa ktorej pozorovania zoradíme. Tento nedostatok je riešený práve Breusch – Paganovým testom, ktorý testuje nasledujúci model heteroskedasticity

$$\sigma_i^2 = \sigma_i^2 f(\alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{z_i}) \tag{7}$$

V nasledujúcej časti príspevku aplikujeme uvedené metódy a testy na konkrétne empirické dáta.

3. Empirické výsledky

Za predbežný model závislosti cien bytov v Bratislave, sme zvolili:

$$\log(cena_i) = 12.57 + \underset{(0,047)}{0.672}\log(ROZ_i) - \underset{(0,025)}{0.199}\log(POL_i) + \underset{(0,026)}{0.115}REK_i + \underset{(0,042)}{0.103}PAR_i + u_i \quad (8)$$

R^2=0,803

teda cena bytov v Bratislave závisí predovšetkým od rozlohy (*ROZ*), polohy (*POL*), stavu bytu (*REK*) a toho, či je k bytu pridelené parkovacie miesto (*PAR*). Naproti tomu v Prahe sa závislosť cien od parkovania nepotvrdila. Za predbežný model sme preto zvolili:

$$\log(cena_i) = 11.38 + \underset{(0.159)}{0.978}\log(ROZ_i) - \underset{(0.038)}{0.208}\log(POL_i) + \underset{(0.045)}{0.133}REK_i + u_i \quad (9)$$

R^2=0.754

Keďže v prípade oboch miest bol koeficient viacnásobnej determinácie vyšší než párové koeficienty determinácie medzi jednotlivými vysvetľujúcimi premennými, na základe Kleinovho kritéria sme nepotvrdili prítomnosť multikolinearity. Tento výsledok sa nám potvrdil aj v prípade testu Farrara a Glaubera, kde na bežných hladinách významnosti nemôžeme zamietnuť nulovú hypotézu.

V prípade heteroskedasticity sme však takéto uspokojivé výsledky nedosiahli. Všetky tri prevedené testy nám potvrdili jej prítomnosť.

Whiteov test :

- Bratislava: p-hodnota = 0,056 (hodnota je na hranici zamietnutia H_0 na 5 % hl. význ.)
- Praha : p-hodnota = 0,003

Goldfeld – Quandtov test

- Bratislava : F-štatistika = 6,599 $f_{0,05;26;26}$ = 1,929
- Praha : F-štatistika = 7,532 $f_{0,05;27;27}$ = 1,904

Breusch – Paganov test

- Bratislava : Θ = 17,092 $\chi^2_{0.05;4}$ =9,487
- Praha : Θ = 86,158 $\chi^2_{0,05;3}$ = 7,815

Aby sme odstránili heteroskedasticitu, rozhodli sme sa na odhad modelov použiť metódu vážených najmenších štvorcov, pričom za váhy sme zvolili $w_i = 1/\hat{\sigma}^2$. Odhadnuté modely majú nasledujúce parametre:

Bratislava:

$$\log(cena_i) = 12,64 + \underset{(0,001)}{0,656}\log(ROZ_i) - \underset{(0,00007)}{0,201}\log(POL_i) + \underset{(0,0002)}{0,115}REK_i + \underset{(0,0006)}{0,107}PAR_i + u_i \quad (10)$$

Whiteov test: p-hodnota = 0,583

Praha:

$$\log(cena_i) = \underset{}{11,447} + \underset{(0,011)}{0,961}\log(ROZ_i) - \underset{(0,001)}{0,205}\log(POL_i) + \underset{(0,003)}{0,125}\log(ROZ_i) + u_i \qquad (11)$$

Whiteov test: p-hodnota = 0,337

Ako môžeme vidieť, použitím metódy vážených najmenších štvorcov sa nám parametre modelov zmenili iba minimálne (avšak došlo k výraznému zlepšeniu ich smerodajných odchýlok). Zároveň môžeme konštatovať, že na základe Whiteovho testu môžeme zamietnuť prítomnosť heteroskedasticity. Po odlogaritmovaní nadobudnú dané modely tento tvar:

Bratislava: $$cena_i = 308171 + ROZ_i^{0,656} - POL_i^{0,201} + e^{0,115*REK_i} + e^{0,107*PAR_i} + u_i \qquad (12)$$

Praha: $$cena_i = 93630 + ROZ_i^{0,961} - POL_i^{0.205} + e^{0.125*REK_i} + u_i \qquad (13)$$

4. Záver

Na základe odhadnutých modelov môžeme konštatovať, že v Bratislave závisí výška ceny bytov predovšetkým od rozlohy bytu, jeho vzdialenosti od centra, stavu a toho, či má alebo nemá parkovacie miesto. Pri bytoch v Prahe sme zaznamenali iba jediný rozdiel v tom, že parkovacie miesto nehrá až tak významnú úlohu. Ani v jednotlivých regresných koeficientoch nebadať medzi zvolenými mestami veľké rozdiely. Kým v Bratislave spôsobí rast rozlohy bytu o 1% rast jeho ceny o 0,656%, v Prahe je to 0,961%, pri ostatných koeficientoch sú rozdiely minimálne.

Avšak, ako sme už spomínali, do úvahy musíme brať aj množstvo iných (aj subjektívnych) faktorov, ktoré naše modely neobsiahli, čo sa môže prejaviť vo väčšej odchýlke modelovaných cien od tých skutočných.

5. Použitá literatúra

GUJARATI, D.N.: Basic Econometrics, New York: The McGraw-Hill, 2003.

HUŠEK, R.: Ekonometrická analýza, Bratislava: EKOPRESS, 1999.

HUŠEK, R.: Aplikovaná ekonometrie. Teorie a praxe, Praha: Prof. Publishing, 2003.

Adresa autora:
Miroslav Kello
Ekonomická fakulta
Univerzita Mateja Bela
Tajovského 10
975 90 Banská Bystrica
miroslavkello@yahoo.com

Analýza zamestnanosti v krajinách Európskej únie

Magna Marián

Abstract: This project appear from basic purposes of Lisbon Strategy of European Union (EU) countries. It discuss about following goals: 1. Overall economy growth and 2. Employment growth in EU. The object of this project is to analyze employment in EU countries on the basis of more indicators simultaneously. Because of more criteria problem, we use principal components analysis.

Key words: *PCA (Principal Components Analysis), European Union, Lisbon Strategy, SAS Enterprise Guide*

1. Úvod

V marci 2000 vtedajší európski lídri poverili Európsku úniu stať sa do roku 2010 „najdynamickejším a najviac konkurencieschopným, na najnovších poznatkoch založeným hospodárstvom na svete, schopným udržateľného rozvoja, s viacerými a lepšími pracovnými príležitosťami a väčšou sociálnou kohéziou a rešpektom k životnému prostrediu“. Tento plán sa všeobecne volá Lisabonská stratégia a je v podstate súborom navzájom sa ovplyvňujúcich reforiem.

Ak chce Európa zvýšiť rast zamestnanosti a produktivity, potrebuje naplniť ciele v piatich oblastiach politiky:

• Znalostná spoločnosť
• Vnútorný trh
• Podnikateľské prostredie
• Trh práce
• Environmentálna udržateľnosť

Jednotlivé členské štáty síce dosiahli pokrok v jednej alebo vo viacerých z týchto prioritných oblastí politiky, ale žiadny z nich neuspel úplne vo všetkom. Spomedzi piatich oblastí politiky sme sa v tomto projekte zamerali na zamestnanosť.

2. Výsledky analýzy zamestnanosti v krajinách EÚ

Na analýzu zamestnanosti sme použili nasledujúce ukazovatele za roky 2003 a 2005:

- **Celková zamestnanosť** (**zam03, zam05)** – zamestnané osoby vo veku od 15-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.
- **Zamestnanosť žien (zeny03, zeny05)** – zamestnané ženy vo veku od 15-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.

- **Zamestnanosť mužov** (**muzi03, muzi05)** – zamestnaní muži vo veku od 15-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.
- **Zamestnanosť starších osôb (star03, star05)** – zamestnané osoby vo veku od 55-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.
- **Zamestnanosť starších žien (star_zeny03, star_zeny05)** – zamestnané ženy vo veku od 55-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.
- **Zamestnanosť starších mužov** (**star_muzi03, star_muzi05)** – zamestnaní muži vo veku od 55-64 rokov ako podiel na celkovej populácií v rovnakej vekovej skupine.

Všetky dáta boli získané z oficiálnych stránok EUROSTATu, sú vykazované 12 týždňov po skončení príslušného roka ako ročný priemer, sú harmonizované s dátami obyvateľstva a národnými účtami.

V prvom kroku analýzy sme vypočítali korelačné matice vstupných ukazovateľov zamestnanosti v sledovaných rokoch. Zistili sme silnú závislosť medzi premennými. Na základe toho sme mohli konštatovať, že dáta sú vhodné pre analýzu hlavných komponentov. Použili sme PCA z kovariančnej matice, lebo ukazovatele zamestnanosti sú vyjadrené v rovnorodých meracích jednotkách. Pri voľbe počtu významných hlavných komponentov sme vychádzali z Kaiserovho pravidla: významné hlavné komponenty sú tie, ktorých vlastné číslo je vyššie ako priemerná hodnota všetkých vlastných čísel. V našom prípade je to len prvý hlavný komponent v obidvoch rokoch (Tabuľka 1 a Tabuľka 2), ktorý vysvetľuje vyše 79% variability dát.

Tabuľka 1*: Vlastné čísla kovariančnej matice, rok 2003*

Eigenvalues of the Covariance Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	**478,0683**	**409,0609**	**0,7915**	**0,7915**
2	69,0073	20,6627	0,1142	0,9057
3	48,3446	39,8121	0,0800	0,9858
4	8,5325	8,4713	0,0141	0,9999
5	0,0612	0,0542	0,0001	1,0000
6	0,0070		0,0000	1,0000

Tabuľka 2*: Vlastné čísla kovariančnej matice, rok 2005*

Eigenvalues of the Covariance Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	**438,0521**	**375,5623**	**0,7931**	**0,7931**
2	62,4898	19,5028	0,1131	0,9062
3	42,9870	34,2244	0,0778	0,9841
4	8,7626	8,7266	0,0159	0,9999
5	0,0360	0,0310	0,0001	1,0000
6	0,0051		0,0000	1,0000

Tabuľka 3*: Pearsonove korelačné koeficienty pre roky 2003 a 2005*

Pearson Correlation Coefficients, N = 27 Prob > |r| under H0: Rho=0						
	zam03	**zeny03**	**muzi03**	**star03**	**star_zeny03**	**star_muzi03**
Prin1 2003	0,8219	0,7754	0,5851	0,9887	0,9438	0,8707
	<,0001	<,0001	0,0013	<,0001	<,0001	<,0001
Prin1 2005	0,8182	0,8173	0,5124	0,9891	0,9465	0,8386
	<,0001	<,0001	0,0063	<,0001	<,0001	<,0001

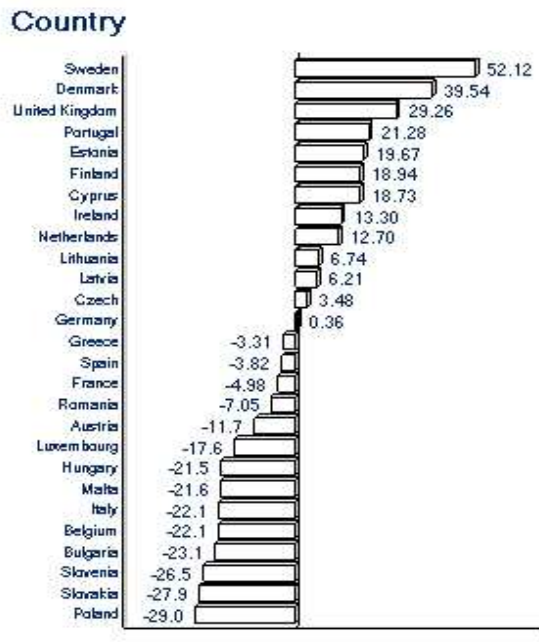
Následne sme overili korelovanosť vybraných ukazovateľov zamestnanosti pre roky 2003 a 2005 s prvým hlavným komponentom (PRIN1). Predpoklad, že ukazovatele sú silne korelované s prvým hlavným komponentom sa potvrdil. Dokazujú to vysoké hodnoty

Pearsonových koeficientov korelácie hlavného komponentu PRIN1 so všetkými ukazovateľmi zamestnanosti (Tabuľka 3).

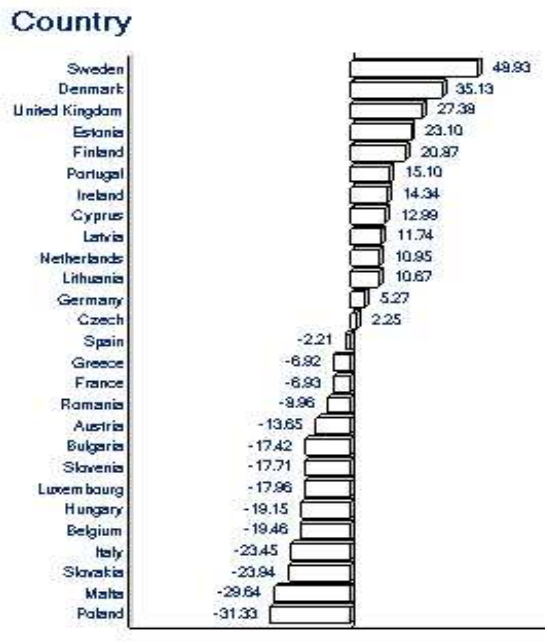
Hodnoty komponentných skóre prvého hlavného komponentu (PRIN1) vyjadrujú pre jednotlivé krajiny EÚ hodnoty syntetického ukazovateľa "integrálnej" zamestnanosti. Vysoké hodnoty PRIN1 znamenajú priaznivú situáciu v oblasti zamestnanosti v krajine a opačne nízke, resp. záporné hodnoty znamenajú nepriaznivú situáciu. Zostavili sme "rebríček" krajín EÚ podľa hodnôt PRIN1. Histogramy vyjadrujú, ktorá krajina EÚ dosahuje najvyššiu a ktorá najnižšiu "integrálnu" zamestnanosť v sledovaných rokoch 2003 a 2005 (Obrázok 1 a Obrázok 2).

Rebríčky pre roky 2003 a 2005 sú veľmi podobné. Na prvých troch miestach sa v obidvoch rokoch umiestnili Švédsko, Dánsko a Veľká Británia. Zhoršenie zamestnanosti (pokles v rebríčku) v roku 2005 oproti roku 2003 zaznamenali krajiny ako Cyprus, Portugalsko, Holandsko, Litva, Česká republika, Grécko, Luxembursko, Maďarsko, Malta a Taliansko. Naopak, zvýšenú zamestnanosť a vzostup v rebríčku zaznamenali krajiny ako Estónsko, Fínsko, Írsko, Lotyšsko, Nemecko, Španielsko, Bulharsko, Slovinsko a Slovensko.

***Obrázok 1**:*
Histogram "integrálneho" ukazovateľa zamestnanosti pre jednotlivé krajiny EÚ (rok 2003)



***Obrázok 2**:*
Histogram "integrálneho" ukazovateľa zamestnanosti pre jednotlivé krajiny EÚ (rok 2005)

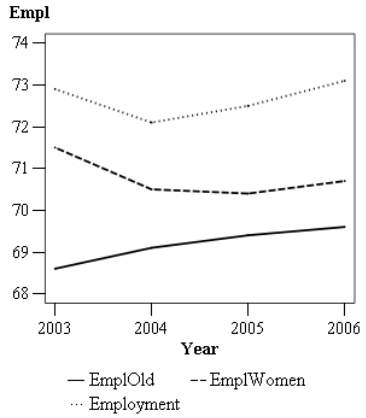


3. Záver

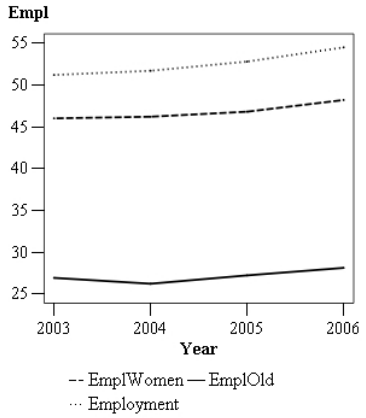
Na základe histogramov znázornených na Obrázku 1 a Obrázku 2 sme vytvorili grafy vývoja zamestnanosti, zamestnanosti žien a zamestnanosti starších osôb troch vybraných krajín: našej krajiny –Slovenska, najlepšej krajiny v rebríčkoch – Švédska a najhoršej krajiny

– Poľska. Všetky krajiny EÚ majú na základe Lisabonskej stratégie stanovené spoločné ciele pre rok 2010: pre celkovú zamestnanosť na úrovni 70%, pre zamestnanosť žien na úrovni 60% a pre zamestnanosť starších osôb na úrovni 50%. Obrázky 3, 4 a 5 znázorňujú vývoj celkovej zamestnanosti, zamestnanosti žien a zamestnanosti starších osôb v rokoch 2003 až 2006. Obrázok 3 prislúcha Švédsku. Táto krajina je už v súčasnosti ďaleko za stanovenými cieľmi, napríklad celková zamestnanosť bola v roku 2006 na úrovni 73,1%. Obrázok 4 prislúcha najhoršej krajine, Poľsku. Táto krajina vykazuje najhoršie hodnoty zamestnanosti. Napríklad celkovú zamestnanosť mala v roku 2006 na úrovni 54,5%. Obrázok 5 prislúcha našej krajine, Slovensku. V roku 2006 sa umiestnilo na 21. mieste v celkovej zamestnanosti spomedzi 27 krajín EÚ, s hodnotou zamestnanosti rovnou 59,4%. Táto hodnota nedosahuje želateľných 70%, avšak do roku 2010 by sa Slovensko k méte 70% mohlo priblížiť.

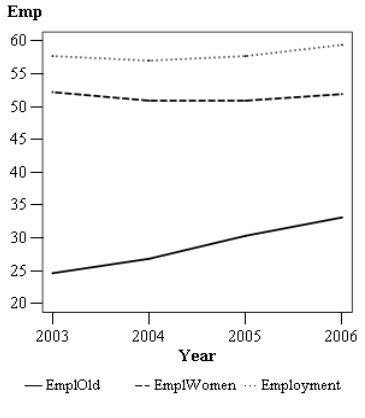
Obrázok 3: *Švédsko*



Obrázok 4: *Poľsko*



Obrázok 5: *Slovensko*



4. Literatúra

EUROPEAN COMMISSION: Facing the challenge. The Lisbon strategy from growth and employment. Luxembourg: Office for Official Publications of the EC. November 2004. ISBN 92-894-7054-2. http://ec.europa.eu/growthandjobs/pdf/kok_report_en.pdf

STANKOVIČOVÁ, IVETA*:* Viackriteriálne hodnotenie zamestnanosti členských krajín EÚ na základe vybraných ukazovateľov Lisabonskej stratégie, In: Zborník z medzinárodnej konferencie STAKAN 2007. Česká statistická společnost, Praha, 2007.
http://www.statspol.cz/stakan/index.htm

VOJTKOVÁ, MÁRIA*:* Hľadanie podobnosti krajín EÚ podľa vybraných ukazovateľov Lisabonskej stratégie, In: Zborník z medzinárodnej konferencie STAKAN 2007. Česká statistická společnost, Praha, 2007. http://www.statspol.cz/stakan/index.htm

Adresa autora:

Marián Magna, Bc. Bc.
Fakulta managementu, Univerzita Komenského
Odbojárov 10, 820 05 Bratislava
E-mail: Marian.Magna@st.fm.uniba.sk

POPULAČNÝ VÝVOJ ŠTÁTOV EURÓPSKEJ ÚNIE V ROKOCH 1960- 2003

Margetínová Alexandra

Abstract: Population Development of the European Union States in 1960- 2003. The main aim of this paper is the explanation of the population dynamics in the states of the European union in 1960 to 2003. The paper pay attention to the comparison of the EU states (EU 15 - original members and EU 12 - new members). It is possible to identify four groups of states with the similar population dynamics.
Key words: population development, states of the EÚ, demographic revolution,

ÚVOD

Pri celkovom hodnotení populačného vývoja sa v rozvinutých krajinách najčastejšie identifikujú znaky, ktoré sú charakteristické pre demografický prechod (revolúciu). Často sa rozdiely prejavujú len v období, kedy tento priebeh začal a akým rozsahom prebehli dané zmeny. Pod pojmom „2.demografická revolúcia“ K. Pastor rozumie komplex zmien v správaní a hodnotovom systéme populácie, ktoré nadhodnocujú individualizmus a osobnú slobodu, oslabujú funkciu manželstva a rodiny a redukujú pôrodnosť na úroveň nezaručujúcu sebareprodukciu populácie (Pastor, 2002).

CELKOVÉ ZHODNOTENIE POPULAČNÉHO VÝVOJA V ŠTÁTOCH EÚ

Viac rozvinuté štáty sveta (kde zaraďujem EÚ 15) s rozvinutejšou ekonomikou, väčšou podporou štátu, zaznamenávajú priebeh vlny zmien skôr. Menej rozvinuté štáty (EÚ 12) s horšou ekonomikou, zmenenou politickou situáciou sa k prechodu a zmene správania dostávajú neskôr. V celej únii sa rozpätie obdobia pohybuje v rokoch 1960- 2000. V západnej časti únie priebeh začal zhruba v 60-tych rokoch, v južnej časti o 20 rokov neskôr, v bývalých socialistických štátoch najväčšie zmeny sledujeme v 90-tych rokoch 20.storočia. Tento neskorší nástup bol však kompenzovaný rýchlejším a intenzívnejším priebehom.

V oblasti pôrodnosti a plodnosti sa vyskytlo viacero zmien spojených so sústavným poklesom živonarodených detí, kde výnimnkou zostávajú povojnové roky, v ktorých nastal „babyboom“ a tým aj nárast miery pôrodnosti.

Úmrtnosť v priebehu obdobia klesla o 0,6 ‰, čo možno vidieť ako pozitívny fakt vývoja. Po zoradení štátov do rebríčka sa na jeho začiatku (s najvyššími hodnotami) nachádzajú nové členské štáty z bývalého východného bloku. V celkovom chápaní v únii dochádza k znižovaniu úmrtnosti a predlžovaniu strednej dĺžky života a to zhruba o 7,7 rokov.

Najvýraznejší pokles sobášnosti nastal v skupine EÚ 12 a tým súvisí aj zvyšovanie priemerného veku pri prvom sobáši. Spravidla je vek muža pri sobáši vyšší ako u žien a zároveň vyšší aj v skupine EÚ 15. Pokles je teda zaznamenávaný v dôsledku prejavovania sa čŕt demografickej revolúcie.

Rozvodovosť, ako negatívny prvok, zaznamenala takmer štvornásobný nárast. Napriek poklesu sobášnosti neustále stúpa, čo poukazuje na nestabilnosť vytvorených manželstiev v celej únii (Podmanická, 2002). Rozhodujúcimi zostávajú kultúrne vpyvy, religiozita, legislatíva.

Potratovosť zaznamenala veľký „boom“, ktorý sa ale neprejavoval dlhé obdobie. V súčasnosti sa potraty začínajú stabilizovať na nižšej úrovni aj na základe zvyšujúcej sa úrovne v oblasti zdravotníctva, rozšírenia antikoncepcie a poznatkov v oblasti ochrany pred počatím. Jediná klesajúca tendencia je zaznamenávaná v postkomunistických krajinách. V ostatných sa už stabilizovala, prípadne niekde mierne narastá.

KLASIFIKÁCIA (TYPOLÓGIA) ŠTÁTOV NA ZÁKLADE PODOBNOSTÍ A ROZDIELOV POPULAČNEJ DYNAMIKY (V ROKOCH 1960- 2003)

Cieľom bolo vytvoriť klasifikáciu štátov na základe podobnosti či rozdielov v ich populačnom vývoji, spôsobu reakcie na demografický prechod. Okrem demografických znakov boli zohľadnené i ďalšie faktory, ktoré vplývali na demografické správanie obyvateľstva (religiozita, národnostná štruktúra, politický, ekonomický, kultúrny pohľad a i.).

Väčšina autorov hodnotí európske štáty ako skupinu pôvodných členských štátov (EÚ 15) a štátov novopričlenených (EÚ 12). Následne ich charakterizujú ako rozvinutejšie a menej rozvinuté. Pri podrobnejšom spôsobe skúmania sa však vyskytujú aj určité skupiny štátov, ktoré tvoria výnimku, respektektíve sa určitým spôsobom odlišovali od ostatných. Pomocou skúmania hodnôt jednotlivých ukazovateľov v rokoch 1960 až 2003, ich vývoja a zároveň aj zohľadnením iných faktorov (politických, ekonomických, kultúrnych) som dospela k vytvoreniu štyroch základných skupín. Členenie je nasledovné:

1. Belgicko, Dánsko, Francúzsko, Holandsko, Luxembursko, Nemecko, Rakúsko, Taliansko, Veľká Británia, Švédsko, Fínsko
2. Bulharsko, Česko, Maďarsko, Rumunsko, Slovensko, Slovinsko
3. Cyprus, Grécko, Írsko, Malta, Poľsko, Portugalsko, Španielsko
4. Estónsko, Lotyšsko, Litva

Prvá skupina

Túto skupinu štátov tvoria pôvodné členské štáty (EÚ 15), charakterizované ako krajiny s vyšším potenciálom, úrovňou ekonomiky, lepšou politickou pozíciou a statusom. Vo vytvorenej skupine sa zoskupili typy štátov s nižšou mierou úmrtnosti už v začiatkoch sledovaného obdobia. Tento proces zaznamenával pozvoľný pokles z 13 ‰ na dnešných 8 ‰. Najväčší pokles pôrodnosti vidieť do 90.tych rokov a odvtedy si zachováva približný status od 8 do 13 ‰. Podobný vývoj sa odohral aj v prípade dojčenskej úmrtnosti s tým rozdielom, že dnešná hodnota sa zastavila na hranici 8 ‰. Úhrnná plodnosť sa zmenšila z do 3 na dnešnú 1,2- 2. Dôsledky demografickej revolúcie však nespočívali len v poklesoch hodnôt jednotlivých ukazovateľov, ale u niektorých bol zaznamenaný i nárast. Napríklad koeficient mimomanželskej pôrodnostii sa do 80.tych rokov udržoval na hranici 20 %, dnes je to v rozpätí 20- 56 %. V minulosti sa hrubá miera sobášnosti pohybovala na úrovni 5- 10 ‰, dnes je to len do 7 ‰. Prirodzený prírastok v tejto skupine štátov poklesol asi o 1 ‰ a to na -0,2 ‰ až 0,4 ‰. Vek žien a mužov pri prvom sobáši postupne stúpal, čo možno zdôvodniť odďaľovaním inštitúcie manželstva spolu s uprednostňovaním kariéry pred rodinou. Vek mužov to bol posun o 2- 3 roky na dnešných 31 až 32 rokov. Ženy zaznamenali posun hranice na 26 až 28 rokov.

Rozvodovosť, ako negatívny prejav,stúpala od 80.tych rokov na 3 ‰. Postupný pokles sa dá potvrdiť aj čo sa týka potratovosti, čo možno pripísať pokroku v oblasti zdravotníctva. Táto skupina štátov je opísaná ako zoskupenie, v ktorých sa prechod vyskytol skôr, hodnoty sa stabilizovali . Obyvateľstvo v týchto štátoch pomerne rýchlo zareagovalo na zmeny v postoji a správaní sa.

Druhá skupina

Do danej skupiny sa zaradili štáty novopričlenené. Mnohí z autorov im dávajú prívlastok menej rozvinutých štátov únie. V tejto časti únie sa demografický prechod zaznamenal o niečo neskôr ako tomu bolo v predošlej skupine. Napriek tomu zmeny boli o to rýchlejšie a intenzívnejšie. Najväčší pokles úmrtnosti bol zrejmý do 90.tych rokov a od toho obdobia si udržuje hodnotu do 14 ‰. Pôrodnosť reagovala na jednotlivé zmeny o niečo prudšie. V priebehu sledovaného obdobia poklesla z 13-14 ‰ na 11 ‰ (v roku 2003). Úmrtnosť dojčiat sa líši od prvej skupiny štátov, zaznamenáva od počiatku obdobia vyššie hodnoty a to zhruba do 20 % (v minulosti až do 60 %). Úhrnná plodnosť začala poklesávať od 90.tych rokov až po hodnotu 1,2, čo je už kritická hodnota z hľadiska reprodukcie obyvateľstva. Kohabitácia, ako možnosť spolužitia partnerov, aj v tejto skupine štátov postupne nadobudla svoj význam a tým sa následne prejavila aj na koeficiente mimomanželskej pôrodnosti, ktorý sa dostal na 40 %. Začiatkom 90.tych rokov sa poklesávajúca pôrodnosť a stabilizujúca úmrtnosť prejavili na prirodzenom prírastku. Na konci sledovaného obdobia sa pohyboval v rozpätí -0,5 až + 0,5 ‰. Novoprichádzajúce myšlienkové postoje sa prejavili na sobášnosti a tým sa ustálila na 3 až 5 ‰. Potratovosť sa dá charakterizovať vysokými hodnotami v začiatku sledovaného obdobia (až do 150 %) a postupným menej výrazným poklesom.

Druhá skupina štátov sa dá teda charakterizovať ako štáty s výraznou reakciou na zmeny, neskorším nástupom demografického prechodu (revolúcie) s o to výraznejšími reakciami.

Tretia skupina

Tretí typ štátov je spomedzi skupín EÚ 15 a EÚ 12. Týchto štátov sa zmeny dotkli osobitne. V každej etape sa za významnú rolu považuje religiozita. Tá vplýva na správanie sa obyvateľstva, populačný vývoj, myšlienkové postoje a vnímanie. Úmrtnosť si udržuje hodnotu do 11 ‰. Úmrtnosť dojčiat je podobná ako u predošlej skupiny. Naopak pôrodnosť zaznamenáva od počiatku pomerne vysoké hodnoty do 12 ‰ (na začiatku obdobia až do 24 ‰). Spolu s úhrnnou plodnosťou si teda zachovávajú jedny z najvyšších hodnôt. Religiozita ako taká tu zaručuje minimálne prejavy kohabitácii. Podstata je v tradícii a postavení náboženstva, ktoré vplývajú na nevýraznosť prejavu mimomanželskej pôrodnosti, ktorá sa zastavila na hranici 20-30 % (v minulosti len do 8 %). Prirodzený prírastok je pozitívne ovplyvňovaný dobrou situáciou v oblasti pôrodnosti a nižšej úmrtnosti. Pohybuje sa v rozpätí 0,8 až 0,2 ‰. Takisto sa charakteristické črty skupiny odzrkadlili aj na vývoji sobášnosti , ktorá od začiatku zaznamenáva pomerne vysoké hodnoty 12 ‰ až dnešných 5-7 ‰. Menej výrazne sa prejavuje rozvodovosť, ako negatívny spoločenský jav a to v rozpätí 2-4 ‰. Negatívny postoj si štáty zachovali aj k potratovosti (rozpätie 20-40 ‰).

Táto skupina štátov sa teda vyznačuje výraznou religiozitou, tradičným chápaním rodiny, manželstva a tým aj odlišnými odozvami na priebeh demografického prechodu.

Štvrtá skupina

Posledná vyčlenená skupina obsiahla Pobaltské štáty. V mnohých otázkach sa Litva prikláňala aj k ostatným skupinám, ale nakoniec najviac sa podobala vývoju tejto časti únie.

Skupina štvrtá združuje štáty, ktoré sa líšia svojim myšlienkovým vnímaním a pochodom, sú voľnejšie a menej obmedzené. Vývoj prebehol s výraznými zmenami. Úmrtnosť sa stabilizovala na 14 ‰ s najvýraznejšími zmenami v 75 až 80.tych rokoch. Úmrtnosť dojčiat sa v tejto skupine začala postupne stabilizovať neskôr z hodnoty do 30 % na dnešnú hranicu do 15 %. Dôvodom je pokrok v oblasti prenatálnej starostlivosti a zároveň aj v samotnom zdravotníctve a situácii v jednotlivých štátoch. Pôrodnosť si udržovala hranicu 22 ‰ a ústálila sa na hodnote do 9 ‰ (v roku 2003). Pokles zaznamenala aj úhrnná plodnosť o 1 a to z 2,5 na dnešých 1,5 (počet detí pripadajúcich na 1 matku). Kohabitácia ako taká sa v tejto skupine prejavuje v najvýraznejšej miere. Samotné Pobaltské a severské štáty majú najvýraznejšie zastúpenie tohto spôsobu mimomanželského spolužitia v celej únii. V priebehu sledovaného časového úseku sa koeficient mimomanželskej pôrodnosti zvýšil o viac ako 45 % (z 15 % v roku 1960 na 60 % v roku 2003). Tento výkyv sa odohral hlavne od 90.tych rokov. Vývoj prirodzeného prírastku sa odvíjal od stabilizujúcej sa miery úmrtnosti a poklesávajúcej sa miery pôrodnosti. V 60.tych rokoch sa pohyboval v rozpätí 1,5 až 0,2 ‰ a v posledných rokoch je to od -0,5 ‰ až -0,3 ‰ (prirodzený úbytok).

Sobášnosť ako taká sa výraznejšie neprejavovala ani na začiatku obdobia. Výrazné zmeny prišli až v 90.tych rokoch kedy sa stabilizovala na hranici 2,5 až 3 ‰. Nárast zaznamenal proces rozvodovosti a to zhruba do 90.tych rokov a odvtedy sa ustálila na hodnote 3 ‰.

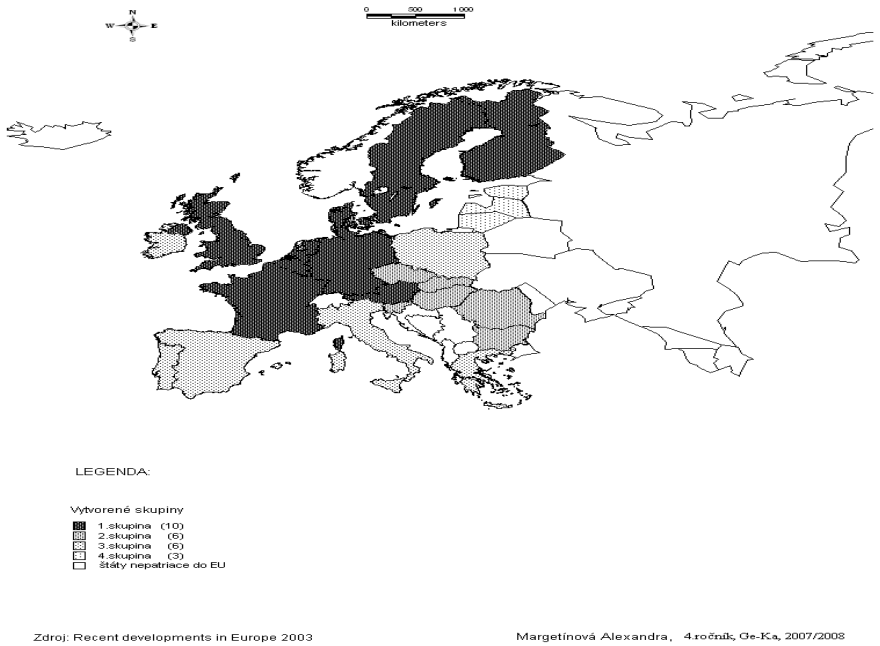
Celkovo možno túto skupinu charakterizovať ako štáty, ktoré sa odlišujú samotnou populačnou dynamikou (vývojom) od ostatných členských prvkov EÚ 12 a preto považujem za dôležité vyčleniť ich z nej.

ZÁVER

Pre celý populačný vývoj zostáva charakteristickou črtou 2.demografický prechod (revolúcia), ktorá sa postupom času odráža na viacerých procesoch populačnej dynamiky jednotlivých krajín. Na základe toho môžeme charakterizovať znaky 2. demografickej revolúcie:

- Pokles úhrnnej plodnosti pod úroveň záchovnej reprodukcie (2,1)
- Pokles natality a čistej miery reprodukcie
- Rast úmrtnosti a následný pokles
- Zvýšenie priemerného veku pri prvom sobáši
- Nárast rozvodovosti
- Popularita a rozšírenie kohabitácii
- Rozšírenie modernej hormonálnej antikoncepcie
- Pozitívny pokles v potratovosti (samotných potratov)

Klasifikácia štátov Európskej únie na základe populačného vývoja



Zdroj: Recent developments in Europe 2003

Margetínová Alexandra, 4.ročník, Ge-Ka, 2007/2008

ZOZNAM POUŽITEJ LITERATÚRY

- Bernardo, C., Demeny, P., Perutz, F. (1996) : Resources and Population. Oxford, Clarendon Press Oxford, pp.11-16, 25-47, 305-309.
- Marenčáková, J., Švajková, J. (2002): Typológia krajín Európy podľa pôrodnosti a plodnosti. Geografické informácie 7/1. s. 140-146.
- Marenčáková, J. (2003): Theoretical and methodological Aspects of Analysis of natality and fertility. In: Kowalcyzk, A. ed.: Geographical Space at the Turn of the Century. Warszaw University. Warszawa, s. 29-37.
- Michalski, T.(2006): The Geographical Aspects of the Transformation Process in Central and East- Central Europe.Gdynia- Pelpin, pp.45-57,79
- Podmanická, Z. (2002): Obyvateľstvo Európskej únie z pohľadu demografickej štatistiky v období 1960- 2002. In: 12.Slovenská štatistická konferencia- Štatistika a integrácia, Bardejovské Kúpele 2002, Slovenská štatistická a demografická spoločnosť, Bratislava, pp.71-80 [ISBN 80-88946-37-9]
- Podmanická, Z. (2005): Pôrodnosť a plodnosť v krajinách Európskej únie v období 1993-2003. In: 10 Slovenská demografická konferencia. Smolenice, Slovenská štatistická a demografická spoločnosť, pp.159- 161. [ISBN 80-88946-39-5]
- Salt, J. (1993) : Migration and Population change in Europe. New York, United Nations Institute for Disarmament Research, 2006.s.11-13.
- Vaňo, B. (2002) :Slovensko a Európska únia- demografické rozdiely a podobnosti, In: 12.Slovenská štatistická konferencia- Štatistika a integrácia, Bardejovské Kúpele 2002, Slovenská štatistická a demografická spoločnosť, Bratislava, pp.47-51 [ISBN 80-88946-37-9]
- Dáta : Recent demographic developments in Europe 2003, Council of Europe 2003, Interné dáta VDC, Interné dáta ŠUSR, Demographic Yearbook 2003

Vrodené vývojové chyby, závažný problém morbidity obyvateľstva Slovenska

Jana MARTINÁKOVÁ

Abstract

Congenital malformations as an important topic of the population morbidity in Slovakia
Incidence of congenital malformations is monitored as an indicator of a health aspect of population. I analyzed historical developemend and spatial distribution of congenital malformations of the live-born neonates on Slovakia in the period 1997 to 2005. In this period frequency of registered CM had stabile tendency. In 1997 – 2005 the average proportion of live-born newborns with CM in Slovakia was 2.492 %. These values were in accordance with European registry EUROCAT.
Key words: congenital malformations, health, teratology, medical geography

Úvod

Medicínska geografia ako veda sa nachádza na rozhraní medicíny a geografie. V kontexte humánnej geografie využíva súbor metód a techník ako i poznatky nielen geografie obyvateľstva (demogeografie), geografie sídiel (urbánnej geografie), sociálnej geografie, ekonomickej geografie, behaviorálnej geografie, ale aj kultúrnej a politickej geografie.
Mimoriadna aktuálnosť a široké spektrum problematiky ktorými sú morbidita a mortalita obyvateľstva ale i štúdium kvality života sú predmetom záujmu a výskumu medicínskej geografie.
Podstatný problém tohto záujmu predstavuje zdravie obyvateľstva. Zdarvie a zdravotný stav obyvateľstva majú v súčasnostui široké spektrum používania a interpretácie.Zdravie je podľa WHO úplný stav úplnej telesnej a duševnej pohody, neznamená len absenciu choroby či oslabenia. Zdravie a zdravotný stav obyvateľstva sa identifikujú pomocou jednotlivých indikátorov, ktoré vyjadrujú kvalitatívnu a kvantitatívnu časovú mieru toho, k čomu sa vzťahujú. Dosť často sa zdravie hodnotí pomocou negatívnych vyjadrení (ochorenia, príčiny smrti).
Geografické faktory rozšírenia niektorých chorôb sú známe prakticky odvtedy, odkedy existuje medicína ako veda.

1. Vrodené vývojové chyby

V zmysle klasifikácie príčin úmrtí a ochorení sú jednou z 20 kapitol i vrodené vývojové chyby.
Vrodená vývojová chyba (VVCh) je postnatálne zistená odchýlka vnútromaternicového vývoja, ktorá presahuje normálne rozmedzie fenotypovej variability danej štruktúry alebo jej funkcie. Je jedným typom chorôb a príčim smrti, ktoré udáva ich klasifikácia (príčina číslo 17- Vrodené chyby, deformácie a chromozómové anomálie).
V klinickej praxi sa stretávame s veľmi rôznorodým pomenovaním VVCh: malformácie, anomálie, atypický rast, vady (správne - chyby), znetvorenia, odchýlky vývoja atď.
Výskyt VVCh sa sleduje len u živo narodených novorodencov. Registrujú sa len tie z nich, ktoré diagnostikuje neonatológ. Snahy o zavedenie plošnej registrácie VVCh u mŕtvo narodených novorodencov sa v minulosti stretávali s nepochopením (Böhmer, 2002). V súčasnosti sa VVCh u mŕtvo narodených novorodencov síce registrujú, ale úplnosť hlásení je sporná.
Hlásenie o novorodencovi s vrodenou vývojovou chybou vykonáva neonatológ na tlačive „Hlásenie vrodenej chyby“, ktoré sa dostáva cez Oddelenia klinickej genetiky do Centra lekárskej genetiky FN v Bratislave, kde je Register, ktorý údaje odosiela na Štatistický úrad SR (ŠÚ SR) a do Národného centra zdravotníckych informácií.
Faktor vyvolávajúci znetvorenie plodu počas vývoja sa nazýva teratogén. (Böhmer, 2004).
Ak sa závažná porucha vývoja objaví už na začiatku gravidity, je tehotenstvo zvyčajne ukončené samovoľným potratom.
Poruchy v neskorších štádiách embryogenézy zaraďujeme k vývojovým chybám, pričom platí selekčná logika, že čím závažnejšie je postihnutie zárodku (plodu) a nepriaznivejšie podmienky pre jeho vývoj, tým skôr sa gravidita ukončí.

Vrodené vývojové chyby (VVCh) sú závažným spoločenským a zdravotným problémom. Časový a priestorový aspekt výskytu VVCh na území Slovenska je možné sledovať aj metódami medicínskej geografie.
Pre hodnotenie zdravotného stavu populácie a kontrolu nepriaznivých vplyvov vonkajšieho prostredia na reprodukciu je kvalitné a priebežné sledovanie výskytu vrodených vývojových chýb (ďalej aj VVCh) jedným zo základných faktorov. (Böhmer, Šticová, 1997).
Registrácia vrodených vývojových chýb začínala pôvodne ako osobná iniciatíva lekárov, neskôr sa stala súčasťou rôznych registračných systémov.
Získané údaje sú zaťažené rôznymi metodickými nedostatkami, ale aj napriek tomu sú často jediným zdrojom informácií o výskyte porúch zárodočného vývoja, ktoré sú dostupné.
Zo štatistických prehľadov vyplýva, že vývojové anomálie sú v posledných rokoch jednou z najčastejších príčin novorodeneckej úmrtnosti. Mnohé z nich nie sú riešiteľné chirurgicky a ich následky nie je možné odstrániť. Prežívajúce deti sú ťažko postihnuté a odkázané na pomoc iných osôb. To má výrazný negatívny vplyv na rodinu postihnutého dieťaťa a tým tieto ochorenia nadobúdajú charakter celospoločenského problému. V prípade, že sa jedinec dožije pôrodu nazývame dôsledky porušenia intrauterinnej ontogenézy vrodenými vývojovými chybami.

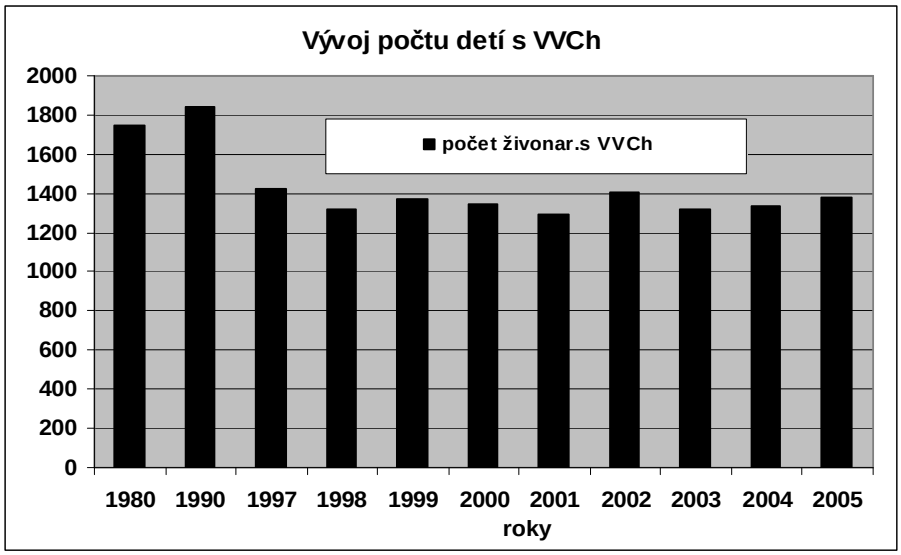
2. Časopriestorová charakteristika VVCH na Slovensku

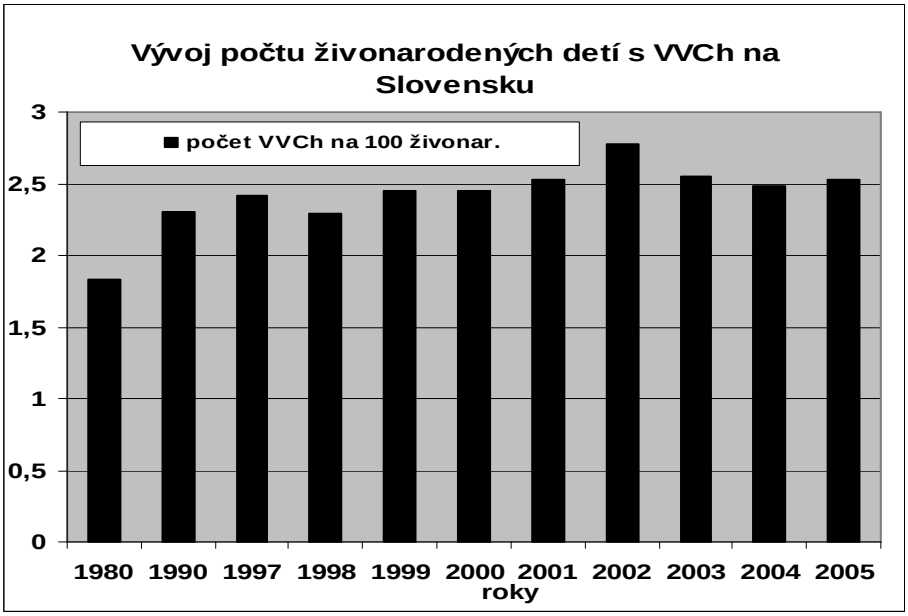
Problematika priestorového rozšírenia vrodených vývojových chýb na Slovensku patrí medzi mladé oblasti skúmania. Evidencia tohto ochorenia sa týka iba živonarodených detí, chýbajú údaje o výskyte VVCh u spontánnych potratov, umelých prerušení ťarchavosti a u mŕtvo narodených detí na Slovensku.
Analýzy sa týkali vývoja a priestorového rozmiestnenia VVCh na Slovensku u novorodencov v rokoch 1997- 2005. Použili sa oficiálne údaje Národného centra zdravotníckych informácií (do 1.2.2006 Ústav zdravotníckych informácií a štatistiky ÚZIŠ). V týchto údajoch boli také znaky ako sú počty hlásených živo narodených s vrodenou vývojovou chybou (VVCh), vek ich matky, číselný kód trvalého bydliska matky (obce), výskyt a typy VVCh a údaje o celkovom počte živo narodených za roky 1997 až 2005.

Vývoj počtu živonarodených detí s VVCh

Roky	1980	1990	1997	1998	1999	2000	2001	2002	2003	2004	2005
Poč. živonar.s VVCh	1747	1844	1427	1322	1374	1349	1294	1409	1320	1335	1377
Poč. VVCh na 100 živonar	1,837	2,305	2,414	2,296	2,444	2,446	2,531	2,771	2,553	2,484	2,530

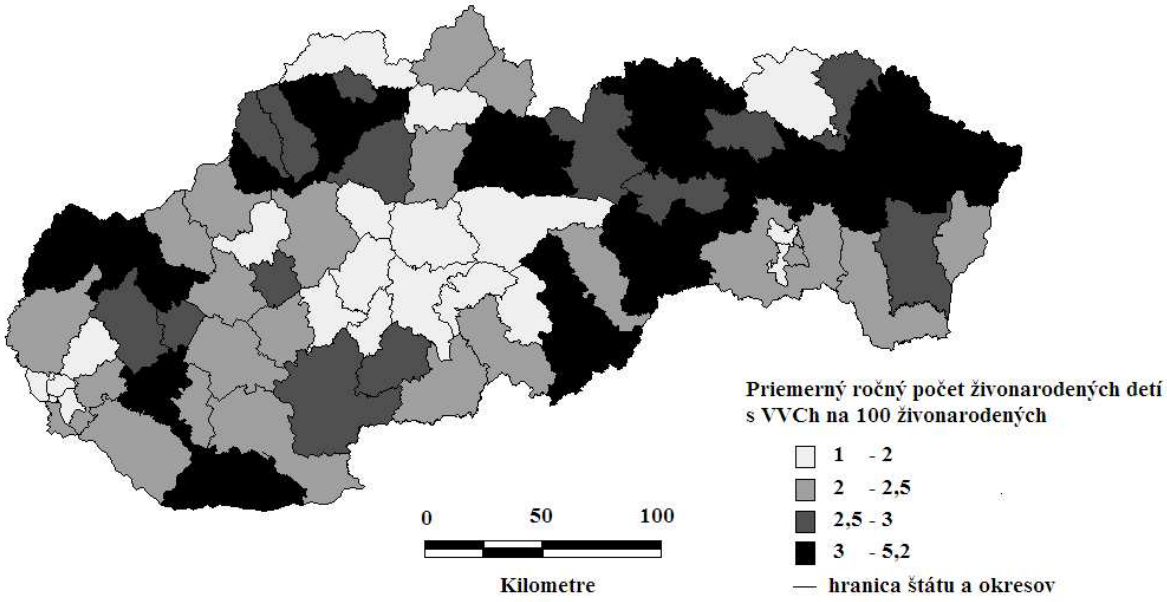
Zdroj: Zdravotnícka ročenka SR 2005





Priestorové rozmiestnenie vrodených vývojových chýb na Slovensku v rokoch 1997- 2005

Autor: Jana Martináková



Zdroj: Národné centrum zdravotníckych informácií
Štatistický úrad SR

Získané údaje o počte hlásených živonarodených s VVCh v danom roku boli roztriedené podľa jednotlivých okresov Slovenska. Frekvencia VVCh sa prepočítala pre každý okres na 100 živonarodených a zobrazila kartogramom. V sledovanom období mal výskyt hlásených VVCh za celé Slovensko ustálenú tendenciu, pri hodnotení relatívnych početností sa ukazuje mierna tendencia rastu. V roku 1980 bol výskyt VVCh na 100 živonarodených 1,837 a do roku 2005 vzrástol na 2,530. V rokoch 1997 až 2005 bol priemer percentuálneho podielu hlásených VVCh v SR 2, 492 %. Tieto hodnoty sú v zhode s európskym registračným systémom EUROCAT.

Ak hodnotíme rozmiestnenie sledovaného javu na úrovni okresov potom územie Slovenska predstavuje pestrú mozaiku. Rozdiely vo výskyte VVCh v jednotlivých okresoch sú dosť veľké. Pozoruhodné sú nižšie hodnoty výskytu v dvoch veľkomestských aglomeráciách- bratislavskej

a košickej. V okolí Brotislavy vykazujú nižšiu úroveň výskytu VVCh aj priľahlé okresy. Ďalšia oblasť nízkeho výskytu VVCh sa sformovala na strednom Slovensku, pričom jej jadro predstavujú okresy Banská Bystrica a Zvolen, ku ktorým sa priraďuje ďalších 8 okresov. Na druhej strane sú oblasti s vyššími hodnotami výskytu VVCh. Najväčšiu z nich predstavujú takmer všetky okresy východného Slovenska s výnimkou spomenutých košických okresov a Bardejova. Dosť výrazná je i vyššia koncentrácia v okresoch severného Slovenska a Stredného Považia. Pozoruhodný je aj výskyt vyšších hodnôt v páse okresov od Skalice a Myjavy cez Galantu po Komárno.

Záver

Explanácia časového vývoja a priestorových rozdielov výskytu VVCh je mimoriadne náročnou úlohou. Príčina je v mnohostranných väzbách tohto javu s celým súborom biologických, kultúrnych, ekonomických, historických faktorov, ktoré sa diferencovane prejavujú v časovom vývoji a vo veľkej regionálnej diferencovanosti.
Ako príklad možno uviesť, že mnohé výskumy ukázali, že vrodené vývojové chyby sa objavujú častejšie u detí žien, ktoré sa liečili pre neplodnosť a mali samovoľné i umelé potraty. Vplyv na výskyt vrodených chýb má i vek matky, najmä v prípadoch, kedy ide o tehotenstvo žien mladších ako 17 a starších ako 35 rokov.
Za rizikové sa považujú aj viacpočetné tehotenstvá, niektoré regionálne vplyvy, nevhodné ekonomické podmienky, tehotenstvá slobodných žien, nedostatočné vzdelanie a negatívne vplyvy rodinného a pracovného prostredia. Tieto tvrdenia nemôžem potvrdiť, pretože som pri spracovávaní údajov neskúmala vplyv veku matky pri pôrode a vplyv iných faktorov na výskyt VVCh.
A práve táto problematika bude predmetom nášho ďalšieho poznávania.

Použitá literatúra

1. Böhmer, D.: Analýza výskytu vrodených vývojových chýb u živonarodených novorodencov v Slovenskej republike, Gynekológia pre prax, Roč. 2, č. 1, 2004, s. 34-36
2. Böhmer, D.: Autoreferát dizertačnej práce, Ústav lekárskej biológie a genetiky, Lekárska fakulta UK v Bratislave, Bratislava, 2002
3. Böhmer, D.: Samovoľné reprodukčné straty a vrodené vývojové chyby ako súčasť reprodukčnej biológie človeka, habilitačné práca, Bratislava, 2003, s. 177.
4. Böhmer, D., Šticová, A.: Priestorová diferencovanosť územia Východného Slovenska z hľadiska výskytu vrodených vývojových chýb, 6. Demografická konferencia Slovenskej štatistickej a demografickej spoločnosti: Pôrodnosť a vybrané aspekty reprodukcie obyvateľstva, Domaša, 1997, Zborník prednášok in extenso, str. 118-123
5. Kandráčová, V.: Medicínska geografia- nová geografická disciplína so starou tradíciou, BIGECHE č.7, Odborno- metodický občasník pre učiteľov biológie, geografie a chémie na základnej a strednej škole, Metodicko. Pedagogické centrum v Prešove, 2005, str. 49- 55
6. Zdravotníctvo v štatistických údajoch, vrodené chyby na Slovensku v roku 1997, Ústav zdravotníckych informácií a štatistiky v Bratislave, 1998
7. www.bio-medical.ac.uk/eurocatlive/search, 21.11.2007, 18:20

Jana Martináková, 4. roč. Demogeografia, Prírodovedecká fakulta UK v Bratislave
m.janka.11@gmail.com

Vzdelanie žien na Strednom Východe a v Severnej Afrike

Viera Onderišinová

Education is a key factor which improves individuals' status in a society. In the region of Middle East and North Africa (MENA),* we observe a dramatic improvement in access to education in past few decades, and also in girls' and women's education.

This article briefly offers an overview of education's benefits to women, families, economies, and societies and highlights the ongoing concerns about education in MENA countries. It also looks at education's links with fertility and employment, two important elements in women's empowerment.

Vzdelanie je kľúčovou časťou stratégie zlepšujúcej životnú úroveň jednotlivcov a podporujúcej ekonomický a sociálny rast spoločnosti. Na Strednom Východe a v Severnej Afrike sa v posledných dekádach dramaticky zlepšil prístup k vzdelaniu. Rovnako tak tu bolo zaznamenaných mnoho povzbudivých trendov vo vzdelávaní žien a dievčat. Návštevnosť základných škôl je v regióne vysoká a rozdiely medzi pohlaviami na stredných školách už v niektorých krajinách úplne vymizli.

Avšak najväčšia výzva stále zostáva. Mnoho detí- prevažne dievčat- je stále vynechávaných zo vzdelávacieho procesu, a ešte viac ich je v ňom zapojených, no naučia sa príliš málo nato, aby boli pripravené vstúpiť na pracovný trh 21. storočia. V niektorých krajinách, prístup k stredoškolskému a vyššiemu vzdelaniu, ktoré pomáha vytvoriť kvalifikovaných a dobre informovaných pracovníkov, je naďalej limitovaný; a aj na miestach kde prístup nie je problémom, kvalita vzdelávania je veľmi nízka.

Vzdelanie: Sociálne právo a smernica vývoja

Dôležitosť vzdelávania bola mnohokrát vyzdvihovaná na medzinárodných kongresoch, vrátane deklarácie o ľudských právach a v akčnom programe medzinárodnej konferencie o populácii a rozvoji z roku 1994. Štvrtá svetová konferencia o ženách, ktorá sa uskutočnila v roku 1995 v Pekingu uznala, že vzdelanosť žien je kľúčom k posilneniu ich účasti na rozhodovacom konaní v spoločnosti a na zlepšení prosperity rodiny.

Vo všeobecnosti, prínos edukačného procesu žien na ich postavenie v spoločnosti a rovnosť pohlaví je jasne rozoznateľný.

- Ak vzdelanie žien stúpa, fertilita, populačný rast, dojčenská a novorodenecká úmrtnosť klesá a zlepšuje sa zdravie rodiny
- Nárast počtu stredoškolsky vzdelaných dievčat je spojený s nárastom participácie žien na pracovnom trhu a ich prispievaniu do rodinného a národného príjmu
- Nárast pracovného príjmu u žien má pozitívny vplyv na výživu dieťaťa

Kultúrne a ekonomické faktory podporujúce nerovnosť pohlaví

Krajiny Stredného Východu a Severnej Afriky dosahujú nižšiu úroveň vo vzdelávaní žien a ich zapojení do pracovného trhu, ako v ostatných regiónoch s podobnou úrovňou príjmov. Interakcia medzi ekonomickou štruktúrou regiónu a konzervatívnou kultúrou,

v ktorej je tradičné rozdelenie úloh medzi pohlaviami striktne presadzované, je zato v značnej miere zodpovedná.

V Islame, vzťahy medzi pohlaviami sú riadené skôr princípom doplňovania sa, než princípom rovnosti. V Islamských spoločnostiach sú úlohy mužov a žien dlhodobo spoločensky rozdelené, vytvárajúc pre ženy miesto pôsobnosti v súkromnej sfére života a pre mužov vo verejnej sfére.

Túto nerovnosť badáme aj v rozdielnej vzdelanostnej úrovni mužov a žien. Takmer jedna dospelá žena z dvoch, verzus menej než 1 chlap zo 4, v regióne MENA nevie ani čítať, ani písať. Najviac negramotných žien je v sub- regióne Severnej Afriky, najmenej v sub- regióne Západnej Ázie. Najvyššiu početnosť má prvý spomínaný región vďaka Egyptu a Maroku, kde každý druhý dospelý človek nevie ani písať, ani čítať, pričom rozdiely medzi pohlaviami sú obrovské. Najmenej negramotných je v štátoch Izrael, Libanon, Turecko a Jordánsko.

Rozdielnu zapojenosť dievčat a chlapcov do vzdelávacieho procesu môžeme vysledovať aj z pomeru návštevnosti základných škôl v roku 2000 (podľa UNESCO je základom pomeru 100% návštevnosť chlapcov). V Severnej Afrike činí tento pomer 91%, v Západnej Ázii 84% a na Arabskom Polostrove 65%. Severoafrické krajiny (okrem Maroka) sa v súčasnosti snažia doplniť medzeru vo vzdelanosti, ktorá sa vytvorila počas 60., 70. a 80.- tych rokov – obzvlášť vo vzdelanosti žien- implementovaním dynamického edukačného postupu. Spomínaná medzera je značne reflektovaná v miere negramotnosti dospelých.

Muži v MENA majú teda jednoduchší prístup nielen k vzdelaniu, ale aj k zamestnaniu, správe majetku, zatiaľ čo ženy sú v značnej miere ekonomicky závislé na mužských členoch rodiny. A aj napriek nárastu v ich zamestnanosti, existujú rôzne druhy regulácií, presne vytyčujúce spoločensky akceptovateľné pracovné pozície.

- Práca by nemala vyžadovať od ženy porušiť Islamský zákon (napr. podávať alkoholické nápoje) a mala by mať na pamäti jej bezpečie.
- Ak práca vyžaduje aby žena opustila svoj dom, musí si ona strážiť a zachovať čistotu
- Práca by nemala ovplyvniť dôležité poslanie ženy, a to starostlivosť o svoju rodinu.

(V štátoch Perzského Zálivu boli „ostatné“ pracovné pozície (napr. čašníčky) doplnené importom ženských pracovných síl z Južnej a Východnej Ázie)

Aj napriek tomu sa situácia v regióne pomaly zlepšuje. Zvyšujúce sa náklady na živobytie nútia rodiny byť závislé aj od príjmu žien. Rovnako tak, politický nátlak k reformám a ekonomické zmeny v krajinách poskytujú ženám možnosť aktívne sa zapojiť do života mimo rodiny.

Vplyv vzdelania na reprodukciu a zamestnanosť

Sobáš je zlomovým bodom v živote mladej ženy na Strednom Východe a v Severnej Afrike. Manželstvo a pôrod definujú život takmer všetkých žien v tomto regióne. Až do posledných dekád sa ženy z celého regiónu vydávali v teenagerskom veku alebo v skorých 20- tich rokoch svojho života. Súčasné transformačné procesy sú spojené

s odkladaním sobáša do neskoršieho veku a manželstvom s mužom ktorý je žene vekovo bližšie.

Vzdelanie je jediný a najdôležitejší determinant pre faktory ako vek vstupu do manželstva a vek pri 1. pôrode v krajinách MENA, keďže ženy v regióne zvyknú porodiť svoje 1. dieťa krátko po uzavretí manželstva.

Napr. vydaté Egyptské ženy vo veku 25-29 rokov, ktoré nemali žiadne vzdelanie sa vydávali v priemere vo veku 18 rokov a prvé dieťa porodili vo veku 20 rokov. Naopak ženy, ktoré mali stredoškolské alebo vyššie vzdelanie sa vydali v priemere vo veku 23 rokov a prvé dieťa priviedli na svet vo veku 25 rokov. V Turecku demografický výskum z roku 1998 ukázal, že 22% dievčat vo veku 15-19 rokov, ktoré nemali žiadne vzdelanie, alebo nedokončili základnú školu, bolo v čase výskumu už matkami, alebo boli tehotné, v porovnaní s 2% dievčat, ktoré mali ukončené stredoškolské vzdelanie.

Vzdelané ženy majú vo všeobecnosti menšie rodiny a viac využívajú informácie a služby plánovaného rodičovstva. Ako príklad poslúži Maroko (TAB 1), kde ženy so stredoškolským alebo vyšším vzdelaním majú o polovicu menej detí ako negramotné ženy.

Rozdiely medzi ženami s rozdielnym stupňom vzdelania badáme aj v prípadoch, ak hodnotíme úmrtnosť dojčiat, alebo používanie modernej antikoncepcie.
V Egypte bola ženám s nižším vzdelaním v menšej miere poskytnutá prenatálna starostlivosť (34% žien), v porovnaní so 75% tých so stredoškolským alebo vyšším vzdelaním.

TAB 1 Plodnosť vo vybraných štátoch MENA podľa dosiahnutého vzdelania matky

krajina	stupeň vzdelania			ÚP
	bez vzdelania	základné	stredné a vyššie	
Egypt (2000)	4,1	3,8	3,2	3,5
Jordánsko (2002)	3,7	3,8	3,7	3,7
Maroko (2003)	3	2,3	1,8	2,5
Turecko (1998)	3,9	2,4	1,6	2,6
Yemen (1997)	6,9	4,7	3,1	6,5

Zdroj: Country reports from latest Demographic and Health surveys

Zamestnanosť

Po tom, čo v krajine vzrástla vzdelanostná úroveň žien, viac z nich sa zapojilo do pracovného trhu. Avšak ich spoluúčasť je stále nízka: iba 30% žien vo veku 15 a viac rokov v krajinách MENA pracuje, čo je najnižšia úroveň zo všetkých svetových regiónov. Najväčší podiel tvoria v Libanone, Maroku, Turecku a Jemene. Pre porovnanie, vo Francúzsku pracuje 45% žien; v Indonézii, ktorá je domovom najpočetnejšej Moslimskej populácie, ženy tvoria 38% podiel pracovných síl. Najnižšiu mieru zapojenosti žien nachádzame v krajinách GCC (Gulf Cooperation Council), teda v skupine 6 konzervatívnych monarchií.

Percentuálny podiel platených zamestnaných žien vzrástol z 25% v roku 1980 na 30% v roku 2006. Ak participácia žien bude aj naďalej narastať a priblíži sa bližšie k celosvetovému priemeru 52%, potom ženy budú predstavovať obrovský potenciál pracovných síl v regióne Stredného Východu a Severnej Afriky.

Ženy s nižším stupňom vzdelania boli väčšinou zamestnané v poľnohospodárstve alebo vo fabrikách. Tie, ktoré boli zamestnané mimo poľnohospodárstva, mali zväčša ukončené stredoškolské alebo vyššie vzdelanie a zastávali funkcie vo verejnom sektore.

Záver:

Zlepšovanie prístupu k vzdelaniu a kvality vzdelávania je najužitočnejšia a najlepšia investícia, akú krajina môže spraviť. Investovanie do vzdelania žien v regióne MENA urýchli jeho ekonomický a sociálny rast tým, že dôjde k zdokonaleniu ľudského kapitálu, spomaleniu populačného rastu a zmierneniu chudoby.

Rôzne štúdie poukazujú nato, že región Stredného Východu a Severnej Afriky sa plne ekonomicky nerozvinie dovtedy, pokiaľ nebudú hrať ženy v ekonomike a v spoločnosti dôležitejšiu úlohu. Navzdory pôsobivému vzrastu vo vzdelaní a zdravotnej starostlivosti, ženy v regióne stále čelia pohlavnej diskriminácii, ktorá im zabraňuje v dosiahnutí vlastného potenciálu.

Pozn.: Región Stredného Východu a Severnej Afriky (Middle East and North Africa-MENA) zahŕňa 20 krajín, rozprestierajúcich sa v Západnej Ázii a v Severnej Afrike. Area pozostáva z 3 subregiónov: Severnej Afriky, Západnej Ázie a Arabského polostrova.

Použitá literatúra

Moghadam, V. M., Roudi-Fahimi, F. (2005): *Reforming Family Laws to Promote Progress in the Middle East and North Africa.* Washington, DC Population Reference Bureau.

Rashad, H., Osman, M., Roudi-Fahimi, F. (2005): *Marriage in the Arab World*. Washington, DC Population Reference Bureau.

Roudi-Fahimi, F., Moghadam, V. M. (2003): *Empowering Women, Developing Society: Female Education in the Middle East and North Africa.* Washington, DC Population Reference Bureau.

Roudi-Fahimi, F., Kent, M. M. (2007): *Challenges and opportunities- The Population of the Middle East and North Africa*. Population Bulletin, Vol. 62, No. 2, June 2007.

Tabutin, D., Schoumaker, B. (2005): *The Demography of the Arab World and the Middle East From the 1950s to the 2000s.* Population-E 60, nos. 5-6 528-29.

Viera Onderišinová, PRIF UK Bratislava, 5. roč., e-mail: onderisinova@yahoo.com

Vybrané modely oceňovania akciových trhov

Tomáš Osička

Abstract: From among many approaches determining whether the stock market is priced correctly, I have chosen two (E-Star model, Asness model). They compare real E/P ratio with E/P ratio of a correctly priced market, which is the result of an econometric model. Furthermore, I have designed my own model, trying to introduce a slightly different concept. Finally I have concentrated on verifying these models and examining their characteristics.

Key words: E-STAR model, Asness model, potential output based model.

1. Úvod

V histórii sú známe prípady, keď investori tak horlivo investovali do akcií, že vlastnými nákupmi dvíhali ich cenu. To vyvolávalo predstavu, že investícia je zisková a oplatí sa do titulov investovať ďalšie prostriedky. Zmena cien je výsledkom pôsobenia ponuky a dopytu po statkoch, teda aj po akciách. Prečo by potom malo byť na tom, že investori investovaním ovplyvňujú ceny titulov, niečo zlé? Racionálne je kúpiť podhodnotenú akciu, teda takú ktorej trhová hodnota je nižšia ako vnútorná. Vnútorná hodnota akcie môže byť daná ziskovosťou spoločnosti a stabilitou jej hospodárskych výsledkov. Hoci tento pojem budí dojem, akoby malo ísť o niečo objektívne, nemusí to tak byť. Investori pri rozhodovaní popri objektívnych informáciách o stave firmy používajú aj svoju skúsenosť či očakávania. Ak sú potom hodnotenia väčšieho množstva investorov nesprávne, ich očakávania prehnané, môže nastať jav známy ako cenová bublina. Účastníci trhu očakávajú, že ceny pôjdu hore, čo sa nejaký čas aj prejaví. Ale nárast bude „priveľký“ a skutočné výnosy neuspokoja očakávania investorov, ktorí začnú sťahovať svoje prostriedky, čím spôsobia pokles ceny. Ten vystraší ďalších investorov, ktorí sa pokúsia tiež čo najskôr predať a nastane reťazová reakcia podobná tej pri nafukovaní bubliny. Bublina spľasne.

Aby sa investori vyhli podobným omylom, ale aj aby centrálne banky dokázali zabrániť šokom, ktoré môžu mať citeľné makroekonomické dôsledky, pokúšajú sa ekonómovia kvantifikovať spomínaný „priveľký“ nárast. Ich cieľom je nájdenie objektívnej vnútornej hodnoty nezaťaženej individuálnymi očakávaniami. V tomto článku budem analyzovať tri podobné prístupy k oceňovaniu akciových trhov z ekonometrického hľadiska. Dva z nich sú používané v praxi, tretí som zostrojil s využitím skúseností zo skúmania predchádzajúcich dvoch.

2. O modeloch

Všetky modely majú spoločný základný rys. Správnosť ocenenia identifikujú na základe porovnania dosiahnutého ziskového výnosu (E/P, recipročná hodnota P/E ratia) so ziskovým výnosom, ktorý odpovedá objektívnej vnútornej hodnote akciového trhu. Takýto ziskový výnos zvykne byť označovaný ako fair E/P. Odlišnosti nastávajú v spôsobe, akým fair E/P vypočítavajú. Modelmi budem skúmať správnosť ocenenia indexu S&P 500. Analyzujem 189 pozorovaní štvrťročných dát od prvého kvartálu roku 1960 po prvý kvartál roku 2007. Historické ceny indexu S&P 500 a jeho P/E ratio som získal zo stránok Standard & Poor’s (http://www2.standardandpoors.com/). Makroekonomické dáta týkajúce sa americkej ekonomiky som čerpal zo stránok banky Federálneho rezervného systému v St. Louis (https://research.stlouisfed.org/fred2/).

E-STAR model

Tento model publikovala investičná banka Phillips & Drew v roku 1999. Autori Cambellovou-Shillerovou log-linearizáciou Gordonovho dividendového diskontného modelu (podrobne Tregler) odvodili Gordonov model v tvare:

$$\ln(\delta_t) = \alpha_0 + \alpha_1 \ln(\phi_t) + \alpha_2 \ln(\theta_t) + \alpha_3 \ln(g_t)$$

kde δ_t je dividendový výnos, ϕ_t je riziková prémia akciového trhu, θ_t je reálna bezriziková úroková sadzba, g_t je reálne tempo rastu dividend, α_0 je úrovňová konštanta a α_1, α_2, α_3 sú koeficienty.

Tieto premenné nie sú v modeli počítané priamo, ale nahradené premennými, ktoré sú ľahšie kvantifikovateľné.

Premenná	Použitá premenná	Značenie
d_t	podiel ročných výnosov vypočítaných ako 10-ročný kĺzavý priemer ku cene indexu S&P 500	*E10_p*
ϕ_t	4-ročná kĺzavá smerodajná odchýlka štvrťročnej miery inflácie	*stdev4_i*
θ_t	štvrťročná miera inflácie (% zmena štvrťročných úrovní CPI)	*i*
g_t	očakávaný rast produktivity práce, podiel HPP_{t+1}/HPP_t, vyhladený Hodrick Prescottovým filtrom	*E_HPP_hp*

Tabuľka 1: Zoznam použitých premenných v E-Star modeli

Asnessov model

Model Clifforda Asnessa, prezentovaný v roku 2000, skúma rozdiel medzi výnosom štátnych dlhopisov a dividendovým, resp. ziskovým výnosom. Predpokladal, že očakávaný výnos akcií *E(A)* je funkciou dividendovému výnosu (*D/P*) a že očakávaný výnos štátnych dlhopisov *E(D)* je funkciou aktuálneho výnosu štátnych dlhopisov (*Y*):

$$E(A) = a + b\left(\frac{D}{P}\right)$$

$$E(D) = Y$$

kde *E(A)* je očakávaný výnos akcií, *E(D)* je očakávaný výnos štátnych dlhopisov, *D/P* je dividendový výnos, *Y* je aktuálny výnos štátnych dlhopisov a *a*, *b* koeficienty.

Rozdiel medzi očakávaným výnosom akcií a dlhopisov vyjadril prostredníctvom dlhodobých volatilít akcií a dlhopisov. Po úpravách (podrobne Tregler) odvodil tvar

$$\frac{E}{P} = \gamma_0 + \gamma_1 Y + \gamma_2 \sigma_A + \gamma_3 \sigma_D$$

kde *E/P* je ziskový výnos akciového trhu, *Y* je aktuálny nominálny výnos štátnych dlhopisov, σ_A je dlhodobá volatilita výnosu akcií, σ_D je dlhodobá volatilita výnosu dlhopisov, γ_0 až γ_3 sú koeficienty.

Premenná	Použitá premenná	Značenie
E/P	podiel ročného výnosu k cene indexu S&P	*E_P*
Y	ročný výnos štátnych pokladničných poukážok s konštantnou dobou splatnosti 10 rokov	*Y10r*
σ_A	20-ročná kĺzavá smerodajná odchýlka ziskového výnosu S&P 500	*stdev20_A*
σ_D	20-ročná kĺzavá smerodajná odchýlka výnosu dlhopisov s ratingom AAA a dobou splatnosti < 30d	*stdev20_D*

Tabuľka 2: Zoznam použitých premenných v Asnessovom modeli

Model odvodený od potenciálneho výkonu ekonomiky

Tento model je výsledkom mojej snahy o korekcie už spomenutých modelov smerom k lepšiemu plneniu ekonometrických predpokladov. Hoci sa mi tento zámer nepodarilo splniť, zostavil som vlastný model, na ktorý sa ďalej v texte budem odkazovať stručne ako na tretí model, s využitím premenných z predchádzajúcich modelov. Predpokladal som, že na tempo rastu výnosov z kapitálu (*E*) by malo byť rovnaké ako rast produktivity práce zodpovedajúci úrovni ekonomiky na svojom potenciáli. Rýchlejší rast by sa dal totiž označiť za prirýchly a vytvárajúci možnosť vzniku bubliny. Pomalší rast na druhej strane môže spôsobiť podhodnotenie akciového trhu. Výnosy z kapitálu v čase *t*, môžem potom vyjadriť pomocou výnosov v čase *t-1* a tempa rastu produktivity práce.

$$E_t = E_{t-1} g_t \qquad (1)$$

kde E_t, E_{t-1} sú výnosy akciového trhu v čase *t*, resp. *t-1* a g_t je tempo rastu produktivity práce odpovedajúce výkonu ekonomiky na úrovni potenciálu.

Ďalej som predpokladal, že rastu ceny akcie zodpovedá miera inflácie. Za indikátor inflačného rastu cien som si zvoli CPI (index spotrebiteľských cien). Treba ale upozorniť, že akcie nie sú súčasťou spotrebiteľského koša, a nie sú zachytené v hodnotách indexu. Nie je teda dôvod predpokladať, že inflácia meraná indexom spotrebiteľských cien dokáže vypovedať o pohybe cien akcií. Tento predpoklad potvrdí, respektíve vyvráti až kvantifikácia modelu. Na základe predpokladu môžem písať

$$P_t = P_{t-1} \pi_t \qquad (2)$$

kde P_t, P_{t-1} sú ceny indexu akciového trhu v čase *t*, resp. *t-1*, a π_t je miera inflácie odpovedajúca výkonu ekonomiky na svojom potenciáli.

Dám rovnice *(1)* a *(2)* do podielu a získam vzťah pre ziskový výnos

$$\frac{E_t}{P_t} = \frac{E_{t-1}}{P_{t-1}} \frac{g_t}{\pi_t}$$

Model zlinearizujem logaritmizáciou.

$$\ln\left(\frac{E_t}{P_t}\right) = \alpha_1 \ln\left(\frac{E_{t-1}}{P_{t-1}}\right) + \alpha_2 \ln\left(\frac{g_t}{\pi_t}\right)$$

kde α_1 a α_2 budú koeficienty.

Teoreticky by bolo možné pridať do modelu aj úrovňovú konštantu. To by ale znamenalo, že ziskový výnos v čase multiplikatívne rastie či klesá o pevnú konštantu, čo nie je v súlade s predpokladom. Zavedenie konštanty zhoršilo výdatnosť modelu.

Premenná	Použitá premenná	Značenie
$\frac{E_t}{P_t}$	ako závislú premennú som zvolil podiel ročných výnosov vypočítaných ako 10-ročný kĺzavý priemer ku cene indexu S&P 500, čo zníži náchylnosť modelu na výkyvy spôsobené cyklickým pohybom ekonomiky	$E10_P_t$
$\frac{E_{t-1}}{P_{t-1}}$	ziskový výnos vystupujúci ako nezávisle premenná je podiel ročných výnosov vypočítaných ako ročný kĺzavý priemer ku cene indexu S&P 500	$E1_P_{t-1}$
g_t	rast produktivity práce, podiel HPP_t/HPP_{t-1}, vyhladený Hodrick Prescottovým filtrom	HPP_hp_t
π_t	štvrťročná miera inflácie, podiel CPI_t/CPI_{t-1} vyhladená Hodrick Prescottovým filtrom	i_hp_t

Tabuľka 3: Zoznam použitých premenných v modeli odvodenom od potenciálneho výkonu ekonomiky

3. Kvantifikácia modelov

Modely som kvantifikoval za pomoci programu EViews 4.1 Student version.

E-STAR model

$$\log(E10_P_t) = -1,58(0,275) + 0,252(0,0555)\log(stdev4_i_t) + 10,05(2,80)\log(i_t) - \\ -48,16(14,72)\log(E_HPP_hp_t)$$

R^2=0,417 a hodnota F štatistiky je 44,15 čomu odpovedá prakticky nulová p-hodnota.

Asnessov model

$$E_P_t = 0,0488(0,00531) + 1,24(0,0565)Y10r_t - 0,986(0,176)stdev20_A_t - 2,48(0,225)stdev20_D_t$$

R^2=0,739 a hodnota F štatistiky je 174,32 čomu odpovedá prakticky nulová p-hodnota.

Model odvodený od potenciálneho výkonu ekonomiky

$$\log(E10_P_t) = 1,06(0,00538)\log(E1_P_{t-1}) + 10,91(1,75)\log\left(\frac{HPP_hp_t}{i_hp_t}\right)$$

R^2=0,747 a hodnota F štatistiky je 547,83 čomu odpovedá prakticky nulová p-hodnota.

4. Verifikácia modelov

Ekonomická verifikácia

V E-Star modeli je kladná závislosť medzi volatilitou inflácie a ziskovým výnosom, čo je pozitívne, pretože nárast volatility inflácie predstavuje nárast rizikovej prirážky. Ďalej kladná závislosť medzi infláciou a ziskovým výnosom rešpektuje skutočnosť, že na nárast inflácie odpovie investor vyššou požadovanou mierou výnosnosti, aby vyvážil nominálne výkyvy a reálne zostal na tom rovnako. Záporné znamienko pri poslednom koeficiente finančná ekonómia vysvetlí rastom ceny akcií pod vplyvom očakávaného nárastu výnosnosti, čo zníži ziskový výnos (E/P). Kvantifikácia E-Star modelu je v súlade s teóriou finančnej ekonómie.

V Asnessovom modeli sú výnosy desaťročných dlhopisov chápané ako bezrizikový výnos. Kladný vzťah medzi výnosmi dlhopisov a ziskovým výnosom súhlasí teóriou finančnej ekonómie. Asnessov model sa trochu ponáša na model CAPM. V ňom je k bezrizikovému výnosu pripočítaný betanásobok rizikovej prirážky. Tú dostaneme ako rozdiel rizikovosti akcií a dlhopisov. Navyše, ak rizikovosť akcií rastie, investor požaduje vyšší výnos. Ak rastie rizikovosť dlhopisov, investori sa obracajú vo väčšej miere k akciám a preto ich výnosnosť klesá. Túto logiku porušuje koeficient volatility akcií, ktorý by mal byť kladný, ale je záporný. Znamená to, že koeficienty Asnessovho modelu nie sú plne v súlade s predpokladmi finančnej teórie.

V treťom modeli sú oba koeficienty kladné. Znamená to, že nárast ročného kĺzavého priemeru ziskových výnosov povedie k nárastu 10-ročného kĺzavého priemeru ziskových výnosov, čo je v súlade s teóriou finančnej ekonómie. V prípade, že rast produktivity práce je vyšší ako rast cenovej hladiny, mala by vzrásť hodnota ziskového výnosu oproti predchádzajúcemu obdobiu. To vyžaduje aby bol koeficient $\alpha 2$ kladný, čo je splnené.

Kladné znamienko pri druhom koeficiente znamená, že zlomok HPP_HP/I_HP nevystupuje v modeli ako prevrátená hodnota, čo spĺňa pôvodné predpoklady.

Štatistická verifikácia

Koeficient viacnásobnej determinácie vyšiel pri E-Star modeli dosť nízky. To ale nepovažujem za závažný problém. Pretože cieľom týchto modelov nie je čo najvernejšie

vystihnúť vývoj ziskového výnosu, ale naznačiť aký vývoj ziskového výnosu odpovedá objektívnej vnútornej hodnote akcií.

Ekonometrická verifikácia

Rezíduá žiadneho z modelov nemajú normálne rozdelenie.

Kleinov test naznačuje, že v modeloch sa nevyskytuje multikolinearita. Ale v korelačných maticiach modelov vidno silné závislosti medzi niektorými premennými modelov, čo sa prejavilo na testoch Farrara a Glaubera, ktoré jednoznačne potvrdili prítomnosť multikolinearity vo všetkých modeloch. Netreba to brať mimoriadne tragicky, nakoľko všetky parametre sú významné aj s multikolinearitou.

Heteroskedasticitu som testoval Spearmanovým koeficientom korelácie. Na všetkých bežných hladinách významnosti a pre všetky premenné nezamietame nulovú hypotézu, teda zdá sa že je dodržaný predpoklad homoskedasticity. Iba v Asnessovom modeli sa na 10% hladine významnosti javí, že volatilita dividend vplýva na veľkosť rezíduí.

Podľa Durbin Watsonovej štatistiky je vo všetkých troch modeloch výrazná autokorelácia prvého stupňa. Tá sa prejavuje aj v korelogramoch a zvyšuje hodnoty autokorelačnej funkcie.

5. Zhrnutie

V článku som skúmal možnosti oceňovania akciových trhov. Odhadol som modely bežne používané pri oceňovaní týchto trhov, konkrétne E-STAR model a Asnessov model. Ale čitateľa môžu zaujímať aj iné trhy okrem akciového, ale problémom pri detekcii cenových bublín na nich je práve nedostatok dát. Tento problém by mohol vyriešiť model, ktorý som odvodil, ktorý zachytáva súvislosť medzi zmenou potenciálu ekonomiky a výnosom akciových trhov. Verifikácia ukázala, že hoci má problémy so splnením niektorých ekonometrických predpokladov, nie je na tom horšie než spomínané dva modely, ktoré sú aj napriek tomu v praxi využívané. Môj model však prináša odlišný pohľad, pretože spravodlivý ziskový výnos neodhaduje na základe súboru viacerých makroekonomických premenných, ale vynásobením ziskového výnosu z predchádzajúceho obdobia podielom rastu produktivity a rastu cenovej hladiny. Tento rekurzívny vzťah nepotrebuje toľko premenných ako už existujúce modely. Navyše premenné, ktoré používa, sú ľahko dostupné vo viacerých variantoch, čo umožňuje vziať do modelu taký rast výnosu, a cenovej hladiny ktorý je vzhľadom na daný trh relevantný. To z neho robí dobrý nástroj, alebo aspoň podklad, pre analýzu trhov ku ktorých nemáme dostatok dát.

6. Literatúra

TREGLER, K. 2005. *Oceňování akciových trhů*. Praha: C.H. Beck. 164 s. ISBN 80-7179-439-2

HUŠEK, R. 1993. *Ekonometrická analýza*. Praha: Ekopress. 303 s. ISBN 80-86119-19-X

Adresa autora:

Tomáš Osička
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
tomas.osicka@gmail.com

Vývoj pôrodnosti a plodnosti podľa vzdelanie matky na Slovensku

Dominika Pavlovčinová

Abstract

In the second half of the 20th century, the important changes in the habitants´ behaviour were done in all European countries and in the last decade they were done in Slovakia as well. Except the changes in the processes of the natural reproduction, the change was also done in the population structures. The new possibilities of employment had appeared, opening the borders allowed movement of the population, the preparation for the job was extended, the women´s emancipation grew. Those facts affected the decision of young people to marry and have children later.

The important stuff in the spread and the intensity of the reproduction behaviour is an education, especially the education of women. Women´s approach to the higher level of education and also their advancement on the employment market are the most important signs of nowadays. Intermarring, children´s bearing, taking care about the family and also household are the main aims in the woman´s life, but it´s not just the only stuff. The level of the finished education has a strong influence on the planning of the first descendant.

Kľúčové slová: vzdelanie obyvateľstva, diferenčná plodnosť, čistá miera plodnosti, regionálna diferencovanosť

Úvod

V druhej polovici 20. storočia dochádza k významným zmenám demografického správania sa obyvateľstva vo všetkých európskych krajinách. Zásluhou značných zmien, ktoré sa prejavili vo vývoji mnohých populačných procesov a vo formovaní populačných štruktúr sa aj na Slovensku štúdiu populačných javov venuje značná pozornosť. Okrem zmien populačných procesov prirodzenej reprodukcie, nastáva zmena aj v populačných štruktúrach. Taktiež sa objavujú aj nové možnosti pracovného uplatnenia sa, otvorenie hraníc umožňuje pohyb obyvateľstva, predlžuje sa príprava na povolanie, vzrastá emancipácia žien. Tieto javy pôsobia na odďaľovanie uzatvárania manželstva a tým aj na narodenie detí. Na Slovensku v súčasnej dobe prebiehajú v oblasti reprodukčného správania zmeny najmä v zmysle poklesu celkovej plodnosti, raste priemerného veku matiek pri narodení prvého dieťaťa, vzraste podielu detí narodených mimo manželstva, časovania narodenia detí vo väčších intervaloch. Tieto zmeny možno považovať za prejav adaptácie sa na nové životné podmienky spojené s rozsiahlymi spoločensko – ekonomickými zmenami posledných rokov.

Dôležitú úlohu pri rozsahu a intenzite zmien reprodukčného správania zohráva vzdelanie obyvateľstva, najmä však vzdelanie žien. Prístup žien k vyššiemu vzdelaniu a ich zvyšujúca sa účasť na trhu práce je jedným z najvýznamnejších znakov súčasnosti. Sobášenie sa, rodenie detí a starostlivosť o rodinu a domácnosť je cieľom v živote ženy, nie však jediným. Úroveň vzdelania má silný vplyv na časovanie prvého pôrodu. Ukončené vysokoškolské vzdelanie ide väčšinou ruku v ruke s vyššími preferenciami stať sa súčasťou pracovného trhu, takže stupeň ukončeného vzdelania môže byť dôležitý faktor pre vysvetlenie odkladania prvého pôrodu. Rastúca ekonomická aktivita žien môže takto spôsobovať väčšiu nezhodu medzi prácou a rodinou.

Cieľom tohto príspevku je vzhľadom na vyššie uvedené skutočnosti, poukázať na väzbu medzi zvyšujúcou sa vzdelanostnou úrovňou u žien a vývojovými zmenami niektorých populačných procesov (pôrodnosti a plodnosti). Z časového hľadiska môžeme sledovanie vývoja pôrodnosti podľa vzdelania rozdeliť na dve etapy.

Vývoj pôrodnosti podľa vzdelania matky v období socializmu

V období socializmu boli možnosti zárobku pre ženy starajúce sa o dieťa pomerne malé. Vyšší plat často krát záležal od veku a od „služobného veku“ ženy. Takže vo všeobecnosti bolo pre ženy lepšie mať deti skôr, v období keď boli ich možnosti zárobku nízke. Preto bol aj vek matky pri prvom pôrode u nás

veľmi nízky a počas 60.,70.,80. rokov sa vlastne ani nemenil (v priemere kolísal stále okolo 22 rokov). Rodičovstvo v tomto období nebolo finančne tak zaťažujúce ako v súčasnosti. Výdaje na dieťa boli znižované rôznymi spôsobmi. Na jednej strane to boli ceny za detské oblečenie, topánky, jedlo, na druhej strane boli rodiny podporované relatívne vysokými sociálnymi dávkami na deti. Na pracovnom trhu existovali len nízke motivácie a nabádania, ktoré by nútili ženy uvažovať nad časovaním ich pôrodu, tak aby sa prispôsobili svojej kariére v práci. Verejná politika ponúkala rôzne výhody a stimuly pre matky pomerne nízkeho veku. Napríklad, jednoduchšia dostupnosť bývania pre páry s deťmi, znižovanie splátok mladomanželských pôžičiek pri narodení každého dieťaťa atď. Určitý vplyv zohrávala i neznalosť a malá dostupnosť moderných antikoncepčných metód.

Ukazovatele pôrodnosti a plodnosti majú teda dlhodobo klesajúci charakter. Nie však, ak sledujeme hodnoty u jednotlivých úrovní vzdelanosti u žien. Ako môžeme vidieť z grafu 1 najväčší pokles čistej miery plodnosti (počet živonarodených detí k strednému stavu žien v reprodukčnom veku)zaznamenávajú ženy s najvyšším ukončeným vzdelaním základným - až 20‰ (za obdobie 1976-1989). Aj pri takomto obrovskom poklese ukazovateľa plodnosti, majú ženy so základným vzdelaním vyššie hodnoty ako ženy s vysokoškolským vzdelaním v tomto období. Tieto naopak zaznamenávajú zvyšovanie hodnôt až do konca 80. rokov. Do roku 1985 si ženy so stredným vzdelaním bez maturity a s maturitou udržiavajú vysoké a podobné hodnoty čistej miery plodnosti (24 - 26‰). Od tohto roku dochádza v oboch skupinách pokles hodnôt tohto ukazovateľa, radikálnejší je však u žien s najvyšším ukončeným vzdelaním stredoškolským bez maturity.

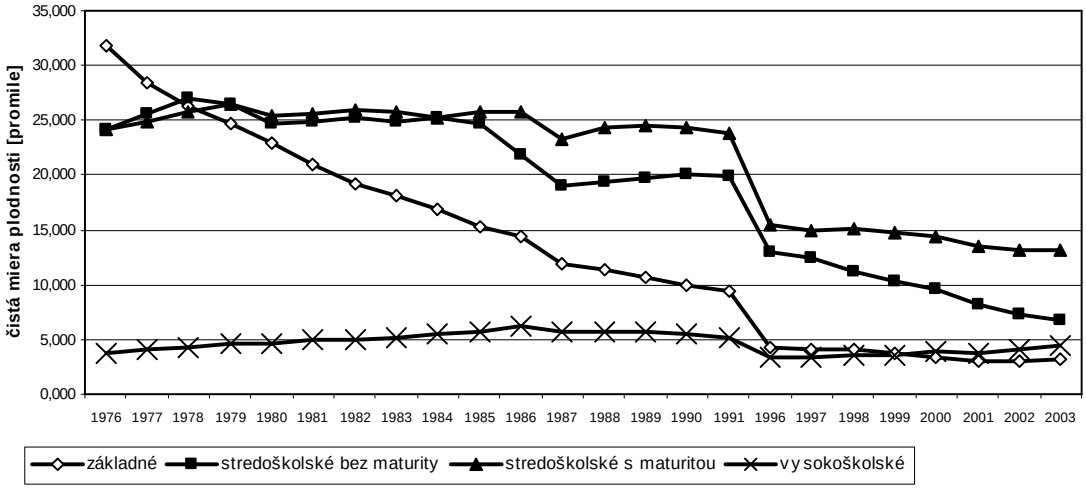
Iná však bude situácia, keď budeme sledovať počet narodených v prepočte na 1000 žien určitého vzdelania. Tieto diferenčné miery sa dajú vypočítať len za roky sčítaní, pretože vzdelanostná štruktúra obyvateľov sa zisťuje len pri sčítaní. V roku 1980 dosahovala táto miera najvyššie hodnoty u žien s najvyšším ukončeným vzdelaním bez maturity – rodilo sa im 103 detí na 1000 žien tohto vzdelania. Aj keď v absolútnych aj relatívnych počtoch živonarodených vykazovali v tomto roku najvyššie hodnoty ženy so stredoškolským vzdelaním s maturitou. Najväčší rozdiel však zisťujeme u žien s najvyšším základným a vysokoškolským. vzdelaním U žien so základným vzdelaním je to veľmi nízka úroveň diferenčnej miery – len 25 detí na 1000 žien tohto vzdelania. Pritom podiel z celkového počtu živonarodených ženám s týmto stupňom vzdelania bol v roku 1980 skoro 30%, naopak ženy s vysokoškolským vzdelaním sa na pôrodnosti v tomto roku podieľali len 6% z celkového počtu živonarodených detí., avšak na 1000 žien tohto vzdelania pripadá až 79 detí. Teda o veľa vyšší počet ako u žien so základným vzdelaním (Graf 2).

Vývoj pôrodnosti podľa vzdelania matky po roku 1989

Počas 90. rokov sa v Slovenskej republike odohrávajú dramatické zmeny. Prechod od k demokratickému spôsobu života vplýva na zmeny aj v reprodukčnom správaní obyvateľstva. Možnosti žien zapojiť sa pracovného trhu s rôznym stupňom vzdelania, sa veľmi zmenili. V trhovej ekonomike sa viac berie do úvahy ukončené vzdelanie. Ženy s vyšším stupňom vzdelania majú viac podnetov pre odkladanie pôrodov do vyššieho veku a vzdelanostné rozdiely pri časovaní prvého pôrodu sa zvyšujú. Je však pravda, že hodnota ženského vzdelania býva ešte stále ovplyvňová časom potrebným na výchopvu dieťaťa.

Pri analýze počtu a podielu živonarodených podľa vzdelania matky je evidentný nárast počtu pôrodov v kategóriách vyššieho vzdelania (stredoškolské s maturitou a vysokoškolské). Najväčší podiel pôrodov v poslednej dekáde realizujú ženy so stredoškolským vzdelaním s maturitou. V roku 2004 to bolo až 48,9%. Výrazný nárast bol zaznamenaný v kategórií žien s vysokoškolským vzdelaním. Podiel detí narodených ženám so základným vzdelaním sa, neustále znižuje. V roku 2004 už to bolo len niečo málo cez 10%. Podobná je situácia u žien s najvyšším ukončeným stupňom vzdelania stredoškolským bez maturity (v roku 2004 už to bolo 10%).

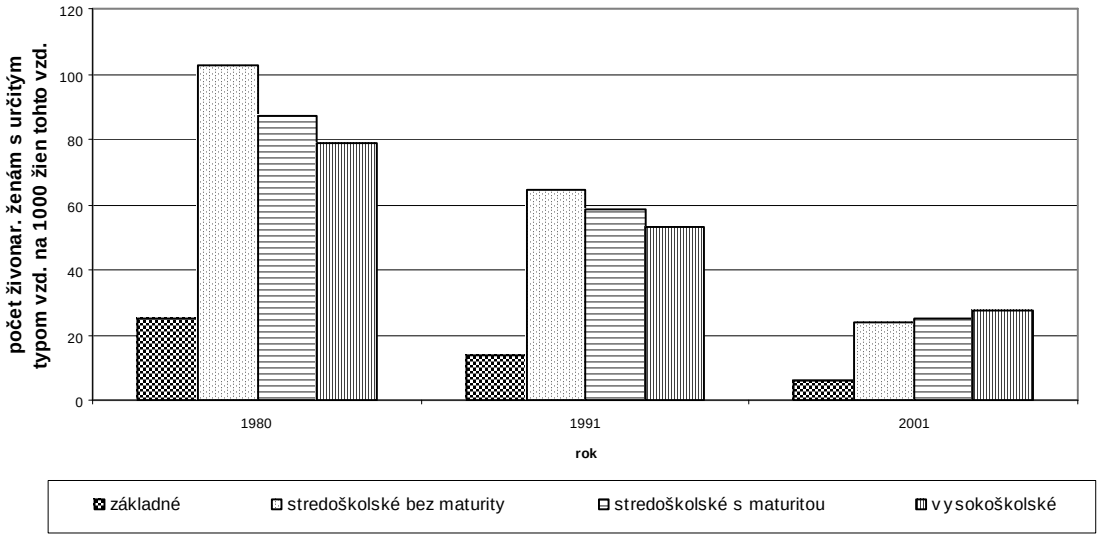
Graf 1 Čistá miera plodnosti podľa vzdelania na Slovensku



Prameň: Pohyb obyvateľstva SR 1976-2004

Zaujímavá je však situácia, keď sledujeme počet živonarodených v prepočte na celkový počet žien s jednotlivými úrovňami vzdelania (Graf 2). V roku 2001 sa situácia výrazne mení a na prvé miesto v počte živonarodených vo „svojom" vzdelaní sa dostávajú vysokoškolsky vzdelané ženy. Rodia v priemere 27 detí na 1000 vysokoškolsky vzdelaných žien, čo je oproti roku 1991 skoro dvojnásobný pokles, ale stále najvyšší počet. Ženy so stredoškolským vzdelaním bez maturity (z celkového počtu žien s týmto vzdelaním), ale v roku 2001 sú až na treťom mieste, za ženami so stredoškolským vzdelaním s maturitou. Ženy so základným vzdelaním už rodia v priemere len 6 detí na 1000 žien so základným vzdelaním.

Graf 2 Počet živonarodených ženám s určitým typom vzdelania na 1000 žien tohto vzdelania na Slovensku v rokoch 1980, 1991, 2001



Prameň: SODB 1980,1991,2001, Pohyb obyvateľstva SR 1980,1991,2001

Vnútorne je Slovensko v rámci pôrodnosti podľa vzdelania matky značne diferencované. Najvyššie absolútne hodnoty počtu narodených dlhodobo vykazuje východné Slovensko, konkrétne Prešovský kraj. V porovnaní s ostatnými krajmi sa tu v roku 2004 rodí najviac detí ženám so stredoškolským vzdelaním s maturitou. Vysokoškolsky vzdelaným ženám sa najviac detí narodí v Bratislavskom kraji, ženám so základným vzdelaním zase v kraji Košickom. Je to logické, pretože práve v týchto krajoch sa nachádza najviac žien s tým ktorým vzdelaním. Pri sledovaní počtu živonarodených ženám s jednotlivými typmi vzdelania v prepočte na celkový počet týchto žien (podľa Sčítania obyvateľov 2001) vidíme, prvenstvo držia vysokoškolsky vzdelané ženy. Najnižšie hodnoty tejto diferenčnej miery dosahujú ženy so základným vzdelaním – a to skoro na celom Slovensku. Úplne opačná je však situácia v dvoch východoslovenských krajoch. V Košickom kraji ženy so základným vzdelaním rodia dokonca najviac so všetkých úrovní vzdelania - 30,16 detí na 1000 žien zo základným vzdelaním.

Záver

Z časového hľadiska sledujem výrazný pokles pôrodnosti u žien s nižším vzdelaním (základné, stredné bez maturity) a naopak nárast u žien s vyšším vzdelaním (stredné s maturitou, vysokoškolské). Najvyššiu úroveň pôrodnosti a plodnosti vykazujú ženy so stredným vzdelaním s maturitou, ale v prepočte na celkový počet žien s týmto vzdelaním sa na Slovensku v roku 2001 dostávajú až na druhé miesto za ženy vysokoškolsky vzdelané. Týchto žien je síce na Slovensku podľa Sčítania obyvateľov 2001 len 8,7%, ale v prepočte na 1000 vysokoškolsky vzdelaných žien, rodia najvyšší počet detí. Ženy s vyšším vzdelaním mávajú väčšinou aj partnerov s podobným vzdelaním a tiež lepšie životné podmienky, preto je pre nich jednoduchšie mať väčšiu rodinu, i keď založenú neskôr ako ženy s nižšími úrovňami vzdelania.

Druhou hlavnou črtou pôrodnosti podľa vzdelania je jej priestorová diferencovanosť. Je viditeľná aj v rámci subregiónov Slovenska. V rámci krajov, je základnou črtou priestorovej diferencovanosti nárast ukazovateľov u žien s vyššou úrovňou vzdelania od východu na západ a zároveň zo severu na juh. Extrém pri ženách s vysokoškolským vzdelaním je Bratislavský kraj. U žien s nižšími stupňami vzdelania vidíme opačný trend. Extrémne sú najmä hodnoty počtu pôrodov a ukazovateľov pôrodnosti a plodnosti pre ženy so základným vzdelaním na východnom Slovensku. Dôvodom je najmä rómske etnikum, u ktorého je typický nízky vek matky pri pôrode, nízka úroveň vzdelania a vysoký počet detí. Na severnom Slovensku má vplyv aj vysoký podiel obyvateľov s rímskokatolíckym vierovyznaním.

Literatúra

BEETS, G. (1999). Education an Age at First Birth. DEMOS, Vol. 15, Special Issue. s.5-8

KANTOROVÁ, V. (2002). Education and entry into motherhood: the Czech republic during state-socialism and the transition period (1970-1997). Paper prepared for Workshop on „Education and postponement of Maternity“. Amsterdam, Netherlands

MARENČÁKOVÁ, J., 2003b, Pôrodnosť obyvateľstva Slovenska a jej vzťah s vybranými demografickými a spoločenskými javmi. *Acta Facultatis Rerum Naturalium Universitatis Comenianae*. *Geographica*, 44, 3-89.

MARENČÁKOVÁ, J., 2004b, Zmeny v úrovni plodnosti na Slovensku po roku 1989. [The Changes of the Fertility in Slovakia after 1989.] *Geografické informácie* 8, 302-309

PIKÁLKOVÁ, S. (2003). Reprodukční chování a postojové orientace žen v závislosti na vzdelání. Demografie. 45 s.

Dominika Pavlovčinová, dominika.pavlovcinova@gmail.com

Demografické problémy Číny

Petra Pohanková

Demographic problems of China

Main purpose of this thesis is to provide a basic overview of population policy in China and its impacts to the society. It describes consequences of this policy and very important China's demographic problems.

The document is divided into three parts. First one consists of China's population development and population measures after 1950. The second one is focused to explain one-child policy and its targets. The last part describes demographic problem of China, such as: sex-selection abortion, infanticide of girls, insufficient work forces and population aging.

Demografické problémy Číny

V minulosti patrila k mocenským centrám sveta a v súčasnosti sa zaraďuje medzi rozvojové krajiny tohto sveta. Má niekoľko tisícročnú históriu a predsa sa za tak dlhý čas neprebojovala k vlastnej demokracii. V čase, keď sa celý svet otvára, ona zostáva uzavretá pred svetom. Svojimi objavmi a vynálezmi vždy fascinovala svet a v posledných desaťročiach sa napriek tomu potýka s pomerne vysokou negramotnosťou. Je to, ale krajina, ktorá je každopádne osobitá svojou, ľudnatosťou, rozmanitosťou, priepastnou rozdielnosťou medzi jej obyvateľmi, režimom, religiozitou aj spôsobom života. Je to Čína – krajina kontrastov a „obmedzených možností".

Vývoj obyvateľov Číny a populačná politika:

Čína je krajina, o ktorej dodnes nevieme s určitosťou povedať, koľko mala obyvateľov nielen v stredoveku, ale ani v novoveku. Odhady demografov sa výrazne líšia a často nevedia zdôvodniť značné výkyvy v predpokladanom vývoji čínskeho obyvateľstva.

Aj keď Čína patrí medzi najstaršie civilizácie, najľudnatejšia republika sveta bola formálne založená až 1. októbra 1949 pod názvom Čínska Ľudová Republika. Jej populácia vtedy dosahovala „len" asi pol miliardy obyvateľov. Od tohto roku dochádzalo k rapídnemu rastu populácie, keď v tých rokoch dosahovala hrubá miera živorodenosti až 44 ‰.

V roku 1956 sa začala séria kampaní, ktoré mali viesť k jedinému cieľu a to zníženie plodnosti. V tom roku neboli prijaté žiadne konkrétne nástroje, išlo len o snahy znížiť negramotnosť obyvateľstva a zavádzali sa prvé programy sexuálnej výchovy a propagácia antikoncepcie. Taktiež došlo k liberalizácii potratového zákona.

Po roku 1963 v Číne prebiehala druhá kampaň na zníženie úrovne plodnosti. Cieľovou skupinou boli najmä obyvatelia miest, kde sa jednej žene mali narodiť maximálne tri deti. Taktiež sa zvyšoval vek pri sobáši a začala sa prejavovať liberalizácia potratovej legislatívy. Oficiálnym cieľom populačnej politiky malo byť zníženie ročného prirodzeného prírastku na hodnotu jedného percenta pred koncom dvadsiateho storočia. (Vtedy sa táto hodnota pohybovala okolo 2,6 %).

Sedemdesiate roky predstavovali tretiu kampaň na zníženie plodnosti, ktorá sa niesla v duchu „later, longer, lower"- „neskôr, dlhšie a menej". Štátna politika sa snažila pôsobiť na zvýšenie priemerného sobášneho veku pri uzavretí prvého manželstva. Napriek všetkým týmto snahám a populačným opatreniam sa v roku 1985 teda za 35 rokov počet obyvateľov Číny približne zdvojnásobil a prekročil magickú hranicu jednej miliardy. V roku 2000 už bolo na svete 1,2 miliardy Číňanov. Dnes má Čína približne 1,322 miliardy obyvateľov. Je najľudnatejšou krajinou sveta, žije tu pätina svetovej populácie.

One-child policy:

Na konci sedemdesiatych rokov si čínska vláda uvedomila, že rapídne rozrastajúca sa populácia môže v budúcnosti znamenať pre krajinu veľký problém. Od roku 1979 je v Číne praktikovaná politika jedného dieťaťa, ktorá mala zabezpečiť zadržanie populačnej explózie známa aj pod názvom one-child policy. Ide o reguláciu a kontrolu pôrodnosti obyvateľstva zo strany štátu pričom sa riadi heslom: Jedno dieťa je akurát, dve sú akceptovateľné, tri sú príliš. Každá rodina žijúca v meste smie mať najviac jedno dieťa. Na vidieku, kde je situácia voľnejšia a ťažšie kontrolovateľná, je povolené mať deti dve, a to tiež len v prípade, ak sa prvé narodí dievča avšak s minimálnym rozdielom štyri až päť rokov. Za každé ďalšie "nepovolené dieťa" platí rodina štátu pokutu. Ak rodina na platenie pokuty finančne prostriedky jednoducho nemá, je jej istá suma strhávaná pravidelne z platu. Celkový cieľ jednodetného systému bolo udržať celkovú populáciu na hranici 1,2 mld. do roku 2000. Tento ambiciózny program spájal demografické a ekonomické faktory. Systém odmeňoval rodiny, ktorý tento program dodržiavali a pokutoval, tých ktorý ho porušovali. Pokutami za porušenie limitu jedného dieťaťa bolo zníženie platu, znížené šance pri vybavovaní bývania, drahšia zdravotná starostlivosť. Na podporu jednodetných rodín bol zavedený celý balík výhod a odmien. Výhodami boli peňažné prémie, školské vzdelanie pre deti zdarma, lepšia starostlivosť o dieťa, väčšia nádej pri riešení bytovej otázky, dlhšia materská dovolenka, rôzne príspevky štátu. (Na deti narodené mimo toho zákonu štát neprispieval nič). Kľúčom k úspechu alebo zlyhaniu tohto systému boli považované vidiecke obce, keďže na vidieku je koncentrovaných až 60 % všetkých obyvateľov. Táto politika sa však uplatňovala najmä na obyvateľoch miest, kde je jednoduchšie kontrolovateľná a kde sa striktne dodržiavala a dodržiava.

Politika jedného dieťaťa mala dva hlavné ciele. Prvým bolo zníženie pôrodnosti ako takej a tým aj celkový rast populácie. Druhou tendenciou bolo zvyšovať sobášny vek mužov aj žien a tiež vek pri pôrode.

Regulácia pôrodnosti sa uskutočňovala zvyšovaním povedomia obyvateľov ohľadne antikoncepcie a ochrany pred tehotenstvom. Ničím výnimočným neboli ani sterilizácia, či vazektómia po splodení prvého potomka. Z výskumov vyplýva, že až 80 % žien a asi 13% mužov využíva určitý typ antikoncepcie.

V Číne existujú tzv. špeciálne pracovníčky pre ženské veci, ktoré kontrolujú dodržiavanie zásad populačnej politiky, užívanie antikoncepcie, počet detí v rodinách. Neoprávnene tehotné ženy musia často podstúpiť potrat a to aj vo vysokých štádiách tehotenstva, mnohé boli po pôrode prvého potomka nútené násilne podstúpiť sterilizáciu.

Predpokladá sa, že implementáciou jednodetnej populačnej politiky sa podarilo zabrániť až 300 mil. pôrodom.

Podľa štatistík v roku 1950 dosahovala miera úhrnnej plodnosti (počet detí na jednu ženu počas jej reprodukčného obdobia) asi 6,22 dieťaťa Táto hodnota sa výrazne znižovala, až roku 1980 dosiahla hodnotu len 2,55. V roku 2005 ešte poklesla na úroveň 1,73 dieťaťa na matku. Taktiež sa posunula veková hranica prvého pôrodu matky z vekovej kategórie 20-24 rokov do kategórie 25-29 rokov.

Aj keď sa vplyvom zavedenia jednodetnej politika dosiahol želaný cieľ zníženia prirodzeného prírastku, klesol počet narodených detí a vek matky pri pôrode sa zjavne zvýšil program si vyslúžil obrovskú kritiku a je považovaný za kombináciu propagandy, sociálneho nátlaku a porušovania ľudských práv. Tieto praktiky boli vždy odsudzované humanitárnymi organizáciami z celého sveta. Podľa nich je obmedzenie počtu detí diskriminačné a vedie k porušovaniu elementárnych práv obyvateľov a k rastu interrupcií.

Demografické problémy:

Extrémna populačná politika Číny a implementácia jednodetnej politiky nespôsobila len zníženie rastu populácie ale zodpovedá aj za vážne demografické problémy, s ktorými sa dnes Čína potýka a ktoré budú mať na vplyv na jej ďalší vývoj hlavne v budúcich rokoch.

Ženy v ohrození

Obrovským problémom čínskej spoločnosti je diskriminácia žien a uprednostňovanie mužov. Preferencia chlapcov je historická, odvodená z konfuciánskeho učenia. Už v starej Číne bol syn väčšmi cenený ako dcéra, pretože ostávajú v rodine a starajú sa o svojich rodičov, keď zostarnú a zachovávajú rod. Dodržovanie obetných rituálov podľa predpisov bolo tiež pokladané za najdôležitejšiu povinnosť syna. Nemať syna, ktorý by pokračoval v rodovej línii bolo teda najväčším nešťastím. Dcéra po vydaji už patrila do inej rodiny, a naviac ako žena nebola „spôsobilá“ vykonávať predpísané obeti predkom. Nemala právo na dedičstvo ani na rodové meno.

Rodičia sa v spoločnosti, ktorá už tradične preferuje mužov, často rozhodnú pre potrat, ak zistia, že ich dieťa bude dievča. V Číne sú selektívne potraty síce zakázané, ženy však majú zadarmo nárok na vyšetrenie a môžu si pohlavie plodu zistiť. Mnohé sa rozhodnú pre potrat. Krutou skutočnosťou je však aj infanticída (zabitie dieťaťa po narodení). V súčasnosti sa preto na každých 100 narodených dievčat rodí približne 119 chlapcov. V celosvetovom meradle však na 100 novonarodených dievčat pripadá 103-107 chlapcov.

Prísna jednodetná politika v Číne teda spôsobila vážne narušenie pohlavnej štruktúry a tým aj rovnováhu medzi pohlaviami. Predpokladá, že čo sa týka osôb v „sobášnom veku“, počet mužov v roku 2020 prekročí počet žien o 30 miliónov. Táto skutočnosť môže v blízkej budúcnosti viesť k rozsiahlym nepokojom v spoločnosti. Zdeformovaný pomer medzi pohlaviami mal už za následok zvýšenie počtu únosov a obchodov so ženami. Sociológovia varujú, že priveľa neženatých mužov môže zvýšiť riziko nesociálneho a násilníckeho správania.

Milióny dôchodcov

Za najvážnejší a zároveň najrizikovejší dôsledok súčasného populačného vývoja Číny možno označiť zmeny vo vekovej štruktúre obyvateľstva, to znamená jeho starnutie. Starnutie populácie zaťažuje sociálny a zdravotnícky systém, zasahuje vzťahy medzi generáciami a narúša sociálnu harmóniu. Preto si tento problém bude vyžadovať zásadné zmeny v sociálnom, zdravotnom a dôchodkovom zabezpečení obyvateľstva. Závažným javom sú tiež vzťahy medzi generáciami. Pre porovnanie, zatiaľ čo v roku 1965 mala Čína vo vekovej kategórii do 15 rokov asi 40 % obyvateľstva, v dnešných dňoch je to už len polovica a podľa prognóz by sa mal tento podiel mal ešte znížiť a to na približne 15% do roku 2050. Ľudí vo vekovej kategórii 60 a viac však neúprosne pribúda. V roku 1970 tvorila táto skupina len okolo 6,8 %. Predlžovaním dĺžky života, zlepšením zdravotnej starostlivosti ale hlavne zavedením jednodetnej politiky sa tento podiel zvýšil v roku 2005 až na 11%. Rapídny nárast však Čína pocíti až v najbližších rokoch, kde sa predpokladá, že do tejto skupiny bude v roku 2020 spadať už 17 % a v roku 2050 enormných 31% percent populácie, čo činí asi 438 miliónov dôchodcov.

Kto bude pracovať?

So starnutím obyvateľstva súvisí aj nedostatok pracovných síl v krajine. Čína sa tak pomaly posúva z éry prebytku pracovných síl do éry ich nedostatku. Vysoký počet lacnej pracovnej sily v krajine s 1,3 miliardy obyvateľov sa však považuje za hlavný motor hospodárskeho rozmachu tejto krajiny. Čína preto musí svoje hospodárstvo preorientovať z produkcie náročnej na prácu na výrobu výrobkov s pridanou hodnotou, aby mohla zostať konkurencieschopná. Relatívnou novinkou je zavedenie nariadenia, ktoré povoľuje mať dve deti rodičom, ktorí sami pochádzajú z jednodetných rodín. Čínske úrady veria, že sa tým pomôžu vyriešiť problémy s nedostatkom pracovných síl a zároveň by to prispelo k omladeniu populácie. V najbližších rokoch pokles počtu pracovných síl nepochybne povedie k zvýšeniu dôchodkového veku ako pre mužov tak aj pre ženy. Je dosť pravdepodobné, že to so sebou prinesie aj jednu veľkú zmenu a možno aj úplný koniec politiky jedného dieťaťa.

Záver:

Čínska spoločnosť prechádza od polovice minulého storočia významnými ekonomickými, spoločenskými aj demografickými zmenami. Nemáme na mysli len politiku jedného dieťaťa, ktorá bola implementovaná na konci sedemdesiatych rokov, ale rozsiahly komplex ekonomických a sociálnych opatrení, ktoré mali za účel znížiť vysoký prirodzený prírastok a zvýšiť životnú úroveň jej obyvateľov.

Medzi najväčšie a najzávažnejšie problémy čínskeho obyvateľstva rozhodne patria selektívne interrupcie, infanticída dievčat, nepomer medzi pohlaviami, rozpad tradičného modelu čínskej rodiny a problematika sociálneho a ekonomického zabezpečenia seniorov v dôchodkovom veku. V horizonte dvadsať až štyridsať rokov sa čínska spoločnosť začne potýkať s dôsledkami starnutia populácie a je len otázkou akým spôsobom bude na ne reagovať.

Prognózy vývoja čínskej populácie jasne hovoria o tom, že rast obyvateľstva sa ustáli na určitej hranici. Z terajších 1,322 miliardy bude rásť len pomaly a zvýši sa približne na 1,46 miliardy do roku 2032. Od tohto roku začne postupne klesať a do roku 2050 sa dostane na hodnotu asi 1,4 miliardy obyvateľov.

Približne v roku 2025 by podľa svetových prognóz mala Čínu na pozícii najľudnatejšieho štátu vystriedať ďalšia rozvojová krajina India.

Veľkým problémom pre čínsku spoločnosť môže byť demografický vývoj u národnostných menšín, pri ktorých neboli uplatňované výrazné antipopulačné nástroje a ktoré dnes tvoria asi 8-10 % populácie Číny.

Najľudnatejšia krajina sveta je pre nás príliš veľkou hádankou na to, aby sa o nej dalo niečo tvrdiť s úplnou určitosťou. Iste bude pozoruhodné sledovať smerovanie a napredovanie Číny v budúcich rokoch, ktorá aj napriek svetovej kritike zotrváva pri svojom tvrdom politickom režime. Bude zaujímavé pozorovať jej určite nie ľahkú cestu za demokraciou, jej snahy o zvýšenie gramotnosti, povedomia, ekonomickej a životnej úrovne obyvateľov. Zaujímavý bude tiež spôsob akým a či vôbec sa jej podarí vyrovnať sa s obrovskými problémami, ktoré momentálne má alebo naozaj zostarne skôr ako zbohatne.

Použité zdroje:

Populační vývoj: Čína ve druhé polovině 20. století
http://www.demografie.info/?cz_detail_clanku=&artclID=443

Populační vývoj: Čína - reprodukce v kontextu regionů a etnik
http://www.demografie.info/?cz_detail_clanku&artclID=449

Uprednostňovanie synov zapríčiňuje vo svete celkový úbytok žien
http://www.velkaepocha.sk/content/view/2220/41/

World Population Prospect: The 2006 Revison

http://www.iiasa.ac.at

http://www.cpirc.org.cn/en/eindex.htm

https://www.cia.gov/cia/publications/factbook/

http://www.medizin-ethik.ch/publik/family_planning.htm

Petra Pohanková
Prírodovedecká fakulta UK, Bratislava, 5. ročník
e-mail: **petra.pohankova@gmail.com**

Analýza infraštruktúrnej vybavenosti rómskych osád v Banskobystrickom kraji

Jana Pukačová

ABSTRAKT

This work in her first part is focused on analyze of species gipsies village. Key area of interest this research , whom primary target is mapping of infrastructural accessories in gipsies villages in Banská Bystrica area. This analyse is searching relationship of gipsies inhabitants at electric energy, gas, drainage and a water – supply. On the other side is trying to compare possibility to have entrance at this network in gipsies villages and in villages of majority citizens.

KĽÚČOVÉ SLOVÁ: rómska osada, rómske osídlenie, infraštruktúra

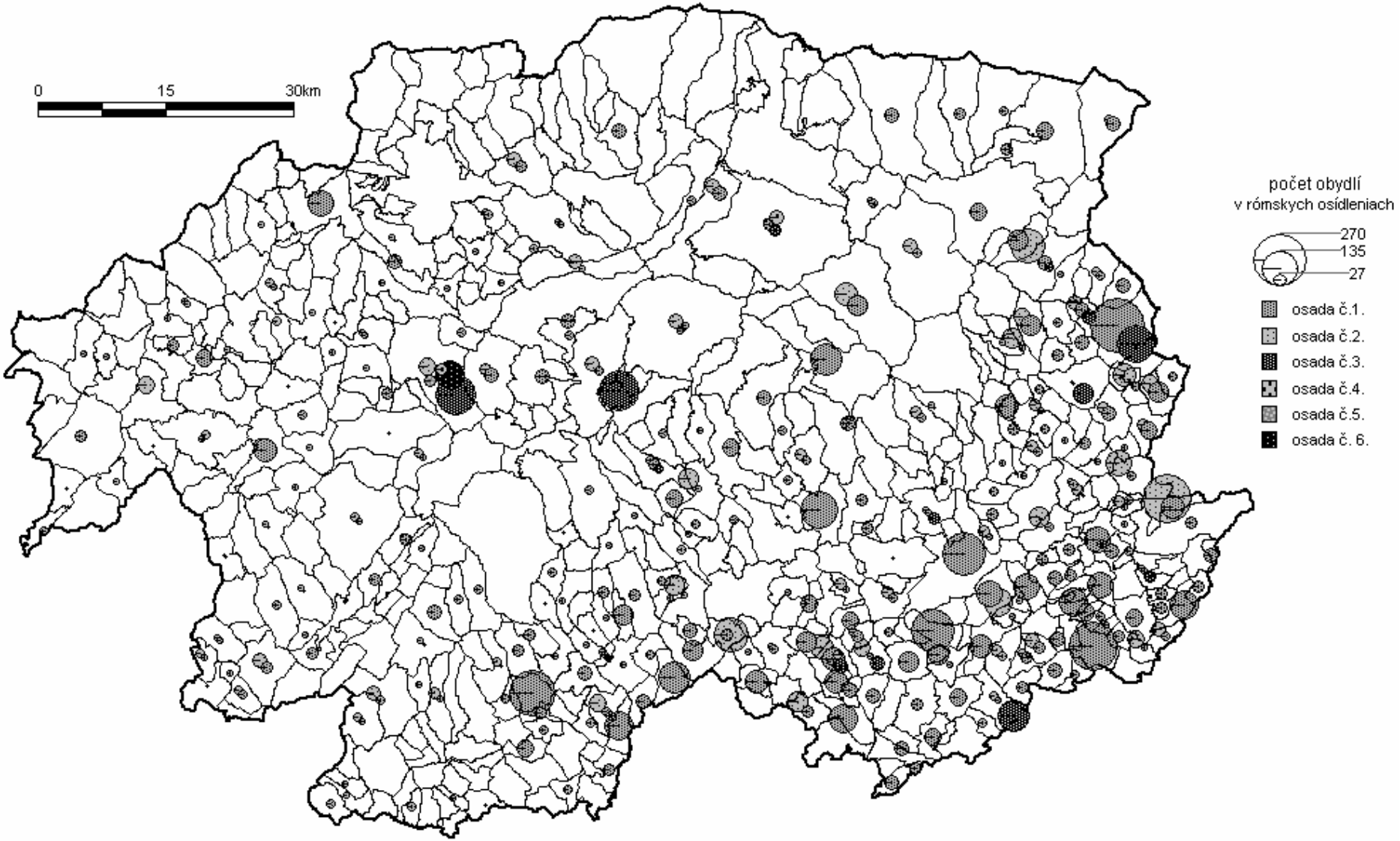
ÚVOD

Hodnotiť stav kvality života rómskych komunít, resp. spôsob ich bývania je značne problematické. Najvýznamnejším faktorom, ktorý na Slovensku na to vplýva je, že v rámci výskumu v tejto oblasti je veľký nedostatok relevantných údajov, na základe ktorých by bolo možné sformulovať adekvátne závery. Primárnym problémom je samotné definovanie pojmu rómska osada. V období deväťdesiatych rokov boli za rómske osady označované osídlenia, v ktorých žila rómska populácia, bez ohľadu na typ osídlenia, kvalitu života v ňom a mieru jeho integrácie. V prvej polovici deväťdesiatych rokoch bol problém rozrastajúcich sa rómskych osád vnímaný len veľmi okrajovo, a tak najmä kvôli chýbajúcim informáciám bol v roku 1997 vypracovaný *Ročný výkaz o obytných zoskupeniach na nízkom sociokultúrnom stupni (rómske osady).* V ňom sa prioritný dôraz kládol na stav osídlení bez presnejšieho vymedzenia miery integrácie, pozorovali sa teda osady ako súčasti obce, osady na okraji obce a osady vzdialené od obce. V roku 1999 sa prestali výkazy robiť len pre osídlenia na nízkom sociokultúrnom stupni, ale do zisťovania boli zaradené aj ľubovoľné typy osídlenia, v ktorých žili Rómovia. V týchto prípadoch sa jednalo predovšetkým o osídlenia, ktoré boli postavené počas socialistického riešenia likvidácie rómskych osád. Do zoznamov sa tak dostávali plne integrované a infraštruktúrne vybavené osídlenia na vlastných pozemkoch bez akýchkoľvek chatrčí, patrili sem tiež sídliska a rôzne bytovky či rodinné domy. V roku 2000 bolo evidovaných 620 rómskych osád rôzneho typu.[1] Týchto 620 rómskych osád sa nachádzalo v 536 obciach Slovenska. V Banskobystrickom kraji sa v tomto období v 86 obciach koncentrovalo 110 rómskych osád. Podľa Amnesty International možno za rómsku osadu považovať vidiecke osady a oddelené časti menších miest a dedín, ktoré obývajú rómske komunity.[2] V roku 2004 sa pre potreby poznania celkového rozloženia rómskeho etnika vzájomne s identifikáciou základných problémov tejto komunity, predovšetkým vo vzťahu k bývaniu, realizovalo sociografické mapovanie. Mapovenie sa uskutočnilo na celom území Slovenska s výnimkou obcí, v ktorých bol počet obyvateľov rómskeho osídlenia nižší ako 15. V prípade, že v obci bola situácia rómskej komunity problematická (z hľadiska životnej úrovne, bývania, vzťahu s majoritou) bol výskum realizovaný aj v týchto obciach. Rómska komunita bola definovaná ako zoskupenie, ktoré majoritné obyvateľstvo subjektívne vníma ako odlišné (v pozitívnom alebo v negatívnom zmysle) na základe antropologických znakov, spôsobu života, životného a kultúrneho štýlu. Pre potreby výskumu bolo osídlenie definované ako zoskupenie minimálne troch domov obývaných komunitou, ktorú majorita označila za rómsku.[3]

VÝSLEDKY SOCIOGRAFICKÉHO MAPOVANIA

V Banskobystrickom kraji sa nachádza v 274 obciach 392 osídlení, v ktorých žije 52 548 obyvateľov. Najviac osídlení je koncentrovaných na juhovýchode kraja, pričom na území jednej obce sa môže nachádzať viac osídlení. (mapa č.1.)

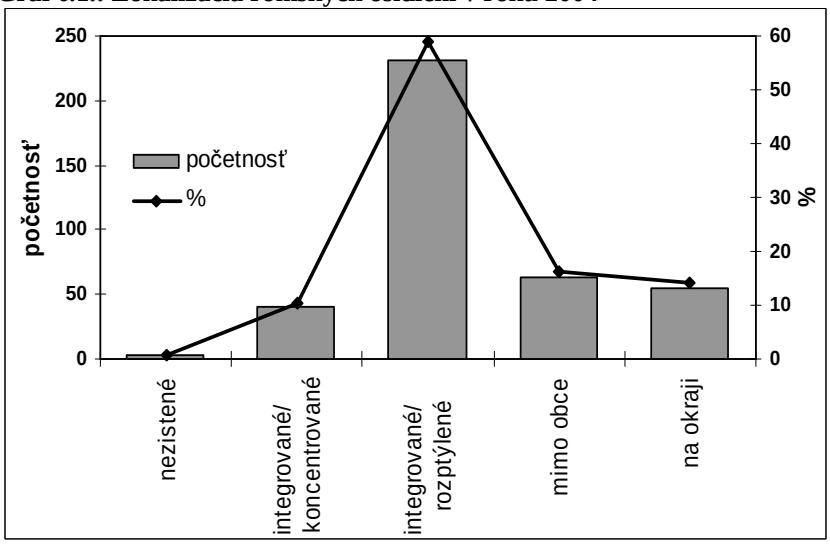
Mapa č.1.: Počet obydlí v rómskych osídleniach v BB kraji v roku 2004



Zdroj: Atlas rómskych komunít, 2004

V prípade kraja je to maximálne šesť osídlení v obci (Zvolen), päť osídlení je v dvoch obciach (Fiľakovo a Hriňová), štyri osídlenia sú v troch obciach (Hodruša-Hámre, Revúca, Čierny Balog), tri osídlenia sa nachádzajú v 14 obciach, v 69 obciach sú lokalizované dve rómske osídlenia a v 185 obciach jedno rómske osídlenie.

Graf č.1.: Lokalizácia rómskych osídlení v roku 2004

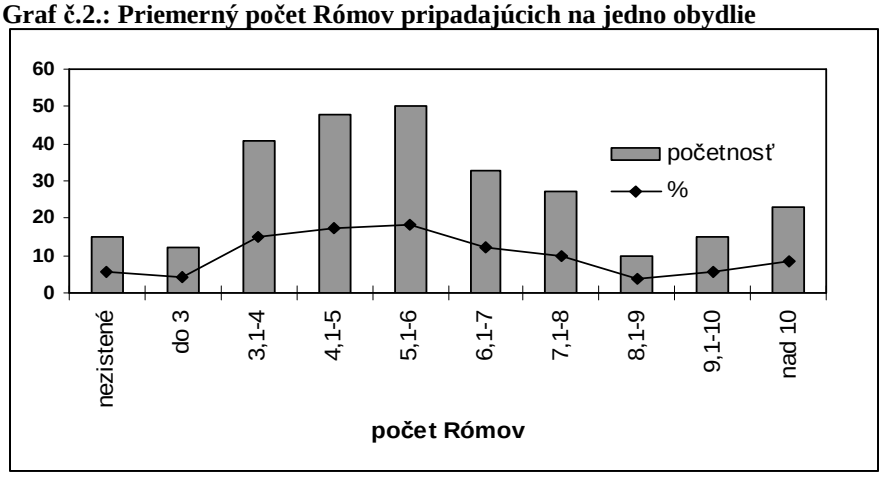


Z hľadiska polohy osídlenia je 231 osídlení koncentrovaných v obci, rozptýlených medzi majoritou je 40 osídlení, na okraji obce (na hranici zastavanej časti obce) je situovaných 55 osídlení a mimo obce je to 63 osídlení. (graf č.1.)

Zdroj: Atlas rómskych komunít, 2004

Pri sociografickom mapovaní sa zisťoval aj počet obydlí v jednotlivých rómskych osídleniach. Obydlie bolo pre potreby výskumu definovaný ako akýkoľvek objekt, ktorý je obývaný. V Banskobystrickom kraji sa vyskytuje 8 871 rómskych obydlí rôzneho typu. Väčšina rómskeho obyvateľstva žije v bytových domoch a legálne postavených rodinných domoch. Taktiež bolo zaregistrovaných 8 legálne postavených drevených domov, 297 nelegálne postavených rodinných domov, 262 chatrčí, 31 unimobuniek, 34 nebytových budov a 20 iných objektov. Medzi iné objekty sa zaraďujú napríklad stany, autobusy, fóliovníky a podobne.
Priemerný počet obyvateľov pripadajúcich na jedno obydlie je značne rozdielny. Najpočetnejšie je to 6 obyvateľov (18,2%) a 5 obyvateľov (17,5%). Vyskytujú sa tu však aj obydlia, kde je priemerný počet obyvateľov vyšší ako 10, je to 23 obydlí (8,4%).(graf č.2.)

Graf č.2.: Priemerný počet Rómov pripadajúcich na jedno obydlie



Najviac obyvateľov v jednom obydlí žije v obci Horná Strehová (18 osôb), v obci Telgárt (17,1 osôb), v obci Hontianske Moravce (17 osôb) a v obci Hontianske Nemce (16,5 osôb).

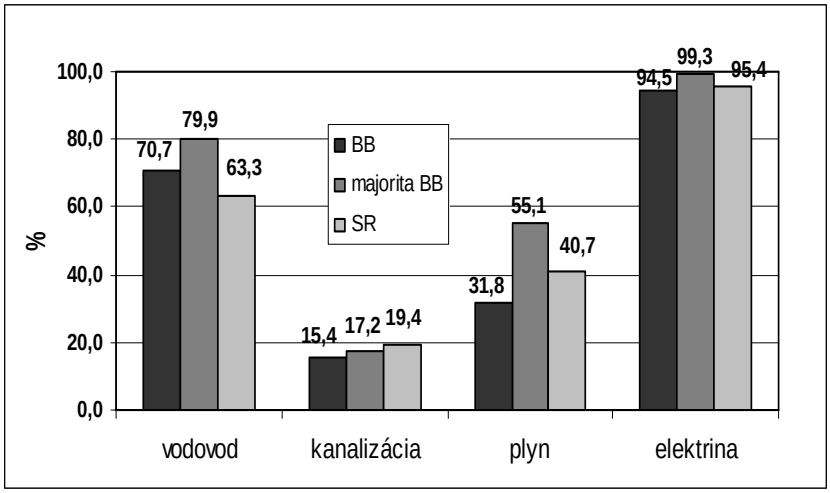
Zdroj: Atlas rómskych komunít, 2004

Výskum sa zameral aj na zisťovanie úrovne rómskych osídlení so zreteľom na možnosti využívania základnej infraštruktúry. Jednotlivé osídlenia sú veľmi rozdielne vybavené infraštruktúrou, pričom jedným z hlavných faktorov je lokalizácia tohto osídlenia. Najlepšie možnosti využívať infraštruktúru majú osídlenia, ktoré sú lokalizované priamo v obci. Čím je však osídlenie ďalej od obce, predovšetkým sa jedná o osídlenia lokalizované mimo obce, často oddelené aj nejakou bariérou (napr. pole, les, potok,

kopec), štandard života sa výrazne znižuje. Pri vybavenosti infraštruktúrou sa brali do úvahy dve charakteristiky.

Prvou bola dostupnosť inžinierskych sietí v rómskych osídleniach. Táto podáva informáciu o tom, či obyvatelia týchto osídlení majú možnosť napojiť sa na jednotlivé siete (vodovod, elektrina, plyn, kanalizácia), a následne ich aj využívať. Okrem dostupnosti inžinierskych sietí pre Rómov sa zisťovala

Graf č.3: Dostupnosť infraštruktúry v rómskych osídleniach v roku 2004



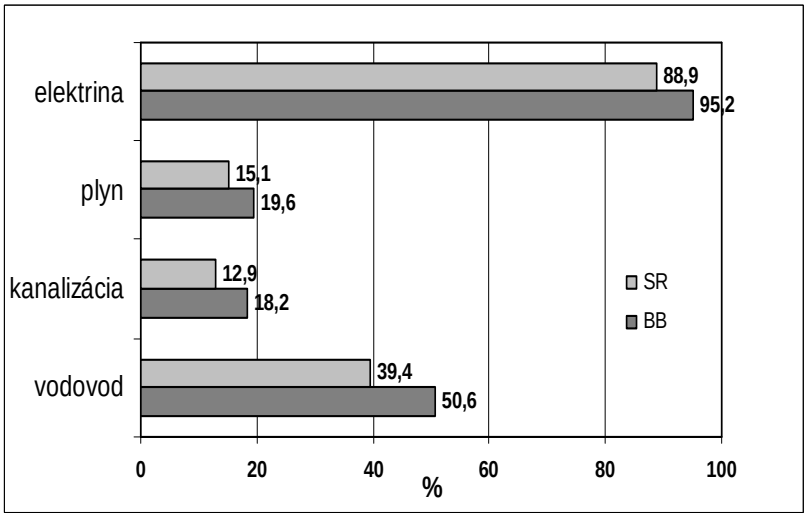
táto dostupnosť aj pre majoritné obyvateľstvo v danej obci, teda či má aj majorita možnosť sa napojiť na tieto siete. Z výsledkov výskumu vyplýva, že Rómovia v Banskobystrickom kraji majú dostupnosť na úrovni porovnateľnej s majoritou. (graf č.3.)

Zdroj: Atlas rómskych komunít, 2004

Výnimku tvorí len dostupnosť sietí plynu, kde je rozdiel medzi majoritou a Rómami najvýraznejší (55% dostupnosť zo strany majority a len 32% zo strany rómskych osídlení). Aj napriek tomu, že plyn je v mnohých obciach zavedený, Rómovia si ho nemôžu dovoliť, nakoľko prípojky sú drahé a tak kúria radšej drevom. Žiadna z inžinierskych sieti nie je dostupná pre 8 rómskych osídlení. Sú to osídlenia v obci Cerovo, Horné Mladonice, Valice, Kameňany, Kyjatice, Sása, Stará Kremnička, ktoré sú situované mimo obce vo vzdialenosti od 70 metrov (Horné Mladonice), 200metrov (Valice), 500 metrov (Cerovo, Kameňany, Stará kremnička – tu je aj bariéra potok), 890 metrov (Kyjatice) až po 8 kilometrov (Sása, kde je bariérou les). V meste Fiľakovo je rómske osídlenie situované na okraji mesta. Spolu v týchto osídleniach bez inžinierskych sieti žije v 39 obydliach 243 osôb. Na druhej strane má všetky siete dostupné 20 rómskych osídlení, v týchto žije v 964 obydliach 5 353 obyvateľov.

Pri druhej charakteristike vybavení rómskych osídlení infraštruktúrou sa do úvahy bral počet obydlí napojených na tieto siete. Využívanie infraštruktúry zo strany Rómov je síce v Banskobystrickom kraji vyšší ako na území celého Slovenska, ale až tak výrazne sa hodnoty nelíšia.(graf č.4.)

Graf č.4.: Podiel rómskych obydlí napojených na inžinierske siete v roku 2004



Používanie inžinierskych sieti však nebolo možné zistiť pri všetkých obydliach kraja. Elektrinu má zo 7 333 obydlí 6 980 obydlí (95,2%), vodovod využíva zo 7 045 obydlí len 3564 obydlí (50,6%), na kanalizáciu je napojených zo 7 267 obydlí iba 18,5% obydlí a na plyn je zo 6 942 obydlí napojených len 1 360 (19,6%). Z uvedeného vyplýva, že Rómovia najviac využívajú

Zdroj: Atlas rómskych komunít, 2004

elektrinu, tí ktorí ju majú dostupnú, ju aj využívajú. Vodovod nevyužíva 20% obydlí, ktoré ho majú dostupný. Kanalizáciu nevyužíva 4% obydlí a na plyn nie je napojených cca 12% obydlí pre ktoré je dostupný.

ZÁVER

Výsledky sociografického mapovania na Slovensku potvrdili, že rómske osídlenia sú veľmi rozdielne z hľadiska množstva ukazovateľov. Rómske obydlia v Banskobystrickom kraji dosahujú porovnateľné hodnoty so Slovenskom, pričom pri využívaní infraštruktúry dokonca tieto priemerné hodnoty prevyšujú. Aj napriek tomu, že mimo obce je lokalizovaných až 63 rómskych osídlení, infraštruktúra úplne absentuje len v ôsmych osídleniach. V podmienkach bývania je značne viditeľný rozdiel medzi rómskymi domácnosťami, ktoré žijú integrovane/koncentrovane s majoritou a tými, ktoré žijú segregovane.[4] Práve vzdialenosť od obce výrazne znižuje možnosť využívať inžinierske siete. V súčastnosti sa pri riešení bytovej otázky v rómskych komunitách sústreďuje pozornosť práve na podporu zlepšenia zlej infraštruktúry obydlí, ale aj na podporu stavieb bytov nižšieho štandardu pre Rómov. Tieto výsledky mapovania poukazujú na osídlenia, v ktorých je infraštruktúra a tým aj životná úroveň najhoršia, a na ne by sa mala zamerať podpora štátu.

POUŽITÁ LITERATÚRA:

[1] Jurová, A.: *Rómovia v procesoch transformácie, problémy, ohrozenia*. In: Šutaj, Š. (ed.): Národ a národnosti na Slovensku v transformujúcej sa spoločnosti – vzťahy a konflikty. Prešov, Universum 2005, pp. 257-271.

[2] *Stále segregovaní, stále v nerovnom postavení: Porušovanie práva na vzdelanie rómskych detí na Slovensku.* [Citované 16. november 2007]. Dostupné na internete:

< http://www.amnesty.sk/article_files/AI_SK_Sprava_vzdelavanie_na_Slovensku.pdf >

[3] Radičová, I. (ed.): *Atlas rómskych komunít na Slovensku*. Bratislava, Inštitút pre verejné otázky 2004

[4] Radičová, I.: *Hic Sunt Romales*. Bratislava, Interlingua 2001

[5] Kadlečíková, J., Kriglerová, E.: *Rómovia*. In: Kollár, M., Mesežnikov, G., Bútora, M. (ed.): Slovensko 2006. Súhrnná správa o stave spoločnosti. Bratislava, Inštitút pre verejné otázky 2006, pp. 213-230.

Adresa:
Bc. Jana Pukačová, Prírodovedecká fakulta UK BA, 2.roč. humánna geografia a demografia
jankap83@seznam.cz

Empirický model miery nezamestnanosti v Slovenskej Republike

Juraj Varga

Abstract: The aim of this paper is the identification of the factors influencing rate of unemployment in The Slovak Republic. We will estimate statistically and economically significant model of unemployment using Polynomial Distributed-Lag model. We suppose the unemployment to be significantly determined by gross domestic product and labour productivity, which stands for economical growth. We further suggest the unemployment to be influences by the direct foreign investments and the index of economic sentiment published by Slovak Statistical Office. We provided several tests and modifications to ensure the correct specification of the model. The final results suggest an interesting contribution to unemployment studies.

Keywords: *unemployment, time lag, Polynomial Distributed-Lag, ARMA models, foreign direct investments, index of economic sentiment*

1. Úvod

Skoro v každej ekonomike sa nachádza skupina ľudí, ktorú môžeme klasifikovať ako nezamestnaných. Ide o skupinu ľudí, ktorí sú nezamestnaní ale proklamujú, že pracovať chcú na pozíciách podobných ako sú pracovné pozície obsadené inými ľuďmi a za rovnakú mzdu, ktorú poberajú za danú prácu títo ľudia. Nezamestnanosť, ako makroekonomická veličina, môže byť ovplyvnená jednak zo strany ponuky, ako aj dopytu. Podľa ekonomickej teórie základnými činiteľmi ovplyvňujúcimi mieru nezamestnanosti sú *reálna mzda* a *vládne výdavky.* V situácii, keď zamestnanosť (resp. nezamestnanosť) určuje strana ponuky, je agregátny dopyt obmedzený nedostatočnou ponukou. Firmy považujú mzdy za príliš vysoké a ceny produkcie za príliš nízke, nasleduje korekcia miezd. Následný pokles miezd vedie k zvýšeniu ziskovosti podnikov. Podniky sú následne stimulované k zvýšeniu objemu produkcie, a tým aj objemu zamestnanosti, uplatňuje sa hospodárska politika klasického typu. V situácii, keď zamestnanosť určuje strana dopytu, je ponuka obmedzená nedostatočným dopytom, ide o keynesiánsky prístup k fungovaniu ekonomiky. Zníženie reálnej mzdy nevedie k návratu do stavu rovnováhy, vyvoláva rast obmedzení, na ktoré naráža dopyt. Nástrojom na obmedzenie tohto javu sú práve vládne výdavky. Tento príspevok sa však nebude venovať procesu dosahovania stavu rovnováhy, ale identifikuje determinanty miery nezamestnanosti v podmienkach Slovenskej republiky. Motívom k napísaniu tohto príspevku je súčasný a minulý stav nezamestnanosti v SR, ktorý má chronický charakter. Vysoká miera nezamestnanosti je veľkým jednak makroekonomickým, ako aj socioekonomickým problémom, ktorý významne pôsobí na prehlbovanie regionálnych rozdielov na Slovensku. Pri modelovaní miery nezamestnanosti budú použité softvérové balíky SPSS 13.0 a EViews 4.1 Student Version.

2. Postup a použité metódy

Predpokladajme, že v slovenských podmienkach je nezamestnanosť ovplyvnená rastom ekonomiky. Model pracuje s kvartálnymi dátami hrubého domáceho produktu meraného v bežných cenách (*HDP*), produktivitou práce (*PP*), priamymi zahraničnými

investíciami (*PZI*) a indikátorom ekonomického sentimentu (*IES*), ktorý hodnotí psychologické faktory. Odhadovaný model predpokladá použitie lineárneho viacnásobného regresného modelu s oneskorenými hodnotami regresorov. Pri zohľadnení hĺbky časových oneskorení, a vzhľadom na perzistenciu miery nezamestnanosti, predpokladáme rozdelenie s vysokou hĺbkou pamäte, model PDL, vďaka čomu sme zredukovali počet odhadovaných parametrov.
Model PDL (*Polynomial Distributed-Lag, Almonová 1965*) predpokladá spojitý priebeh váh oneskorení v čase a nahrádza ich priebeh polynómom rádovo nižšieho stupňa. Predpokladáme teda nasledujúci formálny zápis vplyvu oneskorení na závislú premennú:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_n X_{t-n} + \varepsilon_t = \alpha + \sum_{i=0}^{n} \beta_i X_{t-i} + \varepsilon_t \qquad (1)$$

Pričom pre jednotlivé parametre platí, že ležia na polynóme, ktorý má nasledujúci tvar:

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \ldots + a_r i^r, \qquad t = 1, 2, \ldots n \qquad (2)$$

Apriórne obmedzenie váhových koeficientov je možné zapísať nasledujúcim spôsobom:

$$\boldsymbol{\beta} = \mathbf{Qa} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \\ \beta_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2^1 & \cdots & 2^r \\ 1 & 3^1 & \cdots & 3^r \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n & \cdots & n^r \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_r \end{pmatrix}. \qquad (3)$$

Model hodnotí štatistickú významnosť indikátorov, predpokladaný smer vplyvu a veľkosť oneskorenia indikátorov *PIZ* a *EIS.* Overíme závislosť miery nezamestnanosti na vývoji *PZI* a *IES*, reprezentujúceho nálady v danej ekonomike. Problém pri stanovení vhodného modelu sa javí v časových oneskoreniach vplyvu identifikovaných indikátorov. Vhodnosť modelu vylučuje autokoreláciu rezidií. Kvalita modelu bude overená testami multikolinearity a stability.

3. Dáta a empirický model nezamestnanosti

Použité dáta sú štvrťročné údaje za prvý kvartál roku 1997 až druhý kvartál roku 2007 (v Slovenskej republike), časové rady predstavujú rad štyridsiatich dvoch údajov.
Nezamestnanosť (*u*) vyjadrená ako percentuálna miera nezamestnanosti podľa metodiky VZPS[1].

Hrubý domáci produkt je vyjadrený v bežných cenách. *HDP* je časový rad pozostávajúci z kvartálnych dát o vývoji HDP[2] v Slovenskej republike. Predpokladáme nepriamy vplyv rastu *HDP* na mieru nezamestnanosti.

Produktivita práce je reprezentovaná produktivitou práce z tržieb za vlastné výkony a tovar podľa ekonomických činností (OKEČ) v Sk[3]. Vplyv regresného

[1] Zdroj: Štatistický úrad Slovenskej republiky
[2] merané metodikov ESNÚ 95, podľa štvrťročných národných účtov (v mil. Sk) . Zdroj: Štatistický úrad Slovenskej republiky
[3] Zdroj: Štatistický úrad Slovenskej republiky

koeficientu je nejednoznačný. Jeho smer závisí od kvalitatívneho výstupu ekonomiky a časového vplyvu indikátora.

Priame zahraničné investície predstavujú štvrťročný kumulatívny stav (saldo) k ultimu obdobia. Táto položka je vypočítaná ako súčet účtov Majetková účasť a reinvestovaný zisk a Ostatný kapitál V SR (podnik priamej investície = rezident) položiek Platobnej bilancie SR[4]. Predpokladáme záporný regresný koeficient a značné oneskorenie vplyvu (lagovanie) tohto regresora.

Indikátor ekonomického sentimentu[5] (východisková hodnota je hodnota roku 1995) je vážený aritmetický priemer z indikátora dôvery v priemysle (40%), indikátora dôvery v stavebníctve (5%), indikátora dôvery v maloobchode (5%), indikátora dôvery v službách (30%) a spotrebiteľského indikátora dôvery (20%). Predpoklad záporného znamienka regresného koeficientu a oneskorenia vplyvu (*lag*).

Výsledný model (graf č.1) je odhad medziročnej zmeny miery nezamestnanosti U_t z kvartálnych dát.

$$U_t = u_t - u_{t-4}, t=1,2,...n \quad (4)$$

$$U_t = \alpha_0 + \beta_0 pzi_t + \beta_1 pzi_{t-1} + \beta_2 pzi_{t-2} + \beta_3 pzi_{t-3} + ies_{t-2} + e_t \quad (5)$$

Pričom pre hodnoty parametrov PDL modelu platí rovnica (2) a teda

$$PZI_0 = -4,66887e^{-6}, PZI_1 = -6,1972e^{-6}, \\ PZI_2 = -7,7256e^{-6}, PZI_3 = -9,2539e^{-6} \quad (6)$$

Výsledný tvar má po doplnení hodnôt regresných koeficientov nasledujúci tvar:

$$U_t = 3.9887 - 4,66887e^{-6}PZI_t - 6,1972e^{-6}PZI_{t\text{-}1} - 7,725e^{-6}PZI_{t-2} - \\ -9,2539e^{-6}PZI_{t-3} - 0,02738IES_{t-2} + e_t; R^2 = 0,932525, DW = 2,198233 \quad (7)$$

Ako štatisticky významné sa ukázali indikátory *PZI* a *IES*. Pri odhade koeficientu *PZI* sa použila metóda PDL premennej vysvetľujúcej mieru nezamestnanosti, kde predpokladáme, že regresné koeficienty ležia na priamke. Ukazuje sa, že zmena *IES* ovplyvňuje mieru nezamestnanosti o dva štvrťroky (oneskorením dvoch štvrťrokov).

Modelová konštanta a regresný koeficient *IES* sú významné na 10 % hladiny významnosti, ostatné regresory sú významné na všetkých bežných hraniciach významnosti. Chyby prvotného modelu boli autokorelované, čo sme odstránili pridaním parametra autoregresného modelu prvého stupňa a parametra modelu kĺzavých priemerov prvého stupňa do modelu. Pre chybu modelu platí nasledujúci vzťah:

[4] Zdroj: Národná banka Slovenska (www.nbs.sk)

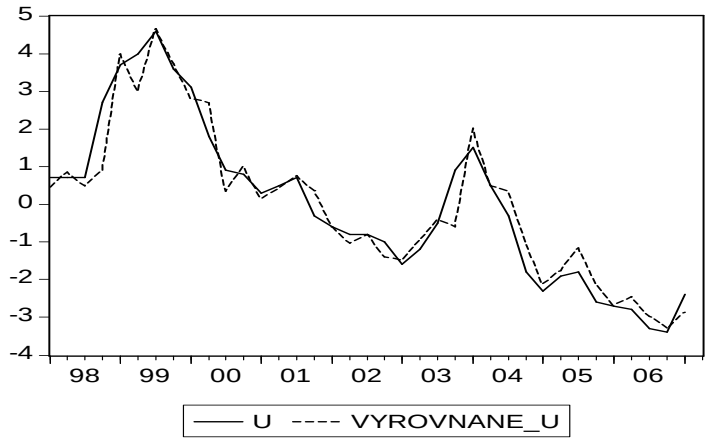
[5] Zdroj: Štatistický úrad Slovenskej republiky. Indikátor ekonomického sentimentu (IES) je zložený ukazovateľ, ktorý má za cieľ zobrazovať aktuálny stav očakávaní všetkých účastníkov ekonomického prostredia. Pozostáva z agregovaných údajov z výsledkov konjunkturálnych prieskumov v priemysle, stavebníctve, maloobchode, službách a z výsledkov prieskumu o názoroch spotrebiteľov na aktuálnu ekonomickú situáciu.

$$(1-\rho_1 L)e_t = (1-\theta_1)\varepsilon_t$$
$$e_t - \rho_1 e_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (8)$$
$$e_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} + \rho_1 e_{t-1}$$

Kde po dosadení hodnôt odhadnutých parametrov AR a MA procesov môžeme chybu modelu zapísať ako:

$$e_t = \varepsilon_t - 0.95398\varepsilon_{t-1} + 0.7727 e_{t-1} \quad (9)$$

Graf č. 1: graf pôvodných a odhadnutých hodnôt



Zdroj: Výstup Eviews 4.1. Student Version

Odhadovaný model sa javí ako štatisticky významný. Vzhľadom na Durbin-Watsonovú štatistiku (2.198) môžeme zamietnuť výskyt sériovej korelácie v modeli. Významnosť tohto testu má vzhľadom na nízky počet testovaných dát iba akýsi informačný význam. Tvrdenie o nekorelovanosti reziduí nám potvrdzuje aj graf ACF autokorelačnej funkcie, náhodné chyby reziduí sú nezávisle.

Stabilitu modelu otestujeme pomocou Chowoho testu bodu zlomu (Chow breakpoint test). Nulová hypotéza predpokladá, že medzi podmodelmi (*subsample*) nie sú štatistický významné rozdiely. Zamietnutie H_0 indikuje štrukturálne zmeny v dátach v analyzovanom období. Problémom testu je, že ak niektoré z testovaných období (podmodel) je kratšie ako počet parametrov, potom nie je možné tento test použiť. Na základe poznatkov o vývoji nezamestnanosti v Slovenskej republike a v súlade podmienkou dĺžky testovaných období, bol odhadnutý rad rozdelený na tri podmodely, podmodel do roku 2000, do roku 2004 a podmodel do druhého štvrťroku roku 2007.

Na základe *p* hodnoty F-štatistiky ($probability = 0,791577$) testu hodnotíme model ako stabilný, nezamietame nulovú hypotézu, (vzťah 11). Test je používaný aj v tvare vierohodnostného pomeru ($LR = 12,40489; \chi^2_{0,05;12} = 21,026$), vzťah (12).

$$F = \frac{\left(\mathbf{e}^{\mathrm{T}}\mathbf{e} - \left(\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 + \mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2\right)\right)/k}{\left(\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 + \mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2\right)/n-2k} \sim F_{k,n-2k} \quad (11)$$

$$LR = -2\left(l_e - l_{e_1} - l_{e_2} - \ldots - l_{e_m}\right) \sim \chi^2_{(m-1)k} \qquad (12)$$

Multikolinearitu regresorov testujeme Kleinovým testom[6] založeným na porovnaní koeficientu viacnásobnej determinácie s párovými koeficientmi determinácie medzi jednotlivými vysvetľujúcimi premennými. V našom prípade ak:

$$R_U^2 < r_{pzi,ies}^2 \qquad (13)$$

Potom je opodstatnený predpoklad multikolinearity. Párové koeficienty determinácie odhadnutého modelu majú hodnotu $r_{pzi,ies}^2 = 0.221627$, predpoklad multikolinearity regresorov teda nie je opodstatnený.

4. Diskusia a záver

Na základe použitých metód a aplikovaných ekonometrických testov sme zostavili model medziročnej zmeny miery nezamestnanosti v Slovenskej republike. Výsledkom je model s vysokou hĺbkou pamäte indikátorov. Z empirických poznatkov a z tvaru odhadnutej funkcie medziročnej zmeny nezamestnanosti by medziročný pokles nezamestnanosti o 0,02738 p. b. (v čase *t*) bol spôsobený rastom *IES* z obdobia druhého štvrťroka (*t-2*) pred sledovaným obdobím zmeny miery nezamestnanosti o jednotku, za predpokladu ostatných premenných nezmenených. Vplyv zmeny *PZI* na zmenu miery nezamestnanosti sa rozkladá na vplyv štvrťročných zmien *PZI*, počas obdobia troch štvrťrokov pred sledovaným obdobím (čas *t*) a zmenou *PZI* v čase *t* sledovanej medziročnej zmeny miery nezamestnanosti. Jednotlivé vplyvy sú kvantifikované regresnými koeficientmi podľa vzťahu (6), za predpokladu podmienok nezmenených (*ceteris paribus*).

5. Literatúra

GUJARATI, D.N. 2003. Basic Econometrics, New York: The McGraw-Hill, 2003.

HUŠEK, R. 1999. Ekonometrická Analýza, Bratislava: EKOPRESS, 1999.

PŘIVAROVÁ, M. 2001. Rovnováha a nerovnováha v ekonomickej teórii, Bratislava: IURA Edition, 2001.

Adresa autora:
Juraj Varga
Ekonomická Fakulta
Univerzita Mateja Bela
Tajovského 10
975 90 Banská Bystrica
juraj.varga@yahoo.com

[6] Kleinov test nie je štatistickým testom v pravom zmysle slova. Toto pomenovanie je ale zaužívane v praxi.

Využitie Brassovho modelu pri vyhodnotení vývoja úmrtnosti v SR

Vojteková Monika

Abstract: The article describes Brass model as one of the fundamental mortality models. The model connects standard life tables and mathematic rule of relation of any life table to the standard. This information is applied on demagraphic data of Slovak Republic and then the mortality is analyzed for the period 1980-2005 by the estimated parameters. Brass model is also used as the prediction tool of mortality.

Key words: *Brass model, life tables, logit transformation, survival function*

1 Úvod

Brassov model úmrtnosti (1971, pozri napr. [2], [4], [6], [7] a [8]) predstavuje parametrizáciu úmrtnostných tabuliek. Pomocou dvoch parametrov porovnáva úmrtnosť v skúmanej populácii s nejakou tabelovanou štandardnou úmrtnosťou. Na Slovensku sa danou problematikou zaoberala S. Sandtnerová vo svojej dizertačnej práci a odhadla parametre modelu pre populáciu SR za obdobie 1990-2000. Vo svojej práci som rozšírila toto časové obdobie na 1980-2005, ktoré zahŕňa aj politické zmeny v roku 1989. Mojím hlavným cieľom je analyzovať parametre modelu a zistiť, či v skúmanom období došlo k nejakým úrovňovým alebo štrukturálnym zmenám úmrtnosti slovenského obyvateľstva.

2 Brassov model

William Brass prišiel s myšlienkou nemodelovať priamo priebeh úmrtnosti s vekom, ale zmeny úmrtnosti v čase vzhľadom k nejakému štandardnému priebehu. Predpokladal, že krivka funkcie prežívania $l(x)$ v strednom veku (pre $25<x<60$) sa podobá na logistickú funkciu a zmeny tejto funkcie je možné popísať parametrami α a β.

Brass ďalej zistil, že vzťah medzi vekom x a funkciou $logit(l(x))$ je v stredných vekoch približne lineárny. Preto ak zoberieme dve úmrtnostné tabuľky a logity ich funkcií prežívania $l(x)$, tak vzťah medzi týmito dvomi množinami logitov je tiež približne lineárny. To možno nasledovne zapísať:

$$logit\big(l(t,x)\big) = \alpha(t) + \beta(t)\cdot logit\big(l_S(x)\big) + \varepsilon(t,x) \tag{1.1}$$

pričom platí:

$$\operatorname{logit}\big(l(x)\big) = 0.5\ln\left(\frac{1-l(x)}{l(x)}\right) \tag{1.2}$$

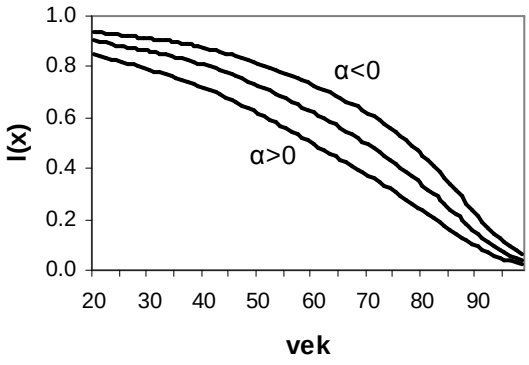
Parametre α a β môžeme získať napríklad pomocou metódy najmenších štvorcov. Ich význam sa dá interpretovať nasledovne:

- α: je ukazovateľom úrovne úmrtnosti ovplyvňujúcim úmrtnosť vo všetkých vekoch v rovnakom smere: vyššia hodnota α naznačuje vyššiu úmrtnosť (nižšiu pravdepodobnosť dožitia) vo všetkých vekoch

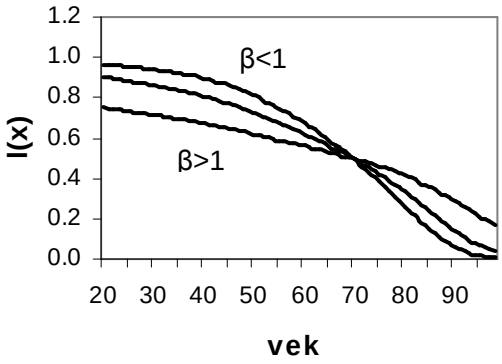
- β: hodnoty $\beta > 1$ implikujú, vzhľadom na štandard, nižšiu úmrtnosť v mladších vekoch, ale vyššiu úmrtnosť v starších vekoch. Hodnoty $\beta < 1$ majú opačný efekt. Zvyšovanie β znamená presúvanie úmrtnosti smerom k starším vekom

Vplyv konkrétnych parametrov α a β na krivku funkcie prežívania a tým jej následnú zmenu zobrazuje Graf 1 a Graf 2.

Graf 1. Funkcia prežívania, zmena α



Graf 2. Funkcia prežívania, zmena β



Odhadovanie parametrov Brassovho relačného modelu nespočíva len v určení parametrov α a β, ale aj v zvolení a odhadnutí samotného štandardu., napr. ako akýsi „priemer“ funkcií prežívania za dané obdobie. Ak predpokladáme, že odchýlky týchto funkcií majú náhodný charakter a normálne rozdelenie, môžeme využiť metódu najmenších štvorcov (regresnú analýzu).

Za prvotný odhad štandardu je možné zvoliť aritmetický priemer (Θ predstavuje množinu časových okamihov):

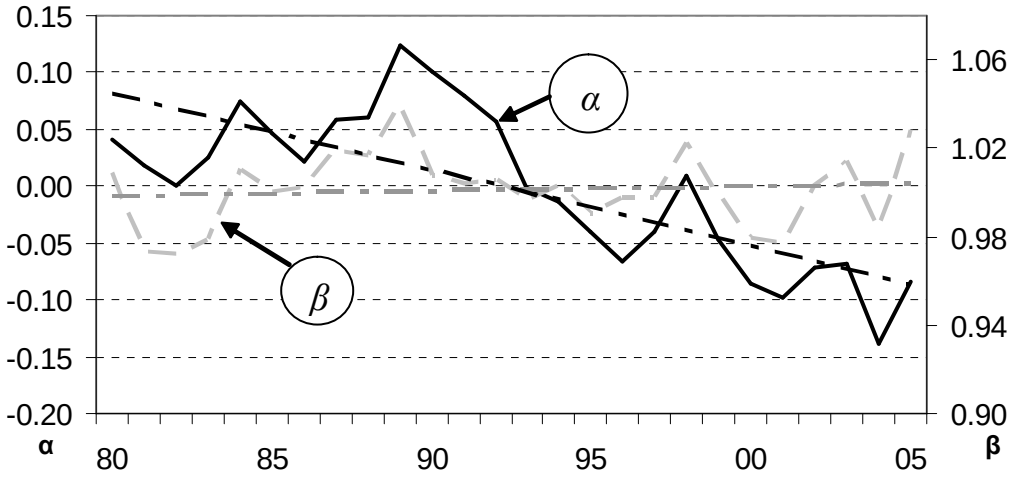
$$l_S(x) = \frac{1}{card\Theta} \sum_{t \in \Theta} l(t, x) \tag{1.3}$$

Následne sa tento štandard môže vylepšiť metódou najmenších štvorcov. Bližší popis odhadnutia a spresňovania štandardu je možné nájsť v [2], [6] a [8].

3 Aplikácia na dáta SR

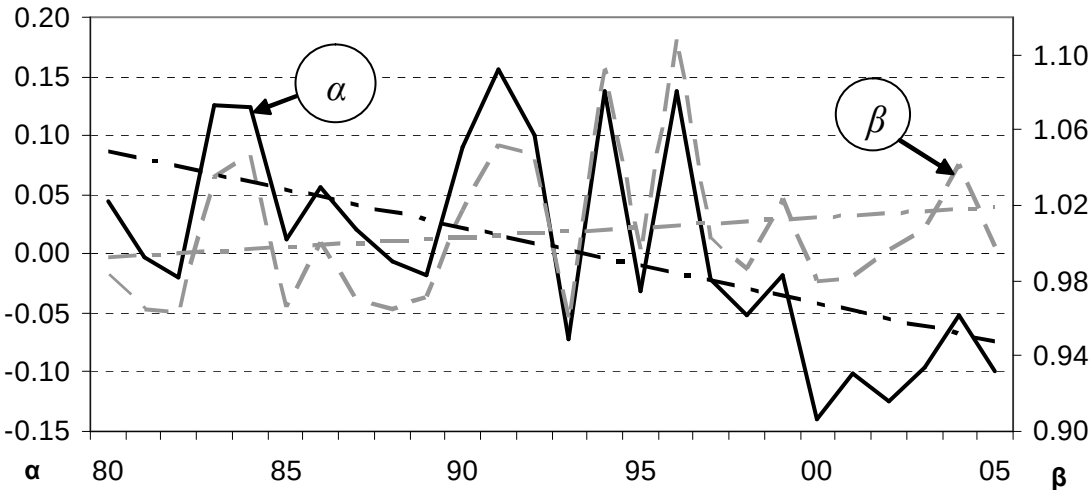
Brasovú relačnú metódu som aplikovala na dáta SR v období 1980-2005 na vekový interval 26-60 ročných obyvateľov. Výsledkom boli dve sady parametrov $\alpha(t)$ a $\beta(t)$ a vyhladené počty dožívajúcich. Keďže absolútne hodnoty parametrov nie sú zaujímavé, znázorňujem v grafoch ich priebeh v čase. Takto je možné pozorovať ako sa menila úroveň úmrtnosti (parameter α), resp. aký bol jej rozptyl oproti štandardu (parameter β).

Graf 3. Parametre α a β - muži



U mužov možno badať výraznú zmenu po roku 1989 u parametra *α*. Zo stúpavého trendu nastal v roku 1989 akoby bod zlomu a odvtedy parameter klesal. To znamená zvyšovanie úmrtnosti v období 1980-1990 a znižovanie od roku 1990 doteraz. Na overenie, či skutočne došlo k zlomu po roku 1989 som použila štatistické testy, ktoré potvrdili, že naozaj došlo k bodu zlomu v roku 1989. Pre parameter *β* nie je možné určiť jednoznačný trend, teda štruktúra úmrtnosti sa výrazne nezmenila.

Graf 4. Parametre α a β - ženy



U žien v období 1980-2005 môžeme vidieť výrazné výkyvy oboch parametrov. Trend parametra *α* za celé obdobie klesá, čo znamená, že sa úroveň úmrtnosť znižovala (aj keď s rôznymi odchýlkami). Parameter *β* v trende mierne stúpal, a to značí pozvoľné presúvanie úmrtnosti k starším vekom. Keďže zmena trendu parametrov po roku 1989 nie je viditeľná, pokúsila som sa, rovnako ako u mužov, použiť štatistické testy, ktoré následne potvrdili hypotézu, že nedošlo k zlomu v 1989.

4 Predikcia vývoja úmrtnosti:

Parametre α a β sú vhodné aj na predpovedanie vývoja úmrtnosti. Už Brass odporúčal jednoduchým predĺžením lineárneho trendu predpovedať vývoj úmrtnosti. Ak parametre vykazujú trend, môžeme ich interpolovať do budúcnosti a následne vypočítať zodpovedajúce krivky počtu dožívajúcich.

Podľa tejto predikcie očakávame u mužov aj naďalej pokles parametra α, teda pozitívny trend znižovanie úmrtnosti vo všetkých vekoch. Parameter β by sa nemal zmeniť, z toho dôvodu aj veková štruktúra úmrtnosti by mala zostať približne konštantná.

Rovnako aj pri ženách použitím lineárneho trendu očakávame pokles parametra α, teda znižovanie úmrtnosti a možno mierne stúpavý trend parametra β, teda presúvanie úmrtnosti k starším vekom.

5 Záver

Hlavným cieľom Brassovho modelu je analýza úmrtnosti za zvolené časové obdobie, ktoré v mojej práci predstavovalo rozmedzie rokov 1980-2005. Vývoj parametrov u mužov ukázal, že došlo k zmene charakteru úmrtnosti po roku 1989. U žien v tomto období nenastali také dynamické zmeny. Pri oboch pohlaviach nastal pozitívny trend, a to znižovanie úmrtnosti vo všetkých vekoch. U žien sa objavil aj trend presúvania úmrtnosti k starším vekom. Podrobnejšie výsledky sú uvedené v práci [8].

Predĺžením lineárnej regresie možno naznačiť budúci trend parametrov pre roky 2006 a 2007. Vo všeobecnosti očakávame u oboch pohlaví pokračovanie trendu z posledného obdobia, teda opätovné znižovanie úmrtnosti a nie výraznú zmenu štruktúry úmrtnosti.

[1] BENJAMIN B., SOLIMAN A. S. *Mortality on the move*. Napier House, Oxford, 1993

[2] KOSCHIN, F. – FIALA, T. – LANGHAMROVÁ, J. – ROUBÍČEK, V. *Úmrtnost v českých zemích v devadesátych letech*. VŠE, Praha, 1998

[3] MITRA, S. A simple model for linking life tables by survival-mortality ratios. In *Demography*. ISSN 00703370, 1983. vol. 20, no. 2, p. 227 – 234.

[4] EWELL C. *Methods and Models in Demography*. The Guilford Press, New York, 1988

[5] PRESTON S.H., HEUVELINE P., GUILLOT M. *Demography*. Brackwell, Oxford, 2001

[6] SANDTNEROVÁ S. *Úmrtnosť v Slovenskej republike v období 1991 až 2000*. Dizertačná práca. Ekonomická univerzita v Bratislave, 2005

[7] United Nations – Department of Economic and Social Affairs. Manual X : Indirect Techniques for Demographic Estimation [online]. (New York) : United Nations, 1983 [cit. 11.4.2007]. Kap. 1. Demographic Models. Dostupné v PDF verzii na internete: <http://www.un.org/esa/population/publications/Manual_X/Manual_X_Chapter_1.pdf>

[8] VOJTEKOVÁ M. *Brassov model a jeho aplikácia*. Bakalárska práca. Univerzita Komenského v Bratislave, 2007.

Adresa autora:
Monika Vojteková
M. Rázusa 30
010 01 Žilina
monika.vojtekova@post.sk



Riešenia pre poľnohospodárov

Agrospra je profilovaná ako firma,
ktorá dodáva kompletný sortiment hnojív,
s dôrazom na:

- **Amofos**
- **Draselnú soľ**
- **Dusičnan amónny**
- **Liadok amónny**
- **Síran amónny**
- **Hakofyt** – kvapalné hnojivo

Agrospra zabezpečuje doplnenie chýbajúcich živín
podľa pôdnych rozborov. Pričom plasticky
zohľadňuje nutnosť zmeny pomeru živín,
a v dodávaných hnojivách.

Agrospra, spol. s r. o.
Podzáhradná 70, Bratislava
Tel.: 02/ 4552 0851 E-mail: info@agrospra.sk

OBSAH

Chajdiak Jozef, Luha Ján: Z histórie seminárov Výpočtová štatistika	2
Bílková Diana: Modely mzdových rozdělení	4
Boďa Martin: Estimating the parameter θ of the Poisson distribution based on the divergence criteria	9
Boďa Martin, Čižmárik Pavol, Gavliak Rudolf: Analýza vybraných problémov reálnej konvergencie Slovenska k EMÚ	18
Boďa Martin, Gavliak Rudolf: Value at risk III. Indexný VCV model a diagnostika modelu value at risk$^{\infty}$	24
Boháčová Hana, Heckenbergerová Jana: Oblasti necitlivosti pro parametry střední hodnoty ve smíšeném lineárním regresním modelu s podmínkami typu I a s nimi spojené výpočetní problémy	31
Fiala Tomáš: Demographic Prognosis of the Future Development of the Number of Information Scientists in the Czech Republic	36
Fischer Jakub, Mazouch Petr: Ukazatele vzdělanosti populace: využitelnost odhadů z různých datových zdrojů	41
Heckenbergerová Jana, Boháčová Hana: Testování hypotéz o paralelnosti silnic aneb problémy satelitního mýtného systému	46
Hurbánková Ľubica: Priame metódy merania úrovne regiónov	52
Chajdiak Jozef, Luha Ján: Názory vybraných odborníkov na problematiku imigrácie	56
Kačerová Eva: Age structure of the inhabitants of the estates of the Eggenbergs in 1651 according to the List of Serfs according to Faith	64
Kanderová Mária: Optimalizácia portfólia investičných nástrojov v prostredí MS Excel	68
Korónyi Samuel, Prokešová Roberta, Medveďová Alžbeta: Exploračná analýza meraní hĺbky vodnej hladiny v hydrogeologických vrtoch na zosuve v Ľubietovej	74
Langhamrová Jitka, Fiala Tomáš: The Ageing of the Population and its Consequences	79
Löster Tomáš: Vliv znečištěného životního prostředí na vrozené vady a samovolné potraty v České republice	84
Majorová Martina: FDH DEA model ako alternatívna metóda benchmarkingu FDH DEA model as an alternative benchmarking method	89
Marek Lubos, Vrabec Michal: Confidence interval for relative frequency	95
Mazouch Petr, Fiala Tomáš, Fischer Jakub: Prognóza struktury vzdělanosti v České republice do roku 2050	103
Nemcová Anna: Gender equality indikátory a databázy vedomostnej spoločnosti	108
Nováková Gabriela: Prognóza populačného potenciálu Slovenska v roku 2025 (Aplikácia modelu populačného potenciálu na územie Slovenska)	113
Odehnal Jakub, Michálek Jaroslav: Konkurenceschopnost vybraných regionů EU a jejich klasifikace	118
Otáhal Miroslav: Počítačové zpracování a statistická analýza rozsáhlých souborů dat	123
Pecáková Iva: Nabídka a poptávka po IT odbornících v ČR 2007	129
Poměnková Jitka, Kašparovská Vlasta: Ověřovací studie přístupu hodnocení finanční výkonnosti stavebních spořitelen v ČR	137
Semerák Pavel, Kubanová Jana: Marketingový průzkum zájmu o produkty poskytovatelů služeb v oblasti cestovního ruchu	142
Sipková Ľubica: Nový prístup – nové možnosti štatistického modelovania, alebo ako „ušiť pravdepodobnostný model na mieru“	147

Stankovičová Iveta, Žambochová Marta: Odhad cenovej kategórie ojazdených motorových vozidiel pomocou rozhodovacích stromov v SASe	152
Tartaľová Alena: Využitie štatistického systému Stagraphics pre modelovanie poistných škôd	159
Úradníček Vladimír, Gavliak Rudolf: Simulácia vývoja úrokových sadzieb pomocou programov SPSS a MS® Excel	165
Vavrová Marianna: Štatistická analýza vývoja vyšetrení na kardiovaskulárne ochorenia v rokoch 1997 - 2006 vo vybranom regióne Slovenska	169
Vavrová Marianna: Štatistická analýza vývoja medikamentóznej liečby kardiovaskulárnych ochorení v sledovanom období vo vybranom regióne Slovenska	174
Marek Luboš, Vrabec Michal: Estimation of absolute and relative frequency for the sampling wi-thout replacement	180
Žváček Jiří: Otevřená statistika 2007	184

Prehliadka prác mladých štatistikov a demografov	191
Čupeľová Katarína: Demografické osobitosti urbánnych a rurálnych štruktúr Slovenska	192
Gurník Ivan: Kritéria rozhodovania investorov pri nákupe podielových fondov	196
Hoffmanová Lucia: Sledovanie časového a priestorového šírenia prvých ľudí druhu Homo sapiens sapiens pomocou najnovších vedeckých výskumov analýzy DNA	201
Horváth Beáta: Why seasonal adjustment?	205
Hrubina Peter: Vývoj akciových trhov krajín V4 a parciálna kointegrácia	210
Chromeček Andrej: Populačná explózia v rozvojových krajinách od roku 1950	214
Kello Miroslav: Analýza faktorov determinujúcich ceny bytov	218
Magna Marián: Analýza zamestnanosti v krajinách Európskej únie	223
Margetínová Alexandra: Populačný vývoj štátov Európskej únie v rokoch 1960- 2003	227
Martináková Jana: Vrodené vývojové chyby, závažný problém morbidity obyvateľstva Slovenska	232
Onderišinová Viera: Vzdelanie žien na Strednom Východe a v Severnej Afrike	236
Osička Tomáš: Vybrané modely oceňovania akciových trhov	240
Pavlovčinová Dominika: Vývoj pôrodnosti a plodnosti podľa vzdelanie matky na Slovensku	244
Pohanková Petra: Demografické problémy Číny	248
Pukačová Jana: Analýza infraštruktúrnej vybavenosti rómskych osád v Banskobystrickom kraji	252
Varga Juraj: Empirický model miery nezamestnanosti v Slovenskej Republike	257
Vojteková Monika: Využitie Brassovho modelu pri vyhodnotení vývoja úmrtnosti v SR	262
Obsah	266

FORUM STATISTICUM SLOVACUM

vedecký 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
02/63812565

e-mail
chajdiak@statis.biz
Jan.Luha@statistics.sk

Registráciu vykonalo
Ministerstvo kultúry Slovenskej republiky

Registračné číslo
3416/2005

Tematická skupina
B1

Dátum registrácie
22. 7. 2005

Objednávky
Slovenská štatistická a demografická spoločnosť
Miletičova 3, 824 67 Bratislava 24
Slovenská republika
IČO: 178764
Číslo účtu: 0011469672/0900

ISSN 1336-7420

Redakčná rada
RNDr. Peter Mach – *predseda*

Doc. Ing. Jozef Chajdiak, CSc. – *šéfredaktor*

RNDr. Ján Luha, CSc. – *tajomník*

členovia:
Ing. Mikuláš Cár, CSc.
Ing. Ján Cuper
Ing. Pavel Fľak, DrSc.
Ing. Edita Holičková
Doc. RNDr. Ivan Janiga, CSc.
Ing. Anna Janusová
RNDr. PaedDr. Stanislav Katina, PhD.
Prof. RNDr. Jozef Komorník, DrSc.
RNDr. Samuel Korónyi
Doc. Ing. Milan Kovačka, CSc.
Doc. RNDr. Bohdan Linda, CSc.
Prof. RNDr. Jozef Mládek, DrSc.
Doc. RNDr. Oľga Nánásiová, CSc.
Doc. RNDr. Karol Pastor, CSc.
Prof. RNDr. Rastislav Potocký, CSc.
Doc. RNDr. Viliam Páleník, PhD.
Ing. Iveta Stankovičová, PhD.
Doc. RNDr. Beata Stehlíková, CSc.
Prof. RNDr. Michal Tkáč, CSc.
Ing. Vladimír Úradníček, PhD.
Ing. Boris Vaňo
Doc. MUDr Anna Volná, CSc., MBA.
Ing. Mária Vojtková, PhD.
Prof. RNDr. Gejza Wimmer, DrSc.
Mgr. Milan Žirko

Ročník
III.

Číslo
6/2007

Cena výtlačku 500 SKK / 20 EUR
Ročné predplatné 1500 SKK / 60 EUR