

Štatistické modelovanie príjmov domácností v Slovenskej republike

Viera PACÁKOVÁ – Ľubica SIPKOVÁ – Eva SODOMOVÁ*

Statistics Modelling of Household's Incomes in the Slovak Republic

Abstract

This article contents the methods and the results of distribution modelling of household's income in the Slovak Republic based on sampling data of yearly real net income of 1 566 households in year 2002.

Goodness of fit tests confirms the Weibull distribution with maximum likelihood estimation of parameters as the best-fit model, except the intervals of the lowest and the highest values of incomes. Two Weibull-Pareto quantile functions have been found as good fit models for the lowest and the highest incomes too.

Keywords: *the income of households, distribution models, Weibull distribution, Weibull-Pareto distribution, quantil functions*

JEL Classification: C13, C16, C51, C52

Úvod

Analýza rozdelenia príjmov obyvateľstva je základom rozhodovania v oblasti rozpočtovej a sociálnej politiky a východiskom odhadu spotreby domácností v budúcnosti, ako aj dôsledkov rôznych sociálnych a ekonomických opatrení a reforiem.

Štatistickú analýzu rozdelenia príjmov domácností, ktorá je obsahom tohto príspevku, možno považovať za kľúčovú v komplexnej analýze životnej úrovne obyvateľov a za východiskovú pre rozhodnutia v sociálnej oblasti. Jej cieľom je prezentovať možnosti aplikácie štatistických metód pri analýze príjmovej diferenciácie a nájsť vhodný model rozdelenia pravdepodobnosti príjmov domácností v Slovenskej republike. Základom analýzy sú výberové údaje zo štatistiky

* doc. RNDr. Viera PACÁKOVÁ, PhD. – Ing. Ľubica SIPKOVÁ, PhD. – doc. Ing. Eva SODOMOVÁ, PhD. Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra štatistiky, Dolnozemska cesta 1, 852 35 Bratislava 5; e-mail: <pacakova@euba.sk>; <lusipko@msn.com>; <sodomova@euba.sk>

rodinných účtov v Slovenskej republike v roku 2002,¹ poskytnuté Štatistickým úradom SR.

Presvedčenie viacerých odborníkov z oblasti analýzy rozdelenia príjmov reprezentuje názor J. S. Chipmana [6], že v súčasnej heterogénnej populácii nie je možné dostatočne charakterizovať príjmové rozdiely len použitím niekoľkých súhrnných charakteristík rozdelenia (stredných hodnôt, mier variability, šikmosti a špicatosti). Z pohľadu ekonomickej aj štatistickej teórie vystupuje do popredia potreba poznať model rozdelenia pravdepodobnosti príjmov pre celú populáciu, ako aj pre subpopulácie, resp. špecifické skupiny populácie. Znalosť takýchto modelov umožní získať lepšie odhady ich charakteristík, ale hlavne komplexnejšie popísať a porovnávať rozdelenia príjmov.

Preto je základom analýzy príjmovej diferenciácie obyvateľstva určitého územného celku nájdenie funkčného vyjadrenia hustoty alebo distribučnej funkcie rozdelenia príjmov. Optimálnym je flexibilný analytický tvar, vhodný aj na modelovanie štruktúrnych rozdielov podľa sociálno-ekonomických alebo geografických atribútov, pritom však čo najjednoduchší a vhodne definovaný, aby bol možný jednoduchý výpočet charakteristík rozdelenia príjmov a príjmovej nerovnosti.

1. Grafické metódy štatistickej analýzy empirických údajov

Predmetom nášho záujmu je náhodná premenná *CP* – čistý ročný príjem domácností v Slovenskej republike v roku 2002. Vlastnosti tejto premennej úplne charakterizuje jej zákon rozdelenia pravdepodobnosti. Na určenie jeho analytických tvarov sme využili pokročilé metódy štatistickej indukcie, a to testy dobrej zhody a modelovanie pomocou kvantilových funkcií.

Vstupnými empirickými údajmi štatistickej analýzy sú údaje o čistých reálnych ročných príjmoch v roku 2002 (*CP*) vybraných 1 566 domácností SR, zistených Štatistickým úradom SR zo štatistiky rodinných účtov. Empirické rozdelenie týchto údajov a jeho vlastnosti, analyzované metódami deskriptívnej štatistiky, sú východiskom modelovania pravdepodobnostného rozdelenia.

Prvým krokom deskriptívnej štatistiky je grafická analýza. Charakteristické vlastnosti rozdelenia príjmov, odhalené v grafoch, dopĺňa interpretácia charakteristík empirického výberového súboru.

Symetriu, resp. asymetriu rozdelenia príjmov je možné posúdiť pomocou tzv. krabicového grafu (*box-plotu*) na obrázku 1, a to na základe postavenia kvartilov² empirického rozdelenia a dĺžky tzv. fúzov, siahajúcich do 1,5-násobku

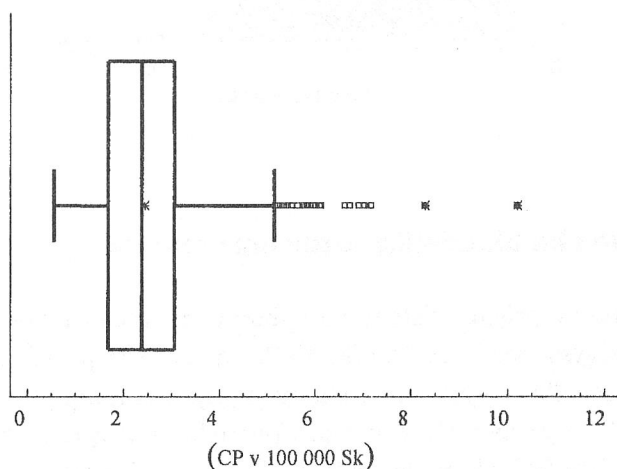
¹ Vhodnejším výberovým súborom by boli údaje z mikrocenzu, ktoré však v čase analýzy ešte neboli k dispozícii.

² Definícia kvartilov a kvartilových charakteristík je uvedená napríklad v [14, s. 59].

kvartilového rozpätia. Znázornené sú aj extrémne vysoké príjmy, presahujúce 1,5-násobok, resp. trojnásobok kvartilového rozpätia. Pozoruhodná je malá odlinosť dvoch častí krabicky, predstavujúcich 1/4 príjmov domácností medzi dolným kvartilom (166 163 Sk) a mediánom (237 407 Sk), ako aj 1/4 príjmov medzi mediánom a horným kvartilom (306 433 Sk), s nevelkým posunom priemeru (244 284 Sk) oproti mediánu. To naznačuje takmer symetrické rozdelenie polovice hodnôt príjmov domácností, ak neuvažujeme 1/4 najnižších a 1/4 najvyšších príjmov.

Obrázok 1

Krabicový graf (box plot) čistých príjmov domácností



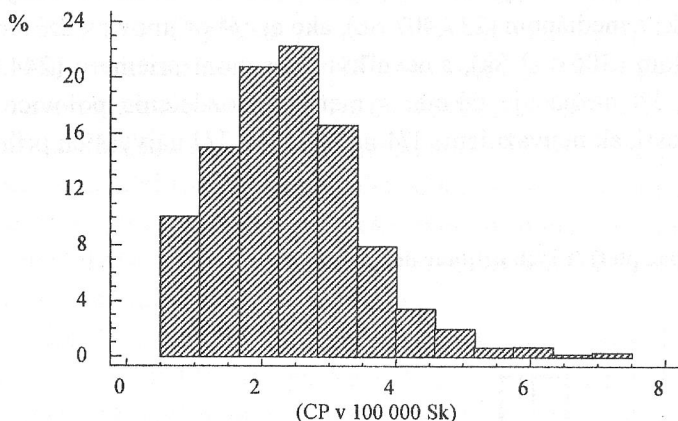
Skutočnosť, že dĺžka pravého „fúza“ je približne dvojnásobkom ľavého, 29 vysokých hodnôt premennej *CP* a dve hodnoty presahujúce trojnásobok kvartilového rozpätia, svedčí o výraznom predĺžení jeho pravého konca, teda existencii aj extrémne vysokých príjmov.

Grafické zobrazenie rozdelenia početností príjmov domácností pomocou histogramu (obr. 2) potvrdzuje skutočnosti, zistené v krabicovom grafe a zvyrazňuje akoby useknutú dolnú časť rozdelenia príjmov.

V ľavej časti histogramu rozdelenia príjmov – od čistého ročného príjmu domácností 100 tis. Sk po vrchol rozdelenia – znázorňujúceho počet domácností z typického (najčastejšieho) intervalu čistých ročných príjmov (250 000; 300 000 Sk), sú prírastky stĺpikov histogramu výrazne menšie ako úbytky početností v časti za vrcholom, po hodnotu čistých ročných príjmov cca 400 tis. Sk. Táto časť rozdelenia nie je pravostranne zošíkmená, vyznačuje sa skôr slabou ľavostrannou (zápornou) asymetriou.

Obrázok 2

Histogram rozdelenia relatívnych početností príjmových intervalov



2. Výberové charakteristiky rozdelenia príjmov³

Najnižší ročný príjem, zistený vo výberovom súbore 1 566 domácností, bol 54 325 Sk, najvyšší bol 1 020 600 Sk. Veľké variačné rozpätie,⁴ rovné 966 275 Sk, mohlo byť dôsledkom výskytu extrémnych príjmov. Rozptyl príjmov ($Variance = 1,2011810^{10}$) však naznačuje aj vysokú variabilitu zistených hodnôt. Štandardná odchýlka 109 598 Sk predstavuje až 44,87 % priemerného príjmu, ktorý vo výberovom súbore domácností na Slovensku v roku 2002 bol 244 284 Sk. Pearsonov koeficient šikmosti (*Skewness*), rovný približne 1,05, naznačuje pravostrannú asymetriu celého rozdelenia príjmov. Koeficient špicatosti 2,997 (*Kurtosis*) vyjadruje výraznú špicatosť empirického rozdelenia príjmov v porovnaní s normálnym rozdelením.

Uvedená zostava hodnôt Pearsonových momentových charakteristík sa bežne používa v štatistickej praxi na posúdenie základných vlastností rozdelenia pravdepodobnosti. Ak by takejto kvantitatívnej analýze nepredchádzala dôsledná grafická a kvantilová analýza empirického rozdelenia, pohľad na skutočné vlastnosti rozdelenia príjmov len na základe momentových charakteristík by mohol byť značne skreslený a mohol by viesť k nevhodným postupom v pravdepodobnostnom modelovaní.

³ Výberové charakteristiky sú uvedené v prílohe článku ako výstup procedúry *Summary Statistics* programového balíka *STATGRAPHICS Plus*. Väčšinu z nich nájdeme aj v tabuľke 2, v stĺpci *empirické rozdelenie*.

⁴ Definície a výpočtové vzorce výberových charakteristík nájdeme napríklad v [14, kap. 2].

Podľa Moorsom [11] odporúčanej štvorice kvantilových charakteristík,⁵ adekvátnych uvedeným momentovým mieram, je variabilita polovice hodnôt empirického rozdelenia veľmi malá. Charakterizuje ju kvartilové rozpätie, rozdiel medzi hodnotou horného a dolného kvartilu, ktoré je len 139 770 Sk. Polovica vybraných domácností sa líšila hodnotou čistých ročných príjmov najviac o cca 140 000 Sk. Dokonca ani rozdiel príjmov 80 % domácností, vyjadrený decilovým rozpätím, nepresiahol 270 000 Sk. Rozdiel ročných príjmov 98 % domácností, ak neuvažujeme 1 % najnižších a 1 % najvyšších príjmov, teda percentilové rozpätie, je len 518 791 Sk, čo je nízka hodnota v porovnaní s variačným rozpätím, ktoré je 966 275 Sk.

Rozdiel horného kvartilu a mediánu, nazývaný hornou kvartilovou diferenciou ($uqd = 69\,026$ Sk), je menší ako dolná kvartilová diferencia ($lqd = 70\,744$ Sk). Z toho vyplýva záporná hodnota Galtonovho koeficientu šikmosti ($g = -0,123$), ktorá naznačuje slabé, ale záporné, t. j. ľavostranné zošikmenie prostrednej polovice rozdelenia príjmov (ak neuvažujeme štvrtinu najnižších a štvrtinu najvyšších príjmov).

Na základe uvedených charakteristík možno konštatovať, že úroveň príjmov domácností v Slovenskej republike, vyjadrená priemernou hodnotou čistých ročných príjmov 244 284 Sk, je dostatočne reprezentatívnou strednou hodnotou. Relatívne vysokú mieru variability, ako aj jej percentuálny podiel na priemernom príjme spôsobilo umocnenie malého počtu extrémne vysokých odchýlok od priemeru. Potvrzuje to malá odchýlka mediánového príjmu od priemeru, malý rozdiel medzi horným a dolným kvartilom, ako aj horným a dolným decilom.

Vysoká hodnota Pearsonovej špicatosti je spôsobená značným predĺžením pravého konca rozdelenia za 97. percentilom v spojení s krátkym hrubým ľavým koncom empirického pravdepodobnostného rozdelenia, teda s vysokým percentom pravdepodobnosti výskytu hodnôt v oboch koncoch. Obe dve tieto faktory výrazne prispievajú k neúmernemu rastu štvrtej mocniny odchýlok od priemeru a vysoká hodnota Pearsonovej špicatosti sa tak stáva mierou koncov rozdelenia, a nie mierou výšky jeho vrcholu.

3. Modelovanie rozdelenia príjmov pomocou testov dobrej zhody

Pri modelovaní rozdelenia náhodnej premennej hľadáme vhodný analytický tvar rozdelenia pravdepodobnosti, vyjadrený najčastejšie pomocou funkcie hustoty, prípadne distribučnej funkcie, ktorý dostatočne presne (resp. dostatočne tesne) vystihuje empirické rozdelenie údajov výberového súboru. Tvar teoretického rozdelenia a odhadnuté hodnoty jeho parametrov by mali zodpovedať základným

⁵ Prehľad a definície kvantilových charakteristík nájdeme v [10].

vlastnostiam empirického rozdelenia, čo sa týka polohy, variability, šikmosti a špicatosti, pri dodržaní zásady jednoduchosti a čo najmenšieho počtu parametrov s ich dobrou interpretáciou.

Z empirickej analýzy výberového súboru vyplýva, že náhodná premenná CP – *čisté ročné príjmy domácností*, má nesymetrické, pravostranne zošikmené rozdelenie s charakteristickým predĺženým pravým koncom. Preto sa v pravdepodobnostnom modelovaní príjmov najčastejšie používajú kladne asymetrické rozdelenia, akými sú Paretovo, Weibullovo, gama, lognormálne a inverznenormálne rozdelenie.

Zhodu s týmito pravdepodobnostnými modelmi, s parametrami, odhadnutými metódou maximálnej virohodnosti (pozri napr. [13, kap. 2.3.1], sme overovali pomocou štatistického programového balíka STATGRAPHICS *Plus* χ^2 -testom a Kolmogorovovým-Smirnovovým (K-S) testom dobrej zhody [13, kap. 2.3.2 a 2.3.3]. Výsledky testov potvrdili, že empirické údaje o príjmoch vybraných 1 566 domácností najlepšie modeluje Weibullovo rozdelenie s maximálne virohodnými odhadmi parametrov.

Túto zhodu môžeme posúdiť na základe výstupnej tabuľky procedúry *Distribution Fitting* systému STATGRAPHICS *Plus* (tab. 1).

Tabuľka 1

Goodness-of-Fit Tests for CP

Chi-Square Test					
	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	50000.0	0	27.86	27.86
	50000.0	95833.3	122	96.97	6.46
	95833.3	141667.0	143	169.90	4.26
	141667.0	187500.0	239	225.08	0.86
	187500.0	233333.0	256	248.72	0.21
	233333.0	279167.0	273	237.75	5.23
	279167.0	325000.0	133	199.73	5.54
	325000.0	370833.0	129	148.54	2.57
	370833.0	416667.0	72	98.12	6.95
	416667.0	462500.0	39	57.61	6.01
	462500.0	508333.0	23	30.07	1.66
	508333.0	554167.0	14	13.93	0.00
above	554167.0		20	8.72	14.59
Chi-Square = 82.2094 with 10 d. f.			P-Value = 1.84963E-13		

Estimated Kolmogorov Statistic DPLUS = 0.0415752

Estimated Kolmogorov Statistic DMINUS = 0.0270346

Estimated Overall Statistic DN = 0.0415772

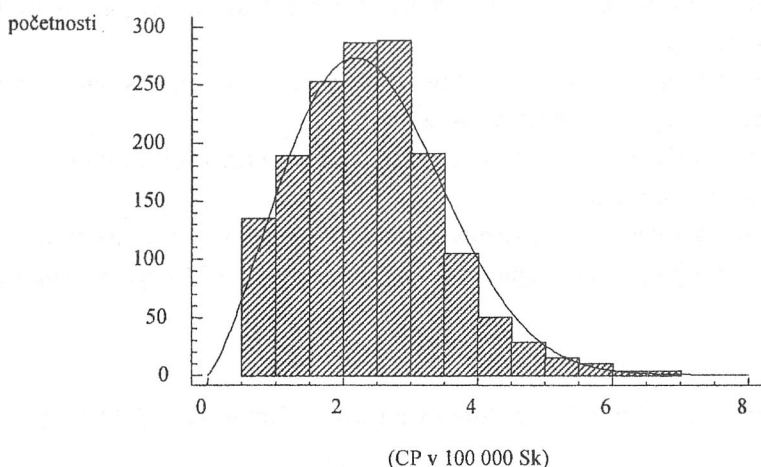
Approximate P-Value = 0.00899895

Pri K-S teste môžeme prijať hypotézu o predpokladanom type Weibullovo rozdelenia na hladine významnosti približne 0,01, lebo hodnota p-Value = 0,00899895.

Výsledky χ^2 -testu, vysoká hodnota testovacej charakteristiky $\chi^2 = 82,2094$ a veľmi nízka p-hodnota ($p\text{-Value} = 1,84963\text{E-}13$) na prvý pohľad jednoznačne vedú k zamietnutiu predpokladaného rozdelenia. Ak však vo výpočtovej tabuľke 1 porovnáme empirické (*observed*) a očakávané (*expected*) početnosti, vidíme, že dôvodom vysokej hodnoty χ^2 charakteristiky, a tým aj zamietnutia predpokladaného rozdelenia, sú veľké rozdiely skutočných a očakávaných početností v intervale najnižších aj najvyšších príjmov.

Obrázok 3

Grafické overenie výsledku testu dobrej zhody s Weibullovým rozdelením



Túto skutočnosť potvrdzuje aj graf na obrázku 3. Umožňuje vizuálne posúdenie zhody empirického rozdelenia, znázorneného histogramom, a teoretického Weibullovho rozdelenia, vyjadreného jeho hustotou.

Vhodný flexibilný tvar rozdelenia pravdepodobnosti, ktoré na rozdiel od Weibullovho rozdelenia dobre modeluje aj intervaly najnižších a najvyšších príjmov, nájdeme metódami modelovania pomocou kvantilových funkcií.

4. Modelovanie rozdelenia príjmov pomocou kvantilových funkcií

Kvantilová funkcia⁶ náhodnej premennej je menej známym spôsobom definovania rozdelenia pravdepodobnosti ako *distribučná funkcia* a *funkcia hustoty* [13, s. 19 – 20]. Hodnotu reálnej premennej X , označenú x , pre ktorú platí

⁶ Kvantilová funkcia (*Quantile Function* – QF); prvá práca, ktorá systematicky spracovala kvantilové funkcie a ich vlastnosti, je od Parzena [16].

$$F(x) = P(X \leq x) = p$$

nazývame *p*-kvantilom. Funkcia $F(x)$ je zrejme distribučná funkcia náhodnej premennej X .

Kvantilová funkcia spojitej náhodnej premennej X je vyjadrená vzťahom:

$$Q(p) = x \quad \text{pre každé reálne } p, 0 \leq p \leq 1.$$

Kvantilová funkcia $Q(p)$ je neklesajúcou funkciou a definuje sa aj ako inverzná funkcia k distribučnej funkcii.

Matematicky upravovať a kombinovať kvantilové funkcie možno len pri dodržaní určitých základných pravidiel tak, aby výsledný tvar bol znovu kvantilovou, teda neklesajúcou funkciou, so štatisticky interpretateľnými parametrami (podrobnejšie pozri napr. [9; 15]):

- *Pravidlo o sčítaní* hovorí, že súčtom dvoch neklesajúcich kvantilových funkcií je znovu neklesajúca kvantilová funkcia.

- Pri použití *pravidla o násobení* dvoch kladných kvantilových funkcií dostaneme opäť kvantilovú funkciu.

- *Pravidlo Q-transformácie* má rozsiahle použitie a podľa neho, ak transformujúca funkcia $z = T(x)$ je neklesajúcou funkciou x , potom $T(Q(p))$ je kvantilovou funkciou:

$$Q_{T(x)}(p) = T(Q_x(p))$$

Ak je $T(x)$ nerastúcou funkciou, potom kvantilovou funkciou je $T(Q(1-p))$:

$$Q_{T(x)}(p) = T(Q_x(1-p))$$

- Podľa *p-transformačného pravidla* funkcia $Q(H(p))$ je kvantilovou funkciou s tým istým oborom funkčných hodnôt ako $Q(p)$, ak $H(p)$ je neklesajúcou štandardizovanou funkciou p , $0 \leq p \leq 1$, t. j. platí $H(0) = 0$ a $H(1) = 1$.

Na modelovanie rozdelenia príjmov domácností v SR sme na základe predchádzajúcich analýz a pri použití uvedených pravidiel modifikácie kvantilových funkcií vytvorili zložené štvorparametrické nelineárne *Weibull-Pareto* rozdelenie, symbolicky označované **WPD**(λ , η , β , γ). Jeho kvantilová funkcia je vyjadrená analytickým vzťahom

$$Q(p) = \lambda + \eta \left\{ (1-p) [-\ln(1-p)]^\beta + p \left[\frac{1}{(1-p)^\gamma} \right] \right\}, \quad \text{kde } 0 < p < 1, \beta > 0, \gamma > 0 \quad (1)$$

pričom

- λ – parameter polohy,
- η – parameter variability,
- β a γ – parametre tvaru.

Mimoriadne náročnou etapou modelovania pomocou kvantilovej funkcie je odhad jej parametrov. Odhadom parametrov kvantilovej funkcie (1) Weibull-Paretovho rozdelenia aproximatívnu Öztürk a Dale metódou, zovšeobecnenou pre kvantilové rozdelenia [9, s. 198], pri použití optimalizačného programu Solver tabuľkového procesora Excel, sme získali tvar pravdepodobnostného rozdelenia

$$\text{WPD}_{\text{OD}}(51700, 6361247; 194314, 5349402; 0,6872262; 0,2122157).$$

Po dosadení odhadnutých parametrov do vzťahu (1) dostaneme analytické vyjadrenie

$$Q(p) = 51700,63612470 + 194314,53494018 \left\{ (1-p) [-\ln(1-p)]^{0,68722619} + p \left[\frac{1}{(1-p)^{0,21221575}} \right] \right\}$$

Metódou minimalizácie absolútnych distribučných rezíduí [9, s. 203], opäť pomocou optimalizačného programu Solver, sme získali tvar

$$\text{WPD}_A(53935,0833985; 186704,8127754; 0,656852; 0,2329894)$$

s analytickým vyjadrením podľa (1) v tvare

$$Q(p) = 53935,08339853 + 186704,81277537 \left\{ (1-p) [-\ln(1-p)]^{0,65685205} + p \left[\frac{1}{(1-p)^{0,23298934}} \right] \right\}$$

5. Posúdenie kvality kvantilových modelov

Obidva získané tvary Weibull-Paretovho kvantilového rozdelenia dobre modelujú rozdelenie empirických hodnôt náhodnej premennej *CP* – čistých ročných príjmov domácností v SR, pričom v závislosti od použitej metódy odhadu parametrov dosahujú lepšiu zhodu v rozdielnych častiach rozdelenia, ako vidno z Q-Q grafov porovnania empirických (výberových) a teoretických hodnôt kvantilov rozdelení WPD_A aj WPD_{OD} (obr. 4).

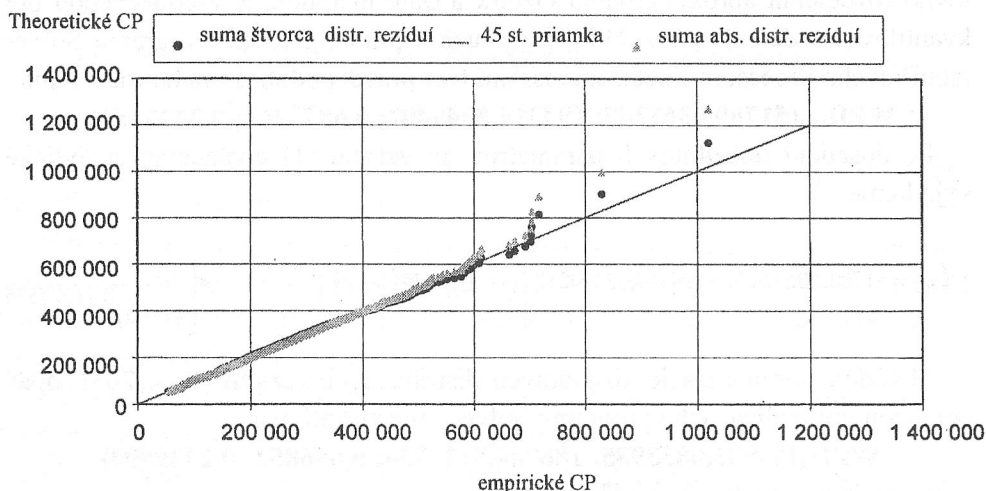
Ak sa Q-Q graf na koncoch odchyľuje od lineárneho tvaru, svedčí to o dlhšom, resp. príliš krátkom konci zvoleného teoretického rozdelenia. Odchylenie na ľavom konci k nižším hodnotám výberových kvantilov naznačuje, že teoretické rozdelenie má mať dlhší ľavý koniec a odchylenie na pravom konci k vyšším hodnotám znamená adekvátne dlhší pravý koniec teoretického rozdelenia.

Mieru lineárnej závislosti hodnôt kvantilov teoretického a empirického rozdelenia v Q-Q grafoch možno vyjadriť koeficientom korelácie. Pri obidvoch odhadnutých tvaroch WPD sú ich hodnoty vysoké (0,9963 a 0,9982) a vyjadrujú veľmi silnú lineárnu závislosť.

Vhodnosť teoretického modelu sa posudzuje hlavne na základe jeho dobrej zhody s rozdelením empirických údajov, napríklad χ^2 testom dobrej zhody.

Obrázok 4

Q-Q grafy nájdených kvantilových modelov



Testovacie kritérium χ^2 testu dobrej zhody pri definovaní teoretického rozdelenia pomocou kvantilovej funkcie vyjadruje vzťah⁷

$$\chi^2_{m-1} = \sum_{j=1}^m \frac{\left(n_j - \frac{n}{m}\right)^2}{\frac{n}{m}} \quad \text{pre } j = 1, 2, \dots, m.$$

Jeho hodnota pre tvar WPD_A je 28,99 a 28,64 pre tvar WPD_{OD} . Výsledné hodnoty χ^2 testovacieho kritéria po porovnaní s kritickou hodnotou χ^2 rozdelenia ($\chi^2_{0,95}(19) = 30,14$) potvrdili na hladine významnosti $\alpha = 0,05$ platnosť hypotézy o dobrej zhode pre obidva kvantilové tvary WPD_A aj WPD_{OD} .

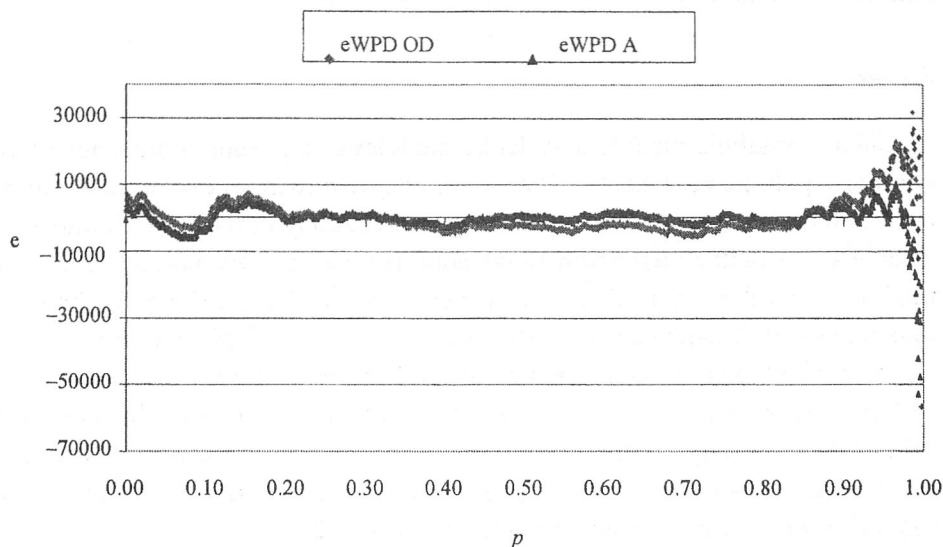
Model WPD_A tesnejšie modeluje 80 % stredných čistých príjmov domácností SR v roku 2002 než model WPD_{OD} , čo názorne vyjadruje graf reziduálnych odchýlok s empirickými údajmi pre obidva nájdené modely. Dôležité je zistenie, že pri modeli WPD_A iba v koncových horných 2 % hodnôt rezíduá presahujú veľkosť 10 tisíc Sk. Je to pri najvyšších hodnotách čistých príjmov, pri ktorých je úroveň aj variabilita hodnôt najväčšia.

Ak do analytického vyjadrenia kvantilovej funkcie $Q(p)$ modelu WPD_A , uvedeného v časti 4, dosadíme za pravdepodobnosť p ľubovoľné číslo od cca 0,1 do 0,9, dostaneme veľmi presnú hodnotu p -kvantilu x_p čistých ročných príjmov všetkých domácností v SR.

⁷ Vzťah odvodila Ľ. Sipková v [18].

Obrázok 5

Graf reziduálnych odchýlok pre nájdené kvantilové modely



Napríklad dosadením za $p = 0,1$ dostaneme dolný decil $x_{0,1} = 111\,364$ Sk, ktorý vyjadruje, že 1/10 domácností v SR mala čisté ročné príjmy v roku 2002 nižšie ako 111 364 Sk. Dosadením za $p = 0,75$ dostaneme horný kvartil $x_{0,75} = 305\,196$ Sk, teda len 1/4 domácností v SR mala čisté ročné príjmy vyššie ako 305 196 Sk.

Číselné posúdenie kvality zhody viacerými štatistickými mierami, ako aj kvantilové charakteristiky empirického a teoretických rozdelení WPD_A a WPD_{OD} sú na porovnanie uvedené v tabuľke 2.

Tabuľka 2

Charakteristiky teoretických a empirického rozdelenia a ich zhody

Názov štatistiky	Empirické rozdelenie	WPD_{OD}	WPD_A
Minimálna hodnota	54 325	52 532	54 938
Dolný percentil	65 388	61 722	64 733
Dolný decil	107 956	108 792	111 364
Dolný kvartil	166 663	165 241	165 621
Medián	237 407	239 865	237 112
Horný kvartil	358 324	308 089	305 196
Horný decil	377 394	371 291	373 608
Horný percentil	584 179	569 548	600 799
Maximálna hodnota	1 020 600	1 124 060	1 271 774
SSE	x	6.9029E+10	1.6290E+11
Rozptyl rezidui	x	44 080 072	104 020 307
Štd. odchýlka rezidui	x	6 639	10 199
SAE	x	6 026 843	3 611 875
Priemer SAE	x	3 849	2 306
Koeficient korelácie	x	0.9982	0.9963

Najnižšiu priemernú absolútnu distribučnú odchýlku (priemer SAE) má tvar WPD_A , len 2 306 Sk. Pri kvantilovom modeli WPD_{OD} sa teoretické hodnoty od empirických odchýľujú v priemere len o $\pm 6\,639$ Sk.

Záver

Článok obsahuje metódy a výsledky modelovania príjmov domácností v Slovenskej republike na základe výberových údajov o ročnom čistom reálnom príjme 1 566 domácností v roku 2002. Testy dobrej zhody potvrdili Weibullovo rozdelenie s maximálne vierohodnými odhadmi parametrov ako najlepší model s výnimkou intervalov najnižších a najvyšších príjmov. Dva zložené Weibull-Paretové modely boli nájdené ako modely s dobrou zhodou aj pre najnižšie a najvyššie príjmy metódami modelovania pomocou kvantilových funkcií.

Treba zdôrazniť, že charakteristiky takého teoretického modelu, ktorý spĺňa kritériá dobrej zhody s empirickým rozdelením, sú charakteristikami základného, nie výberového súboru. V našom prípade teda charakterizujú rozdelenie čistých ročných príjmov všetkých domácností v SR v roku 2002.

Kvantilové modelovanie má viaceré výhody, napríklad veľký výber funkčných tvarov, široké možnosti ich modifikácie pomocou matematického aparátu, a hlavne veľkú flexibilitu komplexných kvantilových rozdelení. To umožňuje ich úspešnú aplikáciu aj v prípadoch, keď klasické postupy modelovania pomocou hustoty pravdepodobnosti alebo distribučnej funkcie neprinášajú uspokojivé výsledky.

Kvantilové modelovanie rozdelenia pravdepodobnosti, rozvíjané vo svete hlavne v posledných desaťročiach, je náročné z teoretického aj aplikačného hľadiska. Teoretické poznatky sú rozpracované v mnohých, takmer výlučne zahraničných odborných článkoch, no len v niekoľkých súhrnnejších knižných publikáciách. Štatistické programové balíky zatiaľ neobsahujú zodpovedajúce programy pre náročné výpočtové postupy kvantilového modelovania.

Literatúra

- [1] AZUDOVÁ, L.: Vývoj sociálnej štruktúry a postavenie domácností v transformačnom procese. Ekonomický časopis/Journal of economics, 53, 2005, č. 1, s. 79 – 91.
- [2] ARNOLD, B. C. – BALAKRISHNAN, N. – NAGARAJA, H. N.: A First Course in Order Statistics. New York: John Wiley and Sons 1992.
- [3] BALAKRISHNAN, N. – COHEN, A. C.: Order Statistics and Inference. San Diego: Academic Press 1991.
- [4] CHAMPERNOWNE, D. G.: A Model of Income Distribution. Economic Journal, 63, 1953, jún, s. 318 – 351.
- [5] CHENG, Ch. – PARZEN, E.: Unified Estimators of Smooth Quantile and Quantile Density Functions. Journal of Statistical Planning and Inference, 59, 1997, s. 291 – 307.
- [6] CHIPMAN, J. S.: The Theory and Measurement of Income Distribution. Advances in Econometrics, 4, 1985, s. 135 – 165.

- [7] DAGUM, C.: Income Distribution Models. The Encyclopedia of Statistical Science, 4, 1984, s. 27 – 34.
- [8] FOWLKES, E. B.: A Folio of Distributions, A Collection of Theoretical Q-Q Plots. New York: Marcel Dekker 1987.
- [9] GILCHRIST, W.G.: Modelling with Quantile Distribution Functions. Journal of Applied Statistics, 24, 1997, č. 1, s. 113 – 122.
- [10] GILCHRIST, W.G.: Statistical Modelling with Quantile Functions. London: Chapman & Hall 2000.
- [11] MOORS, J. J. A – WAGEMAKERS, R. Th. A. – COENEN, V. M. J. – HEUTS, R. M. J. – JANSSENS, M. J. B. T.: Charakterizing Systems of Distributions by Quantile Measures. Statistica Neerlandica, 50, 1996, č. 3, s. 417 – 430.
- [12] OZTÜRK, A. – DALE, R. F.: A Study of Fitting the Generalized Lambda Distribution to Solar Radiation Data. Journal Appl. Meteor., 21, 1982. s. 995.
- [13] PACÁKOVÁ, V.: Aplikovaná poistná štatistika. Bratislava: ELITA 2000.
- [14] PACÁKOVÁ, V. a kol.: Štatistika pre ekonómov. Bratislava: IURA EDITION 2003.
- [15] PACÁKOVÁ, V. – SODOMOVÁ, E.: Modelling with Quantile Distribution Functions. Ekonomika a informatika, 2003, č. 1, s. 30 – 44.
- [16] PARZEN, E.: Nonparametric Statistical Data Modelling. Journal of Amer. Statist. Assoc., 74, 1979, s. 105 – 121.
- [17] RAMBERG, J. – DUDEWICZ, E. – TADIKAMALLA, P. – MYKYTKA, E.: A Propability Distribution and its Uses in Fitting Data. Technometrics, 21, 1979 č. 2, s. 201 – 214.
- [18] SIPKOVÁ, E.: Štatistická analýza rozdelenia príjmov domácností v Slovenskej republike. [Doktorská dizertačná práca.] Bratislava: Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave 2004.
- [19] SIPKOVÁ, E.: Zovšeobecnené lambda rozdelenie a odhad jeho parametrov. Ekonomika a informatika, 2004, č. 1, s. 107 – 128.
- [20] SODOMOVÁ, E.: Štatistická analýza úrovne, variability a diferenciácie miezd zamestnancov v Slovenskej republike. Slovenská štatistika a demografia, 1998, č. 4, s. 48 – 56.

Príloha

Výberové charakteristiky rozdelenia príjmov

Výstup kvantitatívnej empirickej analýzy zo systému STATGRAPHICS

Count = 1566
 Average = 244284,0
 Median = 237407,0
 Mode =
 Variance = 1,20118E10
 Standard deviation = 109598,0
 Minimum = 54325,0
 Maximum = 1,0206E6
 Range = 966275,0

Quantiles for CP

1.0 % = 65235,0
 5.0 % = 85176,0
 10.0 % = 107901,0
 25.0 % = 166663,0
 50.0 % = 237407,0
 75.0 % = 306433,0
 90.0 % = 377556,0
 95.0 % = 432913,0
 99.0 % = 584366,0

Lower quartile = 166663,0
 Upper quartile = 306433,0
 Interquartile range = 139770,0
 Skewness = 1,04975
 Std. skewness = 16,9592
 Kurtosis = 2,99699
 Std. kurtosis = 24,2089
 Coeff. of variation = 44,8652%
 Sum = 3,82549E8

Octiles for CP

12.5 % = 125460,0
 25.0 % = 166663,0
 37.5 % = 202124,0
 50.0 % = 237407,0
 62.5 % = 270816,0
 75.0 % = 306433,0
 87.5 % = 358387,0
 95.0 % = 432913,0
 99.0 % = 584366,0