

MODELOVÁNÍ MZDOVÝCH ROZDĚLENÍ POSLEDNÍCH LET V ČESKÉ REPUBLICE S VYUŽITÍM L-MOMENTŮ A PREDIKCE MZDOVÝCH ROZDĚLENÍ PODLE ODVĚTVÍ

Diana Bílková

Úvod

Užitečnost odhadů mzdových rozdělení v podmínkách tržního hospodářství je poutána metodami umožňujícími odhadovat spotřebu obyvatelstva či poptávku obyvatelstva po zboží a službách, viz [3]. Je však třeba poznamenat, že predikce poptávky není zdaleka jediným účelem modelování mzdových rozdělení. Užitečnost odhadů vývoje mzdových rozdělení umožňuje spojit úvahy o úrovni a diferenciaci mezd s úvahami sociálně politickými. Znalostí modelů mzdových rozdělení se dále využívá např. při hodnocení životní úrovně obyvatelstva či při mezioblastním či mezinárodním srovnávání životní úrovně. V této souvislosti nestačí sledovat pouze úroveň mezd a mzdovou diferenciaci vyzkoušenou některou vhodnou mírou variability, ale podrobné hodnocení mzdové stránky životní úrovně obyvatelstva vyžaduje znalost úplného rozdělení mezd v příslušném období směřující k odhadu podílů pracovníků s nízkými, středními a vysokými mzdami, zejména potom k odhadu podílu osob se mzdami pod určitou hranicí, viz [11].

V oblasti statistiky se můžeme setkat s dalším nepřímým využitím znalosti mzdových rozdělení. Jedná se o zdokonalení postupů při výběrovém zjišťování týkajícím se znaku, který se mzdami značně korelován. Jednat se může např. o výdaje jednotlivců a jejich domácností, ale i vybavenost jejich domácností, využití času, nákupní úmysly a znaky vyjadřující názory a postoje jednotlivců při různých sociologických výzkumech, viz [5].

Běžně používaný statistický postup při popisu pozorovaných statistických souborů

spočívá ve využití jejich konvenčních momentů či kumulantů. Rovněž při výběru vhodného parametrického rozdělení pro daný datový soubor se obvykle odhadují parametry tohoto parametrického rozdělení s využitím momentové metody spočívající ve vytvoření rovnic, ve kterých se výběrové konvenční momenty položí do rovnosti odpovídajícím momentům příslušného teoretického rozdělení. Nicméně momentová metoda odhadu parametrů není vždy vyhovující, a to především v případě malých výběrů. Z [6] plyne, že momentová metoda odhadu parametrů je často méně přesná než jiné metody odhadu parametrů, jakou je např. metoda maximální věrohodnosti.

Alternativní přístup je založen na využití dalších charakteristik, které nazýváme L-momenty, které jsou analogií ke konvenčním momentům, ale jsou založeny na lineárních kombinacích pořádkových statistik., tj. L-statistik. Použití L-momentů je teoreticky výhodnější oproti konvenčním momentům spočívající v tom, že L-momenty charakterizují širší rozpětí rozdělení. Při odhadování z výběru jsou více robustní k přítomnosti odlehklých hodnot v datech. Zkušenosti dále ukazují, že ve srovnání s konvenčními momenty jsou L-momenty méně náchylné ke zkreslení odhadu a v konečných výběrech se přibližují více asymptotickému normálnímu rozdělení. Odhady parametrů získané s využitím L-momentů jsou především v případě malých výběrů často dokonce přesnější než odhady parametrů pořízené metodou maximální věrohodnosti, viz [2].

Tento text se zabývá využitím L-momentů v případě větších datových souborů a porovnáním přesnosti metody L-momentů s přesností

dalších metod (momentová a kvantilová metoda) odhadu parametrů pravděpodobnostních parametrických rozdělení v případě větších datových souborů. Na základě těchto analýz využívajících tříparametrické logaritnicko-normální rozdělení byly vytvořeny předpovědi mzdových rozdělení na období let finanční krize, která nastoupila na podzim roku 2008, ale za předpokladu zachování dosavadního vývoje. Jedná se o mzdová rozdělení podle jednotlivých odvětví. Data pro tento výzkum pocházejí od Českého statistického úřadu a jedná se o „Podíly zaměstnanců v pásmech hrubých měsíčních mezd podle odvětví“ za období let 2002 až 2008 v České republice a údaje o rozsazích výběrů počtů zaměstnanců jednotlivých odvětví v České republice v daném období. Výchozí data představovala 84 souborů mzdových rozdělení ve formě intervalového rozdělení četností s krajními otevřenými intervaly.

Zkoumanou proměnnou je hrubá měsíční mzda v Kč. Data byla zpracovávána s využitím programu Microsoft Excel a statistických programů SAS a Statgraphics.

1. Metodologie

Cílem daného příspěvku není konstrukce nových metod, nýbrž zpracování dat z oblasti mzdových rozdělení s využitím již známých metod popsaných např. v [8] či v [7], [9] a [10], viz kapitoly 1.1 až 1.3.

1.1 L-momenty pravděpodobnostních rozdělení

Budeme předpokládat, že X je reálná náhodná veličina s distribuční funkcí $F(x)$ a kvantilovou funkcí $x(F)$ a $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ jsou pořádkové statistiky náhodného výběru rozsahu n pořízeného z rozdělení veličiny X . Potom r -tý L-moment náhodné veličiny X je definován

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r=1, 2, 3, \dots \quad (1)$$

Symbol „L“ v názvu „L-momenty“ je pro zdůraznění, že r -tý L-moment λ_r je lineární funkcí očekávaných pořádkových statistik. Přirozený odhad L-momentu λ_r založený na pozorovaném

výběru dat je navíc lineární kombinace uspořádaných datových hodnot, tj. tzv. L-statistika. Střední hodnota pořádkové statistiky má tvar

$$EX_{j:r} = \frac{r!}{(j-1)! \cdot (r-j)!} \int_0^1 x [F(x)]^{j-1} \cdot [1-F(x)]^{r-j} dF(x). \quad (2)$$

Dosadíme-li vztah (2) do rovnice (1), získáme po úpravě

$$\lambda_r = \int_0^1 x(F) P_{r-1}^*(F) dF, \quad r=1, 2, 3, \dots, \quad (3)$$

kde

$$P_r^*(F) = \sum_{k=0}^r p_{r,k}^* F^k, \quad (4)$$

a

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}. \quad (5)$$

Výraz $P_r^*(F)$ představuje r -tý posunutý Legendův polynom mající vztah k obvyklým Legendovým polynomům. Posunutý Legendovy

polynomy jsou ortogonální na intervalu $(0; 1)$ s konstantní váhovou funkcí. První čtyři L-momenty mají tvar

$$\lambda_1 = EX_{1:1} = \int_0^1 x(F) dF, \tag{6}$$

$$\lambda_2 = \frac{1}{2}(EX_{2:2} - EX_{1:2}) = \int_0^1 x(F) \cdot (2F - 1) dF, \tag{7}$$

$$\lambda_3 = \frac{1}{3}(EX_{3:3} - 2EX_{2:3} + EX_{1:3}) = \int_0^1 x(F) \cdot (6F^2 - 6F + 1) dF, \tag{8}$$

$$\lambda_4 = \frac{1}{4}(EX_{4:4} - 3EX_{3:4} + 3EX_{2:4} - EX_{1:4}) = \int_0^1 x(F) \cdot (20F^3 - 30F^2 + 12F - 1) dF. \tag{9}$$

Podrobněji se L-momentům věnuje např. [8] nebo [10]. Definovány jsou tzv. koeficienty L-momentů, pro které platí

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, 5, \dots \tag{10}$$

L-momenty $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_r$ a koeficienty L-momentů $\tau_1, \tau_2, \tau_3, \dots, \tau_r$ lze použít jako charakteristiky rozdělení. L-momenty jsou v určitém směru podobné konvenčním centrálním momentům a koeficienty L-momentů jsou analogií k momentovým poměrům, viz [12] a [15]. Zejména L-momenty λ_1 a λ_2 a koeficienty L-momentů τ_3 a τ_4 jsou považovány za charakteristiky po řadě polohy, variability, šikmosti a špičatosti.

S využitím vztahů (6) až (8) a vztahu (10) získáváme první tři L-momenty tříparametrického logaritnicko-normálního rozdělení $LN(\mu, \sigma^2, \xi)$, které je popsáno např. v [4] nebo [6]. Pro tyto L-momenty platí

$$\lambda_1 = \xi + \exp\left(\mu + \frac{\sigma^2}{2}\right), \tag{11}$$

$$\lambda_2 = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \operatorname{erf}\left(\frac{\sigma}{2}\right), \tag{12}$$

$$\tau_3 = \frac{6\pi^{-1/2}}{\operatorname{erf}\left(\frac{\sigma}{2}\right)} \cdot \int_0^{\sigma/2} \operatorname{erf}\left(\frac{x}{\sqrt{3}}\right) \cdot \exp(-x^2) dx, \tag{13}$$

kde $\operatorname{erf}(z)$ je tzv. chybová funkce tvaru

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \cdot \int_0^z e^{-t^2} dt. \tag{14}$$

1.2 Výběrové L-momenty

Budeme předpokládat, že x_1, x_2, \dots, x_n je náhodný výběr a $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ je uspořádaný výběr. Potom r -tý výběrový L-moment je definován

$$l_r = \binom{n}{r}^{-1} \cdot \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_r \leq n} r^{-1} \cdot \sum_{k=0}^{r-1} (-1)^k \cdot \binom{r-1}{k} \cdot x_{i_r-k:n}, \quad r = 1, 2, \dots, n. \tag{15}$$

Speciálně pro první čtyři výběrové L-momenty platí

$$l_1 = n^{-1} \cdot \sum_i x_{i:n}, \quad (16)$$

$$l_2 = \frac{1}{2} \cdot \binom{n}{2}^{-1} \cdot \sum_{i>j} \sum_j (x_{i:n} - x_{j:n}), \quad (17)$$

$$l_3 = \frac{1}{3} \cdot \binom{n}{3}^{-1} \cdot \sum_{i>j>k} \sum_j \sum_k (x_{i:n} - 2x_{j:n} + x_{k:n}), \quad (18)$$

$$l_4 = \frac{1}{4} \cdot \binom{n}{4}^{-1} \cdot \sum_{i>j>k>l} \sum_j \sum_k \sum_l (x_{i:n} - 3x_{j:n} + 3x_{k:n} - x_{l:n}). \quad (19)$$

Výběrové L-momenty mohou být využity obdobně jako konvenční výběrové momenty, neboť charakterizují základní vlastnosti výběrového rozdělení, tj. polohu, variabilitu, šikmost i špičatost, a odhadují odpovídající vlastnosti pravděpodobnostního rozdělení ze kterého výběr pochází. Mohou být tedy použity k odhadu parametrů tohoto výchozího rozdělení. V těchto případech jsou L-momenty často upřednostňovány před konvenčními momenty, neboť jakožto lineární funkce dat jsou méně citlivé na výběrovou variabilitu či velikost chyb v případě extrémních hodnot v datech než konvenční momenty, a proto se počítá s tím, že poskytují přesnější a robustnější odhady charakteristik nebo parametrů základního pravděpodobnostního rozdělení.

1.3 Odhad parametrů

Označíme-li distribuční funkci normovaného normálního rozdělení Φ , potom Φ^{-1} představuje kvantilovou funkci normovaného normálního rozdělení. Pro distribuční funkci tříparametrického logaritmicke-normálního rozdělení rozdělení $\text{LN}(\mu, \sigma^2, \xi)$ platí

$$F = \Phi \left[\frac{\ln(x - \xi) - \mu}{\sigma} \right]. \quad (20)$$

Koeficienty L-momentů (10) se obvykle odhadují pomocí

$$t_r = \frac{l_r}{l_2}, \quad r = 3, 4, 5, \dots \quad (21)$$

Odhady parametrů tříparametrického logaritmicke-normálního rozdělení nyní pořídíme jako

$$z = \sqrt{\frac{8}{3}} \cdot \Phi^{-1} \left(\frac{1+t_3}{2} \right), \quad (22)$$

$$\hat{\sigma} \approx 0,999\,281z - 0,006\,118z^3 + 0,000\,127z^5, \quad (23)$$

$$\hat{\mu} = \ln \left[\frac{l_2}{\text{erf} \left(\frac{\sigma}{2} \right)} \right] - \frac{\hat{\sigma}^2}{2}, \quad (24)$$

$$\hat{\xi} = l_1 - \exp \left(\hat{\mu} + \frac{\hat{\sigma}^2}{2} \right), \quad (25)$$

více např. v [2].

1.4 Vhodnost zkonstruovaného modelu

Při posuzování vhodnosti zkonstruovaného modelu je třeba využít některého z kritérií, jakým může být např. součet absolutních odchylek pozorovaných a teoretických četností za všechny intervaly S , případně známé kritérium χ^2 , viz [6]. Otázka vhodnosti dané křivky jako modelu mzdových rozdělení při takto velkých rozsazích výběrů, s jakými se v případě mzdových rozdělení

setkáváme, je vysvětlena v [6]. Určitou představu z hlediska přesnosti dané metody odhadu parametrů může vnést graf představující vývoj výběrového mediánu ve sledovaném období a mediánu teoretického rozdělení při použití dané metody odhadu parametrů, viz Obr. 11, kde si můžeme např. všimnout, že průběh mediánu teoretického tříparametrického logaritnicko-normálního rozdělení, jehož parametry byly odhadnuty kvantilovou metodou odhadu parametrů, splývá s průběhem výběrového mediánu v daném období.

2. Výstupy

V Tabulkách 1–2 jsou uvedeny první tři výběrové L-momenty pro jednotlivá mzdová rozdělení let 2002–2008 za odvětví „Zemědělství a rybolov“ a za odvětví „Průmysl“, odhadnuté hodnoty parametrů tříparametrického logaritnicko-normálního

rozdělení metodou L-momentů ke každému z těchto rozdělení. Odhadnutá hodnota parametru ξ je v řadě případů záporná. Zpočátku svého průběhu se tedy logaritnicko-normální křivka dostává do záporných hodnot. Protože však zpočátku svého průběhu má tato křivka velice těsný kontakt s vodorovnou osou, nemusí být záporné hodnoty parametru ξ na úkor dobré shody modelu se skutečností. V takovémto případě nelze ale parametru ξ přisuzovat žádnou interpretaci. V Tabulkách 1–2 dále nalezneme predikce hodnot prvních tří výběrových L-momentů na roky 2009 a 2010 získané na základě předpokladu zachování dosavadního vývoje mzdových rozdělení a na základě těchto predikcí odhadnuté hodnoty parametrů tříparametrického logaritnicko-normálního rozdělení pro roky 2009 a 2010 podle odvětví.

Tab. 1: Výběrové L-momenty a odhadnuté parametry tříparametrického logaritnicko-normálního rozdělení metodou L-momentů za odvětví „Zemědělství a rybolov“

Rok	Výběrové L-momenty			Odhadnuté parametry		
	l_1	l_2	l_3	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\xi}$
2002	13 793,44	2 802,19	835,70	8,820 749	0,388 624	5 567,394 9
2003	14 318,48	2 949,76	908,52	8,826 973	0,415 751	5 928,041 2
2004	15 246,70	3 147,21	1 006,26	8,838 238	0,449 734	6 615,790 7
2005	15 921,75	3 250,71	1 042,09	8,866 779	0,452 233	7 029,863 3
2006	17 372,84	3 498,06	1 173,87	8,872 590	0,498 093	8 221,676 3
2007	19 321,90	4 096,13	1 409,25	8,992 663	0,525 123	8 862,828 9
2008	20 460,99	4 311,81	1 226,74	9,316 088	0,352 241	7 204,946 5
2009	22 614,00	4 837,10	1 420,66	9,387 782	0,376 389	8 199,698 0
2010	24 803,90	5 375,18	1 504,31	9,558 609	0,340 382	8 009,509 2

Zdroj: vlastní výpočty

Tab. 2: Výběrové L-momenty a odhadnuté parametry tříparametrického logaritnicko-normálního rozdělení metodou L-momentů za odvětví „Průmysl“

Rok	Výběrové L-momenty			Odhadnuté parametry		
	l_1	l_2	l_3	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\xi}$
2002	17 170,55	4 068,31	1 181,06	9,230 592	0,367 373	4 908,336 2
2003	18 272,83	4 267,22	1 151,73	9,375 726	0,315 710	4 456,839 5
2004	19 186,39	4 608,13	1 435,52	9,256 943	0,425 785	6 223,637 8
2005	20 041,87	4 798,90	1 470,66	9,320 713	0,411 403	6 324,553 4
2006	21 126,08	4 998,89	1 551,02	9,344 033	0,422 220	7 008,961 8
2007	22 937,41	5 775,70	2 133,64	9,224 189	0,611 250	9 173,375 6
2008	24 635,76	6 190,73	2 351,80	9,247 609	0,649 126	10 276,173 6
2009	26 423,10	6 823,06	2 812,72	9,203 193	0,775 110	11 794,293 8
2010	28 474,70	7 523,44	3 314,78	9,176 112	0,897 438	13 338,704 5

Zdroj: vlastní výpočty

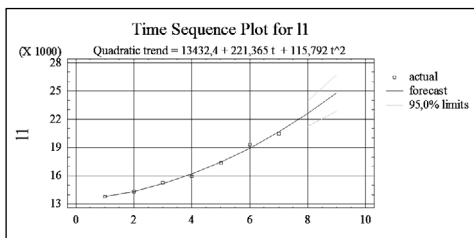
Předpovědi prvních tří výběrových L-momentů příjmových rozdělení byly konstruovány na základě trendové analýzy, která je podrobněji rozebrána např. v [1]. Na základě interpolačních kritérií, především potom MSE (střední kvadratická chyba odhadu) a MAPE (střední absolutní procentní chyba odhadu) byla vybrána nejvhodnější trendová funkce. V naprosté většině případů byla takto zvolena kvadratická trendová funkce, případně lineární trendová funkce, výjimečně trendová funkce exponenciální, viz také [13] a [14]. Jistou představu o průběhu takto zkonstruovaných trendových funkcí prvních tří výběrových L-momentů pro odvětví „Zemědělství a rybolov“ a pro odvětví „Průmysl“ poskytují Obrázky 1–6 (včetně předpovědí na následující dva roky). Z těchto obrázků je kromě toho možné získat určitou představu o tom, jak přesně daná trendová funkce příslušnou časovou řadu výběrových L-momentů vystihuje. Parametry trendové funkce byly odhadnuty metodou nejmenších čtverců, viz opět Obrázky 1–6 pro odvětví „Zemědělství a rybolov“ a pro odvětví „Průmysl“. Na základě takto zkonstruovaných trendových funkcí byly vytvořeny predikce prvních tří výběrových L-momentů na následující dva roky dopředu, tj. pro roky 2009 a 2010. S využitím vztahů (22) až (25) byly odhadnuty parametry mzdových rozdělení pro další dva roky 2009 a 2010.

Jistou představu o odhadnutých hustotách pravděpodobnosti modelovaných mzdových rozdělení a jejich předpovědi poskytují Obrázky 7–10 pro odvětví „Finanční zprostředkovatelství“ – tj. odvětví s nejvyšší úrovní mezd, pro odvětví „Stavebnictví“ – tj. odvětví se střední úrovní mezd a dále pro dvě v poslední době z hlediska mezd hodně diskutovaná odvětví, a to odvětví „Vzdělávání“ a „Zdravotnictví“. Z těchto obrázků je rovněž patrný vývoj mzdových rozdělení sledovaného období posledních let i vývoj charakteristik polohy a diferenciacie mzdových rozdělení. Je zde patrný růst úrovně mzdových rozdělení v letech 2002–2008 doprovázený růstem diferenciacie mzdových rozdělení, sílí zde skupiny s výrazně vysokými mzdami. S tím souvisí rostoucí šikmost mzdových rozdělení (rozdělení se protahují stále více doprava). Klesá špičatost těchto rozdělení.

Na základě takto vytvořených modelů mzdových rozdělení pro roky 2009 a 2010 byly zkonstruovány předpovědi mzdových rozdělení pro roky 2009 a 2010 podle odvětví v Tabulkách 3 a 4 (pomocí distribuční funkce tříparametrického logaritnicko-normálního rozdělení). Je třeba si však uvědomit, že tyto předpovědi mzdových rozdělení byly konstruovány za předpokladu zachování stávajícího vývoje mzdových rozdělení. Je otázka, do jaké míry zasáhnou do vývoje mzdových rozdělení důsledky finanční krize mající svůj počátek právě na podzim posledního roku zkoumaného období, tj. roku 2008. Může se paradoxně například stát, že dojde k rychlejšímu narůstání úrovně mezd, ale způsobenému tím, že v období finanční krize se do skupiny nezaměstnaných dostávali z velké části především pracovníci s velmi nízkými mzdami. „Podíly zaměstnanců v pásmech hrubých měsíčních mezd“ v České republice, mimo jiné i podle odvětví, budou zveřejněny na oficiálních stránkách Českého statistického úřadu přibližně na přelomu listopadu a prosince roku 2010. Porovnáním těchto výsledků s predikcemi mzdových rozdělení lze do určité míry získat představu o vlivu finanční krize na vývoj rozdělení mezd.

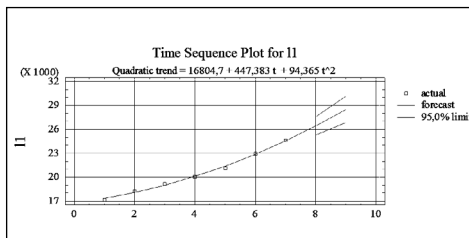
Přesnost použitých metod odhadu parametrů byla ověřena pomocí kritéria χ^2 . Hodnoty tohoto testového kritéria vedou ve všech případech mzdových rozdělení k zamítnutí testované hypotézy o předpokládaném logaritnicko-normálním rozdělení, jak je v případě mzdových a příjmových rozdělení obvyklé. Při výpočtu součtu absolutních odchylek S a testového kritéria χ^2 bylo patrné, že nepřesnosti vznikají především na obou koncích rozdělení („těžké konce“), pokud bychom neuvažovali krajní (otevřené) intervaly, potom by metoda L-momentů (ale i kvantilová a momentová metoda) přinesla podstatně přesnější výsledky. Přesnost metody L-momentů byla zde porovnávána s momentovou metodou odhadu parametrů a s kvantilovou metodou. Celkově lze z těchto výsledků konstatovat, že v případě takto velkých rozsahů výběrů, s jakými se v případě mzdových rozdělení setkáváme, přinesla metoda L-momentů srovnatelně přesné výsledky jako momentová a kvantilová metoda.

Obr. 1: Trendová funkce 1. výběrového L-momentu – „Zemědělství a rybolov“



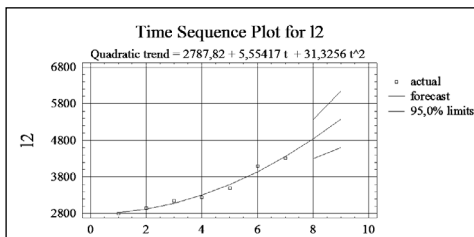
Trend = $13432,4 + 221,365 t + 115,792 t^2$
Zdroj: vlastní výpočty

Obr. 4: Trendová funkce 1. výběrového L-momentu – „Průmysl“



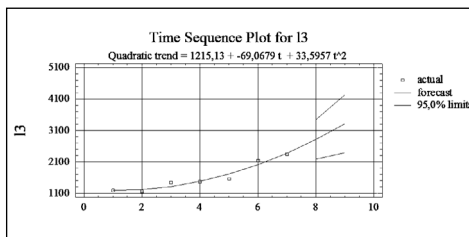
Trend = $16804,7 + 447,383 t + 94,365 t^2$
Zdroj: vlastní výpočty

Obr. 2: Trendová funkce 2. výběrového L-momentu – „Zemědělství a rybolov“



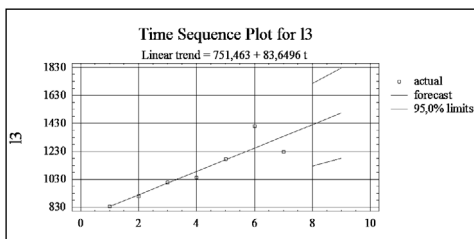
Trend = $2787,82 + 5,55417 t + 31,3256 t^2$
Zdroj: vlastní výpočty

Obr. 5: Trendová funkce 2. výběrového L-momentu – „Průmysl“



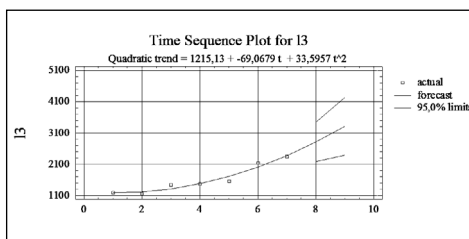
Trend = $4030,21 + 36,8645 t + 39,0302 t^2$
Zdroj: vlastní výpočty

Obr. 3: Trendová funkce 3. výběrového L-momentu – „Zemědělství a rybolov“



Trend = $751,463 + 83,6496 t$
Zdroj: vlastní výpočty

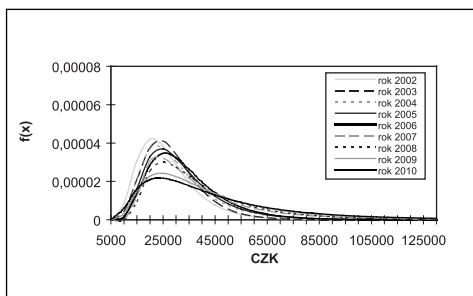
Obr. 6: Trendová funkce 3. výběrového L-momentu – „Průmysl“



Trend = $1215,13 - 69,0679 t + 33,5957 t^2$
Zdroj: vlastní výpočty

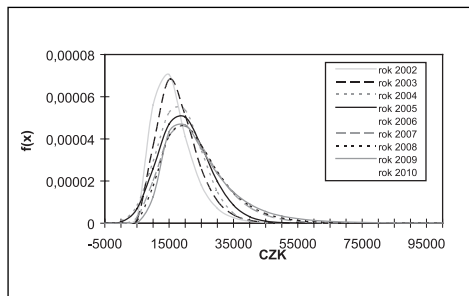
Obr. 7:

Hustoty pravděpodobnosti
logaritmicko-normálních křivek
s parametry odhadnutými
metodou L-momentů za odvětví
„Finanční zprostředkovatelství“



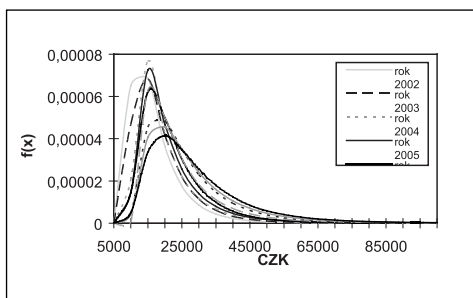
Obr. 9:

Hustoty pravděpodobnosti
logaritmicko-normálních křivek
s parametry odhadnutými
metodou L-momentů za odvětví
„Vzdělávání“



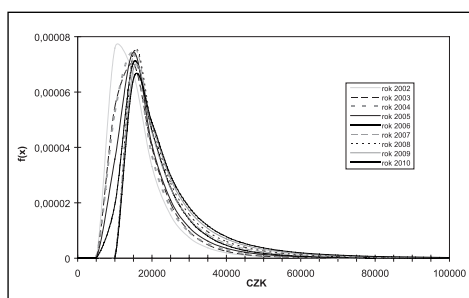
Obr. 8:

Hustoty pravděpodobnosti
logaritmicko-normálních křivek
s parametry odhadnutými
metodou L-momentů za odvětví
„Stavebnictví“



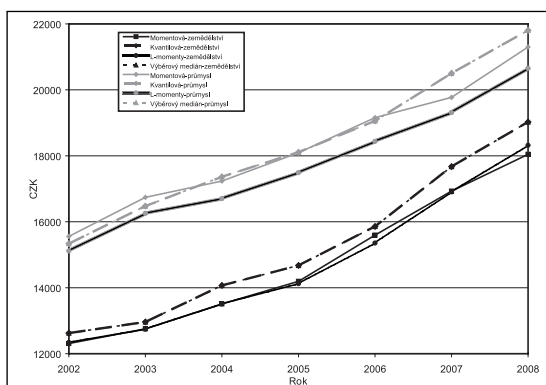
Obr. 10:

Hustoty pravděpodobnosti
logaritmicko-normálních křivek
s parametry odhadnutými
metodou L-momentů za odvětví
„Zdravotnictví“



Obr. 11:

Vývoj teoretického mediánu při použití momentové a kvantilové metody
a metody L-momentů při odhadu parametrů tříparametrického logaritmicko-
normálního rozdělení a vývoj výběrového mediánu v České republice v letech
2002–2008 pro odvětví „Zemědělství a rybolov“ a pro odvětví „Průmysl“



Zdroj: vlastní výpočty

Tab. 3: Odhadnuté podíly zaměstnanců v pásmech hrubých měsíčních mezd podle odvětví v roce 2009 (v %)

Intervaly mezd	Odvětví												
	Zem.	Prům.	Stav.	Obch.	Ubyt.	Dopr.	Fin.	Nem.	Veř.	Vzd.	Zdr.	Ost.	
do 14 000	11,96	4,37	8,07	6,60	41,54	0,28	3,58	22,62	2,28	12,14	11,70	14,56	
14 001 – 16 000	12,42	12,09	8,07	16,41	8,40	7,33	3,03	8,57	5,14	8,63	13,52	14,30	
16 001 – 18 000	12,99	13,21	9,21	13,67	7,64	12,56	3,79	7,68	7,92	9,52	12,64	12,23	
18 001 – 20 000	11,86	11,76	9,30	10,55	6,76	12,58	4,33	6,74	9,51	9,54	10,69	9,89	
20 001 – 22 000	10,09	9,82	8,77	8,18	5,86	10,97	4,66	5,87	9,93	8,97	8,75	7,91	
22 001 – 23 000	4,36	4,22	4,08	3,41	2,61	4,79	2,40	2,64	4,85	4,16	3,73	3,35	
23 001 – 24 000	3,92	3,81	3,85	3,04	2,40	4,35	2,42	2,47	4,68	3,92	3,36	3,00	
24 001 – 25 000	3,51	3,43	3,62	2,72	2,21	3,94	2,42	2,30	4,46	3,67	3,02	2,70	
25 001 – 26 000	3,13	3,09	3,38	2,44	2,03	3,56	2,41	2,15	4,22	3,41	2,72	2,43	
26 001 – 27 000	2,79	2,79	3,15	2,20	1,85	3,22	2,39	2,01	3,95	3,16	2,45	2,20	
27 001 – 28 000	2,48	2,52	2,92	2,00	1,69	2,91	2,36	1,87	3,68	2,91	2,21	1,99	
28 001 – 29 000	2,20	2,28	2,71	1,81	1,55	2,64	2,32	1,75	3,41	2,68	2,00	1,81	
29 001 – 30 000	1,96	2,06	2,50	1,65	1,41	2,39	2,28	1,64	3,15	2,46	1,81	1,64	
30 001 – 32 000	3,28	3,57	4,44	2,89	2,45	4,14	4,40	2,98	5,56	4,30	3,13	2,86	
32 001 – 34 000	2,59	2,95	3,78	2,44	2,03	3,42	4,17	2,63	4,65	3,57	2,59	2,39	
34 001 – 36 000	2,04	2,46	3,20	2,07	1,67	2,85	3,92	2,32	3,87	2,96	2,15	2,02	
36 001 – 38 000	1,62	2,06	2,71	1,78	1,38	2,39	3,67	2,06	3,21	2,44	1,80	1,71	
38 001 – 40 000	1,28	1,73	2,3	1,54	1,13	2,01	3,42	1,83	2,65	2,01	1,52	1,46	
40 001 – 50 000	3,46	5,49	7,24	5,27	3,28	6,36	13,65	6,67	7,70	5,80	4,79	4,82	
50 001 – 60 000	1,22	2,66	3,30	2,95	1,26	3,08	9,10	4,04	2,99	2,21	2,31	2,52	
60 001 – 80 000	0,67	2,20	2,38	2,98	0,70	2,55	10,03	4,29	1,73	1,24	1,90	2,30	
80 001 – 100 000	0,13	0,78	0,65	1,37	0,13	0,90	4,55	2,03	0,34	0,23	0,66	0,92	
100 001 a více	0,66	0,34	2,03	2,03	0,77	4,72	4,70	0,11	0,07	0,54	0,99	0,99	
Celkem	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	
Průměrné mzdy	22 614	26 423	27 366	28 434	18 524	28 200	43 456	30 313	28 082	24 920	24 780	25 393	
Směr. odchylka	9 744	15 829	14 286	26 950	11 782	16 392	29 397	29 451	11 783	11 614	15 121	18 690	
Variač. koeficient	0,43	0,60	0,52	0,95	0,64	0,58	0,68	0,97	0,42	0,47	0,61	0,74	

Zdroj: vlastní výpočty

Tab. 4: Odhadnuté podíly zaměstnanců v pásmech hrubých měsíčních mezd podle odvětví v roce 2010 (v %)

Intervaly mezd	Odvětví													Zdr.	Vzd.	Ost.
	Zem.	Prům.	Stav.	Obch.	Ubyt.	Dopr.	Fin.	Nem.	Veř.	Vzd.	Zdr.	Ost.				
do 14 000	7,01	0,23	7,09	0,69	35,38	0,00	5,12	23,19	0,80	8,60	8,61	6,30				
14 001 – 16 000	9,31	8,44	6,70	12,95	7,39	3,02	3,19	8,22	4,53	9,27	12,75	14,30				
16 001 – 18 000	11,15	13,41	7,89	14,06	7,08	10,93	3,71	7,28	8,34	10,33	12,43	13,08				
18 001 – 20 000	11,28	12,65	8,29	11,45	6,60	12,51	4,07	6,37	10,22	10,11	10,73	10,67				
20 001 – 22 000	10,40	10,67	8,12	9,08	6,02	11,34	4,27	5,54	10,46	9,23	8,91	8,54				
22 001 – 23 000	4,71	4,59	3,89	3,82	2,78	5,02	2,17	2,50	5,01	4,20	3,84	3,62				
23 001 – 24 000	4,36	4,14	3,73	3,42	2,62	4,59	2,17	2,33	4,76	3,91	3,47	3,25				
24 001 – 25 000	4,01	3,73	3,56	3,08	2,46	4,17	2,17	2,18	4,49	3,62	3,14	2,93				
25 001 – 26 000	3,66	3,36	3,38	2,78	2,30	3,79	2,15	2,04	4,19	3,34	2,84	2,64				
26 001 – 27 000	3,33	3,04	3,19	2,51	2,15	3,44	2,13	1,91	3,89	3,07	2,57	2,39				
27 001 – 28 000	3,02	2,74	3,01	2,28	2,01	3,12	2,10	1,79	3,60	2,82	2,33	2,17				
28 001 – 29 000	2,73	2,48	2,82	2,08	1,87	2,84	2,07	1,68	3,31	2,59	2,12	1,98				
29 001 – 30 000	2,47	2,25	2,64	1,90	1,73	2,58	2,03	1,58	3,04	2,37	1,93	1,80				
30 001 – 32 000	4,22	3,90	4,78	3,33	3,09	4,49	3,93	2,88	5,34	4,15	3,36	3,16				
32 001 – 34 000	3,42	3,24	4,15	2,82	2,63	3,74	3,75	2,55	4,46	3,46	2,80	2,66				
34 001 – 36 000	2,76	2,71	3,59	2,41	2,24	3,14	3,55	2,27	3,72	2,89	2,35	2,26				
36 001 – 38 000	2,23	2,28	3,10	2,07	1,89	2,65	3,34	2,03	3,10	2,42	1,98	1,94				
38 001 – 40 000	1,80	1,93	2,67	1,80	1,59	2,24	3,14	1,82	2,58	2,02	1,68	1,67				
40 001 – 50 000	5,01	6,24	8,70	6,19	4,84	7,25	12,90	6,74	7,78	6,20	5,44	5,64				
50 001 – 60 000	1,83	3,14	4,18	3,51	1,97	3,64	9,05	4,22	3,31	2,72	2,73	3,08				
60 001 – 80 000	1,03	2,75	3,15	3,58	1,13	3,16	10,74	4,71	2,24	1,91	2,35	2,99				
80 001 – 100 000	0,20	1,05	0,90	1,66	0,20	1,19	5,39	2,36	0,57	0,51	0,86	1,30				
100 001 a více	1,03	0,48	2,52	2,53	1,14	6,83	3,83	0,28	0,26	0,78	1,66	1,63				
Celkem	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %				
Průměrné mzdy	24 804	28 475	29 182	31 024	20 566	30 184	46 539	32 381	28 800	26 127	26 228	28 211				
Směr. odchylka	10 694	18 247	15 553	29 493	13 255	18 630	35 109	35 504	13 096	13 350	16 933	23 280				
Variač. koeficient	0,43	0,64	0,53	0,95	0,64	0,62	0,75	1,10	0,45	0,51	0,65	0,83				

Zdroj: vlastní výpočty

Závěr

Jedním z nejpoužívanějších rozdělení využívaných při modelování rozdělení mezd a příjmů je logaritmicko-normální rozdělení, a to především rozdělení tříparametrické. K odhadu parametrů tohoto rozdělení je možné využít různých metod, např. momentovou metodu, kvantilovou metodu, Kemsleyovu či Cohenovu metodu odhadu parametrů. Jednou z možností, jak odhadnout parametry tohoto rozdělení, je využití metody L-momentů. Metoda L-momentů většinou poskytuje v případě malých rozsahů výběrů výrazně přesnější výsledky, než jiné metody odhadu parametrů, a to včetně metody maximální věrohodnosti, viz [8]. Ukázalo se však, že v případě tak velkých rozsahů výběrů, s jakými pracujeme v případě mzdových rozdělení, poskytuje metoda L-momentů srovnatelně přesné výsledky jako momentová či kvantilová metoda odhadu parametrů.

Při řešení otázky, která metoda odhadu parametrů tříparametrického logaritmicko-normálního rozdělení je nevhodnější, byla patrná značná závislost hodnoty testového kritéria χ^2 na velikosti rozsahu výběru. Jak je u takto velkých rozsahů výběrů obvyklé, všechny testy vedly k zamítnutí nulové hypotézy o předpokládaném rozdělení, viz výše.

Předpovědi mzdových rozdělení pro jednotlivá odvětví na roky 2009 a 2010 jsou konstruovány za předpokladu zachování dosavadního vývoje mzdových rozdělení. Otázkou zůstává, do jaké míry se projeví vliv finanční krize na vývoj mzdových rozdělení co do jejich úrovně i diferenciací. V rámci finanční krize docházelo např. k značnému propouštění především zaměstnanců s velmi nízkými mzdami. Tato skutečnost se může dokonce projevit např. zvyšováním úrovně mzdové hladiny a zcela určitě ovlivní rovněž vývoj diferenciací mezd.

Podporováno grantem IGA 24/2010 Fakulty informatiky a statistiky Vysoké školy ekonomické v Praze „Analýza vývoje příjmových rozdělení v České republice od roku 1990 do období finanční krize a porovnání tohoto vývoje s vývojem rozdělení příjmů v období finanční krize – podle sociologických skupin, pohlaví, věku, vzdělání, oboru profese a krajů“

Literatura

- [1] ARLT, J., ARLTOVÁ, M. *Ekonomické časové řady*. 1. vyd. Praha: Professional Publishing, 2009. 275 s. ISBN 978-80-86946-85-6.
- [2] ADAMAOWSKI, K. Regional Analysis of Annual Maximum and Partial Duration Flood Data by Nonparametric and L-moment Methods. *Journal of Hydrology*. 2000, Vol. 229, Iss. 3–4, s. 219–231. ISSN 0022-1694.
- [3] BARTOŠOVÁ, J. Logarithmic-Normal Model of Income Distribution in the Czech Republic. *Austrian Journal of Statistics*. 2006, Vol. 35, Iss. 23, s. 215–222. ISSN 1026-597X.
- [4] BÍLKOVÁ, D. Application of Lognormal Curves in Modeling of Wage Distributions. *Journal of Applied Mathematics*. 2008, Vol. 1, Iss. 2, s. 341–352. ISSN 1337-6365.
- [5] BÍLKOVÁ, D. Vývoj příjmových rozdělení v letech 1956–1992 a jejich předpovědi pro rok 1995 a 1997. *Politická ekonomie*. 1995, roč. 43, č. 4, s. 510–531. ISSN 0032-3233.
- [6] BÍLKOVÁ, D. Modelování mzdových rozdělení v České republice v letech 2004 a 2005 s využitím logaritmicko-normálních křivek a křivek Pearsonova a Johnsonova systému. *Statistika*. 2008, roč. 88, č. 2, s. 149–166. ISSN 0322-788X.
- [7] GUTTMAN, N.B. The Use of L-moments in the Determination of Regional Precipitation Climates. *Journal of Climate*. 1993, Vol. 6, Iss. 12, s. 2309–2325. ISSN 0894-8755.
- [8] HOSKING, J.R.M. L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society (Series B)*. 1990, Vol. 52, No. 1, s. 105–124. ISSN 1467-9868.
- [9] HOSKING, J.R.M., WALES, J.R. *Regional frequency analysis: An Approach Based on L-moments*. 1st ed. New York: Cambridge University Press, 1997. 209 s. ISBN 0-521-43045-3.
- [10] KYSELÝ, J., PICEK, J. Regional Growth Curves and Improved Design Value Estimates of Extreme Precipitation Events in the Czech Republic. *Climate Research*. 2007, Vol. 33, No. 3, s. 243–255. ISSN 1616-1572.
- [11] PACÁKOVÁ, V., SIPKOVÁ, L. Generalized Lambda Distributions of Households Incomes. *E+M Ekonomie a Management*. 2007, roč. 10, č. 1, s. 98–107. ISSN 1212-3609.

[12] SMITHERS, J.C., SCHULZE, R.E. A Methodology for the Estimation of Short Duration Design Storms in South Africa Using a Regional Approach Based on L-moments. *Journal of Hydrology*. 2001, Vol. 241, Iss. 1–2, s. 42–52. ISSN 0022-1694.

[13] ŠIMSOVÁ, J. Analýza časové řady míry nezaměstnanosti v regionu Teplice. *E+M Ekonomie a Management*. 2006, roč. 9, č. 1, s. 13–19. ISSN 1212-3609.

[14] TREŠL, J., BLATNÁ, D. Dynamic Analysis of Selected European Stock Markets. *Prague Economic Papers*. 2007, roč. 14, č. 4, s. 291–302. ISSN 1210-0455.

[15] ULRYCH, T.J., VELIS, D.R., WOODBURY, A.D., SACCHI, M.D. L-moments and C-moments. *Stochastic Environmental Research and Risk Assessment*. 2000, Vol. 14, Iss. 1, s. 50–68. ISSN 1436-3240.

[16] ZIMKOVÁ, E., BAROCHOVSKÝ, J. Odhad potenciálního produktu a produkčnej medzery v Slovenských podmienkach. *Politická ekonomie*. 2007, roč. 55, č. 4, s. 473–489. ISSN 0032-3233.

doc. Ing. Diana Bílková, Dr.

Vysoká škola ekonomická

Fakulta informatiky a statistiky

Katedra statistiky a pravděpodobnosti

bilkova@vse.cz

Doručeno redakci: 24. 5. 2010

Recenzováno: 14. 6. 2010, 23. 9. 2010

Schváleno k publikování: 27. 9. 2013

MODELING OF WAGE DISTRIBUTION IN RECENT YEARS IN THE CZECH REPUBLIC USING L-MOMENTS AND THE PREDICTION OF WAGE DISTRIBUTION BY INDUSTRY

Diana Bílková

L-moments are based on the linear combinations of order statistics. The question of L-moments presents a general theory covering the summarization and description of sample data sets, the summarization and description of theoretical distributions, but also the estimation of parameters of probability distributions and hypothesis testing for parameters of probability distributions. L-moments can be defined for any random variable in the case that its mean exists. Within the scope of modelling of wage distributions we currently use the method of conventional moments, the quantile method or the maximum likelihood method. The theory of L-moments parallels the other theories and the main advantage of the method of L-moments over these methods is that L-moments suffer less from impact of sampling variability. L-moments are more robust and they provide more secure results in the case of small samples.

The three-parametric lognormal distribution is one from the most frequent used distributions within the frame of modelling wage and income distribution. In the case of wage distributions we usually work with very large data sets and in such cases the method of L-moments provides say about alike accurate results as for example the method of moment or quantile method. The question of fitness of concrete parametric distribution for model of wage distribution tends to rejection of tested hypothesis about supposed form distribution practically always in the cases of such large samples. In this connection we can see close relationship between sample size and the value of criterion χ^2 , too. The forecasts of wage distributions were constructed based on the observance of previous development. Within the frame of the financial crisis were set free the employees with very low wages above all. The effect of this truth to forecasts of wage distribution will be exactly known in the autumn of this year.

Key Words: *L-moments, linear combination of order statistics, wage distributions, lognormal distribution.*

JEL Classification: *C13, C16, H31.*