

Modelování rozdělení ročních příjmů českých domácností

J. Bartošová¹ M. Forbelská²

¹Katedra managementu informací
Fakulta managementu v Jindřichově Hradci
Vysoká škola ekonomická v Praze

²Ústav matematiky a statistiky
Oddělení aplikované matematiky
Přírodovědecká fakulta MU Brno

Robust 2010
31. ledna – 5. února 2010, Králíky

Zobecňování výsledků šetření příjmů domácností v ČR

- Výběrová šetření Mikrocensus a česká modifikace šetření EU–SILC, Životní podmínky, představují datovou základnu pro získávání informací týkajících se především příjmů českých domácností.
- Důležitou součástí výběrových souborů jsou také kalibrační váhy (veličina PKOEF), které slouží k přepočtu výsledků z výběru na celou populaci a které ovlivňují výsledky prováděných analýz.

Otázkou je:

Jak lze korektně zpracovat kalibrační koeficienty do prováděných analýz?

Konstrukce kalibračních koeficientů domácností v Mikrocensu 2002

Při konstrukci kalibračních koeficientů (veličina PKOEF) byla použita iterační metoda, která minimalizuje rozdíl mezi odhadnutými a přepočítanými výběrovými charakteristikami vybranými pro každý kraj zvlášť, a to s pomocí následujících charakteristik¹:

- počet trvale obydlených bytů
- počet osob bydlících v bytech
- počet důchodců (pracujících i nepracujících)
- počet nezaměstnaných
- počet samostatně činných osob

¹Zdroj: Metodika. Mikrocensus 2002

Konstrukce kalibračních koeficientů domácností v šetření EU-SILC

K výše uvedeným charakteristikám, které byly použity pro přepočítání výsledků z Mikrocensu 2002, přibylo v šetření EU-SILC 2005 ještě další členění domácností, a to podle:

- věku osoby stojící v čele domácnosti
- velikostní skupiny obcí

V šetření také dochází k asi 10%-nímu podhodnocení příjmů, a to proto, že dotazovaní si na všechny své příjmy nevzpomenou a mají snahu udávat nižší příjmy než jaké odpovídají skutečnosti.

Toto zkreslení se velice obtížně kvantifikuje, a proto jsou hodnoty korigovány po porovnání s údaji o průměrných hrubých mzdách. Podobně se postupovalo v případě sociálních dávek, kde uváděné hodnoty naopak překračují skutečnost².

²Někteří respondenti sem nesprávně zahrnují dávky sociální potřeby.

Základní charakteristiky kalibračních vah domácností (EU-SILC 2007)

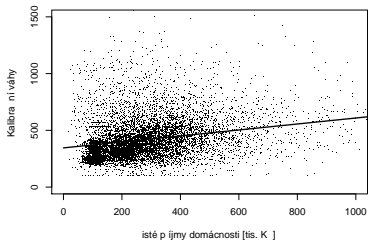
Základní statistické charakteristiky kalibračních vah PKOEF (EU-SILC 2007).

| | | | | |
|---------|------------|-------------|------------|---------|
| minimum | 1. kvartil | medián | 3. kvartil | maximum |
| 100.0 | 294.6 | 369.8 | 493.6 | 3475.0 |
| | průměr | směr. odch. | suma vah | |
| | 417.9 | 205.5 | 4043341 | |

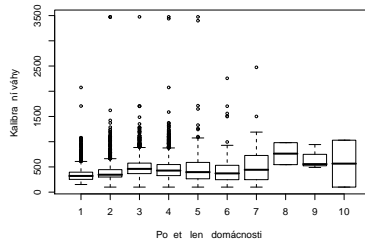
- kalibrační váha PKOEF je (v EU-SILC 2005–2007) vždy větší nebo rovna 100
- součet vah je cca 4 milióny

Ukázka závislosti kalibračních vah na vybraných proměnných (EU–SILC 2007)

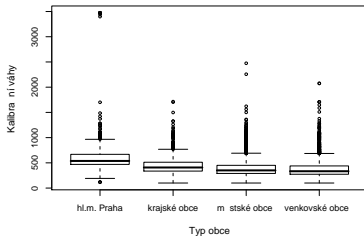
Váhy vs. p íjem



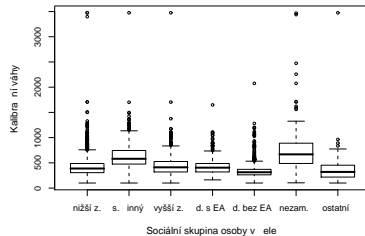
Váhy vs. počet členů domácnosti



Váhy vs. typ obce



Váhy vs. sociální skupina



- **rozdělení příjmů (výdajů) domácností (jednotlivců)**
- **závislosti příjmů (výdajů) domácností (jednotlivců)** na různých demografických a společensko-ekonomických faktorech
- **šancí domácností (jednotlivců)**

Klíčovou otázkou je:

Jak lze korektně zpracovat kalibrační koeficienty do prováděných analýz?

- **normované kalibrační váhy**

$$w_i^* = \frac{w_i}{\sum_{i=1}^n w_i},$$

kde w_i jsou váhy ze souboru EU–SILC (tj. nenormované),

- **výběrový vážený průměr**

$$\bar{x} = \sum_{i=1}^n w_i^* \cdot x_i$$

- **vážený rozptyl**

- maximálně věrohodný odhad

$$s_{\text{ML}}^2 = \sum_{i=1}^n w_i^* \cdot (x_i - \bar{x})^2$$

- nevychýlený odhad výběrového váženého rozptylu

$$s_{\text{UN}}^2 = \frac{n}{n-1} \sum_{i=1}^n w_i^* \cdot (x_i - \bar{x})^2$$

- **vážený $p\%$ -ní kvantil** je nejvyšší hodnota proměnné X (setříděné vzestupně), jejíž kumulativní váha (setříděná a normovaná) je menší nebo rovna p . (V prostředí R na př. v knihovně `Hmisc`)

Porovnání vážených a prostých odhadů základních charakteristik ročních příjmů (EU-SILC 2007)

Prosté a vážené odhady základních charakteristik čistých ročních příjmů domácností (v Kč) podle typu obce.

| typ obce | průměr | | směrodatná odchylka | |
|---------------|----------|----------|---------------------|----------|
| | obyčejný | vážený | obyčejná | vážená |
| Hl. m. Praha | 332170.5 | 349557.6 | 230874.8 | 243909.5 |
| krajská města | 269649.7 | 290622.2 | 183855.6 | 193654.1 |
| městské obce | 268747.6 | 287025.9 | 188883.4 | 212320.2 |
| venkov. obce | 271349.3 | 292450.9 | 172154.0 | 192697.8 |
| bez rozlišení | 275468.1 | 297598.7 | 187442.7 | 208833.5 |

| typ obce | medián | | průměr vah | počet hodnot |
|---------------|----------|----------|------------|--------------|
| | obyčejný | vážený | | |
| Hl. m. Praha | 275645.0 | 302144.0 | 617.3 | 864 |
| krajská města | 222000.0 | 243500.0 | 446.4 | 1423 |
| městské obce | 227354.5 | 243506.0 | 395.6 | 3952 |
| venkov. obce | 235143.5 | 254005.0 | 381.7 | 3436 |
| bez rozlišení | 232864.0 | 253446.0 | 417.9 | 9675 |

Porovnání vážených a prostých odhadů základních charakteristik ročních příjmů (EU-SILC 2007)

Prosté a vážené charakteristiky čistých ročních příjmů domácností (v Kč) podle sociální skupiny.

| sociální skupina | průměr | | směrodatná odchylka | |
|------------------|----------|----------|---------------------|----------|
| | prostý | vážený | prostá | vážená |
| nižší zaměst. | 304047.1 | 314675.0 | 139148.1 | 144471.9 |
| samost. činní | 418465.6 | 435951.2 | 334487.1 | 354580.3 |
| vyšší zaměst. | 375554.9 | 386432.4 | 186078.1 | 189254.2 |
| důchodci s EA | 38958.3 | 353382.5 | 117133.1 | 126796.6 |
| důch. bez EA | 159761.9 | 162135.9 | 70998.9 | 70866.1 |
| nezaměstnaní | 159139.3 | 162382.5 | 118090.8 | 120612.9 |
| ostatní | 171766.3 | 172946.8 | 220989.5 | 217403.1 |
| bez rozlišení | 275468.1 | 297598.7 | 187442.7 | 208833.5 |

| sociální skupina | medián | | průměr vah | počet hodnot |
|------------------|----------|----------|------------|--------------|
| | prostý | vážený | | |
| nižší zaměst. | 284144.0 | 292871.0 | 420.4 | 2385 |
| samost. činní | 349214.0 | 363412.0 | 630.1 | 802 |
| vyšší zaměst. | 345405.0 | 354063.0 | 433.6 | 2279 |
| důchodci s EA | 323381.5 | 337618.0 | 429.2 | 418 |
| důch. bez EA | 142392.0 | 152520.0 | 332.3 | 3423 |
| nezaměstnaní | 138387.0 | 147647.0 | 731.1 | 258 |
| ostatní | 124720.0 | 123432.0 | 380.4 | 110 |
| bez rozlišení | 232864.0 | 253446.0 | 417.9 | 9675 |

Porovnání vážených a prostých odhadů monetární chudoby českých domácností v letech 2002 – 2007

Vliv kalibračních vah na hranici rizika monetární chudoby za použití různých jednotek pro určení příjmů domácnosti.

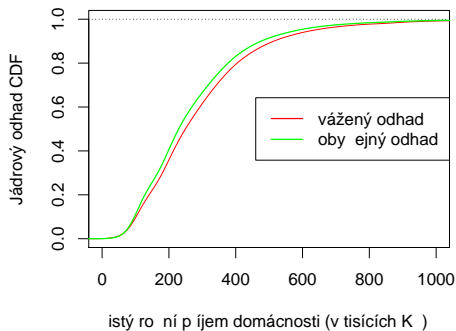
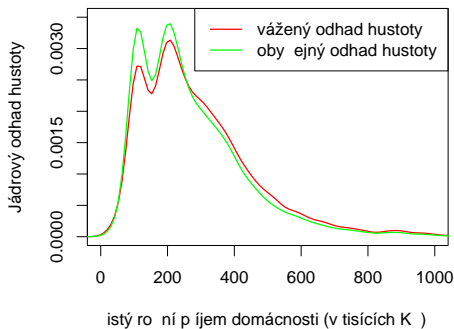
| Rok | Typ spotřební jednotky | Hranice rizika monetární chudoby | |
|------|------------------------|----------------------------------|--------------|
| | | vážený odhad | prostý odhad |
| 2002 | na domácnost | 116909 | 114554 |
| | na osobu | 52000 | 53522 |
| 2005 | na domácnost | 132549 | 123246 |
| | na osobu | 58200 | 58230 |
| | def. EU | 78786 | 76500 |
| | def. OECD | 68223 | 67199 |
| 2006 | na domácnost | 139743 | 128088 |
| | na osobu | 60912 | 60384 |
| | def. EU | 83052 | 79568 |
| | def. OECD | 72000 | 69926 |
| 2006 | na domácnost | 152069 | 139718 |
| | na osobu | 65850 | 65246 |
| | def. EU | 89611 | 86129 |
| | def. OECD | 77662 | 75600 |

Porovnání vážených a prostých odhadů monetární chudoby českých domácností v letech 2002 – 2007

Vliv kalibračních vah na podíl domácností pod hranicí rizika monetární chudoby za použití různých jednotek příjmů.

| Rok | Typ spotřební jednotky | Četnost pod hranicí chudoby | | | | Pearsonův test nez. | |
|------|------------------------|-----------------------------|-----------|--------------|-----------|---------------------|---------------|
| | | vážený odhad | | prostý odhad | | Statistika χ^2 | p-hodnota |
| | | absolutní | relativní | absolutní | relativní | | |
| 2002 | na domácnost | 1833 | 22.99 % | 1782 | 22.35 % | 0.894 | 0.3443 |
| | na osobu | 672 | 8.43 % | 757 | 9.49 % | 5.424 | 0.0199 |
| 2005 | na domácnost | 1095 | 25.17 % | 1012 | 23.26 % | 4.211 | 0.0402 |
| | na osobu | 439 | 10.09 % | 439 | 10.09 % | 0.001 | 0.9716 |
| | def. EU | 331 | 7.61 % | 291 | 6.69 % | 2.634 | 0.1046 |
| | def. OECD | 176 | 4.05 % | 167 | 3.84 % | 0.194 | 0.6594 |
| 2006 | na domácnost | 1878 | 25.10 % | 1691 | 22.60 % | 12.729 | 0.0004 |
| | na osobu | 753 | 10.06 % | 733 | 9.80 % | 0.270 | 0.6035 |
| | def. EU | 570 | 7.62 % | 469 | 6.27 % | 10.343 | 0.0013 |
| | def. OECD | 297 | 3.97 % | 253 | 3.38 % | 3.490 | 0.0617 |
| 2007 | na domácnost | 2409 | 24.90 % | 2193 | 22.67 % | 13.179 | 0.0003 |
| | na osobu | 858 | 8.87 % | 832 | 8.60 % | 0.405 | 0.5244 |
| | def. EU | 697 | 7.20 % | 566 | 5.85 % | 14.315 | 0.0002 |
| | def. OECD | 363 | 3.75 % | 324 | 3.35 % | 2.179 | 0.1400 |

Vliv kalibračních vah na rozdělení příjmů českých domácností (EU–SILC 2007)



Jádrové odhady hustoty pravděpodobnosti (gaussovské jádro) a příslušné empirické distribuční funkce vypočtené s využitím kalibračních vah a bez jejich využití. Jedná se o odhady rozdělení čistých ročních příjmů domácností v roce 2007.

- Výsledky šetření Mikrocensus a EU–SILC nelze přímo zobecňovat na celou populaci, neboť vybrané domácnosti nepředstavují reprezentativní vzorek. K tomuto závěru vede porovnání s výsledky šetření Sčítání lidu, bytů a domů (SLBD). Prosté zobecňování informací získaných z výběrových souborů by proto mohlo vést ke zkreslení, a proto je potřeba při analýzách provádět přepočty pomocí kalibračních vah.
- Ukazuje se, že vliv kalibrace závisí na mnoha faktorech. Velikost kalibračních vah např. roste s růstem příjmů, i když tento růst se u velmi vysokých příjmů zastavuje. Na zesilování či zeslabování vlivu kalibrace má vliv také počet členů domácnosti a další faktory, jako je sociální skupina osoby stojící v čele domácnosti, typ obce, v níž žije apod.
- Z provedeného šetření také vyplývá, že kalibrační váhy mají vliv jak na rozdělení příjmů brané z globálního pohledu (hustotu rozdělení, či distribuční funkci), tak i na jeho souhrnné charakteristiky a další statistiky, které z tohoto rozdělení vycházejí.

Musíme se proto vrátit k otázce:

Jak lze korektně zapracovat kalibrační koeficienty
do prováděných analýz?