



Průběžná zpráva - Analýza dat z cenových šetření CPI v České republice + realizace kvantitativního očištění dat z CPI, doplnění matice regionálních cen, výpočet cenových parit – nevážených

Výstup výzkumné aktivity VA6 (DC3): Sběr vstupních dat pro výpočet RCI

Autoři: Ing. Jana Šimanová, Ph.D., jana.simanova@tul.cz

Ing. Aleš Kocourek, Ph.D., ales.kocourek@tul.cz

Mgr. et Mgr. Jiří Rozkovec., jiri.rozkovec@tul.cz

Zpráva zpracována průběžně za období: 07 - 09/2015



Úvod

V období 07 – 09 pokračuje další očišťování dat – kvantitativní a výpočet nevážených a vážených cenových parit.

Tato podetapa navazuje na rozsáhlou a časově náročnou aktivitu realizovanou v první polovině roku 2015, jejíž výstupem jsou klastry stejných či podobných variet šetřené napříč regiony na podúrovni cenových reprezentantů v *.xls souboru (označené „tagy“). Autoři volí a experimentálně ověřují následující postup, který bude předmětem certifikace MMR ČR:

I. Kvantitativní očištění dat pro odstranění vybočujících hodnot (outlierů)

Vybraná data (výstup z předchozí části) byla na úrovni jednotlivých „tagů“ (podobných variet) dále kvantitativně očištěna tak, aby byly odstraněny extrémně vybočující hodnoty. Aplikováno bylo pravidlo 3σ , tudíž uplatněn předpoklad, že téměř všechny relevantní hodnoty statistického souboru by měly spadat do tří směrodatných odchylek od průměru při normálním rozdělení. Pravidlo 3σ bylo na data aplikováno třikrát.

Výstup: *.xls soubor s kvantitativně očištěnými „tagy“.

II. Výpočet průměrných cen a vytvoření neúplných trojrozměrných regionálních matic cen

Pro každého cenového reprezentanta byla v softwaru eViews 8.1 vytvořena trojrozměrná matice průměrných cen (prostý aritmetický průměr cen) na elementární úrovni jednotlivých tagů v rámci cenových reprezentantů.

Výstup: Matice region (36) × rok (2011, 2012, 2013) × „tagy“ (klastry identických variet v rámci daného cenového reprezentanta).

III. Ekonometrický odhad chybějící cen charakteristických variet

Nejvíce zastoupená varieta je definována jako charakteristická pro všechny regiony. Chybějící průměrné ceny u těchto charakteristických variet byly v maticích aproximovány metodou nejmenších čtverců (LSM). Úloha byla zpracována na panelových datech v prostředí ekonometrického softwaru eViews 8.1 (viz Obrázek 1 a 2)

Obrázek 1: Provedení úlohy v eViews 8.1

Zdroj: vlastní

Obrázek 2: Provedení úlohy v eViews 8.1

Zdroj: vlastní

Výstup: *.xls soubor charakteristických variet v každém regionu s průměrnou regionální cenou (zjištěnou či odhadnutou metodou nejmenších čtverců LSM).

IV. Výpočet tranzitivních, multilaterálních regionálních cenových parit

Cenové parity musí splňovat podmínku multilaterality a tranzitivity. Cenové parity jsou kalkulovány ve dvou fázích.





- I. fáze: Vytvoření úplné matice bilaterálních indexů (cenových poměrů) pro každou jednotlivou charakteristickou varietu. Cenové poměry (nikoli průměrné ceny) jsou zvoleny proto, že chybějící ceny v matici očištěných cenových průměrů byly doplněny odhady (viz bod III metodického postupu).
- II. fáze: Výpočet nevážených parit kupní síly na úrovni cenových reprezentantů dle metody GEKS (Gini, Éltető, Köves, Szulc). Pro každého cenového reprezentanta a region je kalkulováno jedno číslo – tzv. nevážená cenová parita, která vyjadřuje vztah ceny vybraného cenového reprezentanta v daném regionu k cenám zjištěným nebo odhadnutým metodou LSM ve všech ostatních regionech. Paritu regionu A na úrovni daného reprezentanta vyjadřuje vzorec (1). Jde vlastně o geometrický průměr cenových poměrů v řádku matice **Chyba! Nenalezen zdroj odkazů.** ze strany **Chyba! Záložka není definována..**

$$p_A = \prod_{k=1}^{36} (P_{A,k})^{1/36}, \quad (1)$$

kde $P_{A,k}$ je poměr cen regionu A ke k-tému regionu, přičemž $k = 1, 2, 3 \dots 36$; p_A je parita regionu A na úrovni daného cenového reprezentanta (charakteristické variety).

Výstup: Cenové parity (multilaterální tranzitivní cenové koeficienty) pro každý region a každého cenového reprezentanta (nyní již prostorově kvalitativně srovnatelnou charakteristickou varietu), přičemž žádný region není nadřazený, všechny parity jsou mezi sebou navzájem srovnatelné.



Odhad regionálních výdajových vah domácností

V. Shromáždění vstupních dat z databází Českého statistického úřadu

Pro odhad regionálních výdajů na nižší (okresní, LAU 1) úrovni je nutno shromáždit údaje týkající se:

- regionálních výdajů domácností na vyšších regionálních úrovních či v sousedních (souvisejících) oblastech v členění dle CZ-COICOP,
- potenciálních vysvětlujících proměnných, které statisticky významně ovlivňují výši a strukturu výdajů českých domácností v regionech.

Zdroji dat pro odhad regionálních výdajů na spotřebu domácností jsou:

- Statistika rodinných účtů (ČSÚ) – výdaje domácností dle NUTS 2
- Městská a obecní statistika (ČSÚ),
- Statistika Místních akčních skupin (ČSÚ) a
- SLDB 2011 (ČSÚ).

VI. Aplikace metody Small Area Estimation prostřednictvím softwaru eViews 8.1

Aplikace metody Small Area Estimation (viz kapitolu **Chyba! Nenalezen zdroj odkazů.** na straně **Chyba! Záložka není definována.**), tzn. sestavení a kalibrování ekonometrického modelu v softwaru eViews 8.1 probíhá ve dvou fázích:

- I. fáze: **Konstrukce regresního modelu** závislosti spotřebních vydání na vybraných kvantitativních a kategoriálních veličinách. Jako regresní model byl zvolen vícerozměrný lineární model ve tvaru:

$$COICOP_i = \beta_0 + \beta_1 NUTS + \beta_2 VEK_{PRUM} + \beta_3 UCH + \beta_4 BYT + \beta_5 MIG + \beta_6 VEK_{0-14} + \beta_7 VEK_{15-65} \quad (2)$$

Hodnoty závisle proměnné ($COICOP_i$) byly čerpány ze statistiky rodinných účtů ČSÚ podle jednotlivých oddílů spotřebních vydání CZ-COICOP ($i = 01, 02, 03, \dots, 12$) v členění dle regionů soudržnosti NUTS 2 v období 2011–2013. Nezávisle proměnné se dělí na kategoriální a kvantitativní: jako kategoriální veličina byla použita příslušnost k regionu soudržnosti NUTS 2

(označena jako *NUTS*¹). Jako kvantitativní veličiny byly vybrány:

- věkový průměr v daném regionu (VEK_{PRUM}),
- celkový počet uchazečů o zaměstnání vztažený k počtu obyvatel od 15 do 65 let (UCH),
- celkový počet dokončených bytů a dokončených bytů v rodinných domech vztažený na 1000 obyvatel (BYT),
- saldo migrace vztažené na celkový počet obyvatel v % (MIG),
- počet obyvatel do 14 let v % celkové populace (VEK_{0-14}),
- počet obyvatel od 15 do 65 let v % celkové populace (VEK_{15-65}).

II. **fáze: Odhad parametrů modelu** pro všech dvanáct tříd spotřebních vydání CZ-COICOP. Pro výpočty byl použit program pro ekonometrické modelování eViews 8.1, hladina významnosti $\alpha = 5\%$. Obecný postup vyjadřuje následující schéma:

- a) Do modelu bylo zahrnuto všech osm nezávisle proměnných a odhadnuty příslušné regresní parametry.
- b) Podle hodnot *p-value* byla poté v jednotlivých případech vyhodnocena jejich statistická významnost.
- c) Pokud alespoň jeden parametr nebyl statisticky významný, byl z modelu vyloučen, přičemž se vždy jednalo o ten, který měl hodnotu *p-value* nejvyšší.
- d) Poté byly zbývající regresní parametry odhadnuty znovu a postup se buď vrátil do bodu b), nebo již byly všechny parametry statisticky významné a odhad byl tímto ukončen. Výstupem je lineární model sestávající se z 12 rovnic (pro každý oddíl CZ-COICOP).
- e) Vizualizace výsledků modelu: Pomocí Scheffého metody bylo provedeno vícenásobné porovnávání podle jednotlivých tříd kategoriální proměnné NUTS 2. Výstupem jsou podmnožiny regionů NUTS 2 a podmnožiny na úrovni okresů, které lze z pohledu chování závisle proměnné považovat za homogenní.

¹

kde: $I1(1) = 1$ pro NUTS 2=CZ01, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(2) = 1$ pro NUTS 2=CZ02, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(3) = 1$ pro NUTS 2=CZ03, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(4) = 1$ pro NUTS 2=CZ04, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(5) = 1$ pro NUTS 2=CZ05, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(6) = 1$ pro NUTS 2=CZ06, -1 pro NUTS 2=CZ08, 0 jinak,
 $I1(7) = 1$ pro NUTS 2=CZ07, -1 pro NUTS 2=CZ08, 0 jinak.

VII. Odhad výdajů domácností dle CZ-COICOP na úrovni LAU 1

Na základě výsledků lineárního modelu byly vypočteny průměrné výdaje (sub)regionálních domácností (na 1 osobu) na nejvyšší agregované úrovni CZ-COICOP. Pro odhad byly použity statisticky významné regresory z předchozího kroku (2b) týkající se malých oblastí.

Výstup: 36 × 12 odhadnutých výdajových oddílů dle klasifikace CZ-COICOP.

VIII. Odhad výdajů domácností na bydlení ve vlastním

Pro oddíl **CZ-COICOP 04 Bydlení, voda, energie, paliva** byly kalkulovány a včleněny výdaje osob bydlících ve vlastním jako **průměrná výše splátky** hypotečních úvěrů vč. úroků na jednoho obyvatele v daném regionu.² Záměrem je do indexu včlenit rostoucí výdaje na splátky hypoték vč. úroků, které stále významněji ovlivňují koupěschopnost českých domácností. Zdroji dat pro kalkulaci tohoto odhadu jsou:

Generální finanční ředitelství Ministerstva finanční ČR. Objem odpočtu úroků na bydlení, které byly uplatněny jako nezdanitelná položka v rámci daně z příjmů fyzických osob podle § 15 odst. 3 a 4 Zákona o daních z příjmu.³

- ČNB (Arad). *Konečné zůstatky úvěrových účtů klientů v časové řadě.*⁴ Dostupné z WWW: http://www.cnb.cz/cnb/STAT.ARADY_PKG.PARAMETRY_SESTAVY?p_sestuid=13288&p_strid=AABBAD&p_lang=CS
- MMR ČR. *Objem nově poskytnutých hypoték v letech 2011–2013.* Dostupné z WWW: <http://www.mmr.cz/cs/Stavebni-rad-a-bytova-politika/Bytova-politika/Hypotecni-uvery/Hypotecni-uvery-poskytnute-v-krajich-CR>

Rozdílem konečných zůstatků dvou období a přičtením nově poskytnutých hypoték lze určit celkový objem splátek (bez úroků) za celou ČR. Samotná výše úroků v regionálním členění (zdroj MF ČR)

² Snahou je zahrnout do kalkulace regionálního cenového indexu výdaje na bydlení ve vlastním, které odráží ceny nemovitostí vč. pozemku. Výdajovou vahou je objem zaplacených splátek hypotečních úvěrů vč. úroků v daném regionu (subregionu). Přiřazená cenová parita odráží průměrné ceny typových nemovitostí – domů a bytů v poměru odpovídajícím počtu převodů v daném regionu (zdroj cenových dat je uveden v Příloze 2 této metodiky).

V současné době představuje objem rezidentských nesplacených hypoték v ČR více jak 874 mld. Kč. V letech 2011–2013 bylo podle autorských kalkulací vycházejících z údajů ze statistických databází ČNB – Arad, MMR ČR a MF ČR zaplaceno na splátkách hypoték vč. úroků ročně v průměru cca 115 mld. Kč. Detailní členění na nižší regionální úroveň umožňují údaje Generálního finančního ředitelství MF ČR o uplatněných úrocích v rámci přiznání k dani z příjmů fyzických osob.

³ Data byla autorům metodiky elektronicky zaslána na základě žádosti o poskytnutí dat podle zákona č. 106/1999 Sb. o svobodném přístupu k informacím

⁴ Tj. počáteční zůstatky plus čerpání nových úvěrů minus splátky dříve poskytnutých úvěrů.



slouží jako regionalizační klíč za předpokladu konstantní úrokové sazby napříč regiony a podobných úvěrových parametrů (doba fixace, doba splácení, hodnocení bonity klienta).

Vybrané experimentální výpočty navrženého metodického postupu jsou uvedeny ve zprávě k aktivitě VA6 k 09/2015 a budou součástí příkladu uplatnění metodiky v návrhu na certifikaci u orgánu MMR ČR, Odboru regionální politiky.

