

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

4/2019

ročník/volume 29

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

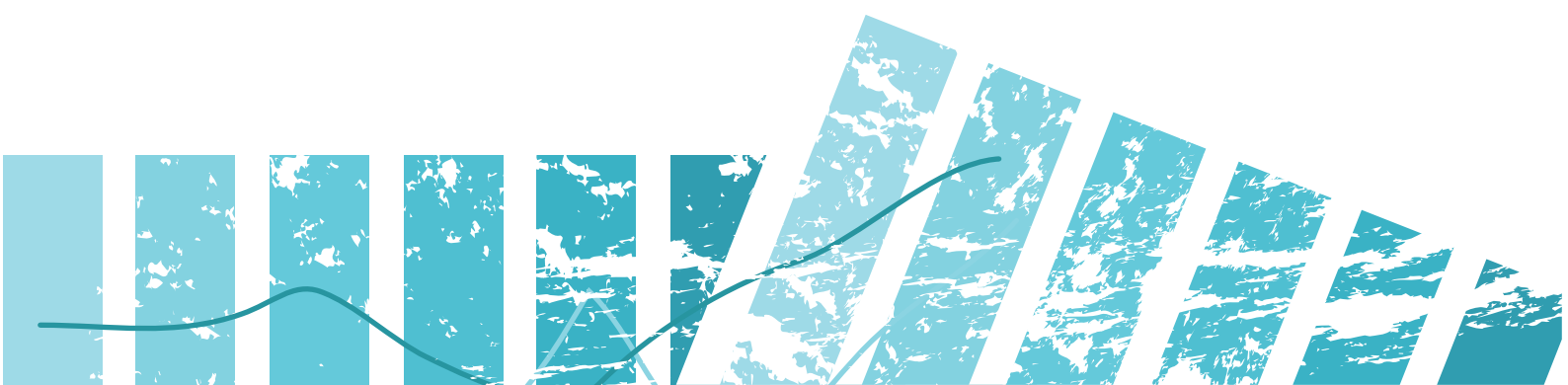
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 4

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 49 – 66

Dátum vydania/Publication date: 15. október 2019/October 15, 2019



Helena GLASER-OPITZOVÁ
Štatistický úrad Slovenskej republiky

NOVÉ ZDROJE ÚDAJOV PRE CENOVÚ ŠTATISTIKU A METÓDY ICH SPRACOVANIA

NEW DATA SOURCES FOR PRICE STATISTICS AND THE METHODS OF THEIR PROCESSING

ABSTRAKT

Implementácia nových zdrojov údajov, tzv. „scanner data“, do produkcie cenovej štatistiky predstavuje významnú metodologickú zmenu, ktorej zavedenie môže významne ovplyvniť tak úroveň cenových indexov, ako aj dynamiku ich zodpovedajúcej mesačnej alebo ročnej miery zmeny. Práca s týmto typom transakčných údajov je pre Štatistický úrad Slovenskej republiky novou neprebádanou oblasťou činnosti a ich využívanie v oficiálnej štatistike prináša mnohé otázky a vyžaduje množstvo analýz s dopadom na zmenu produkčných procesov.

ABSTRACT

Implementation of new data sources, the so-called “scanner data” into the production of price statistics represents a significant methodological change, the introduction of which may significantly affect both the level of price indices and the dynamics of their corresponding monthly or annual rate of change. Working with this type of transaction data is a new unexplored area of the activity for the Statistical Office of the Slovak Republic, their use in official statistics brings many questions, and the issues require a number of analyses with an impact on the change of production processes.

KLÚČOVÉ SLOVÁ

scanner data, cenové indexy, cena za jednotku tovaru, drift v časovom rade reťazených indexov

KEY WORDS

scanner data, price indices, unit value, chain drift

1. ÚVOD

Spoločnosť je čoraz dynamickejšia vďaka všadeprítomným digitálnym technológiám a procesom tzv. datafikácie. Datafikácia je moderný technologický trend, ktorý premieňa mnohé aspekty nášho života na počítačové údaje a tieto transformuje na novú formu hodnoty. K dispozícii je stále viac údajov a preto najväčšou výzvou 21. storočia je ich spracovanie a analýza tak, aby sa stali poznatkami. Táto výzva by sa aj pre štatistikov mala zmeniť na príležitosť. Nové zdroje údajov totiž predstavujú jedinečnú príležitosť na vytvorenie nových štatistík a skvalitnenie existujúcich. Dôveryhodná štatistika, pojem, ktorý sa dnes často skloňuje na pôde Európskeho štatistického systému, je založená na princípoch oficiálnej štatistiky, ktoré sú vymedzené v Kódexe postupov pre Európsku štatistiku, a v celom procese má mimoriadny význam udržanie vysokej kvality a spoľahlivej metodiky. V tejto súvislosti sa Európsky štatistický systém a teda aj Štatistický úrad Slovenskej republiky (ŠÚ SR) musí čo najrýchlejšie prispôbiť novým technológiám a stať sa flexibilnejším – musí neustále svoju produkciu modernizovať.

Jeden z možných smerov modernizácie produkcie ŠÚ SR je modernizácia cenových štatistík, v ktorých sa otvárajú možnosti v podobe nových zdrojov údajov vo forme transakčných údajov, nazývaných aj „scanner data“. Forma a spôsob ich využitia na výpočet Harmonizovaného indexu spotrebiteľských cien (HICP) alebo Indexu spotrebiteľských cien (CPI) je v súčasnosti v štádiu riešenia v rámci interného projektu ŠÚ SR.

HICP, ktorý slúži na medzinárodné porovnanie cenového vývoja v rámci krajín Európskej Únie a Európskeho hospodárskeho priestoru, je jedným z konvergenčných kritérií na vstup do eurozóny. CPI je dôležitým a významným makroekonomickým ukazovateľom nielen pre Národnú banku Slovenska, Ministerstvo financií SR, Ministerstvo práce, sociálnych vecí a rodiny SR a ostatné ústredné orgány štátnej správy, ale je aj kľúčovým ukazovateľom pri posudzovaní vplyvu vývoja cien na životnú úroveň obyvateľstva a taktiež sa využíva pri valorizácii dôchodkov a miezd a pri posudzovaní vplyvu vývoja cien na životné minimum. Výpočet týchto indexov sa do dnešných dní vykonáva z údajov získavaných prostredníctvom terénneho zberu údajov o cenách tovarov a služieb. Zber údajov sa realizuje priamo v prevádzkach a obchodoch a ceny, ktoré sa zisťujú, sú tzv. pultové ceny. Predané množstvá jednotlivých tovarov a služieb nie sú k dispozícii a tradičný spotrebný kôš, na ktorom je založený výpočet indexu spotrebiteľských cien, je relatívne malá vzorka z celej množiny tovarov a služieb. Spôsob získavania vstupných dát, ako aj doterajší spôsob výpočtu je z hľadiska stúpajúcich nárokov na kvalitu a tempo tvorby produktov štátnej štatistiky nepostačujúci.

Tzv. „scanner data“ získavané priamo od obchodníkov obsahujú naproti tomu informáciu o cene, predanom množstve tovarov a tržbách za jednotlivé produkty. Uvedené údaje môžu dopĺňať alebo v budúcnosti úplne nahradiť klasický zber údajov o cenách a poslúžiť ako nový zdroj údajov na výpočet cenových indexov. Práca so „scanner data“ je pre ŠÚ SR novou neprebádanou oblasťou činnosti a ich využívanie v oficiálnej štatistike prináša mnohé otázky a vyžaduje množstvo analýz s dopadom na zmenu produkčných procesov.

2. SÚČASNÝ STAV

V podmienkach Slovenskej republiky sa vývoj indexov spotrebiteľských cien v súčasnosti sleduje na univerzálnom spotrebnom koši založenom na súbore 728 reprezentantov – vybraných druhov tovarov a služieb platených obyvateľstvom (po poslednej revízii od januára 2019). Univerzálny spotrebný kôš sa člení na 12 odborov podľa klasifikácie COICOP (Medzinárodná klasifikácia individuálnej spotreby podľa účelu). Ako cenoví reprezentanti boli vybrané výrobky a služby, ktoré sa významne podieľajú na výdavkoch obyvateľstva a svojím rozsahom reprezentujú celú sféru spotreby. Spotrebiteľské ceny jednotlivých reprezentantov sa zisťujú na celom území SR vo vybranej sieti predajní a prevádzok služieb, kde obyvatelia obvykle nakupujú. Ceny sú zisťované počas prvých 20 dní sledovaného mesiaca. Váhy jednotlivých reprezentantov sa počítajú z údajov zistených štatistikou rodinných účtov, z podkladov štatistiky národných účtov a rôznych administratívnych zdrojov údajov.

Index spotrebiteľských cien je indexom Laspeyresovho typu:

$$I = \frac{\sum \frac{p_1}{p_0} p_0 q_0}{\sum p_0 q_0} 100 \quad (1)$$

kde:

p_1 je cena tovaru alebo služby v sledovanom období,

p_0 je cena tovaru alebo služby v základnom období,

q_0 predstavuje predané množstvo tovaru alebo služby v základnom období,

$\frac{p_1}{p_0}$ individuálny index ceny určitého tovaru alebo služby.

Počnúc januárom 2005 sa realizuje výpočet indexu spotrebiteľských cien k základnému obdobiu (december 2000) reťazením cez december predchádzajúceho roku. Aby bolo zabezpečené, že váhy za základné agregáty odrážajú súčasnú výdavkovú štruktúru a spotrebiteľské správanie, sú váhy každoročne revidované. Dôležitou črtou reťazového indexu je, že táto každoročná aktualizácia váh umožňuje a uľahčuje zavedenie nových položiek a subindexov a odstránenie zastaraných.

Na zabezpečenie prepojenia starých a nových časových radov, je pri počítaní reťazového indexu potrebné prekryvacie obdobie (k), v ktorom je index počítaný, použijúc starú a novú sadu váh. Týmto prekryvacím obdobím môže byť mesiac alebo rok. ŠÚ SR pri reťazení stanovil december predchádzajúceho roka ako prekryvacie obdobie, t. j. v roku 2019 je týmto obdobím december 2018. December 2018 je posledným obdobím, za ktoré sa počíta index na predchádzajúcom spotrebnom koši a systéme váh, a zároveň je referenčným (základným) obdobím pre ceny nového spotrebného koša a nový systém váh. December predchádzajúceho roka tvorí teda prepojenie (prekryvacie obdobie) medzi predchádzajúcim a novým indexom a tie vytvárajú reťazový index. Reťazový index môže mať dve, alebo viac prepojení. Medzi každým prepájacím obdobím je index počítaný ako index s fixnými váhami.

Predpokladajme, že časové rady indexov s fixnými váhami sa vypočítali s obdobím 0 ako referenčným obdobím pre index a ceny a že v období k (prekryvacím obdobím) bola zavedená nová sada váh a aktualizovaný spotrebný kôš.

Reťazový index je potom počítaný ako:

$$I = I^{0:k} \sum W_i^k I_i^{k:t} = I^{0:k} I^{k:t} \quad (2)$$

kde:

0 je referenčné obdobie pre index, k je prekryvacie obdobie a t je bežné obdobie.

S výnimkou výrobkov s volatílnymi cenami, ceny získavané tradičným spôsobom sa zbierajú raz mesačne a len na výber predajní. Ceny, ktoré sa zbierajú, sú pultové ceny – ceny, za ktoré sa tovary ponúkajú.

Aktuálny spotrebný kôš (opis a váhy jednotlivých odborov, skupín, tried, podtried medzinárodnej klasifikácie individuálnej spotreby podľa účelu (COICOP); pod tým opis a váhy jednotlivých reprezentantov) je publikovaný na stránke Štatistického úradu SR: www.statistics.sk. V rámci ponuky hlavnej hornej lišty Štatistiky > Makroekonomické štatistiky > Spotrebiteľské ceny a ceny produkčných štatistik >

Ukazovatele > Spotrebiteľské ceny (inflácia) je možné nájsť spotrebné koše od roku 2013 vrátane súčasného.

3. CHARAKTERISTICKÉ VLASTNOSTI NOVÝCH ZDROJOV ÚDAJOV

„Scanner data“, ktoré generujú terminály predajní v obchodoch, nám poskytujú informácie na úrovni čiarového kódu alebo presnejšie GTIN (Global Trade Item Number)¹. Tento systém okrem iných zahŕňa kódy EAN (the European Article Number) a UPC (the Universal Product Number) a je rozvíjaný a manažovaný súkromnou spoločnosťou GS1.

Platobné terminály maloobchodných predajní zaznamenávajú každú transakciu. „Scanner data“ predstavujú informácie o tržbách a množstve jednotlivých transakcií na úrovni GTIN (ďalej len „kód položky“) za dané obdobie a miesto (predajňa alebo maloobchodník) a ďalšie informácie o produktoch. Tieto informácie umožňujú vypočítať cenu za jednotku tovaru na úrovni každého kódu položky.

„Scanner data“ nie sú vytvárané ani zhromažďované na účely cenovej štatistiky, ale bežne sa používajú napríklad na monitorovanie vývoja trhu. Z pohľadu oficiálnej štatistiky predstavujú sekundárny zdroj údajov, teda zdroj údajov, ktorý bol pôvodne zozbieraný na iný ako štatistický účel. Na porovnanie, pri tradičnom zisťovaní spotrebiteľských cien národnej štatistickej úrady (NŠÚ) určujú a sú zodpovedné za všetky kroky v procese zhromažďovania údajov o cenách, a preto tento zdroj údajov označujeme ako primárny.

Súbory údajov s kódmi položiek a cenami, za ktoré sa tovar ponúka alebo údaje s cenovými ponukami získané z internetu sa nepovažujú za transakčné údaje, hoci ich spracovanie môže byť veľmi podobné.

„Scanner data“ od konkrétneho predajcu a za dané časové obdobie predstavujú vyčerpaný zoznam všetkých kódov položiek, ich tržieb a predaných množstiev. „Scanner data“ umožňujú kompilovať index zo všetkých transakcií predajcu alebo predajne. Obvykle sa v supermarkete používa 10 000 – 25 000 kódov položiek len na pokrytie sortimentu potravín a nealkoholických nápojov. „Scanner data“ umožňujú štatistikom použiť to, čo sa skutočne predalo, a zahrnúť do výpočtu HICP/CPI oveľa viac položiek, ako je to možné pri tradičnom zbere cien. Znamená to tiež, že ak sú k dispozícii informácie o tržbách, jednotlivým položkám možno priradiť váhu. Implementácia „scanner data“ do produkcie cenovej štatistiky môže tiež ušetriť náklady na tradičný zber údajov o cenách.²

Kódy položiek identifikujú tovar veľmi presne, takže dva tovary s rovnakým kódom položky sú z pohľadu spotrebiteľa identické. Výsledná cena za jednotku tovaru na úrovni kódu položky je priemer cien skutočne platených spotrebiteľmi.

„Scanner data“ odrážajú dynamiku skutočných nákupov v každom elementárnom agregáte, pretože je zaznamenaná každá transakcia. V súbore údajov je viditeľný

¹ Globálne číslo obchodnej položky (GTIN) je jedinečné číslo, ktoré identifikuje akúkoľvek vopred definovanú obchodnú položku (produkt alebo službu), ktorá môže byť spoplatnená, objednaná alebo fakturovaná v ktoromkoľvek bode dodávateľského reťazca. Medzi obchodné položky patria všetky služby a výrobky, od surovín až po výrobky koncových používateľov.

² Tradičné zisťovanie cien patrí medzi jedno z najnákladnejších zisťovaní v ŠÚ SR.

vznik nových kódov položiek, zánik kódov položiek a zmeny ich relatívnej dôležitosti. V priebehu roka zaniká 25 % až 60 % kódov položiek, samozrejme, v závislosti od krajiny [3].

Okrem vzniku skutočne nových tovarových položiek sa položky často nahrádzajú novými verziami tovarov nazývanými „relaunches“ alebo „znovuzavedenia“. Ide v podstate o rovnaké tovarové položky s nejakým povrchným rozdielom, ako je napríklad nový obal. Vzniká takto nielen nový kód položky bez zmeny kvality, ale aj problém týkajúci sa spracovania údajov, ak napríklad medzimesačné párovanie tovarov prebieha na základe GTIN kódu. V prípade zľavy (napr. o 20 % viac obsahu za rovnakú cenu) sa položke priradí aj nový kód. V iných prípadoch môže ísť o významnejšie zmeny, ak sú napríklad položky určitej značky nahradené podobnými položkami inej značky.

Jeden zo spôsobov riešenia problému „znovuzavedenia“ tovarových položiek by bolo definovať produkty podľa vlastností, ktoré určujú ich cenu. Alternatívny spôsob riešenia uvedeného problému je použiť interný opis produktov jednotlivých obchodných reťazcov.

Definícia produktu je teda nevyhnutnou súčasťou tvorby cenového indexu. Bez ďalších informácií, ktoré charakterizujú produkt, je ťažké, ak nie nemožné, určiť, či sú dve položky s rôznymi GTIN úplne porovnateľné, t. j. či ide o opätovné znovuzavedenie produktu do predaja alebo je nutné aplikovať úpravu z dôvodu kvalitatívnej zmeny produktu, ak sa produkty líšia.

4. METÓDY SPRACOVANIA

4.1 Statický vs dynamický prístup

Eurostat [3] definuje tri základné kroky pri spracovaní „scanner data“ a dva prístupy k výberu položiek, ktoré budú vstupovať do výpočtov.

Tri kroky sú:

1. klasifikácia kódov položiek na klasifikáciu COICOP,
2. výber kódov položiek a výpočet elementárneho agregovaného indexu,
3. integrácia výsledkov do HICP/CPI.

Dva prístupy k výberu položiek zo „scanner data“ sú statický a dynamický prístup.

Predpokladajme, že súbor údajov je úplný a správny, klasifikácia kódov položiek na COICOP je prvým krokom pri spracovaní „scanner data“. Pod úrovňou podtriedy medzinárodnej klasifikácie COICOP 5 je možné definovať ďalšiu, „umelú“ úroveň klasifikácie, COICOP 6 na účel definovania tzv. homogénnych skupín produktov a výpočtu cenových indexov elementárnych agregátov na nižšej úrovni agregácie ako je COICOP 5.

Po zatriedení kódov položiek podľa klasifikácie nasleduje výber kódov položiek, ak to vybraná metóda vyžaduje.

Pri statickom prístupe sa výber uskutočňuje v roku t a používa sa 12 mesiacov po decembri roka t , v prípade potreby sa realizujú zámény vybraných položiek.

K zámene položky dôjde, ak sa položka stane menej reprezentatívnou alebo sa počas roka stratí. Statický prístup imituje tradičný fixný výber (spotrebný kôš) s tým rozdielom, že dochádza k zmene cenového konceptu. Ceny získavané tradičným spôsobom sú nahradené cenami za jednotku tovaru zo „scanner data“. Koncom roka sa pripravuje výber reprezentantov na nasledujúci rok. Všetko sa riadi tradičnou metodikou, ale s výhodou úplných informácií o skutočných transakciách, na základe ktorých sa realizuje počítačový výber kódov položiek a v prípade potreby ich nahradenie počas roka.

Viacere krajiny ako napríklad Austrália, Belgicko, Dánsko alebo Švajčiarsko implementovali v počítačovej fáze tento spôsob výpočtu CPI a ako väčšina krajín s tradičným zisťovaním cien používali Jevonsov index. Ich metódy sa odlišovali v počte týždňov, ktoré vstupovali do výpočtu ceny za jednotku tovaru, a tým, či bola alebo nebola vykonávaná manuálna zámena reprezentantov a následná úprava z dôvodu ich kvalitatívnych zmien. Táto metóda má výhody najmä vtedy, ak sa „scanner data“ používajú len v limitovanom rozsahu a je potrebné ich kombinovať s údajmi zozbieranými tradičným spôsobom. Uvedený spôsob má aj nevýhody. Je pracovne náročný a dostupné údaje využíva len v obmedzenej miere.

Dynamická metóda automaticky vyberie reprezentatívny výber kódov položiek pre každé dva po sebe nasledujúce mesiace (t a $t + 1$, $t + 1$ a $t + 2$, $t + 2$ a $t + 3$ a tak ďalej) výberom všetkých spárovaných položiek, ktoré majú tržby nad určitú hranicu a budú obsahovať nové a dostatočne dôležité položky, pričom položky, ktoré sú menej dôležité, sa vypustia. Dynamická metóda sa uprednostňuje vtedy, keď je potrebné spracovať veľké množstvo údajov, pretože sa dá ľahko automatizovať. Ak sa však často vyskytujú tzv. zámenny a znovuzavedenia položiek, tie sa musia riešiť separátne, aby sa zabezpečila úprava z dôvodu kvalitatívnych zmien. Základný agregátny index sa vypočíta na základe súboru spárovaných reprezentatívnych kódov položiek pre položky, ktoré sa skutočne predávajú v dvoch po sebe nasledujúcich obdobiach.

Nevážený Jevonsov index [1] sa počíta za aktuálny a predchádzajúci mesiac ako:

$$P_J^{0t,mt} = \frac{(\prod_{k=1}^K p_k^{mt})^{1/K}}{(\prod_{k=1}^K p_k^{0t})^{1/K}} = \left(\prod_{k=1}^K \frac{p_k^{mt}}{p_k^{0t}} \right)^{1/K} \quad (3)$$

kde K označuje súbor spoločných kódov položiek patriacich k základnému agregátu K . Ceny za jednotlivé kódy položiek, ktoré nie sú prítomné v nasledujúcich obdobiach, sú imputované cenovým vývojom základného agregátu na obdobie 14 mesiacov, aby sa zabezpečilo, že sezónne položky sa vrátia do indexu v správnom čase, čo umožňuje zachytiť zmeny medzi rokmi v dôsledku počasia alebo sviatkov ako je Veľká noc.

Reťazený index je potom nasledovný [3]:

$$\begin{aligned} CP_J^{0t,mt} &= P_J^{0t,1t} \cdot P_J^{1t,2t} \cdot \dots \cdot P_J^{(m-1)t,mt} = \\ &= \frac{(\prod_{k=1}^{K_1} p_k^{mt})^{1/K_1}}{(\prod_{k=1}^{K_1} p_k^{0t})^{1/K_1}} \cdot \dots \cdot \frac{(\prod_{k=1}^{K_m} p_k^{mt})^{1/K_m}}{(\prod_{k=1}^{K_m} p_k^{(m-1)t})^{1/K_m}} \cdot \frac{(\prod_{k=1}^K p_k^{mt})^{1/K}}{(\prod_{k=1}^K p_k^{0t})^{1/K}} = P_J^{0t,mt} \end{aligned} \quad (4)$$

kde $K1$ označuje množinu spoločných položiek v období 0 a 1 , $K2$ v období 1 a 2 a tak ďalej. Keďže sa objavujú nové položky a iné zmiznú, reťazený medzimesačný index sa nezredukuje na priamy Jevonsov index. Je preto dôležité, aby sa zistilo, do akej miery sa súbory $K1$, $K2$... K líšia, a vyhodnotiť riziko vychýlenia reťazeného časového radu indexov (chain drift). Podľa [3] by sa tento spôsob nemal používať, ak sú zmeny podstatné.

Príručka síce upozorňuje, kedy sa uvedený postup nemá používať, ale neobsahuje oficiálne odporúčania, ako v uvedenom prípade postupovať. Viaceré národné štatistické úrady preto vyvíjajú, testujú a analyzujú vlastné metódy a postupy a zavádzajú tzv. tieňové spracovanie pre empirické porovnanie jednotlivých metód.

„Scanner data“ musia byť na určitom stupni agregácie kombinované s tradične zozbieranými údajmi.

Štandardne môžu nastať dva prípady:

1. ak „scanner data“ pokrývajú iba časť spotrebných výdavkov na konkrétny COICOP agregát.
2. ak „scanner data“ zo supermarketov pokrývajú všetky výdavky spotrebiteľov na konkrétny agregát.

V prvom prípade váha agregátu nemôže pochádzať len zo „scanner data“, ale aj z tradičných zdrojov (napr. Národné účty).

4.2 Indexy vhodné na statický prístup

Označme množinu tovarových položiek patriacich do určitej kategórie S a počet predaných položiek ako N . Cieľom je skonštruovať indexy, ktoré porovnávajú bázičné obdobie 0 – počiatočné obdobie časového radu indexov s obdobím t . Ceny (ceny za jednotku tovaru) každej položky $i \in S$ v období 0 a t označme p_i^0 a p_i^t . Ak kvantitatívne informácie nie sú k dispozícii [14], odporúča sa priamy alebo bilaterálny pomer nevážených geometrických priemerov alebo Jevonsov cenový index:

$$P_J^{0,t} = \left(\prod_{i=1}^S \frac{p_i^t}{p_i^0} \right)^{1/N} = \frac{(\prod_{i=1}^S p_i^t)^{1/N}}{(\prod_{i=1}^S p_i^0)^{1/N}} \quad (5)$$

Pretože v „scanner data“ sú vždy k dispozícii informácie o váhach, je možné konštruovať aj teoreticky atraktívnejšie formy indexov, a to superlatívne cenové indexy ako napríklad Fisherov alebo tiež Tornqvistov index, ktorého geometrická forma zjednodušuje analýzu rozkladu. Vo väčšine prípadov vedú oba indexy k veľmi podobným výsledkom.

Bilaterálny *Törnqvist cenový index* je daný:

$$P_T^{0,t} = \left(\prod_{i=1}^S \frac{p_i^t}{p_i^0} \right)^{s_i^0 + s_i^t / 2} \quad (6)$$

kde $s_i^0 = p_i^0 q_i^0 / \sum_{i \in S} p_i^0 q_i^0$ a $s_i^t = p_i^t q_i^t / \sum_{i \in S} p_i^t q_i^t$ označujú podiely výdavkov v období 0 a t , q_i^0 a q_i^t sú predané množstvá.

Bázické indexy majú tú výhodu, že neobsahujú drift v časovom rade indexov, ale majú aj nevýhodu. V priebehu času sa objavujú nové výrobky a staré výrobky miznú a je stále ťažšie porovnávať položky, ktoré sú k dispozícii v aktuálnom období, s položkami, ktoré boli dostupné v bázičkom období.

4.3 Indexové metódy na dynamický prístup

Indexy na dynamický prístup k spracovaniu „scanner data“ možno klasifikovať vo vzťahu k novozaradeným položkám do predaja a k miznúcim položkám. Metódy, ktoré s nimi uvažujú, sa označujú ako *metódy upravujúce kvalitu* (z dôvodu kvalitatívnych zmien v sortimente obchodníka). Metódy, ktoré nezohľadňujú nové a miznúce položky sa označujú ako *metódy zhodného modelu* (pracujú so súborom spárovaných kódov). Ak teda zostavujeme HICP/CPI zo „scanner data“, všeobecne sa odporúča použiť reťazové superlatívne indexy a to z dôvodu vyššej miery spárovania jednotlivých kódov položiek medzi dvoma po sebe idúcimi obdobiami a predpokladu menších rozdielov v cene a kvantite. Tento predpoklad však nezohľadňuje existenciu výpredajov a zliav, pri ktorých sa môže výrazne zvýšiť kvantita predaného tovaru a to až niekoľkonásobne. Ak sa obdobie zľavy skončí, cena sa obvykle vráti na pôvodnú hodnotu. Môže sa stať, že sú obyvatelia tak zásobení, že nejaký čas potrvá, kým sa kvantita predávaného tovaru vráti na pôvodnú hodnotu. Za týchto podmienok môžu mať reťazené superlatívne indexy tendenciu k poklesu a len postupnému návratu na pôvodnú úroveň (downward chain drift) v porovnaní s bázičkými indexmi.

Použitie *multilaterálnych metód* maximalizuje množstvo spárovaných položiek v údajoch bez rizika zavedenia driftu do reťazených časových radov [11].

4.3.1 Holandská metóda

Metódu opísanú v tejto podkapitole na základe [7] používa Holandský štatistický úrad, aby tak obišiel problém tzv. driftu reťazených vážených (pokiaľ možno superlatívnych) indexov.

„Scanner data“ ukazujú, že nákupy v supermarketoch sú vysoko dynamické. Veľa „starých“ položiek z predajní mizne a veľa nových sa neustále objavuje. Indexy počítané na statickom princípe spotrebného koša túto dynamiku ignorujú. Stáva sa tiež, že niektoré položky zo súborov údajov zmiznú dočasne, buď preto, že nie sú na sklade, alebo sa nepredávajú. „Scanner data“ odhaľujú aj distribúciu výdavkov za jednotlivé položky v každom mesiaci.

De Haan [8] prezentoval príklad čistiacich prostriedkov, ktorý dokazuje, že nakúpené množstvá sa pri znížení ceny enormne zvýšia. Skutočnosť, že ceny produktov predávané v supermarketoch nie sú vôbec stále, bola pozorovaná a analyzovaná v roku 2006 vo viacerých akademických štúdiách. Následkom toho by mohli byť indexy vypočítané zo „scanner data“ na nízkej úrovni agregácie dosť volatilné. Z empirických štúdií tiež vyplýva, že v mnohých prípadoch 50 – 60 % položiek predstavuje menej ako 10 % celkových výdavkov, preto použitie nevážených indexov by nadhodnotilo význam mnohých málopredajných položiek. Vzhľadom na vysoký počet vstupujúcich a vystupujúcich položiek pozorovaných v „scanner data“, bola voľba Holandského štatistického úradu použiť reťazový index na najnižšej úrovni agregácie celkom prirodzená. Keďže reťazenie vážených indexov môže viesť k podstatnému vychýleniu reťazeného časového radu indexov, na

výpočet cenového indexu elementárneho agregátu bol vybratý nevážený Jevonsov index. Z hľadiska výdavkov sú do výpočtu zahrnuté najdôležitejšie položky, zatiaľ čo menej dôležité položky sú vylúčené. Prahová hodnota na zahrnutie položiek bola zvolená tak, aby bolo vybraných zhruba 50 % položiek, čo predstavuje 80 – 85 % celkových výdavkov. Cenovým konceptom je jednotková hodnota, t. j. cena za jednotku tovaru pre položku v reťazci supermarketu.

Agregácia elementárnych cenových indexov vypočítaných zo „scanner data“ je podobná agregácii cenových indexov vypočítaných z iných zdrojov a veľmi podobná postupu, ktorý v súčasnosti aplikuje aj ŠÚ SR.

Výpočet:

V tejto časti je detailne opísaný výpočet indexov zo „scanner data“ pre konkrétny obchodný reťazec. Použité je nasledovné označenie:

Cena (cena za jednotku tovaru) a podiel položky i na výdavkoch pre elementárny agregát a v mesiaci m roka y sú označené ako $p_{i,a}^{y,m}$ and $s_{i,a}^{y,m}$; $N_a^{(y,m-1)(y,m)}$ označuje počet spárovaných položiek medzi mesiacmi m a $m-1$ roka y (kde $y, m-1$ je rovné $y-1, 12$ pre $m=1$). Každá položka i má priradenú pravdepodobnosť zahrnutia (implicitnú váhu) do vzorky $w_{i,a}^{y,m}$ nasledovne:

$$w_{i,a}^{y,m} = 1 \quad \text{ak} \quad \frac{s_{i,a}^{y,m-1} + s_{i,a}^{y,m}}{2} > \frac{1}{N_a^{(y,m-1)(y,m)} \chi}; \quad (7)$$

$$\text{inak} \\ w_{i,a}^{y,m} = 0 \quad (8)$$

To znamená, že ak podiel položky i na výdavkoch v mesiacoch m a $m-1$ prekročí prahovú hodnotu $\frac{1}{N_a^{(y,m-1)(y,m)} \chi}$, položka bude zahrnutá do vzorky a veľkosť vzorky elementárneho agregátu a je určená súčtom všetkých implicitných váh a je rovná:

$$\sum_{i=1}^{N_a^{(y,m-1)(y,m)}} w_{i,a}^{y,m} = n_a^{(y,m-1)(y,m)} \quad (9)$$

Na základe empirických skúseností sa odporúča nastaviť hodnotu parametra $\chi = 1,25$ pre všetky elementárne agregáty, t.j. pre všetky produktové kategórie na úrovni pseudo 6-miestnej klasifikácie COICOP (národná úroveň). Nastavenie hodnoty parametra $\chi = 1,25$ znamená, že ak napríklad $N_a^{(y,m-1)(y,m)} = 80$, potom položky s priemerným podielom na výdavkoch väčším ako 1% budú zahrnuté do vzorky.

Zmena ceny medzi $y, m-1$ a y, m pre elementárny agregát a je teraz počítaná ako:

$$\pi_a^{y,m/y,m-1} = \prod_{i=1}^{n_a^{(y,m-1)(y,m)}} \left(\frac{p_{i,a}^{y,m}}{p_{i,a}^{y,m-1}} \right)^{1/n_a^{(y,m-1)(y,m)}} \quad (10)$$

Rovnica (1) predstavuje medzimesačný Jevonsov index založený na vzorke. Tieto mesačné indexy sú následne reťazené, aby sme získali dlhodobý časový rad s ľubovoľným referenčným alebo počiatočným mesiacom y_0, m_0

$$P_a^{y,m/y_0,m_0} = P_a^{y,m-1/y_0,m_0} \pi_a^{y,m/y,m-1} \quad (11)$$

Kde $P_a^{y,m/y_0,m_0}$ je reťazený cenový index vypočítaný zo spárovaného súboru reprezentatívnych kódov položiek. Ceny položiek, ktoré nie sú dostupné v mesiaci y, m , ale ktoré sa predávali v predchádzajúcich obdobiach sú imputované a to nasledovne:

$$\hat{p}_{y,m}^{y,m} = p_{i,a}^{y,m-1} \pi_a^{y,m/y,m-1} \quad (12)$$

Na výpočet cenovej zmeny na vyššej úrovni agregácie sa používa index Laspeyresovho typu s fixnými váhami a s referenčným obdobím $y-1$.

$$P_A^{y,m/y-1} = \frac{\sum_{a \in A} w_a^{y-1} P_a^{y,m/y-1}}{\sum_{a \in A} w_a^{y-1}} \quad (13)$$

kde:

$$P_a^{y,m/y-1} = \frac{P_a^{y,m/y_0,m_0}}{\frac{1}{12} \sum_{s=1}^{12} P_a^{y-1,s/y_0,m_0}} \quad (14)$$

Váhy w_a^{y-1} v (13) sú založené na ročných výdavkoch za všetky položky patriace do elementárneho agregátu a bez ohľadu na to, či boli zahrnuté do vzorky. Ďalej sú krátkodobé časové rady reťazené v decembri, ktorý vytvára prepojenie na konštrukciu dlhodobého časového radu s referenčným obdobím 0.

$$P_{ch,A}^{y,m/0} = \left(\frac{P_A^{y,m/y-1}}{P_A^{y-1,12/y-1}} \right) \left[\prod_{\tau=1}^{y-1} \frac{P_A^{\tau,12/\tau-1}}{P_A^{\tau-1,12/\tau-1}} \right] P_A^{0,12/0} \quad (15)$$

Krátkodobé indexy $P_A^{y,m/y-1}$ a reťazené indexy $P_{ch,A}^{y,m/0}$ sú počítané pre všetky úrovne COICOP, pre všetky obchodné reťazce a cez všetky reťazce použijúc vzťahy (13) a (15).

4.3.2 Multilaterálne cenové indexy

Metódy multilaterálnych cenových indexov sa zvyčajne používajú na porovnanie cenových úrovní medzi krajinami a regiónmi tak, aby porovnania boli tranzitívne, t. j. výsledky sú nezávislé na výbere báze krajiny. Najznámejšie metódy sú GEKS [5], [1], [20], Geary-Khamisova metóda [4], [17] a Country-Product Dummy (CPD) metóda navrhnutá v [19].

Multilaterálne priestorové porovnanie cien môže byť jednoducho prispôbené na porovnanie v čase. Multilaterálne indexy spĺňajú požiadavku *cirkularity* alebo „tranzitivity“, to znamená, že sa dopracujeme k rovnakým výsledkom či už porovnáваме entity navzájom priamo, alebo prostredníctvom vzťahov s inými entitami. Štandardné bilaterálne indexy nespĺňajú túto podmienku.

Rozlišujeme medzi dvoma typmi multilaterálnych metód. Prvý typ vychádza zo súboru bilaterálnych indexov počítaných nad databázou spárovaných položiek t. j. porovnanie cien vždy medzi dvoma časovými obdobiami a potom „tranzitizácia“ bilaterálnych cenových indexov (napr. GEKS). Druhý typ multilaterálnych metód dosahuje tranzitivitu iným spôsobom (napr. Geary-Khamis metóda, CPD)

Metóda (RY)GEKS

Predstavuje alternatívu k metóde opísanej v predchádzajúcej podkapitole (4.2.1) a jej riešenie bolo navrhnuté autormi [15]. Autori prispôbili metódu GEKS, známu z literatúry ako metódu na medzinárodné (priestorové) porovnanie cien, na porovnanie cien v čase.

Tranzitívny multilaterálny index GEKS [5], [1], [20] je geometrický priemer pomerov všetkých bilaterálnych (v štandardnej verzii Fisherových) indexov, kde každá entita je uvažovaná ako báza.

Výhodou tohto prístupu oproti priamym indexom založeným na fixnej báze (spotrebiteľskom koši) je, že môžeme použiť flexibilný kôš, na každé z bilaterálnych porovnaní v indexe GEKS. Toto je tiež výhoda použitia reťazových indexov, ktoré nám dovoľujú porovnať údaje všetkých položiek, ktoré sú prítomné v dvoch porovnávaných časových obdobiach.

Nech P^{jl} a P^{kl} sú bilaterálne indexy medzi párom entít j a l ($l = 1, \dots, M$) a medzi entitami k a l . Index GEKS medzi j a k je možno napísať ako:

$$P_{GEKS}^{jk} = \prod_{l=1}^M [P^{jl} / P^{kl}]^{1/M} = \prod_{l=1}^M [P^{jl} \times P^{lk}]^{1/M} \quad (16)$$

Druhý výraz v rovnici (16) platí, ak bilaterálne cenové indexy spĺňajú tzv. „reverzný test entity“ t. j.

$$P^{kl} = 1 / P^{lk} \quad (17)$$

Dá sa dokázať, že:

$$P_{GEKS}^{jk} = P_{GEKS}^{jl} / P_{GEKS}^{kl} \quad (18)$$

Výraz (18) hovorí, že cenový index GEKS spĺňa požiadavku cirkularity alebo transitivity, t. j. rovnaký výsledok sa dosiahne, ak sa entity porovnávajú priamo medzi sebou alebo prostredníctvom vzťahov s inými entitami. Na porovnanie cien v čase sú entity v (16) časovými obdobiami, na účely výpočtu CPI sú to zvyčajne mesiace. Uvažujem rozsah $[0, T]$ z $T+1$ mesiacov, kde T je aktuálny (posledný) mesiac. Index GEKS za mesiac t ($t \leq T$) s referenčným obdobím 0 je potom:

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^T [P^{0,\tau} / P^{t,\tau}]^{1/(T+1)} = \prod_{\tau=0}^T [P^{0,\tau} \times P^{\tau,t}]^{1/(T+1)} \quad (19)$$

O mesiac neskôr je $T+1$ aktuálny mesiac a časové rady sa počítajú pre $[0, T+1]$. Index GEKS za mesiac t ($t \leq T+1$) s referenčným obdobím 0 sa teraz stáva:

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^{T+1} [P^{0,\tau} / P^{t,\tau}]^{1/(T+2)} = \prod_{\tau=0}^{T+1} [P^{0,\tau} \times P^{\tau,t}]^{1/(T+2)} \quad (20)$$

Problémom tohto prístupu je, že výsledky vypočítané pre všetky mesiace $1, \dots, T$ použitím (20) sa budú líšiť od tých, ktoré budú vypočítané použitím (19). To znamená, že keď je časový interval pozorovaní rozšírený a sú pridané nové údaje, GEKS indexy podliehajú nepretržitým revíziám, čo je v prípade výpočtu HICP/CPI neprijateľné. Riešením tohto problému by bolo použitie aktualizácie indexu medzi T a $T+1$, vypočítanej pomocou (20), na aktualizáciu časového radu (19) prostredníctvom reťazenia. Tak ako je konštruovaná, je táto GEKS mesačná zmena ceny ovplyvnená cenami a predanými množstvami všetkých mesiacov $0, \dots, T+1$, teda vrátane $0, \dots, T-1$ mesiacov. Ak by sme takto pokračovali v predlžovaní časových radov, zmena indexu medzi poslednými dvoma mesiacmi by bola ovplyvnená cenami a množstvami, ktoré sa stále viac a viac vzťahujú na mesiace v minulosti. Autori [15] preto navrhli metódu „kízavého roka“ („rolling year“). Metóda kízavého roka na výpočet indexu GEKS používa 13-mesačné kízavé okno, pretože je to najkratšie okno, ktoré umožňuje porovnanie položiek silne zaťažených sezónnosťou. Index GEKS (19) pre obdobia $\{0, 12\}$ zostáva východiskovým bodom. Ako už bolo uvedené, ďalší medzimesačný index je zreťazený s GEKS časovým radom. Tento index sa vypočíta ako pomer indexov GEKS, na základe údajov $\{1, 13\}$ za mesiace 13 a 12 s referenčným obdobím 1, podľa vzťahu:

$$P_{GEKS}^{1,13} = \prod_{\tau=1}^{13} [P^{1,\tau}/P^{13,\tau}]^{1/13} \quad (21)$$

$$P_{GEKS}^{1,12} = \prod_{\tau=1}^{13} [P^{1,\tau}/P^{12,\tau}]^{1/13} \quad (22)$$

Pretože GEKS indexy sú tranzitívne, platí:

$$P_{GEKS}^{12,13} = P_{GEKS}^{12,1}/P_{GEKS}^{13,1} = P_{GEKS}^{1,13}/P_{GEKS}^{1,12} \quad (23)$$

To znamená, že namiesto pomeru (21) a (22) GEKS index medzi mesiacmi 12 a 13 sa rovná:

$$P_{GEKS}^{12,13} = \prod_{\tau=1}^{13} [P^{12,\tau}/P^{13,\tau}]^{1/13} = \prod_{\tau=1}^{13} [P^{\tau,13}/P^{\tau,12}]^{1/13} \quad (24)$$

môže byť použitý na aktualizáciu existujúceho časového radu. Výraz pre „rolling year“ GEKS (RYGEKS) index, týkajúci sa 13. mesiaca je potom nasledujúci:

$$P_{RYGEKS}^{0,13} = P_{GEKS}^{0,12} \prod_{\tau=1}^{13} [P^{12,\tau}/P^{13,\tau}]^{1/13} = \prod_{\tau=0}^{12} [P^{0,\tau}/P^{12,\tau}]^{1/13} \prod_{\tau=1}^{13} [P^{12,\tau}/P^{13,\tau}]^{1/13} \quad (25)$$

Všeobecný výraz pre RYGEKS index týkajúci sa aktuálneho mesiaca T , ($T > 12$) s referenčným obdobím 0 je:

$$P_{RYGEKS}^{0,T} = \prod_{\tau=0}^{12} [P^{0,\tau}/P^{12,\tau}]^{1/13} = \prod_{t=13}^T \prod_{\tau=t-12}^t [P^{t-1,\tau}/P^{t,\tau}]^{1/13} \quad (26)$$

A všeobecný výraz pre medzimesačný index je:

$$\prod_{\tau=t-12}^t [P^{t-1,\tau}/P^{t,\tau}]^{1/13} = \left[\frac{P^{t-12,t}}{P^{t-12,t-1}} \times \frac{P^{t-11,t}}{P^{t-11,t-1}} \times \dots \times \frac{P^{t-3,t}}{P^{t-3,t-1}} \times \frac{P^{t-2,t}}{P^{t-2,t-1}} \times (P^{t-1,t})^2 \right] \quad (27)$$

Vzorec (27) poskytuje užitočné informácie o skutočnom postupe výpočtu. Medzimesačná zmena indexu RYGEKS sa jednoducho rovná (geometrickému)

priemeru priameho indexu medzi $t-1$ a t , $P^{t-1, t}$, ktorý sa počíta dvakrát, a medzi nimi 11 nepriamych mesačných indexov. Autori [15] použili indexy cien Fishera. V ďalšej štúdií [9] boli použité namiesto Fisherových, Törnqvistove indexy. Holandský štatistický úrad tak isto testuje použitie metódy RYGEKS na báze Törnqvistoveých indexov.

Ďalšie metódy, ktoré by mohli byť v centre našej pozornosti sú metódy GEKS-Törnqvist a Geary-Khamis a TPD metóda, ktoré v súčasnosti testuje napr. Belgický štatistický úrad, ten deklaroval postupný prechod na metódu multilaterálnych indexov v časovom horizonte 2022 [18].

GEKS-Törnqvist

Metóda GEKS-Törnqvist [16] využíva všetky možné spárované produkty a počíta cenový index medzi mesiacmi 0 a t ako nevážený geometrický priemer $T+1$ pomerov bilaterálnych cenových indexov vypočítaný na základe spárovaného súboru kódov P^{0l} a P^{lt} pre $l \in (0, T)$:

$$P_{GEKS}^{0,t} = \prod_{l=0}^T [P^{0,l}/P^{l,t}]^{1/(T+1)} = \prod_{l=0}^T [P^{0,l}P^{l,t}]^{1/(T+1)} \quad (28)$$

Indexy $P^{0,l}$ a $P^{l,t}$ sú bilaterálne Törnqvist indexy medzi obdobím 0 a l a obdobím l a t . Törnqvist index je definovaný ako:

$$P_T^{0,t} = \prod_{i=1}^n \left[\frac{p_i^t}{p_i^0} \right]^{0.5(s_i^0 + s_i^t)} \quad (29)$$

s trhovým podielom produktu s_i^t respektíve (s_i^0) v období t (resp. 0).

Geary-Khamis

Metóda Geary-Khamis [13] sa začína konceptom jednotkovej hodnoty (ceny za jednotku tovaru). Produkt i má jednotkovú cenu p_i^t a predané množstvo q_i^t . Agregácia kvantity je náročná, ak skupina výrobkov nie je homogénna. Metóda Geary-Khamis rieši tento problém použitím faktorov upravujúcich kvalitu v_i . Faktory úpravy kvality transformujú predané množstvá v bežných jednotkách $v_i q_i^t$ a ceny sa stanú cenami upravenými o kvalitu $\frac{p_i^t}{v_i}$. Výsledkom je cena za jednotku tovaru upravená o kvalitu (QU) \tilde{p}^t pre skupinu výrobkov v mesiaci t .

$$\tilde{p}^t = \frac{\sum_{i \in G_t} p_i^t q_i^t}{\sum_{i \in G_t} v_i q_i^t} \quad (30)$$

Cenový Geary-Khamis index medzi mesiacmi t a 0 môže byť vyjadrený ako:

$$P^{0,t} = \frac{\tilde{p}^t}{\tilde{p}^0} = \frac{\sum_{i \in G_t} p_i^t q_i^t / \sum_{i \in G_0} p_i^0 q_i^0}{\sum_{i \in G_t} v_i q_i^t / \sum_{i \in G_0} v_i q_i^0} \quad (31)$$

Tento výraz môžeme považovať za pomer indexu tržieb a váženého indexu kvantity. V Geary-Khamisovej metóde sa váhy definujú nasledovne:

$$v_i = \frac{\sum_{z=0}^T q_i^z q_i^z / P^{0,z}}{\sum_{z=0}^T q_i^z} \quad (32)$$

Táto hodnota jednotky upravená o kvalitu používa všetky dostupné údaje mesiacov 0 až T . Pretože cenový index sa používa na výpočet faktorov pre úpravy kvality, ktoré sa samotné používajú na výpočet cenového indexu, obe rovnice (31) aj (32), musia byť riešené súčasne pomocou iteračnej metódy.

TPD metóda

TPD (Time Product Dummy) metóda [12] používa regresný prístup na odhad cenových indexov zo všetkých dostupných údajov v časovom okne. Model s N rôznymi položkami v rámci časového obdobia $[0, T]$ možno napísať ako:

$$\ln p_i^t = \alpha + \sum_{t=1}^T \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \quad (33)$$

Parametre δ^t sú umelé parametre času, γ_i predstavujú fixné efekty položiek. Umelá premenná času D_i^t má hodnotu 1, ak je produkt i k dispozícii v období t , a 0, ak nie je k dispozícii. Umelá premenná D_i má hodnotu 1, ak sa pozorovanie týka položky i , inak má hodnotu 0. Ceny upravené na kvalitu súboru výrobkov G_t v mesiaci t možno písať ako:

$$\tilde{p}^t = \prod_{i \in G_t} \left[\frac{p_i^t}{v_i} \right]^{s_i^t} \quad (34)$$

kde je cena upravená použitím fixných efektov položky: $v_i = \exp(\gamma_i)$, ktoré sú nazývané faktory upravujúce kvalitu. Trhový podiel položky i v mesiaci t je označený ako s_i^t .

TPD index potom môžeme napísať ako:

$$P_{TPD}^{0,t} = \frac{\tilde{p}^t}{\tilde{p}^0} = \frac{\prod_{i \in G_t} \left(\frac{p_i^t}{\exp(\gamma_i)} \right)^{s_i^t}}{\prod_{i \in G_0} \left(\frac{p_i^0}{\exp(\gamma_i)} \right)^{s_i^0}} \quad (35)$$

5. ZÁVER

Implementácia nových zdrojov údajov do produkcie cenovej štatistiky predstavuje významnú metodologickú zmenu, ktorej zavedenie môže významne ovplyvniť tak úroveň cenových indexov, ako aj dynamiku ich zodpovedajúcej mesačnej alebo ročnej miery zmeny.

Prechod od údajov zo zisťovaní ku „scanner data“ všeobecne znamená prechod od fixného spotrebného koša do dynamického sveta, ktorý obsahuje všetok predaný tovar. Kým pre fixný koš môžu byť vhodné klasické bilaterálne indexy, prechod na dynamický prístup vyvoláva otázku, či sú tradične používané metódy ešte stále platné.

Pri použití „scanner data“ na výpočet HICP/CPI je pre štatistické úrady prioritou zavedenie robustných metód, ktoré zabezpečia maximálne spárovanie produktov v čase a elimináciu problému s driftom v časovom rade reťazených indexov.

Autori [7] publikovali empirické porovnanie Holandskej metódy spracovania údajov zo supermarketov s alternatívnou metódou RYGEKS. Z uvedenej štúdie vyplynulo,

že z teoretického hľadiska je metóda RYGEKS nepochybne lepšia ako Holandská metóda. Metóda RYGEKS vytvára vážené indexy, je založená na bilaterálnych hodnotách superlatívnych indexov (ktoré majú základ v ekonomickej teórii a majú uspokojivé axiomatické vlastnosti) a optimálne využíva všetky dostupné informácie o cene a množstve bez driftu v reťazených časových radoch. Ich empirické dôkazy tento názor podporujú.

V Holandskej metóde bola zavedená umelá agregačná 6-miestna úroveň klasifikácie COICOP, aby bolo možné počítať elementárny agregát na nižšej ako 5-miestnej úrovni klasifikácie. Táto úroveň umelej agregácie je odlišná pre každý obchodný reťazec, čo je pre ďalšiu agregáciu trochu nepraktické. Keďže sa používa ročné reťazenie, váhy sa na tejto úrovni musia každý rok aktualizovať.

Metódu RYGEKS je možné aplikovať priamo na 5-miestnej úrovni COICOP – nie je potrebné ďalšie členenie, pokiaľ si používatelia nebudú želať mať v budúcnosti podrobnejšie údaje, čo je nepravdepodobné – a výsledné indexy sa môžu prenášať priamo do agregáčného modulu CPI. Navyše metóda RYGEKS sa môže a pravdepodobne by sa mala uplatňovať na všetky položky patriace do 5-miestnej kategórie COICOP namiesto vzorky (Holandská metóda), t. j. systém by bol jednoduchší. Nevýhodou použitia metódy RYGEKS môže byť zvýšená volatilita časových radov. Na druhej strane zvýšenie volatility nemusí byť problémom, ak opisuje skutočný stav.

Aj keď by bol systém pri používaní RYGEKS vo viacerých ohľadoch jednoduchší, metóda samotného výpočtu by bola zložitejšia, lebo na výpočet aktuálneho indexu sa používajú údaje týkajúce sa 13 mesiacov namiesto 2 mesiacov. Komplikujúcim faktorom môže byť aj to, že nie je možné rozložiť zmenu v indexe RYGEKS presne podľa príspevkov jednotlivých položiek, ako je to možné pri bežných metódach. Cenoví štatistici vo všeobecnosti považujú takýto rozklad za užitočný proces validácie indexov. Autor [6] ukázal, s výnimkou veľmi neobvyklých okolností, že položky s najväčšou váženou zmenou ceny medzi poslednými dvomi mesiacmi majú dominantný vplyv na zmenu indexu RYGEKS. Zdá sa teda, že by nebolo potrebné meniť zaužívaný mesačný proces validácie. Autori experimentálnej štúdie odporúčajú prejsť na metódu RYGEKS, pretože podľa ich názoru metodické výhody jednoznačne prevažujú nad akýmikoľvek praktickými nedostatkami.

Čo sa týka situácie na Slovensku v tejto oblasti, ŠÚ SR má k dnešnému dňu na základe Programu štátnych štatistických zisťovaní a bilaterálnych dohôd s najvýznamnejšími obchodnými reťazcami k dispozícii na týždennej báze niekoľko súborov „scanner data“, na ktorých môže začať analyzovať vplyv použitia rôznych metód na agregované cenové zmeny, t. j. vážené vs nevážené metódy, mesačné reťazené vs priame indexy a bilaterálne vs multilaterálne metódy. Popis jednotlivých metód v tomto dokumente a výsledky výskumu, empirických štúdií a dobrej praxe štatistických úradov môžu slúžiť ako štartovací bod pre realizáciu takýchto empirických štúdií na ŠÚ SR.

Príspevok predstavuje prvú časť konceptuálnej štúdie zameranej na implementáciu nových zdrojov údajov v oblasti cenovej štatistiky. Metodické úvahy v súvislosti s konštrukciou „unit value“ v podmienkach SR ako aj grafické ilustrácie

výsledkov experimentálnych štúdií (takisto v podmienkach SR) budú súčasťou nasledujúcich plánovaných príspevkov.

LITERATÚRA

- [1] ELTETÖ, Ö. – KÖVES, P.: On a Problem of Index Number Computation Relating to International Comparisons (in Hungarian). In.: Statisztikai Szemle, 1964, č. 42, s. 507 – 518.
- [2] EUROPEAN COMMISSION, EUROSTAT: Harmonised Index of Consumer Prices (HICP): Methodological Manual. Luxembourg: Publications Office of the European Union, 2018. 354 s. ISBN 978-92-79-76861-3.
- [3] EUROPEAN COMMISSION, Eurostat (September 2017): HICP, Practical Guide for Processing Supermarket Scanner Data, [online]. [Dostupné na: <https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf>, prístup 13. 5. 2019].
- [4] GEARY, R.C.: A Note on the Comparison of Exchange Rates and Purchasing Power between Countries. In.: Journal of the Royal Statistical Society A 121, 1958, č. 1, s. 97 – 99.
- [5] GINI, C.: On the Circular Test of Index Numbers. In: International Review of Statistics, 1931, s.3 – 25.
- [6] VAN DER GRIENT, H. A.: How to Analyse Changes in Price Indexes when Applying the RYGEKS Formula, The Hague: Statistics Netherlands, 2010.
- [7] VAN DER GRIENT, H. A. – DE HAAN, J.: Scanner Data Price Indexes: The "Dutch" Method versus Rolling Year GEKS. Paper presented at the 12th Ottawa Group meeting, 4 – 6 May 2011. Wellington, New Zealand. [Dostupné na: <https://www.researchgate.net/publication/264881843>, prístup 4. 6. 2019].
- [8] VAN DER GRIENT, H. A. – DE HAAN, J.: Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data, Paper presented at the ninth Economic Measurement Group Workshop, Sydney, 10. – 12. December 2008.
- [9] DE HAAN, J. – VAN DER GRIENT, H. A.: Eliminating Chain Drift in Price Indexes Based on Scanner Data, 2009. [online]. [Dostupné na: [http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1bd88ae9af79cfa1ca257693001bb7fa/\\$FILE/2009%2011th%20meeting%20-%20Jan%20de%20Hann%20Heymerik%20van%20der%20Grient%20\(Statistics%20Netherlands\)_Eliminating%20Chain%20Drift%20in%20Price%20Indexes.pdf](http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1bd88ae9af79cfa1ca257693001bb7fa/$FILE/2009%2011th%20meeting%20-%20Jan%20de%20Hann%20Heymerik%20van%20der%20Grient%20(Statistics%20Netherlands)_Eliminating%20Chain%20Drift%20in%20Price%20Indexes.pdf), prístup 13. 7. 2019]
- [10] DE HAAN, J. – VAN DER GRIENT, H. A.: Eliminating chain drift in price indexes based on scanner data. In: Journal of Econometrics, 2011, č.1, s. 36 – 46.
- [11] DE HAAN, J. – HENDRIKS, R. – SCHOLZ, M.: A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes In: Graz Economics Paper 2016-13.
- [12] DE HAAN, J. – KRSINICH, F.: Scanner Data and the Treatment of Quality Change in Non-Revisable Price Indexes. In.: Journal of Business and Economic Statistics, 2014, č. 3, s. 341 – 358.
- [13] CHESSA, A.: A new methodology for processing scanner data in the Dutch CPI. In.: Eurostat review of National Accounts and Macroeconomic Indicators, 2016, č. 1, s. 49 – 69.
- [14] ILO – IMF – OECD – UNECE – Eurostat – The World Bank: Consumer price index manual: Theory and practice. Geneva: International Labour Office, 2004, 535 s., ISBN 92-2-113699-X

- [15] IVANCIC, L. – DIEWERT, W. E. – FOX, K. J.: Scanner Data, Time Aggregation and the Construction of Price Indexes, Discussion Paper no. 09-09, Vancouver: Department of Economics, University of British Columbia, 2009.
- [16] IVANCIC, L. – DIEWERT, W. E. – FOX, K. J.: Scanner Data, Time Aggregation and the Construction of Price Indexes. In.: Journal of Econometrics, 2011, č. 1, s. 24 – 35.
- [17] KHAMIS, S. H.: A New System of Index Numbers for National and International Purposes, In.: Journal of the Royal Statistical Society A, 1972, č. 1, s. 96 – 121.
- [18] VAN LOOK, K. – ROELS, D: Integrating big data in the Belgian CPI, Meeting of the Group of Experts on Consumer Price Indices. Geneva, 2018. 28 s.
- [19] SUMMERS, R.: International Price Comparisons Based Upon Incomplete Data. In: Review of Income and Wealth, 1973, č. 19, s. 1 – 16.
- [20] SZULC, B.: Indices for Multiregional Comparisons (in Polish). In: Przegląd Statystyczny, 1964, č. 3, s. 239 – 254.

RESUMÉ

V súčasnosti je k dispozícii množstvo prístupov ako konštruovať cenové indexy používajúc „scanner data“. „Scanner data“ sú transakčné údaje, obsahujúce transakcie všetkých tovarov, ktoré boli predané, ceny aktuálne zaplatené spotrebiteľom a predané množstvá za každú tovarovú položku. Naopak, ceny zbierané tradičným spôsobom sú ceny, za ktoré sa tovar ponúka, predané množstvá nie sú k dispozícii a tradičný spotrebný kôš je relatívne malá vzorka kompletnej množiny tovarov. Príspevok je zameraný hlavne na metódy výpočtu indexov, ktoré môžu byť aplikované na „scanner data“ a ktoré využívajú nielen cenové ale aj kvantitatívne informácie obsiahnuté v súbore údajov, t. j. viac na multilaterálne metódy ako na tradičné metódy založené na spotrebnom koši. „Scanner data“ odrážajú dynamiku skutočných nákupov, pretože je zaznamenaná každá transakcia. V súbore údajov je viditeľný vznik nových tovarových položiek, zánik položiek a zmeny ich relatívnej dôležitosti. V závislosti od krajiny v priebehu roka zaniká 25 % až 60 % tovarových položiek.

Pre štatistické úrady je preto prioritou zavedenie robustných metód, ktoré zabezpečia maximálne spárovanie predávaných produktov v čase a elimináciu problému s driftom v časovom rade reťazených indexov. Opis jednotlivých metód v tomto dokumente popri výsledkoch výskumu, empirických štúdií a dobrej praxe iných štatistických úradov môžu slúžiť ako štartovací bod na realizáciu takýchto empirických štúdií na ŠÚ SR a následnú implementáciu tohto zdroja údajov do produkcie cenovej štatistiky.

RESUME

A number of approaches associated with the task of constructing price indices using a "scanner data" are now available. The conception "scanner data" means transaction data, covering transactions of all goods that have been sold, prices actually paid by the consumers and quantities sold for each item. Conversely, prices collected in the traditional way are prices at which the goods are offered, the quantities sold are not available, and the traditional consumer basket is a relatively small sample of the complete set of goods. The paper focuses mainly on index calculation methods that can be applied to the "scanner data" and which use not only

the price but also the quantitative information contained in the data set, i.e., it is focused more on multilateral methods than on the traditional small-sample based methods. The “scanner data” reflect the dynamics of actual purchases, since every transaction is recorded. The emergence of new commodities, the disappearance of items and changes of their relative importance are visible in the dataset. In the course of the year, from 25 % up to 60 % of the commodities are ceased, of course depending on the country. Therefore the introduction of robust methods that ensure maximum matching of products sold over time and eliminate the drift problem in the time series of chained indices has become the priority for the statistical institutions. The description of individual methods in this paper as well as the results of research, empirical studies and good practice of other statistical institutions, can serve as a starting point for the conduct of such empirical studies at the Statistical Office of the SR and the subsequent implementation of this type of data source into the production of price statistics.

PROFESIJNÝ ŽIVOTOPIS

Ing. Helena Glaser-Opitzová je generálna riaditeľka sekcie všeobecnej metodiky a registrov Štatistického úradu SR a členka riaditeľskej skupiny Eurostatu pre metodológiu (DIME), ktorá poskytuje poradenstvo Európskemu štatistickému výboru (ESSC) v strategických otázkach. Zaoberá sa predovšetkým efektívnymi metódami na realizáciu štatistických zisťovaní s dôrazom na štandardizáciu štatistických procesov, systémom monitorovania, reportovania a vyhodnocovania kvality štatistických zisťovaní a ich produktov, sezónnou analýzou časových radov, ochranou dôverných štatistických údajov, využívaním administratívnych zdrojov údajov na štatistické účely. Viedla mnohé modernizačné aktivity úradu alebo sa na nich podieľala. V súčasnosti riadi interný projekt úradu zameraný na modernizáciu cenových štatistík.

KONTAKT

helena.glaser-opitzova@statistics.sk