

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

3/2023

ročník/volume 33

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

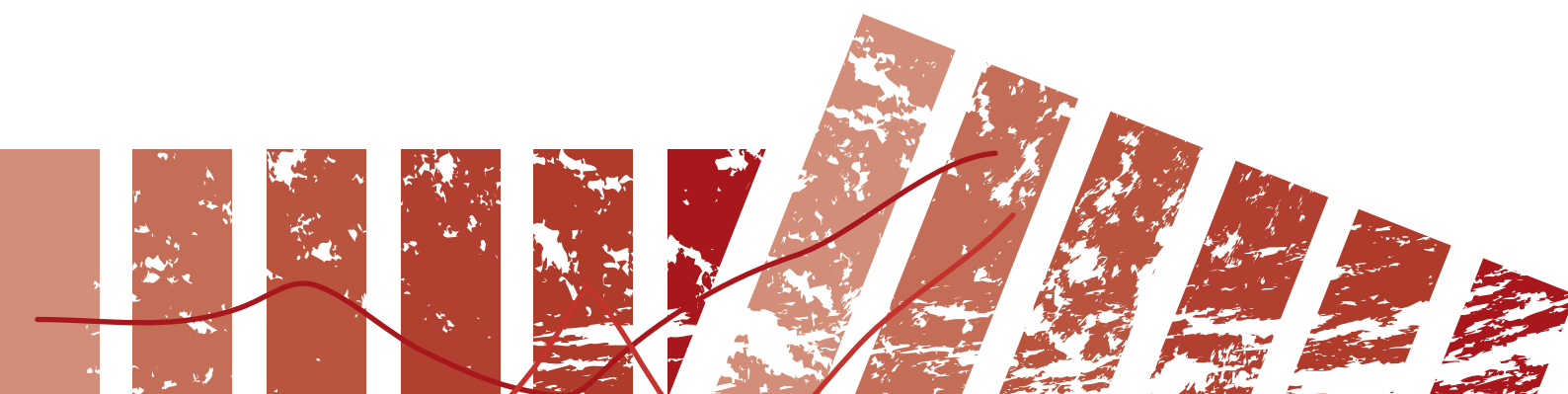
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 4

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

Strany/Pages: 39 – 54

Dátum vydania/Publication date: 15. júl 2023/July 15, 2023



Helena GLASER-OPITZOVÁ, Petra MAZUREKOVÁ
Štatistický úrad Slovenskej republiky

VPLYV PARCIÁLNEHO ČASOVÉHO POKRYTIA ÚDAJOV ZO SKENEROV NA PRESNOSŤ CENOVÝCH INDEXOV

EFFECT OF PARTIAL TIME COVERAGE OF SCANNER DATA ON THE ACCURACY OF PRICE INDICES

ABSTRAKT

V procese modernizácie a skvalitňovania cenových štatistík sa Štatistický úrad Slovenskej republiky snaží o implementáciu nových zdrojov údajov. Jedným z možných nových zdrojov sú transakčné údaje obchodných reťazcov, nazývané aj údaje zo skenerov. Využívanie takýchto zdrojov však predstavuje významnú metodologickú zmenu. Jednou zo zásadných zmien je zmena cenového konceptu. Ceny získavané tradičným spôsobom ako tzv. pultové ceny, ktoré sa zisťujú v konkrétnom čase, sú nahradené priemernými cenami za jednotku tovaru. Na presnosť jednotkovej ceny má vplyv obdobie, ktoré berieme do úvahy pri jej stanovení. Cieľom tohto článku je zistiť, či a aký vplyv má navrhované parciálne časové pokrytie sledovaného obdobia (2 týždne vs. mesiac) na hodnoty mesačných cenových indexov.

ABSTRACT

In the process of modernizing and improving price statistics, the Statistical Office of the Slovak Republic is attempting to implement new data sources. One of the possible new sources is transaction data from retail chains, also known as scanner data. However, the use of such sources represents a significant methodological change. One of the fundamental changes is the shift in the price concept. Prices obtained in the traditional way, known as shelf prices, which are determined at a specific time, are replaced by average prices per unit of goods. The accuracy of the unit price is influenced by the period taken into account when determining it. The aim of this article is to determine whether and what impact the proposed partial time coverage of the observed period (2 weeks vs a month) has on the values of monthly price indices.

KLÚČOVÉ SLOVÁ

cenové indexy, transakčné údaje, bilaterálne indexy, multilaterálne indexy

KEY WORDS

price indices, transaction data, bilateral indices, multilateral indices

1. ÚVOD

Používané zdroje údajov vo všeobecnosti významným spôsobom ovplyvňujú kvalitu štatistického produktu a efektivitu štatistického procesu.

V oblasti cenovej štatistiky sú dátové zdroje o cenách tovarov a služieb v podmienkach Štatistického úradu SR stále získavané prostredníctvom terénneho zberu údajov. Zber údajov sa realizuje priamo v prevádzkach a obchodoch na celom území SR, kde obyvatelia obvykle nakupujú, a ceny, ktoré sa zisťujú, sú tzv. pultové ceny. Ceny sa zisťujú počas prvých 20 dní sledovaného mesiaca a pri tradičnom zbere ide o ceny zisťované v konkrétnom čase [22].

V súčasnosti sa vo svete tradičný zber cien v prevádzkach a obchodoch postupne nahrádza, najmä v oblasti cien potravín, nealkoholických nápojov a tovarov bežnej spotreby údajmi zo skenerov alebo údajmi získanými formou web scrapingu [20]. Meranie cien na základe údajov zo skenerov je už niekoľko rokov aktívnou oblasťou výskumu aj v Štatistickom úrade SR. Ich použitie okrem iných významných metodologických zmien prináša aj zmenu cenového konceptu. Ceny získavané tradičným spôsobom sú nahradené cenami za jednotku tovaru.

Podľa [1] jednotková cena presnejšie odráža ceny, ktoré platia spotrebiteľia počas celého sledovaného obdobia, ako zistenie ceny v konkrétnom čase pri tradičnom zbere. Jednotkové ceny zohľadňujú zľavy a ich vplyv na množstvo predaného tovaru. Na presnosť jednotkovej ceny má vplyv obdobie, ktoré berieme do úvahy pri jej stanovení (napr. mesiac vs len 2 alebo 3 týždne referenčného mesiaca). [4] tvrdia, že jednotkové ceny používané na konštrukciu CPI¹ by mali byť vypočítané za rovnaké obdobie ako je obdobie, za ktoré sa zostavuje index (napr. mesiac), a nie za čiastkové obdobie. V praxi však štatistické úrady zodpovedné za zostavenie HICP²/CPI štandardne používajú čiastkové obdobie referenčného obdobia z dôvodu včasnosti³ a časovej presnosti⁴ poskytovania štatistických produktov. Vzhľadom na termíny poskytovania transakčných údajov zo strany obchodných reťazcov a publikačnú prax Štatistického úradu SR predpokladáme využívať na stanovenie jednotkovej ceny prvé dva kompletne týždne referenčného mesiaca. Vzhľadom na uvedené skutočnosti je nanajviš žiaduce vplyv nedostatočného časového pokrytia kvantifikovať a vyhodnotiť.

2. VLASTNOSTI TRANSAKČNÝCH ÚDAJOV

Transakčné údaje nazývané aj údaje zo skenerov sú pre štatistické úrady relatívne novým zdrojom údajov, ich dostupnosť sa však v posledných rokoch zvyšuje. Sú to údaje, ktoré zaznamenávajú maloobchodníci pri nákupoch spotrebiteľov skenovaním čiarových kódov. Tieto údaje obsahujú pre každý predaný tovar v obchode v danom období predané množstvo a predajnú/realizačnú cenu na úrovni kódu položky.

Transakčné údaje za konkrétneho maloobchodníka a časové obdobie predstavujú tak vyčerpávajúci zoznam všetkých kódov položiek, ktoré boli predané, ich tržby a predané množstvá. Umožňujú zostaviť index zo všetkých transakcií maloobchodníka alebo obchodu, pričom do výpočtu vstupujú realizačné ceny produktov a umožňujú zahrnúť do CPI/HICP oveľa viac položiek v porovnaní s tradičným zberom cien. Na porovnanie môžeme uviesť, že pri tradičnom zbere údajov v súvislosti so spotrebiteľskými cenami tovarov a služieb za oblasť potravín a nealkoholických nápojov je každá z vybraných predajní navštívená jedenkrát v mesiaci a za každého reprezentanta sa v predajni vyberie len jeden produkt (spolu cca 142 konkrétnych cien) [22] Oproti tomu pri údajoch zo skenerov získame informáciu za celý predaj daného

¹ CPI (Consumer Price Index) – Index spotrebiteľských cien meria celkovú zmenu spotrebiteľských cien na základe reprezentatívneho spotrebného koša tovarov a služieb v čase [22].

² HICP (Harmonised Index of Consumer Prices) – Harmonizovaný index spotrebiteľských cien bol vytvorený s cieľom poskytnúť vysokokvalitný, porovnateľný ukazovateľ inflácie spotrebiteľských cien [22].

³ Včasnosť – vzťahuje sa na obdobie medzi dostupnosťou informácie a udalosťou alebo javom, ktorý opisuje.

⁴ Časová presnosť – súvisí s časovým rozdielom medzi termínom zverejnenia údajov a cieľovým termínom, keď mali byť údaje dodané napríklad v súvislosti s termínmi zverejnenými v niektorom oficiálnom kalendári, ustanovenom predpismi alebo predchádzajúcou dohodou medzi partnermi (odchýlka od harmonogramu zverejnenia).

produktu v týždennom agregáte. To však znamená, že pri použití týchto údajov na účely zostavenia HICP/CPI musí dôjsť k zmene cenového konceptu. Do výpočtu cenových indexov tak nebude vstupovať cena zistená v konkrétnom čase, ale priemerná cena za jednotku tovaru za dané sledované obdobie určená takto:

$$\text{priemerná cena} = \text{tržby} / \text{počet predaných kusov}$$

Znamená to tiež, že ak sú k dispozícii informácie o tržbách, resp. predaných množstvách, môžeme každej položke priradiť váhu, čo otvára možnosti použiť na zostavenie cenových indexov na úrovni elementárneho agregátu⁵ aj indexové vzorce pre superlatívne indexy, ktoré sú najpreferovanejšími indexmi na účely merania CPI [3]. Superlatívne indexy využívajú ceny a množstvá (t. j. výdavkové váhy) v oboch porovnávaných obdobiach (referenčné obdobie a bežné sledované obdobie).

V doterajšej praxi váhy výdavkov bežného obdobia, ale ani referenčného obdobia nie sú známe, takže v praxi sa štatistickí pri zostavení CPI spoliehajú na elementárnej úrovni na nevážené indexy a pri agregácii na vyššiu úroveň klasifikácie ECOICOP⁶ využívajú fixné váhy, ktoré sa vzťahujú na predchádzajúci rok.

3. ZDROJE ÚDAJOV A PRVOTNÉ SPRACOVANIE

V súčasnosti Štatistický úrad SR preberá transakčné údaje za oblasť potravín a nealkoholických nápojov od piatich obchodných reťazcov na týždennej báze. Dátové súbory obsahujú tržby a predané množstvá jednotlivých tovarov vo forme týždenného agregátu (od pondelka do nedele). Údaje sa na základe dohody medzi Štatistickým úradom SR a jednotlivými obchodnými reťazcami zasielajú s oneskorením spravidla 3 dni po ukončení referenčného obdobia.

Na vstupe do informačného systému prebieha štrukturálna validácia súboru a pri identifikovaní závažných chýb je poskytovateľ údajov kontaktovaný a požiadaný o poskytnutie opravného súboru.

Následne sa jednotlivé produkty zatriedujú do klasifikácie ECOICOP, ktorá rozdeľuje kôš tovarov a služieb do odborov (2-miestna), skupín (3-miestna), tried (4-miestna) a podtried (5-miestna). Pre divíziu 01 – Potraviny a nealkoholické nápoje klasifikácie ECOICOP bola na účely spracovania údajov zo skenerov, definovaná národná, podrobnejšia 6-miestna úroveň klasifikácie - ECOICOP6, ktorá je spoločná pre údaje všetkých obchodných reťazcov, ktoré v súčasnosti spolupracujú so Štatistickým úradom SR. 6-miestna úroveň bola definovaná tak, aby vznikli homogénne skupiny produktov.

Každý produkt by mal vstupovať do výpočtu indexu na elementárnej úrovni na základe svojej dôležitosti. V usmernení Eurostatu [7] týkajúceho sa spracovania transakčných údajov zo supermarketov, sú odporúčané dva možné prístupy k výberu položiek, ktoré by mali byť zahrnuté do výpočtov, statický a dynamický prístup.

⁵ Elementárny agregát je najmenší a relatívne homogénny súbor tovarov alebo služieb, pre ktorý sa určuje indexové číslo bez explicitných výdavkových váh. Môže byť definovaný nielen na základe charakteristických vlastností tovarov a služieb, ale aj s ohľadom na región a typ odbytiska, v ktorom sa nachádzajú a predávajú.

⁶ ECOICOP je Európska klasifikácia individuálnej spotreby podľa účelu.

Statický prístup napodobňuje tradičný fixný výber (spotrebiteľský kôš) s tým rozdielom, že dochádza k zmene cenového konceptu. Ceny získané tradičným spôsobom sú nahradené cenami za jednotku tovaru z údajov zo skenerov a množina produktov je výrazne väčšia.

V prípade dynamického prístupu sa automaticky vyberajú produkty, ktoré sú v predaji súčasne v oboch po sebe nasledujúcich obdobiach/mesiacoch (t a $t + 1$, $t + 1$ a $t + 2$, $t + 2$ a $t + 3$ atď.) s tržbami nad určitou hranicou. Každý mesiac sa teda súbor jednotlivých výrobkov, ktoré vstupujú do zostavovania indexu, vyberá nanovo. Index elementárneho agregátu sa vypočíta na základe súboru spárovaných reprezentatívnych tovarových položiek, ktoré sa skutočne predávajú v dvoch nasledujúcich obdobiach.

Štatistický úrad SR sa po mnohých analýzach a sledovaní diskusií v rámci európskeho štatistického systému rozhodol ísť cestou dynamického prístupu, ktorý lepšie zohľadňuje aktuálnosť predávaných produktov a je jednoduchší z pohľadu automatizácie procesu spracovania.

Keďže cenové indexy sú publikované s mesačnou periodicitou, je potrebné týždenné súbory agregovať na mesačné. Tieto mesačné súbory obsahujú agregované hodnoty tržieb a predaného množstva jednotlivých tovarových položiek za príslušné vybrané týždne daného mesiaca a následne je vypočítaná priemerná cena produktu/tovarovej položky.

Na takto pripravené údaje sú aplikované filtre v závislosti od použitej metódy na zostavenie indexu, ktoré z výpočtu vylučujú niektoré produkty. Filtrovanie odstraňuje produkty s extrémnou zmenou ceny voči predchádzajúcemu obdobiu a ďalej sú odstránené dopredajové produkty (pokles v cene a súčasne výrazný pokles v tržbách). V prípade využitia bilaterálnych, resp. superlatívnych indexov sa aplikuje aj filter na málo predávané produkty. Viac o použití filtrov v súvislosti s údajmi zo skenerov sa uvádza v [11].

4. VYBRANÉ TYPY INDEXOV

Na výpočet cenových indexov na základe transakčných údajov je možné použiť viaceré indexové metódy (bilaterálne, multilaterálne, vážené, nevážené) a k nim prislúchajúce indexové vzorce. Pri výbere vhodného indexu, ktorý bude použitý na elementárnej úrovni v procese kompilácie CPI/HICP, možno použiť dva hlavné prístupy, axiomatický a ekonomický prístup.

Axiomatický prístup znamená, že index spĺňa určité špecifické axiómy alebo testy, na základe ktorých je možné vybrať vhodný index. Najpreferovanejšími testami sú test proporcionality, test súmernosti, časovo reverzný test a test tranzitivity. [15]. Z axiomatického hľadiska je Jevonsov index jednoznačne indexom s najlepšimi vlastnosťami. Jeho nevýhodou je, že je to nevážený typ indexu, do výpočtu nevstupujú informácie o predaných množstvách resp. tržbách a teda pri použití Jevonsovho indexu plnohodnotne nevyužijeme výhody transakčných údajov.

Jevonsov index možno interpretovať ako geometrický priemer zmien cien takto [7]:

$$I_{Jevons}^{m-1,m} = \prod_{i \in N} \left(\frac{p_i^m}{p_i^{m-1}} \right)^{\frac{1}{N}} \quad (1)$$

kde m je aktuálne obdobie, $m-1$ je predchádzajúce obdobie a p_i je cena i -tého produktu v danom období. Spotrebný kôš je dynamický a tvoria ho produkty, ktoré sú súčasne v aktuálnom a predchádzajúcom mesiaci v predaji a neboli vylúčené filtrami.

Okrem axiomatického prístupu na výber vhodného indexu je možné použiť aj prístup ekonomický, založený na ekonomickej teórii správania sa spotrebiteľa, čomu zodpovedá použitie vážených indexov, kde použitá váha je odrazom predaného množstva jednotlivých tovarových položiek.

Z indexov využívajúcich váhy sú najviac preferované superlatívne indexy Fischera a Törnqvista. Törnqvistov index je definovaný takto [15]:

$$I_{Törnqvist}^{m-1,m} = \prod_{i \in N} \left(\frac{p_i^m}{p_i^{m-1}} \right)^{\frac{s_i^{m-1} + s_i^m}{2}} \quad (2)$$

kde $s_i^{m-1} = p_i^{m-1} q_i^{m-1} / \sum_{i \in N} p_i^{m-1} q_i^{m-1}$ a $s_i^m = p_i^m q_i^m / \sum_{i \in N} p_i^m q_i^m$ sú podiely výdavkov v období $m-1$ a m v danej skupine produktov, pričom q_i sú predané množstvá v danom mesiaci.

Tieto indexy porovnávajú zmenu ceny medzi dvoma po sebe idúcimi obdobiami, avšak pre publikačnú prax Štatistického úradu SR v oblasti cenovej štatistiky sú dôležité indexy k základnému obdobiu 0, t. j. k decembru predchádzajúceho roku. Preto je potrebné medzimesačné indexy reťaziť.

$$I^{0,m} = I^{0,1} \cdot I^{1,2} \dots I^{m-1,m} \quad (3)$$

Ak sa v praxi HICP/CPI zostavujú z transakčných údajov vo všeobecnosti sa odporúča na elementárnej úrovni používať reťazené superlatívne indexy z dôvodu vyššieho stupňa zhody jednotlivých kódov položiek medzi dvoma nasledujúcimi obdobiami a predpokladu menších rozdielov v cenách a množstve. Tento predpoklad však nezohľadňuje existenciu výpredajov a zliav, ktoré môžu výrazne zvýšiť množstvo predaného tovaru, a to až niekoľkonásobne. Po skončení obdobia zľavy sa cena vráti na pôvodnú hodnotu. Obyvateľstvo však môže byť na dostatočne dlhý čas zásobené a bude trvať určitú dobu, kým sa množstvo predaného tovaru vráti na pôvodnú hodnotu. Za týchto podmienok majú reťazené superlatívne indexy v porovnaní s neváženými indexmi tendenciu klesať a len postupne sa vracajú na pôvodnú úroveň. Tento jav sa v odbornej literatúre nazýva *chain drift*. Jedným zo spôsobov, ako sa v dôsledku predzásobenia tomuto javu vyhnúť, je na výpočet cenových indexov

na elementárnej úrovni použiť nevážené indexy, najmä medzimesačný Jevonsov a ten reťaziť, hoci aj reťazený Jevonsov index môže za určitých okolností driftovať.⁷

V [18] navrhli postup, ktorý poskytuje indexy superlatívneho typu bez driftu, prostredníctvom adaptácie teórie multilaterálnych indexov, ktoré merajú agregovanú zmenu cien medzi dvoma obdobiami na základe pozorovaných cien vo viacerých obdobiach vrátane dvoch porovnávaných období. Multilaterálne indexy boli vyvinuté na porovnanie cien medzi krajinami (parita kúpnej sily) a boli prispôsobené na porovnanie cien v čase. Boli navrhnuté najmä na použitie údajov zo skenerov.

Jedným z predstaviteľov tejto skupiny indexov je GEKS-Törnqvistov index. Multilaterálna metóda GEKS používa bilaterálne indexy (Törnqvistov), ktoré sú vypočítané medzi každou dvojicou časových období v definovanom časovom okne.

Pri aplikácii metódy je najprv potrebné určiť dĺžku intervalu (okna), na ktorý sa má metóda aplikovať. Metóda sa spravidla uplatňuje na dĺžku okna jedného alebo dvoch rokov plus jedno obdobie (v prípade mesačných údajov je to 13, resp. 25 mesiacov), aby sa zohľadnili aj produkty zaťažené sezónnosťou. Dĺžku okna budeme označovať W . V rámci okna sa vyberie základné obdobie označené ako l a vypočíta sa bilaterálny index medzi obdobím l a každým nasledujúcim obdobím v okne. Postup sa opakuje pre všetky možné alternatívy l . Týmto spôsobom je možné získať maticu bilaterálnych indexov veľkosti $w \times w$ pre všetky možné dvojice jednotlivých období v rámci časového okna. Multilaterálny GEKS-Törnqvistov index je definovaný takto [8]:

$$I_{GEKS-Törnqvist}^{0,m} = \prod_{l \in W} (I_{Törnqvist}^{0,l} \times I_{Törnqvist}^{l,m})^{\frac{1}{|W|}} \quad (4)$$

Hore uvedené indexy sú aplikované na elementárnej úrovni, čo v našom prípade znamená úroveň ECOICOP6 pre každý obchodný reťazec. Na výpočet cenovej zmeny na vyššej úrovni agregácie klasifikácie ECOICOP sa používa index Laspeyresovho typu s fixnými váhami, ktoré sa vzťahujú na predchádzajúci rok ako referenčné obdobie $(y-1)$ [12]:

$$P_A^{y,m/y-1} = \frac{\sum_{a \in A} w_a^{y-1} I_a^{y,m/y-1}}{\sum_{a \in A} w_a^{y-1}} \quad (5)$$

kde w_a^{y-1} sú váhy založené na ročných výdavkoch za všetky položky patriace do elementárneho agregátu a bez ohľadu na to, či boli zahrnuté do výpočtu po aplikovaní filtrov. Následne sú takto vypočítané indexy reťazené k základné obdobiu.

⁷ O reťazovom indexe sa hovorí, že driftuje, ak sa nevráti k jednotke, keď sa ceny v bežnom období vrátia na úroveň, ktorú dosiahli v základnom období. Reťazové indexy sú náchylné na drift, keď ceny počas období, ktoré pokrývajú, kolíšu.

5. ANALÝZA VPLYVU ČASOVÉHO POKRYTIA

Ako už bolo uvedené, jednotkové ceny používané na konštrukciu CPI by mali byť vypočítané za rovnaké obdobie ako je obdobie, za ktoré sa zostavuje index (napr. v našom prípade mesiac), a nie za limitované parciálne obdobie. V praxi sa však štandardne používajú parciálne obdobia z dôvodu včasnosti a časovej presnosti poskytovania štatistických produktov. Vzhľadom na termíny poskytovania transakčných údajov zo strany obchodných reťazcov a publikačnú prax Štatistického úradu SR predpokladáme využívať na stanovenie jednotkovej ceny prvé dva kompletne týždne referenčného mesiaca.



Naším cieľom bolo formou referenčného porovnávania analyzovať vývoj časových radov cenových indexov a kvantifikovať vplyv parciálneho časového pokrytia na 6-miestnej, národnej úrovni klasifikácie ECOICOP, pretože cenové indexy na tejto úrovni sú základné stavebné prvky na zostavenie HICP/CPI. Porovnanie sme vykonali aj na 3-miestnej úrovni klasifikácie, pre skupinu tovarov Potraviny. Výstupy na tejto úrovni sú pravidelne publikované a vždy budú stredobodom záujmu pre svoj významný vplyv na životnú úroveň obyvateľstva.

Štatistický úrad SR preberá transakčné údaje od poskytovateľov na základe dohody, na týždennej báze vo forme týždenných a nie denných agregátov, čo má svoje konzekvencie. Pri zostavovaní mesačných súborov agregujeme tržby a predané množstvá za vybrané týždne. Pri zostavovaní mesačných súborov z údajov za celé obdobie mesiaca, bolo potrebné stanoviť určité pravidlá. Napríklad, v prípade prelomového týždňa, t. j. časť dní v danom týždni patrilo do jedného mesiaca a časť dní do druhého, bolo zavedené pravidlo, že daný týždeň bude zaradený do mesačného súboru k mesiacu, ku ktorému prislúchal väčší počet dní daného týždňa.

Nasledujúca tabuľka č. 1 porovnáva jednotkové ceny v eurách vypočítané z údajov za kompletne obdobie daného mesiaca a parciálne obdobie 2 týždňov na úrovni vybraných produktov homogénnej skupiny cukor kryštálový.

Modrou a žltou farbou je označená protichodná medzimesačná zmena ceny daného produktu a zelenou farbou takmer 19-percentný pokles priemernej mesačnej jednotkovej ceny produktu spôsobený parciálnym časovým pokrytím.

Tabuľka č. 1 : Porovnanie jednotkových cien tovarov – homogénna skupina cukor kryštálový

Kód produktu	Opis produktu	Mesiac	Jednotková cena v eurách	
			Pokrytie	
			Parciálne – 2T	Kompletné – 4T
4008671013004	Cukor kryštálový 1kg	01	0,7826	0,7864
4008671013004	Cukor kryštálový 1kg	02	0,7814	0,7789
4008671013004	Cukor kryštálový 1kg	03	0,6574	0,7133
4008671013004	Cukor kryštálový 1kg	04	0,7774	0,7793
4008671013004	Cukor kryštálový 1kg	05	0,7813	0,7813
8588000178391	Cukor kryštálový korunný 1kg	01	 0,7057	 0,7139
8588000178391	Cukor kryštálový korunný 1kg	02	0,7178	0,7065
8588000178391	Cukor kryštálový korunný 1kg	03	0,5643	0,6490
8588000178391	Cukor kryštálový korunný 1kg	04	0,7096	0,7059
8588000178391	Cukor kryštálový korunný 1kg	05	0,7139	0,7139
8588000178391	Cukor kryštálový korunný 1kg	06	0,6903	0,6943
8594003781131	Cukor korunný kryštálový 1kg	01	0,7347	0,7334
8594003781131	Cukor korunný kryštálový 1kg	02	0,7395	0,7365
8594003781131	Cukor korunný kryštálový 1kg	03	0,5762	0,7094
8594003781131	Cukor korunný kryštálový 1kg	04	0,7670	0,7644

Zdroj údajov: Štatistický úrad SR, výpočty autoriek

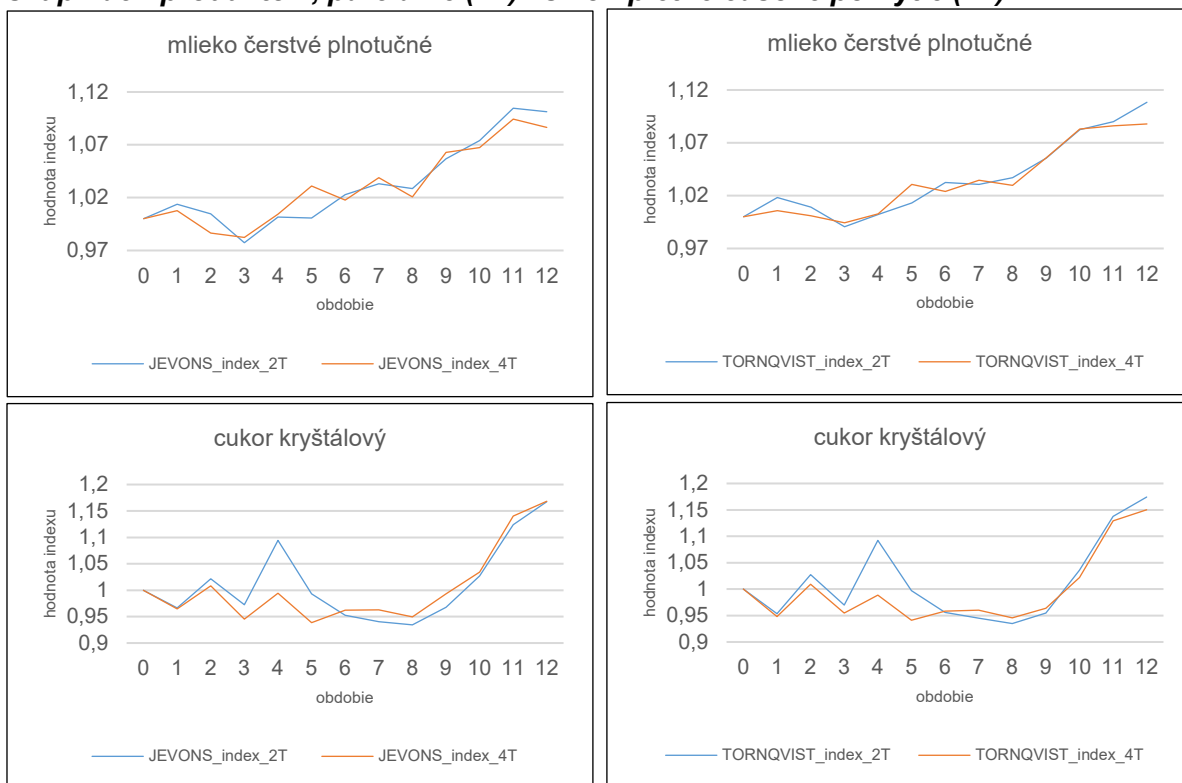
5.1 BILATERÁLNE INDEXY

Najprv sme vplyv časového pokrytia na vývoj cenových indexov analyzovali pre dynamický prístup k výberu produktov a následne sme zmenu ceny na elementárnej úrovni (ECOICOP6 a obchodný reťazec) vypočítali ako bilaterálny nevážený Jevonsov index (1) a vážený Törnqvistov index (2).

Na mesačné súbory údajov tržieb a predaného množstva za všetky tovary predané v jednotlivých obchodných reťazcoch boli aplikované už spomenuté tri typy filtrov na očistenie údajov a to filter na extrémne zmeny ceny, na odstránenie dopredajových a málo predávaných produktov. Indexy vypočítané na elementárnej úrovni boli agregované spolu za všetky dostupné obchodné reťazce použitím vzorca pre cenový index Laspeyresovho typu (5) na úroveň ECOICOP6 alebo ECOICOP3 a následne reťazené k základnému obdobiu.

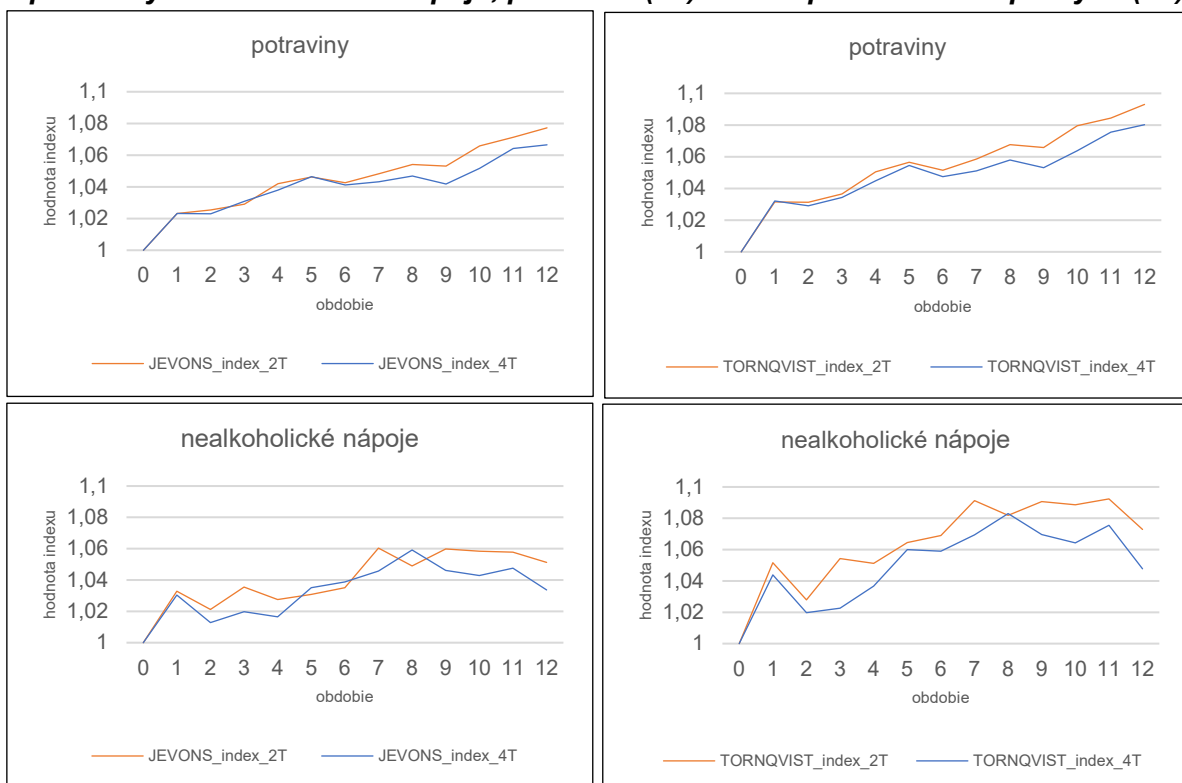
Vplyv parciálneho časového pokrytia na hodnoty a vývoj Jevonsovho a Törnqvistovho cenového indexu, ktoré sú vypočítané na úrovni vybraných homogénnych skupín produktov – mlieko čerstvé plnotučné a cukor kryštálový ilustrujú grafy č. 1 až 4. Modrou čiarou je zobrazený časový rad indexov, ktoré boli vypočítané z mesačných súborov údajov s parciálnym časovým pokrytím 2 týždňov, a oranžová čiara zobrazuje časový rad indexov, ktoré boli vypočítané nad kompletnou databázou údajov, ktorá pokrýva celé časové obdobie.

Grafy č. 1 – 4: Porovnanie vývoja mesačných cenových indexov na homogénnych skupinách produktov, parciálne (2T) vs. úplné časové pokrytie (4T)



Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Grafy č. 5 – 8: Porovnanie vývoja mesačných cenových indexov na skupine produktov – potraviny a nealkoholické nápoje, parciálne (2T) vs. úplné časové pokrytie (4T)



Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Na úrovni homogénnych skupín produktov bol rozdiel hodnôt indexov úplného a parciálneho pokrytia v prípade niektorých produktov výrazný (cukor kryštálový) a v iných prípadoch minimálny. Hodnoty indexov na takejto nízkej úrovni klasifikácie nie sú štandardne publikované, preto sme sa zaoberali porovnávaním aj na vyššej 3-miestnej úrovni klasifikácie ECOICOP, a to konkrétne pre skupinu 01.1 - Potraviny a 01.2 Nealkoholické nápoje. Grafy č. 5 až 8 zobrazujú vplyv parciálneho časového pokrytia na hodnoty a vývoj Jevonsovho a Törnqvistovho cenového indexu.

Možno konštatovať, že parciálne časové pokrytie ovplyvňuje hodnoty jednotkových cien a v niektorých momentoch je tento vplyv aj protichodný, ale—nespôsobuje systematické vychýlenie cenovej úrovne (tabuľka č. 1).

Jednotkové ceny na rozdiel od „pultových“ cien, ktoré sa využívajú pri tradičnom prístupe k výpočtu cenových indexov, zohľadňujú zľavy a ich vplyv na množstvo predaného tovaru, a preto časové pokrytie zohľadnené vo výpočte jednotkovej ceny logicky do určitej miery ovplyvňuje hodnoty cenových indexov. Veľkosť vplyvu závisí od skutočnosti, či zľava predstavovala významnú cenovú zmenu, či sa týkala viacerých produktov patriacich do spoločnej homogénnej skupiny a či boli produkty v zľave v jednom alebo vo viacerých obchodných reťazcoch v približne rovnakom čase. Nemenej dôležitá je aj skutočnosť, či bol tovar ponúkaný v zľave len v priebehu parciálneho časového obdobia alebo práve len mimo neho.

Chybu, akej by sme sa dopustili pri parciálnom časovom pokrytí, t. j. použitím len 2 úplných týždňov na stanovenie jednotkovej ceny tovarov do výpočtu cenových indexov k úplnému časovému pokrytiu, sme kvantifikovali prostredníctvom priemernej absolútnej percentuálnej chyby na dĺžke intervalu 12 mesiacov definovanej ako:

$$MAPE = \frac{1}{12} \sum_{i=1}^{12} \frac{|I_i^{4t} - I_i^{2t}|}{I_i^{4t}} * 100 \quad (6)$$

kde I_i^{4t} je index vypočítaný v i -tý mesiac z kompletných údajov zvyčajne zostavených za obdobie štyroch týždňov a I_i^{2t} je index vypočítaný z údajov za obdobie dvoch kompletných týždňov i -tého mesiaca. Priemerná absolútna percentuálna chyba je miera relatívnej chyby, ktorá používa absolútne hodnoty, aby sa kladné a záporné chyby navzájom nerušili, a možno ju interpretovať ako priemerný percentuálny rozdiel medzi hodnotami z modelu, ktorý využíva parciálne časové pokrytie sledovaného obdobia a hodnotami z modelu, ktorý využíva úplné časové pokrytie a ktorého hodnoty považujeme za referenčné. Výsledky výpočtov sú uvedené v tabuľke č. 2.

Tabuľka č. 2: Porovnanie priemernej ročnej absolútnej percentuálnej chyby

Produktová skupina klasifikácie ECOICOP	Priemerná ročná absolútna percentuálna chyba (MAPE)	
	Jevonsov index	Törnqvistov index
Mlieko čerstvé plnotučné	0,9574%	0,7069%
Cukor kryštálový	2,4894%	2,3794%
Potraviny	0,5189%	0,6626%
Nealkoholické nápoje	1,0272%	1,4786%

Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Na základe výsledkov výpočtov uvedených v tabuľke č. 2 môžeme predpokladať, že ak do štatistickej praxe zavedieme v súvislosti s implementáciou údajov zo skenerov výpočet bilaterálnych indexov, či už vážených, alebo nevážených a parciálne časové pokrytie sledovaného obdobia, dopustíme sa na 3-miestnej úrovni klasifikácie, t. j. na úrovni Potravinový chyby menšej ako 1 % a v prípade Nealkoholických nápojov menej ako 1,5%.

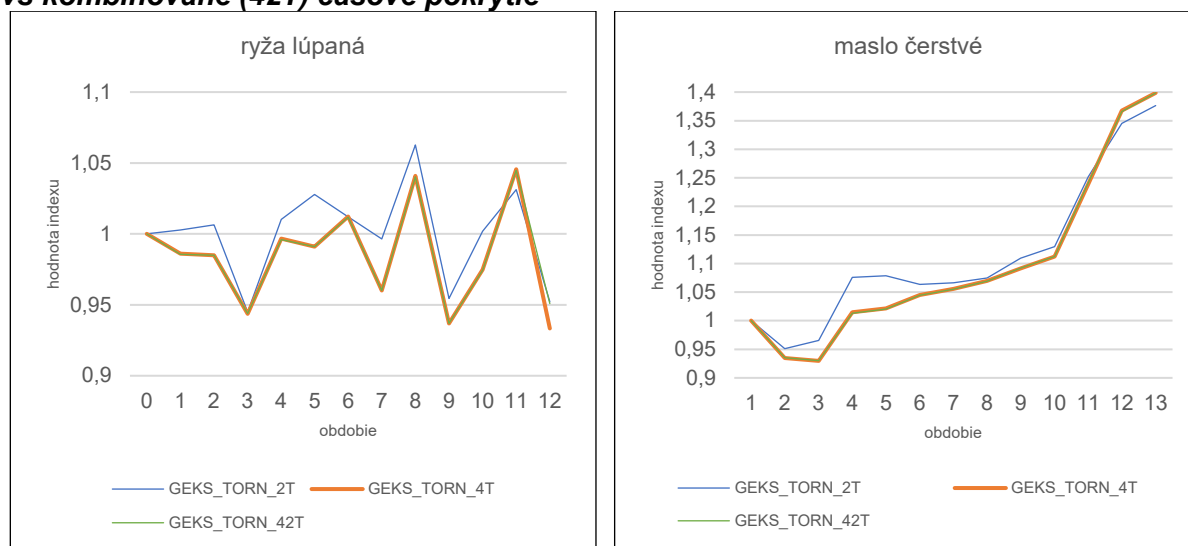
5.2 MULTILATERÁLNY PRÍSTUP

Ako už bolo uvedené, v prípade implementácie transakčných údajov obchodných reťazcov ako dátového zdroja pre výpočet cenových indexov, môžu byť cenové indexy na elementárnej úrovni vypočítané ako multilaterálne indexy, t. j. indexy, ktoré merajú agregovanú zmenu cien medzi dvoma obdobiami na základe pozorovaných cien vo viacerých obdobiach. V prípade multilaterálnych indexov sme vplyv parciálneho časového pokrytia analyzovali prostredníctvom GEKS-Törnqvistovho indexu (4) s dĺžkou časového okna 13 mesiacov.

K referenčným hodnotám, ktoré boli vypočítané nad mesačnými súbormi údajov s úplným časovým pokrytím sme porovnávali hodnoty indexov vypočítané nad parciálnym časovým pokrytím jednotlivých mesiacov, ktoré vstupovali do výpočtu, ako aj hodnoty indexov, ktoré boli vypočítané nad údajmi s kombinovaným pokrytím. V prípade kombinovaného pokrytia, ktoré by sa tiež mohlo použiť v praxi, aktuálny sledovaný mesiac bol pokrytý parciálne (2 týždne) a predchádzajúcich 12 mesiacov bolo pokrytých kompletne.

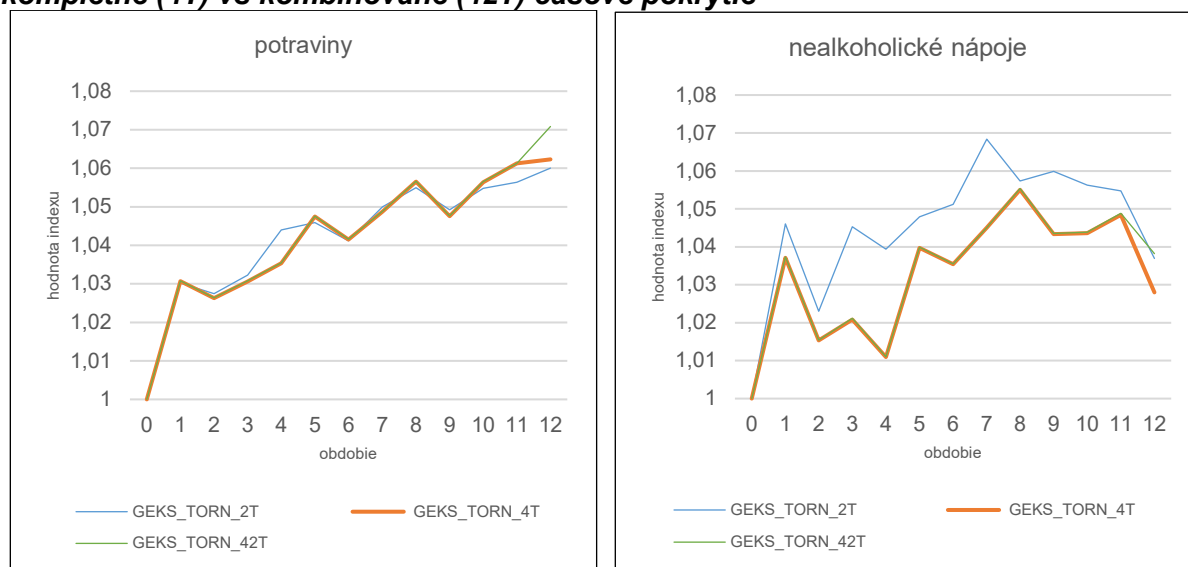
Grafické znázornenie vplyvu časového pokrytia na hodnoty cenových indexov pri použití multilaterálneho prístupu sú uvedené v nasledujúcich grafoch.

Grafy č. 9 a 10: Porovnanie vývoja mesačných cenových indexov (GEKS-Törnqvist) na skupine produktov – ryža lúpaná a maslo čerstvé, parciálne (2T) vs kompletne (4T) vs kombinované (42T) časové pokrytie



Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Grafy č. 11 a 12: Porovnanie vývoja mesačných cenových indexov (GEKS-Törnqvist) na skupine produktov – 01.1. Potraviny a 01.2 Nealkoholické nápoje, parciálne (2T) vs kompletné (4T) vs kombinované (42T) časové pokrytie



Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Podobne ako pri bilaterálnych indexoch grafy č. 9 až 12 ilustrujú vplyv parciálneho časového pokrytia na hodnoty a vývoj GEKS-Törnqvistových multilaterálnych cenových indexov, ktoré sú vypočítané na úrovni vybraných homogénnych skupín produktov – ryža lúpaná, maslo čerstvé a na 3-miestnej úrovni klasifikácie ECOICOP - 01.1-Potraviny a 01.2 Nealkoholické nápoje. Modrá čiara predstavuje hodnoty indexu vypočítaného nad mesačnými údajmi s parciálnym časovým pokrytím 2 týždňov, oranžová čiara predstavuje hodnoty indexov vypočítané nad databázou údajov s úplným časovým pokrytím a zelená čiara predstavuje multilaterálne indexy vypočítané nad kombinovaným časovým pokrytím.

Taktiež pri použití multilaterálneho prístupu sa vypočítala priemerná ročná absolútna percentuálna chyba (6). Hodnoty výpočtov sú uvedené v nasledujúcej tabuľke č. 3.

Tabuľka č. 3: Porovnanie priemernej ročnej absolútnej percentuálnej chyby

Produktová skupina klasifikácie ECOICOP	Priemerná ročná absolútna percentuálna chyba (MAPE)	
	Parciálne pokrytie	Kombinované pokrytie
Ryža lúpaná	1,9177%	0,1902%
Maslo čerstvé	2,3286%	0,0695%
Potraviny	0,2166%	0,0749%
Nealkoholické nápoje	1,3214%	0,1002%

Zdroj údajov: Štatistický úrad SR, výpočty autoriek

Z porovnania voči referenčným hodnotám indexov, ktoré boli vypočítané nad databázou údajov s plným časovým pokrytím, vyplynulo, že aj pri multilaterálnych indexoch sa na úrovni homogénnych produktových skupín pri parciálnom časovom pokrytí dopúšťame porovnateľnej chyby. Na úrovni produktovej skupiny Potraviny (3-miestna úroveň ECOICOP) sú však výsledky priaznivejšie a v prípade multilaterálnych

indexov je chyba spôsobená parciálnym pokrytím nižšia (0,2%) oproti chybe viac ako 0,5 % pri bilaterálnych indexoch.

Či už z grafov alebo tabuľky č. 3 je zrejmé, že ideálnym kandidátom na riešenie problému s parciálnym časovým pokrytím je kombinovaný prístup, ktorý by bol aplikovateľný v praxi. Problémom sú však samotné multilaterálne indexy, ktorých výpočet je náročný tak z pohľadu samotných postupov ako aj času. Ďalšou ich nevýhodou je, že princíp a metodiku ich zostavenia je veľmi náročné vysvetliť bežnej populácii používateľov.

6. ZÁVER

Implementácia transakčných údajov do výpočtu cenových indexov predstavuje významnú metodologickú zmenu jednak v spracovaní zdrojových údajov, ale aj v procese ich kompilácie. Eurostat ako náš partner v európskom štatistickom systéme poskytuje v tomto kontexte len odporúčania a je na národných štatistických úradoch, aby si zvolili čo najvhodnejší postup a metódu spracovania vzhľadom na podmienky v tej ktorej krajine. Pri výbere vhodných pracovných postupov treba zvažovať nielen zabezpečenie kvality štatistického produktu z hľadiska presnosti, ale aj včasnosť a časovú presnosť poskytovania štatistických produktov. V reálnej štatistickej praxi teda hľadáme kompromisné riešenie, ktoré zabezpečí dostatočnú kvalitu produktu a nenaruší publikačnú prax požadovanú používateľmi týchto produktov.

Cieľom článku bolo zistiť, či a aký vplyv na presnosť má kompromisné riešenie týkajúce sa parciálneho časového pokrytia sledovaného obdobia (v našom prípade mesiac vs 2 týždne), ktoré má zabezpečiť včasnosť a časovú presnosť poskytovania cenových indexov. Analýza bola vykonaná pre bilaterálne typy indexov, vážený (Törnqvistov) aj nevážený (Jevonsov) a tiež pre multilaterálne typy indexov, ktoré reprezentoval GEKS-Törnqvist.

Z výsledkov vyplynulo, že vo všetkých analyzovaných prípadoch je na úrovni produktovej skupiny Potraviný priemerný absolútny percentuálny rozdiel medzi hodnotami cenových indexov vypočítaných nad databázou údajov s parciálnym pokrytím oproti výsledkom nad databázou s plným pokrytím menší ako 1 %, čo je viac ako vyhovujúci výsledok. Okrem iného, nárast alebo pokles cien z druhej polovice mesiaca, ktoré nevstupovali do výpočtu cenového indexu, budú zohľadnené v nasledujúcom mesiaci.

Navyše tak ako vyplynulo z analýzy týkajúcej sa multilaterálnych indexov, najlepšie výsledky sme dosiahli kombinovaným prístupom k riešeniu parciálneho pokrytia. Jeho prípadná aplikácia v praxi si však vzhľadom na mnohé komplikácie súvisiace jednak s aplikáciou samotných indexov ako aj špeciálnou prípravou zdrojových údajov bude vyžadovať ďalší výskum v oblasti metodológie a automatizácie spracovania.

LITERATÚRA

- [1] BALK, B. M.: On the Use of Unit Value Indices as Consumer Price Subindices. Paper presented at the Fourth Meeting of the International Working Group on Price Indices, Washington, DC, April 22–24. 1998. <http://www.ottawagroup.org>.
- [2] DE HAAN, J. – KRSINICH, F.: Scanner Data and the Treatment of Quality Change in Rolling Year GEKS Price Indexes. Paper presented at the eleventh Economic Measurement Group Workshop, 21-23 November 2012, Sydney, Australia

- [3] DIEWERT, W.E.: Exact and Superlative Index Numbers. In: Journal of Econometrics, 1976, 4, s. 114 – 145.
- [4] DIEWERT, W. E. – FOX, K. J. – DE HAAN, J. : A Newly Identified Source of Potential CPI Bias: Weekly versus Monthly Unit Value Price Indexes, Economics Letters, 2016, 141, s. 169 – 172.
- [5] ELTETÖ, O. – KÖVES, P.: On a Problem of Index Number Computation Relating to International Comparisons, Statisztikai Szemle 1964, 42, s. 507 – 518 (originál v maďarčine).
- [6] 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.
- [7] EUROPEAN COMMISSION, Eurostat (September 2017): HICP, Practical Guide for Processing Supermarket Scanner Data [cit. 2023-05-02] Dostupné na: <https://circabc.europa.eu/ui/group/7b031f10-ac19-4da3-a36f-58708a70133d/library/8e1333df-ca16-40fc-bc6a-1ce1be37247c/details>
- [8] EUROSTAT: Guide on Multilateral Methods in the Harmonised Index of Consumer Prices, Manuals and Guidelines, Luxembourg: Publication Office of the European Union, 2022. ISBN 978-92-76-44354-4.
- [9] GINI, C.: On the Circular Test of Index Numbers, 1931, Metron 9, s. 3 – 24.
- [10] GLASER-OPITZOVÁ, H.: Nové zdroje údajov pre cenovú štatistiku a metódy ich spracovania. In: Slovenská štatistika a demografia, 2019, roč. 29, č. 4, s. 49 – 66. [cit. 2023-05-02] Dostupné na: https://ssad.statistics.sk/SSaD/wp-content/files/4_2019/4_2019_komplet_cislo.pdf
- [11] GLASER-OPITZOVÁ, H.: Use of Scanner Data in Measuring the Consumer Price Index in the Conditions of the Slovak Republic, 2022. Proceedings from the EDAMBA 2021 conference, s. 92 – 102 [cit. 2023-05-02]. Dostupné na: <https://doi.org/10.53465/EDAMBA.2021.9788022549301.92-102>
- [12] VAN DER GRIET, H. – DE HAAN, J.: The use of supermarket scanner data in the Dutch CPI. [cit. 2023-05-02] Dostupné na: [2010-scanner-data-dutch-cpi](#)
- [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] CHESSA, A.: The QU-method: A new methodology for processing scanner data. Proceedings of Statistics Canada 2016 International Methodology Symposium.
- [15] ILO/IMF/OECD/UNECE/Eurostat/The World Bank. Consumer Price Index Manual: Concepts and Methods. IMF Publications, Washington, DC. 2020. ISBN 978-1-51354-298-0 (PDF).
- [16] ILO/IMF/OECD/UNECE/Eurostat/The World Bank. Consumer Price Index Manual: Theory and Practice. ILO Publications, Geneva. 2004. ISBN 92-2-113699-X.
- [17] IVANCIC L. – DIEWERT W. E. – FOX K. J.: Scanner Data, Time Aggregation and the Construction of Price Indexes, , In: Journal of Econometrics. 2011. vol. 161, no. 1, s. 24 – 35.
- [18] IVANCIC, L. – DIEWERT, W. E., – FOX, K. J. *Scanner data, time aggregation and the construction of price indexes*. Discussion paper 09-09, University of British Columbia, Vancouver, Canada. 2009.
- [19] KNÍŽAT, P. – GLASER-OPITZOVÁ, H.: Index spotrebiteľských cien z webscrapovaných údajov: analýza vybranej produktovej skupiny. In: Slovenská štatistika a demografia, 2023, roč. 33, č.1, s. 37 – 49. [cit. 2023-05-02]. Dostupné na: https://ssad.statistics.sk/SSaD/wp-content/files/1_2023/SSaD_1_2023_kompletne_cislo.pdf

- [20] KNÍŽAT, P.: Web scraped data in consumer price indices. In: Statistical Journal of the IAOS, 2023, vol. 39, no.1, s. 203 – 212.
- [21] RAMON - Reference and Management of Nomenclatures: Europa - RAMON -Classification Detail List [cit. 2022-12-05].
- [22] Spotrebiteľské ceny a ceny produkčných štatistik [cit. 2023-05-02].

RESUMÉ

Súčasná doba ponúka čoraz viac nových zdrojov údajov, ktoré môžu prispieť ku skvalitneniu a modernizácii štatistik. Jedným z takýchto nových zdrojov údajov sú transakčné údaje obchodných reťazcov, nazývané aj údaje zo skenerov. Tieto údaje môžu veľkou mierou prispieť k modernizácii štatistiky spotrebiteľských cien, ktorá patrí medzi významné makroekonomické ukazovatele danej krajiny. Údaje obsahujú úplné informácie o predaných tovaroch a to okrem kódu a stručného popisu hlavne predané množstvo a tržby za dané obdobie. Naproti tomu, ceny zbierané tradičným spôsobom sú ceny vybranej, relatívne malej množiny tovarov, za ktoré je tovar ponúkaný. Informácie o predaných množstvách nie sú dostupné. Implementácia transakčných údajov do produkcie cenovej štatistiky znamená implementáciu mnohých metodologických zmien. Jednou z nich je aj zmena cenového konceptu. Ceny získavané tradičným spôsobom sú nahradené cenami za jednotku tovaru. Jednotkové ceny používané na konštrukciu cenových indexov by mali byť teoreticky vypočítané za rovnaké obdobie ako je obdobie, za ktoré sa zostavuje index (napr. mesiac). Vzhľadom na termíny poskytovania transakčných údajov zo strany obchodných reťazcov, publikačnú prax Štatistického úradu SR a termíny poskytovania štatistických produktov ich používateľom je nevyhnutné nájsť kompromis medzi presnosťou cenových indexov a ich včasnosťou a časovou presnosťou. Takýto kompromis predstavuje parciálne časové pokrytie, t. j. výpočet indexov sa realizuje nad databázou údajov, ktoré nepokrývajú obdobie celého sledovaného mesiaca, ale len obdobie napríklad 2 týždňov. Cieľom príspevku je zistiť či má takéto kompromisné riešenie zásadný vplyv na úroveň a vývoj časových radov indexov a tento vplyv kvantifikovať. Analýza bola realizovaná pre bilaterálne, vážené aj nevážené indexy ako aj pre multilaterálne indexy. Z výsledkov analýzy vyplynulo, že vo všetkých analyzovaných prípadoch bol na úrovni produktovej skupiny Potraviny, t. j. na 3-miestnej úrovni klasifikácie ECOICOP, priemerný absolútny percentuálny rozdiel medzi hodnotami cenových indexov vypočítaných nad databázou údajov s parciálnym pokrytím oproti výsledkom nad databázou s plným pokrytím menší ako 1 %, čo môžeme považovať za viac ako vyhovujúci výsledok.

RESUME

The present time offers an increasing number of new data sources that can contribute to the improvement and modernization of statistics. One such new data source is transaction data from retail chains, also known as scanner data. These data can greatly contribute to the modernization of consumer price statistics, which are among the significant macroeconomic indicators of a country. These data contain comprehensive information about the sold goods, including not only the code and a brief description but also the quantity sold and the turnover for a given period. In contrast, prices collected in the traditional way are prices of selected, relatively small set of goods, for which the goods are offered, and information about the quantities sold is not available. The implementation of transaction data into price statistics production entails several methodological changes, including a change in the price concept. Prices obtained in the traditional way are replaced by prices per unit of goods. Unit prices used to

construct price indices should ideally be calculated for the corresponding period as the period for which the index is compiled (e.g., a month). Given the timing of providing transaction data by retail chains, the publication practices of the Statistical Office of the Slovak Republic, and the deadlines for providing statistical products to the users, it is necessary to find a compromise between the accuracy of price indices and their timeliness and temporal accuracy. Such a compromise is represented by a partial time coverage, which means that the calculation of indices is performed using a database that does not cover the entire observed month but only a period, i.e., the first two weeks of the given months. The aim of this contribution is to determine whether such a compromise solution has a significant impact on the level and trend of index time series and to quantify this impact. The analysis was conducted for bilateral, weighted and unweighted indices, as well as for multilateral indices. The results of the analysis showed that in all the analyzed cases, at the product group level "Food," i.e., at the 3-digit level of the ECOICOP classification the average absolute percentage difference between the values of price indices calculated using a database with partial coverage (2 weeks) compared to the results using a database with full coverage was less than 1%, which can be considered as a highly satisfactory outcome.

PROFESIJNÝ ŽIVOTOPIS

Ing. Helena Glaser-Opitzová je generálna riaditeľka sekcie všeobecnej metodiky, registrov a koordinácie národného štatistického systému Štatistického úradu SR a členka riaditeľskej skupiny Eurostatu pre metodológiu (DIME), ktorá poskytuje poradenstvo Európskemu štatistickému výboru (ESSC) v strategických otázkach. Riadila a podieľala sa na mnohých modernizačných aktivitách úradu. V súčasnosti vedie interný projekt úradu zameraný na modernizáciu cenových štatistík.

RNDr. Petra Mazureková PhD. pracuje ako metodička v sekcii všeobecnej metodiky, registrov a koordinácie národného štatistického systému na odbore metód štatistických zisťovaní Štatistického úradu SR ako členka tímu zodpovedného za efektívne štatistické metódy, analýzy a výpočty s dôrazom na štandardizáciu štatistických procesov, systém monitorovania, vykazovania a hodnotenia kvality štatistických zisťovaní a ich produktov, a využívanie administratívnych zdrojov údajov na štatistické účely. Od mája 2018 je projektovou manažérkou projektu Scanner data.

KONTAKT

helena.glaser-opitzova@statistics.sk

petra.mazurekova@statistics.sk