NÁRODNÁ BANKA SLOVENSKA
EUROSYSTÉM

Sezónne očisťovanie

(jednoduchý sprievodca postupmi, možnosťami, problémami a výsledkami s aplikáciou na očisťovanie HICP)

FEBRUÁR **2014**

© Národná banka Slovenska
www.nbs.sk
Imricha Karvaša 1
813 25 Bratislava

michal.doliak@nbs.sk, juraj.hucek@nbs.sk

február 2014

Práca neprešla jazykovou úpravou.

Prezentované názory a výsledky v tejto štúdii sú názormi autorov a nevyjadrujú oficiálne stanovisko Národnej banky Slovenska.

Všetky práva vyhradené.
Krátke časti textu, nie viac ako dva odseky, môžu byť citované bez predchádzajúceho súhlasu autorov, pokiaľ bude úplne uvedený zdroj.

Sezónne očisťovanie

(jednoduchý sprievodca postupmi, možnosťami, problémami a výsledkami s aplikáciou na očisťovanie HICP)

Juraj Huček a Michal Doliak
Odbor ekonomických a menových analýz
Národná banka Slovenska

„Je ľahké veci urobiť, keď vieme ako na to a ako to má v konečnej podobe vyzerať. Ale ako spraviť veci, pri ktorých len tušíme, čo by sme mali dostať....?“
autori

Kým sa nepristúpilo k využívaniu sezónne očistených dát, analýza časových radov bola jednoduchšia. Prináša to však aj svoje obmedzenia. Pozerať sa na sezónne očistené dáta je korektnejšie, keďže sezónnosť je pravidelné opakovanie udalosti v rovnakých intervaloch. Pre ekonomický vývoj a jeho analýzu je potrebné vedieť identifikovať neštandardné zmeny nevyplývajúce zo sezónnosti. Preto správne porozumenie sezónnosti umožňuje identifikovanie týchto neštandardných zmien v časovom rade. Má to však jedno veľké ALE. Sezónnosť ako taká nie je jednoznačná a definovať, čo je najsprávnejší postup alebo metóda na jej získanie, je pomerne zložité.

V nasledujúcich riadkoch sa pokúsime problematiku sezónneho očisťovania priblížiť pokiaľ možno jednoduchým spôsobom bez použitia jediného vzorca alebo definície. Účelom nie je vytvárať presnú metodiku, ale prezentovať, s akými problémami sa pri sezónnom očisťovaní časových radov najčastejšie môžeme stretnúť a ako sa s nimi vysporiadať. Čitateľ nie vždy dostane jednoznačnú odpoveď či presný návod, ako na to, ale získa ucelenejšiu predstavu o postupoch a možnostiach. Obsah je určený tým, ktorí s problematikou sezónneho očisťovania začínajú, ako aj tým, ktorí už nejaký ten časový rad očistili bez toho, aby sa bližšie zamysleli, akú metódu alebo aké nastavenie pri tom použili. Pre napísanie tohto „bedekra“ sme sa rozhodli pri práci na sezónnom očisťovaní inflácie meranej harmonizovaným indexom spotrebiteľských cien (HICP), preto je doplnený o praktické ukážky a aplikácie poznatkov, ktoré sme využili pri tejto práci, ako aj o výsledky sezónneho očistenia HICP.

Pohľady na sezónnosť - čo je to, ako ju vnímať

Sezónnosť, ako sme uviedli v úvode príspevku, je pravidelné opakovanie udalosti v rovnakých intervaloch a jej správne porozumenie umožňuje skorú identifikáciu neštandardnej zmeny nevyplývajúcej zo sezónnosti. Pri ekonomických ukazovateľoch sa so sezónnosťou najčastejšie stretávame, ak pracujeme s mesačnými alebo štvrťročnými údajmi.

Príklad: *Informácie, že v júni sa oproti máju oteplilo a v novembri sa oproti októbru ochladilo sú štandardné a pravidelné zmeny, ktoré sa každý rok predpokladajú. V prípade ekonomickej veličiny ako sú ceny vieme, že ceny ovocia a zeleniny klesajú v lete a na jeseň a naopak v zime a na jar rastú a teda ide o štandardné a pravidelné zmeny. Po očistení o sezónne výkyvy je možné povedať, že sa drží nejaká priemerná cena.*

Zdroj:Autori

Zdroj: Štatistický úrad SR

V grafe je vidno „neštandardný“ vývoj cien zeleniny v druhej polovici roka 2010 a začiatkom roka 2011. Ten vyplýval z nízkej úrody v roku 2010, čo sa prejavilo v nižšom ako zvyčajnom poklese cien v lete roku 2010 a výraznejšom (ako zvyčajne) raste koncom roka 2010 a začiatkom roka 2011. Zároveň, výraznejší prepad koncom roka 2011 je korekcia vyšších cien v roku 2010.

Pre centrálnu banku, vládu, alebo pre inštitúcie, ktoré analyzujú ekonomický vývoj, je z tohto pohľadu dôležité poznať tieto neštandardné zmeny, lebo môžu mať vplyv na budúci ekonomický vývoj. Včasná reakcia na neštandardný vývoj zo strany tvorcov hospodárskej politiky môže budúci vývoj ovplyvniť správnym smerom. Naopak nesprávne vyhodnotenie sezónnosti môže viesť k nesprávnym záverom a interpretácii. Úlohou sezónneho očisťovania je odčleniť z časového radu pravidelne sa opakujúcu, t. j. sezónnu zložku.

Aplikácia - Sezónne očisťovanie HICP : *Inflácia alebo rast spotrebiteľských cien sa najčastejšie vyjadruje (publikuje) ako medziročná zmena cien (t. j. ako sa ceny zmenili oproti rovnakému obdobiu minulého roka). S vyjadrením pomocou medzimesačných zmien sa stretávame zriedkavo. V komentároch k inflácii sa tieto informácie vyskytujú, ale ide skôr o doplňujúce informácie. Sezónne očistené údaje sa nespomínajú takmer nikde. Jednou z mála inštitúcií je US Bureau of Labor Statistics, ktorý vo svojich správach o vývoji inflácie prezentuje sezónne očistené medzimesačné zmeny. Sezónne očistené inflačné dáta môžu poskytovať cenný prínos k analýzam cenového vývoja, ako aj ku krátkodobej predikcii inflácie aj vzhľadom na to, že medziročné dynamiky „trpia“ tzv. bázickým efektom, teda sú výrazne ovplyvnené vývojom v rovnakom období minulého roka. Medzimesačné zmeny môžu odhaliť zvyčajný cenový vývoj (usual patern), nie sú ovplyvnené každoročnou zmenou váh a reprezentantov (tieto zmeny ovplyvňujú skôr medziročné indexy), a vyjadrujú tak čisté cenové zmeny. Pre správnu interpretáciu medzimesačných zmien je však potrebné časové rady sezónne očisťovať. ECB zostavuje od roku 2000 sezónne očistené časové rady za eurozónu s použitím postupov vypracovaných pracovnou skupinou v rokoch 1999 až 2001. V tomto prípade je sezónne očistený HICP agregáciou sezónne očistených hlavných komponentov (t. j. cenových indexov pre nespracované potraviny, spracované potraviny, priemyselné tovary, služby) a sezónne neočistených energií (ide o tzv. nepriame očistenie). NBS sa pripojila k projektu ECB na sezónne očisťovanie HICP inflácie, aby sa táto „medzera“ vo zverejňovaných dátach vyplnila aj na národnej úrovni.*

Sezónnosť, rozklad časového radu, prístupy k sezónnemu očisťovaniu

Z hľadiska definície môžeme časový rad rozložiť na tri časti: trendovú, sezónnu (pravidelnú) a iregulárnu (nepravidelnú). Očistený časový rad sa teda skladá z trendovej a iregulárnej zložky, vďaka ktorej dáta zbavené sezónnosti nemusia byť nevyhnutne hladké.

Časové rady je možné očistiť od sezónnych vplyvov viacerými spôsobmi. Najjednoduchšie metódy v skutočnosti odstránia okrem sezónnej zložky aj všetku volatilitu, čím sa namiesto sezónne očisteného radu získa iba trend. K takýmto typom by sme mohli zaradiť očistenie lineárnym trendom, kĺzavé priemery, HP filter a iné. K pokročilejším, čisto štatistickým filtrom, ktoré dokážu identifikovať a odstrániť aj napr. vplyv výnimočných, extrémnych, neočakávaných hodnôt alebo zlomov v časových radoch, tzv. outlierov, patrí obľúbená metóda X11. Ide o veľmi dôležitú súčasť dekompozície, ktorá môže výrazne ovplyvniť výsledky sezónne očistených dát.
Sofistikovanejšie metódy využívajú modelový odhad, pri ktorom je možné okrem identifikácie outlierov uvažovať aj s vplyvom pracovných dní, sviatkov a iných dodatočných faktorov. Takéto metódy sú založené na odhade ARIMA modelov, z ktorých najznámejšie a najčastejšie používané sú X12-ARIMA a Tramo/Seats.

Proces sezónneho očisťovania

V praxi sa najčastejšie stretávame s potrebou očistiť jeden, či dva časové rady, ktoré navzájom ekonomicky úzko súvisia, ale z hľadiska sezónnosti sa na ne pozeráme samostatne (napr. HDP a HICP). Voľba metódy a nastavení sa robí zvlášť pre každý rad bez bližšieho vzájomného porovnania. V prípade veľkého počtu časových radov, ktoré treba sezónne očistiť je dobré zjednodušiť si prácu a hneď od začiatku pristúpiť k **selekcii**.

1. Výber časových radov

Na základe **grafickej analýzy** je možné vo veľa prípadoch vylúčiť alebo potvrdiť prítomnosť sezónnosti. Časové rady s evidentnou pravidelnou „pílou“ budú s najväčšou pravdepodobnosťou vykazovať sezónnosť a naopak pri časových radoch hladkých alebo s absolútne nepravidelnými „zubami“ bude sezónnosť štatisticky zamietnutá[1].

Spustením **automatických** (všeobecných, prednastavených) **modulov** (je to v podstate jedno, o aký programový balík[2] ide) je možné spraviť ďalšiu selekciu, poprípade overiť si vhodnosť výberu získaného z grafickej analýzy. Všeobecný modul zväčša testuje vplyv kalendára (obsahuje vplyv pracovných resp. obchodných dní– Trading days - **TD** / Working days - **WD**), pričom sviatky a iné špecifické dni ako také si treba „manuálne“ nastaviť, resp. vybrať z možných volieb (spravidla kalendáre sú pre jednotlivé štáty rôzne – sviatky, prázdniny). Taktiež vplyv veľkonočných sviatkov (Easter effekt) sa zväčša testuje.

Aplikácia – Sezónne očisťovanie HICP : *Vychádzajúc z úrovne tried základnej COICOP klasifikácie (tzv. štvormiestny COICOP), autori pracovali s približne 180 časovými radmi (okrem základných tried, hlavných zložiek a celkového HICP, aj s vybranými časovými radmi, napr. alkoholické nápoje, nealkoholické nápoje a pod)., indexov na báze roku 2005. Na základe grafickej analýzy a spustením automatického modulu sa vyselektovalo 56 časových radov, ktorým sa venovali individuálne a podrobnejšie.*

Pri pohľade na dole uvedené grafy je zrejmé, že v prípade cien počítačov sa sezónnosť nevyskytuje a v podstate sa týmto časovým radom už nie je potrebné zaoberať (štatisticky

[1] *Netvrdíme, že zdanlivú skutočnosť netreba štatisticky overiť, ale pri prvom štatistickom potvrdení, že daný časový rad neobsahuje sezónnu zložku nemusíme sa silou-mocou pokúšať nájsť takú špecifikáciu sezónneho modelu, ktorá nám dokáže akú-takú sezónnosť identifikovať. Prvotné očistenia robíme zvyčajne na základe všeobecného/automatického nastavenia, ktoré obsahuje každý software.*
[2] *Autori používali pri svojej práci software DEMETRA+.*

potvrdené aj spustením automatického modulu programu). V prípade druhého grafu je zjavná sezónnosť (vidieť pôvodné dáta a sezónne upravený časový rad).

Zdroj: ŠÚSR, prepočty autorov

Zdroj: ŠÚSR, prepočty autorov

Vo všeobecnosti platí, že nie je podstatné akú metódu použijeme (X12-ARIMA vs. Tramo/Seats) pretože voľba správnej metódy/nástroja má význam len v prípade problémových radov, t.j. radov, pri ktorých je ťažšie identifikovať sezónnosť resp. sezónnosť sa často mení.

Poznámka: *Rozdiely najpoužívanejších modelových prístupov, resp. vlastností, ktoré druhá metóda nemá sú nasledovné :*

X12-ARIMA

Túto metódu možno považovať za konzervatívnu, pretože zmenu sezónnosti (či už z hľadiska veľkosti alebo lokalizácie) premietne do sezóne očisteného časového radu až po druhom pozorovaní rovnakého druhu/rozsahu, teda zväčša po jednom roku. Teda veľkosti a tvar sezónnych faktorov sa menia často a výsledný sezónne očistený časový rad „trpí" veľkou zotrvačnosťou, čo spôsobuje veľakrát neskorú identifikáciu outliera. Výhodou tejto metódy je možnosť širšieho nastavenia vplyvu veľkonočných sviatkov (Easter effect 1-20 dní).

TRAMO/SEATS

Táto metóda predpokladá, že veľkosť sezónnych faktorov sa v čase výrazne nemení a tým dochádza k „rýchlejšej" zmene v sezónne očistenom časovom rade, teda najmä k skorej identifikácii outliera.

2. Typ sezónnosti

Vybratím metódy na sezónne očistenie a prvotným spustením všeobecného (automatického) modulu sa samotný proces len začína. Pre kvalitné a vo vzťahu k danému časovému radu správne výsledky je potrebné vybrať vhodné nastavenie parametrov, ktoré popisujú a odzrkadľujú ekonomické vlastnosti časového radu. Treba sa zamyslieť, akým časovým radom sa zaoberáme. Či ide o rastúce údaje v čase, teda či s časom rastie aj sezónna zložka alebo ide o časový rad, ktorého hodnoty v čase kolíšu okolo nejakej konštanty, resp. trendu, teda volatilita je zhruba konštantná a tak možno uvažovať o nemennej veľkosti sezónnych zložiek. V prvom prípade ide o multiplikatívnu, časom sa znásobujúcu (Orig=Trend*-Seasonal*Irregular) sezónnosť, pri ktorej volíme v softvéri logaritmickú transformáciu5. V druhom prípade ide o tzv. aditívnu sezónnosť (Orig=Trend+Seasonal+Irregular), kde do procesu sezónneho očistenia vstupujú pôvodné (neupravené) časové rady.

Príklad: *Teoreticky aditívna a multiplikatívna sezónnosť graficky vyzerá takto:*

Zdroj: Autori

Zdroj:Autori

Tieto nastavenia je potrebné si pri automatických moduloch skontrolovať a pri manuálnej špecifikácii správne nastaviť. Opačné nastavenie môžu spôsobiť v prípade aditívnej sezónnosti klesanie významnosti sezónnej zložky a v prípade multiplikatívnej sezónnosti rapídny nárast sezónnej zložky. Taktiež môže spôsobiť falošnú identifikáciu outliera.

Aplikácia – Sezónne očisťovanie HICP : : *Z časových radov, ktoré sme analyzovali v rámci HICP, sú príkladom správne nastavenej multiplikatívnej sezónnosti ceny obuvi, napriek tomu, že je možné vidieť v čase rastúce sezónne zložky. Je to rastúcim trendom sezónnych výpredajov. Naopak príkladom ekonomicky chybne zvolenej aditívnej sezónnosti sú ceny zeleniny, kde sezónnosť je v čase stabilná. V skutočnosti sezónne faktory by mali klesať, kvôli otvorenosti trhu a možnosti doviezť zeleninu v akomkoľvek období roka. Toto sú špecifické príklady, kde ekonomický aspekt je dôležitejší než vizuálny (technický).*

Zdroj: ŠÚSR, prepočty autorov

Zdroj: ŠÚSR, prepočty autorov

3. Vplyv kalendára

K základnému nastaveniu podlieha v ďalšom rade **vplyv kalendára**, ktorý dokáže mnohé vylepšiť aj pokaziť. Výber možností použitia pracovných či obchodných dní (TD/WD) je dôležitý a často môže zabrániť falošnej identifikácii outlierov resp. prispieť k redukcii ich počtu. Vždy si treba položiť otázku, či časový rad, ktorý chceme zbaviť sezónnosti, je alebo môže byť ovplyvnený počtom voľných alebo pracovných dní. Do tejto časti kalendára patria aj tzv. fixné sviatky, ktoré sú pre každú krajinu iné a špecifické (1. január, 6. január, 1. máj, 8. máj atď.).

Príklad: *V prípade väčšiny cenových časových radov je asi jedno, či je v mesiaci 20 alebo 22 pracovných dní, pretože dva pracovné dni navyše s najväčšou pravdepodobnosťou nezvýšia ani neznížia cenu tovaru. V prípade produkcie je však dôležité, koľko pondelkov, stried či piatkov máme za mesiac k dispozícii (najnižšia produkcia je štatisticky potvrdená v pondelok a v piatok a naopak najvyššia v stredu). Inými slovami tvrdenie, že výrobná linka zvýšila svoju mesačnú produkciu v marci oproti februáru o 10 % nie je pravdivé, lebo v skutočnosti výrobná linka pracovala v marci o dva dni viac, čo s istotou môžeme považovať za kalendárny efekt.*

Aplikácia – Sezónne očisťovanie HICP : *Využitím výsledkov predtestov na významnosť pracovných dní sa pristúpilo k úplnému vypnutiu vplyvu pracovných dní vo všetkých časových radoch štruktúry HICP. V niektorých prípadoch by mohlo dôjsť k náhodnému identifikovaniu významnosti pracovných dní, čo by viedlo k odlišným výsledkom. Jedná sa najmä o opätovné očistenie pri využití iného rozsahu dát, či už z dôvodu aktualizácie dát alebo zmeny metodiky výpočtu.*

Ďalšou dôležitou časťou kalendára je **vplyv veľkonočných sviatkov.** Na rozdiel od fixných sviatkov sú v každom roku v inom období, čo každoročne mení počet voľných a pracovných dní v mesiacoch marec a apríl, resp. v 1. a 2. štvrťroku.

Poznámka: *Základné nastavenie softvérov predpokladá vplyv veľkonočných sviatkov už 6 dní pred Veľkonočnou nedeľou. Túto hodnotu je možné zmeniť a testovať v rozmedzí 1-15 dní (TRAMO/SEATS), resp 1-20 dní (X12-ARIMA)). Od vhodného nastavenia závisí štatistická významnosť tohto efektu a najmä kvalita (vyhladenosť) v mesiacoch (marec, apríl) či štvrťrokoch (1-2 štvrťrok). Treba si však uvedomiť, že efekt veľkonočných sviatkov nezahŕňa vplyv jednotlivých sviatočných dní, ale len ovplyvnenie hodnôt dát pred Veľkonočnou nedeľou (napr. zmena cien niektorých potravín alebo cien pohonných hmôt). Teda Veľký piatok, Veľkonočná nedeľa a Veľkonočný pondelok, ktoré majú z hľadiska produkčných časových radov nezanedbateľný vplyv, nie sú zahrnuté. Preto je ich potrebné zadávať do pohyblivých sviatkov kalendára.*

Aplikácia – Sezónne očisťovanie HICP : *Pre potreby korektného sezónneho očistenia bol vytvorený Slovenský kalendár pre všetkých 56 radov – (slovenské sviatky + veľkonočné sviatky (piatok, nedeľa, pondelok)). Efekt veľkonočných sviatkov sa testoval pre každý časový rad zvlášť (použili sme automatické, ako aj manuálne nastavenia dĺžky vplyvu blížiacich sa sviatkov).*

Príklad : *ceny pohonných hmôt rastú viac ako cena ropy, resp. neklesajú ako cena ropy (v dôsledku vyššieho dopytu po pohonných látkach v tomto období).*

Na nasledujúcom grafe sú sezónne faktory[3] pri cenách pohonných látok, kde je vidieť vplyv Veľkej noci ako pohyblivého sviatku na mesiace marec a apríl. Sezónnosť sa v týchto mesiacoch mení podľa toho, či sú v danom mesiaci veľkonočné sviatky, alebo nie.

[3] Časový vývoj úrovne sezónnej zložky pre jednotlivé mesiace (štvrťroky)

Zdroj: prepočty autorov

4. Outliery

Poslednou kľúčovou časťou nastavenia sezónneho modelu je **hľadanie outlierov**. Nevhodný výber, identifikácia a lokalizácia outlierov má často za následok očistenie časových radov, ktoré nezobrazuje skutočný vývoj. Taktiež spôsobuje veľké zmeny a revízie historických časových radov pri nových zverejneniach.

Poznámka: Outlier *je vo svojej podstate, výnimočná, extrémna a hlavne neočakávaná hodnota alebo zlom. Môže byť jednorazový – aditívny outlier (AO - je v iregulárnej zložke), permanentný – zlom – level shift (LS - je v trendovej zložke) alebo dočasný, postupne sa vytrácajúci – transitory change (TC - je v iregulárnej zložke).*
Graficky by sme ich mohli znázorniť asi takto:

Zdroj: Autori *Zdroj: Autori* *Zdroj: Autori*

Príklad: V skutočných slovenských dátach boli všetky tri druhy outlierov identifikované na časovom rade HDP, ktorý je očistený pomocou metódy TRAMO/SEATS.

Zdroj: ŠÚ SR, Autori

Poznámka: *Pri volatilných časových radoch, najmä pri nejasnej alebo nejednoznačnej sezónnosti, je počet identifikovaných outlierov veľmi citlivý od zvolenia kritickej hodnoty. V niektorých prípadoch teda dochádza k falošnej identifikácii outliera, pričom lokálne dochádza k posunu trendu sezónne očisteného časového radu. V takýchto prípadoch je vhodné pomocou zmeny kritickej hodnoty znižovať citlivosť detekcie outlierov. Veľkosť kritickej hodnoty treba nastavovať tak, aby počet a výskyt outlierov mal svoje ekonomické opodstatnenie.*

Automatický modul (pri zapnutom vyhľadávaní a testovaní prítomnosti outlierov) dokáže identifikovať outliery v celom rozsahu dát, t.j. na začiatku aj na konci. I keď identifikácia na konci časového radu je často podmienená výraznou zmenou v časovom rade. Pre zníženie samotnej volatility sezónne očistených časových radov je vhodné zadávať „známe“ outliery fixne[4] . Počet identifikovaných outlierov do veľkej miery závisí aj od nastavenia kritickej hodnoty testovacej štatistiky.

Poznámka: *Software Demetra+ poskytuje možnosť zvoliť si presne špecifikovaný časový úsek na vyhľadanie alebo nevyhľadanie outlierov. Čiže v prípade neželanej identifikácie outliera, ktorý nemá svoje opodstatnenie, je možné toto obdobie vynechať, alebo v prípade potreby hľadať outliery len na konci časového radu.*

[4] Používateľ zadáva len lokalizáciu (obdobie) a typ outliera (additive, transitory change, level shift). Veľkosť a vplyv je automaticky vypočítaný použitou metódou.

Aplikácia – Sezónne očisťovanie HICP : *Detekcia outlierov komplikovala nájdenie stabilného sezónneho modelu v čase (stabilný model je ten, ktorý sa na rôznych časových obdobiach nemení). V tomto prípade pre detekciu outlierov boli nastavené vyššie kritické hodnoty (inak by bolo veľmi veľa identifikovaných outlierov a ich interpretácia by bola diskutabilná). Pre zjednodušenie a opodstatnenosť boli nastavené fixné outliery na obdobia 7/99, 1/03, 1/04, 1/11, ktoré reprezentovali zmeny výšky sadzby DPH.*

Kedy časový rad neočisťovať

Niekedy sa stáva, že očistením sa časový rad „nevyhladí“, ale zostáva dokonca rovnako „zubatý“ ako predtým[5]. V takýchto prípadoch je vhodné radšej časový rad neočisťovať. Pred samotným rozhodnutím odporúčame takéto dáta podrobiť tzv. kvázi real-time simulácii, ktorá odhalí, či takéto „nepekné“ očistenie je aspoň stabilné v čase. Jednoznačne zamietnuť proces sezónneho očistenia odporúčame aj vtedy, ak iregulárna zložka je výraznejšia (väčšie odchýlky a absolútna veľkosť) v porovnaní so sezónnou zložkou.

Aplikácia – Sezónne očisťovanie HICP : *Pre vybrané časové rady (56 – po selekcii na základe grafickej analýzy a spustení automatických modulov) sme testovali a hľadali najvhodnejšie sezónne ARIMA modely s ohľadom na dĺžku časového radu (má význam hlavne pri skracovaní z dôvodu zmeny metodiky výpočtu alebo pri „odstrihnutí“ obdobia, v ktorom bola kvalita dát diskutabilná), detekciu (lokalizáciu) a opodstatnenosť outlierov, efekt Veľkej noci. Dôvodom na vyradenie časového radu zo zoznamu „kandidáta na sezónne očistenie“ nastal najmä v prípade, ak softvér nenašiel sezónny ARIMA model (často to bol časový rad determinovaný skôr veľkým množstvom outlierov než sezónnym modelom).*

Priame, či nepriame očistenie?

Niektoré súbory dát sú priamo previazané buď metodicky, alebo technicky. Ako príklad možno uviesť cenový index HICP a jeho zložky – celkový index sa vypočíta vážením jednotlivých zložiek. Inou skupinou je napr. HDP, ktorý je súčtom jeho zložiek. Pri potrebe očistiť takéto skupiny údajov vzniká otázka, ako technicky očistiť celkový komponent. Či priamo, teda samostatne, alebo nepriamo, pričom celkový komponent vznikne agregáciou subkomponentov, t. j. ako vážený priemer sezónne očistených zložiek (v prípade cenového indexu) alebo ako súčet jeho zložiek (v prípade produkčného ukazovateľa, napr. HDP). Jednoznačná teória či návod, ako sa rozhodnúť, neexistuje. K dispozícii je len všeobecné pravidlo:

– ak komponenty (subindexy) majú rovnaké charakteristiky a rovnaké časovanie pre vrcholy/maximá
a dná/minimá, je vhodné použiť priame očistenie;
– ak komponenty (subindexy) nemajú rovnakú charakteristiku alebo ak sa relatívna dôležitosť komponentov/
subindexov (podľa dôležitosti/váh) mení veľmi rýchlo, malo by sa použiť nepriame očistenie.

Aplikácia – Sezónne očisťovanie HICP : *HICP a jeho zložky sú typickým príkladom výberu priameho alebo nepriameho očistenia. Kritériom výberu priameho či nepriameho prístupu bola „zosúladenosť“ jednotlivých zložiek toho-ktorého komponentu. Pri agregáte* ***nespracované potraviny*** *je z grafu zrejmé, že sezónne faktory jednotlivých zložiek (komponentov agregátu) nemajú rovnaké vrcholy a dná. Pri tomto type dát je vhodnejšie*

[5] V týcho prípadoch nie je odhadnutá sezónna časť ARIMA modelu.

použiť nepriame očistenie, pretože sezónne faktory zložiek majú svoje vrcholy a dná v rôznych obdobiach.

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Druhým príkladom je časový rad cien ***spracovaných potravín****, kde sezónne faktory zložiek majú svoje dná a vrcholy v rovnakých obdobiach (až na cukor). Pri tomto type dát je možné použiť priame očistenie.*

Výrazné rozdiely v priamo a nepriamo očistenom rade môžu poukazovať na zle identifikovanú sezónnosť. Ak komponenty (subindexy) majú opačné, posunuté alebo neproporcionálne sezónne koeficienty (aditívne verzus multiplikatívne), pri priamom očistení môže nastať ich „zatienenie“, anulovanie, t. j. sezónnosť nie je identifikovaná v plnom rozsahu.

Nepriame očisťovanie a výber špecifikácie v rámci skupiny/odvetvia

V prípade, že očisťujeme väčšie množstvo dát (skupinu dát), ktoré spolu nevyhnutne súvisia, je vhodné uvažovať so súborom dát ako celkom a nielen sa pozerať na jednotlivé časové rady samostatne. Je potrebné sa najmä zamerať na najužšie vzťahy medzi radmi. Teda ak sa pri procese sezónneho očistenia uvažuje s vplyvom outlierov a kalendára, tak je potrebné takéto nastavenie použiť na všetky súvisiace časové rady. Minimálne je potrebné používať rovnaké metódy sezónneho očistenia (X12-ARIMA alebo TRAMO/SEATS).

Poznámka: *Ak si zoberieme ktorúkoľvek zložku HDP, napr. spotrebu domácností, tak v rámci metodiky ESA môžeme uvažovať o sezónnosti v nominálnych, reálnych dátach a ich deflátore ako jednom celku. Keďže je medzi nimi priamy vzťah, čo sa týka sezónne neočistených dát, treba rovnako hľadať priamy vzťah aj pri sezónne očistených údajoch. Keďže je jedna zložka dopočet k ostatným (v tomto prípade deflátor), pri sezónne očistených radoch bude rovnako jedna zložka dopočet (deflátor). Ak napr. identifikujeme outlier v nominálnych dátach konečnej spotreby domácností, potom s ním treba uvažovať aj v reálnych dátach (v takom prípade ide o dopytový šok) alebo v nominálnych dátach a v deflátore (cenový šok), alebo vo všetkých troch (dopytovo-cenový šok). Ak je outlier potvrdený len v jednej zložke, nastáva problém s jeho správnou identifikáciou. Takýto výsledok môže poukazovať aj na chybu v pôvodných dátach (sezónne neočistených).*

Parciálne očisťovanie

Pri zmene metodiky zo strany ŠÚ SR, Eurostatu či inej inštitúcie často dochádza k skracovaniu časových radov. Ak je dôležité zachovať údaje aj z časti, ktorá sa už nepublikuje, je možné pristúpiť k parciálnemu očisťovaniu, resp. očisťovaniu po častiach. Ide o špecifický prístup k sezónnemu očisťovaniu s mnohými komplikáciami, ktoré s tým súvisia. Časový rad sa rozdelí na vybrané časti (metodologicky odlišné, alebo časti, ktoré sú

z hľadiska kvality zberu dát alebo samotnej povahy odlišné), ktoré sa očistia samostatne (musia však mať minimálnu dĺžku 4 roky). Hlavný problém nastáva pri spojení takýchto radov. Ak k spojeniu dochádza na ucelených rokoch, problém môže byť v medziročných rastoch v nasledujúcom roku po spojení. Ak k spojeniu dochádza v priebehu roka, komplikácia nastáva aj pri zachovaní ročného súčtu/priemeru. Toto sa rieši určitou formou aproximácie (interpolácie) v danom roku, avšak na úkor správne identifikovaných sezónnych faktorov.

Aplikácia – Sezónne očisťovanie HICP : *Problém spájania parciálne očistených častí jedného časového radu sa vyskytol aj pri očisťovaní HICP inflácie. V grafe vidieť príklad, keď časový rad indexu cien služieb bol rozdelený na tri rôzne obdobia podľa tvaru sezónnej zložky, pre ktoré bol samostatne nájdený vhodný model (jednotlivé časové úseky sa očisťovali samostatne). Výsledky sa porovnali s očistením celého časového radu jedným modelom (konkrétne modelom odhadnutým na dátach z rokov 2005 až 2011). Bod spojenia (na prelome rokov 2004 a 2005) v prípade parciálneho očistenia identifikuje opačné smerovanie medzimesačnej dynamiky v porovnaní so sezónnym očistením celého časového radu. Takýto výsledok na jednej strane môže dávať opticky krajší výsledok, na druhej strane však nemusí odzrkadľovať skutočnosť, čo by v tomto prípade poukazovalo na zle identifikovanú sezónnosť v bode spojenia. Z tohto dôvodu sa uprednostnilo sezónne očistenie celého časového radu pred parciálnym očistením.*

Zdroj: prepočty autorov Zdroj: prepočty autorov

Úvahy o aditivite a ročných priemeroch, resp. súčtoch

Často sa stáva, že po sezónnom očistení dochádza k zmene ročných údajov, t. j. pri objemových radoch nie je rovnaký ročný objem očistených a neočistených dát a pri indexových radoch nesedia ročné priemery očistených a neočistených dát. Z technického hľadiska to spôsobuje samotný proces očistenia, kde k identifikácií sezónneho modelu, trendu, outlierov a kalendárnych efektov (spolu s efektom veľkonočných sviatkov) dochádza v rôznych iteráciách/krokoch. Na jednej strane sa ročný súčet, resp. priemer má zhodovať pri očistených aj pôvodných dátach[6], čím by sa docielilo rovnaké posúdenie celoročných údajov, najmä pri porovnaniach medzi rôznymi rokmi (tento prístup je aj z hľadiska použitia jednoduchší a ľahšie interpretovateľný pri (ekonometrickom) modelovaní a tvorbe prognóz). Na druhej strane je logické, že sezónne očistené dáta nemusia zákonite spĺňať ročné súčty (priemery).

[6] Ecotrim, vyvinutý Eurostatom, je jednou z možností ako upraviť ročné súčty/priemery sezónne očistených dát.

Príklad: *V jednom roku, ktorý je priestupný, došlo k postupnému nárastu produkcie v prvom polroku a veľkonočné sviatky pripadli na apríl. Logicky aj z technického hľadiska je v prvom štvrťroku produkcia nadhodnotená vplyvom jedného pracovného dňa navyše z dôvodu priestupného roka a posunom veľkonočných sviatkov do druhého štvrťroka. Ak sa rozhodneme, aby metóda očistila kalendárne vplyvy, tak dôjde pri sezónne očistených dátach k zníženiu hodnôt produkcie v prvom štvrťroku (v pôvodných, teda sezónne neočistených dátach je z dôvodu priestupného roka produkcia vyššia, pretože máme o jeden deň viac oproti ostatným rokom). Zahrnutie efektu veľkonočných sviatkov v druhom štvrťroku spôsobí v sezónne očistených dátach zvýšenie produkcie (v pôvodných dátach došlo v skutočnosti k poklesu z dôvodu zníženého počtu pracovných dní) v druhom štvrťroku a naopak zníženie v prvom štvrťroku (v pôvodných dátach došlo v skutočnosti k nárastu z dôvodu zvýšeného počtu pracovných dní). Keďže priestupný rok sa opakuje až každé štyri roky a veľkonočné sviatky sú nepravidelné, týmito metodologicko-technickými presunmi dôjde k narušeniu rovnosti ročných dát. V skutočnosti na narušenie rovnosti ročných dát postačuje očistenie od vplyvu priestupného roka, pre zachovanie súčtu (priemeru) dát z dôvodu priestupného roka (celom časovom rade).*

Vplyv kalendára má výraznejší dopad na sezónne očistené dáta produkčných ukazovateľov a naopak nižší vplyv na cenové ukazovatele, teda na premenné, ktoré nie sú ovplyvňované počtom pracovných dní, resp. sviatkov. Na túto skutočnosť treba myslieť aj pri otázke voľby zahrnutia kalendára do sezónneho modelu.

Očisťovanie vplyvu kalendára a použitie dodatočnej úpravy smerom k dodržaniu ročných priemerov pôsobí kontraproduktívne. Kým sezónna metóda očistí dáta o pravidelné vplyvy a kalendárne efekty, tak proces „zosúladenia“ očistených a pôvodných dát spôsobí zmenu identifikovanej sezónnosti.

Podobný problém nastáva pri očisťovaní agregátu a jeho zložiek priamou metódou (t.j. agregát aj zložky sú očistené priamo). Súčet, resp. vážený priemer sezónne očistených zložiek nie rovný sezónne očistenému agregátu. Pre dodržanie rovnosti:

- rozdiel sa rozpočíta medzi jednotlivé zložky na základe váh – spôsobí zmenu identifikovanej sezónnosti vo všetkých zložkách,
- rozdiel sa pripočíta do najmenej významnej zložky (v prípade HDP napr. diskrepancia) – čo samozrejme môže spôsobiť, že táto zložka získa podstatne vyššiu váhu v niektorých obdobiach čím sa stane ekonomicky neinterpretovateľnou,
- rozdiel sa pripočíta do objemovo najväčšej zložky (v prípade HDP napr. spotreba domácností).

Ak problém aditivity nie je prioritou, ale dôraz kladieme na individuálne očistenie zložiek, resp. agregátu tak máme tie to dve možnosti:

- očistený agregát získať nepriamo, teda súčtom resp. váženým priemerom očistených zložiek – výsledné sezónne očistenie nemusí prinášať želané „vyhladenie“,
- všetky zložky aj agregát sa očistia priamo.

Posledná pomoc alebo „keď nevieme ako ďalej“

Veľakrát sa dostaneme do situácie, kedy nevieme, ku ktorému alternatívnemu nastaveniu sa prikloniť, alebo odhadnutý sezónny model sa s pridávaním ďalšieho pozorovania výrazne mení, či sezónnosť je raz štatisticky potvrdená a druhýkrát zamietnutá v závislosti na

identifikácii pracovných dní. V takomto prípade odporúčame využívať prednastavenú špecifikáciu označovanú ako RSA1[7], ktorá v nastavení odhaduje AIRLINE model (ARIMA (011)(011)), testuje časový rad na prítomnosť outlierov, abstrahuje od kalendárnych efektov a vplyvy Veľkonočných sviatkov. Výsledkom je relatívne hladký a stabilný časový rad.

Zmeny sezónneho modelu

Pri potrebe pravidelného očisťovania časových radov z dôvodu aktualizácie alebo zverejnenia nových dát je pre uľahčenie práce možných niekoľko procesov, ktoré sa líšia v hĺbke úrovní nastavení parametrov sezónneho očistenia. Každý proces sa vykonáva bez potreby opätovného nastavenia sezónnej procedúry zo strany používateľa, čo prináša svoje výhody a nevýhody.

Najjednoduchším spôsobom je **bežná a pravidelná aktualizácia (current adjustment)**, kde sa pôvodné nastavenia špecifikácie a parametrov ARIMA modelov, outlierov, kalendárnych regresorov používajú aj na nové dáta (v podstate na celý časový rad) dovtedy kým nedôjde k novému preodhadnutiu zo strany používateľa (zvyčajne raz za rok alebo pri revízií dát). Takýmto spôsobom sa získavajú v čase najstabilnejšie časové rady (medzi jednotlivými aktualizáciami dát).

Ďalším spôsobom ako pristupovať k aktualizáciám je **čiastočná priebežná úprava (partial concurrent adjustment)**. V tomto prípade je možných viacero úrovní, pri ktorých má používateľ na výber, ktorú z časti nastavenia dať nanovo procedúre odhadnúť.

Možnosti partial concurrent adjustment:

- *odhadujú **parametre** ARIMA modelu pri **fixovaní** jeho špecifikácie, outlierov a kalendárnych regresorov,*
- *odhadujú parametre ARIMA modelu pri fixovaní jeho špecifikácie, outlierov a kalendárnych regresorov pričom sa **posledné outliery** preodhadujú,*
- *odhadujú parametre ARIMA modelu pri fixovaní jeho špecifikácie, kalendárnych regresorov pričom sa **všetky outliery** preodhadujú,*
- *odhadujú **parametre aj špecifikácia** ARIMA modelu, pri fixných kalendárnych regresoroch pričom sa **všetky** outliery preodhadujú.*

Najvšeobecnejší proces je **priebežná úprava (concurrent adjustment),** čo znamená, že ARIMA model (tak špecifikácia ako aj parametre), outliery a kalendárne regresory sa identifikujú a odhadujú nanovo. Pri tejto úprave dostávame najaktuálnejšie očistené časové rady, avšak ich stabilita v čase je najhoršia, teda sezónne očistené rady sú z aktualizácie na aktualizáciu pomerne volatilné. Tento prístup sa odporúča použiť, ak ide o zmenu metodiky výpočtu pôvodných dát, alebo o revíziu dát, pri ktorej sa mení časť časového radu.

Výber úpravy je veľmi individuálny a vedie aj k rôznym výsledkom. Stratégiu pri sezónnom očisťovaní si každý musí zvoliť sám. Napríklad priebežná úprava (concurrent adjustment) vedie ku častejším revíziam, ale prináša presnejšie výsledky. Ďalšou dôležitou otázkou je stabilita, čo znamená, že pri preodhadovaní by sme mali zvážiť, či nový model je vhodnejší ako ten predchádzajúci (či sa „oplatí“ vôbec meniť model) a pozerať sa aj na to, ako sa zmenia historické údaje pri každom novom odhade. Vždy v tomto prípade treba hľadať kompromis.

[7] V programoch je často označený ako základný automatický modul.

Aplikácia – Sezónne očisťovanie HICP : *Testovanie stability (stálosti, nemennosti, pokiaľ ide o historické časové rady) alebo inak povedané volatility medzi jednotlivými vydaniami či aktualizáciami) sezónne očistených časových radov je možné robiť pomocou kvázi real-time simulácií bežnej/pravidelnej (current adjustment), čiastočne priebežnej (partial concurrent adjustment) alebo priebežnej (concurrent adjustment) aktualizácie. Príklad, keď priebežná úprava (concurrent adjustment) alebo zafixovanie modelu (current adjustment) nespôsobovalo výrazný posun v historických dátach, je v nasledujúcich dvoch grafoch na časovom rade* ***spracovaných potravín****.*

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Na druhej strane pri iných časových radoch boli viditeľné rozdiely v historických dátach pri zafixovaní modelu a priebežnej úprave. Takýmto príkladom je časový rad cien ***vybavenia pre šport, kempovania a rekreáciu*** *(nasledujúce grafy). Zafixovanie modelu viedlo k stabilnejším historickým radom a naopak, priebežná úprava viedla k výraznejším historickým posunom.*

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Predikcie sezónne očistených časových radov, alebo ako na to s budúcnosťou

Algoritmy na očisťovanie časových radov X12-ARIMA a TRAMO/SEATS dokážu automaticky vygenerovať trend a sezónne faktory až na dva roky dopredu, pričom veľkosť ich sezónnych faktorov je v oboch rokoch rovnaká. Ak abstrahujeme od iregulárnej zložky, vieme takýmto spôsobom získať priamo zo softvéru[8] predikciu sezónne očisteného aj neočisteného časového radu. V prípade dostupnosti predikcie sezónne neočisteného časového radu je možné aj takýto časový rad podrobiť sezónnemu očisteniu. Ak by sme však takýto neočistený

[8] Napr. Demetra+.

časový rad s predikciou podrobili opäť sezónnemu očisteniu, nový sezónne očistený časový rad bude iný ako pôvodný, a to najmä na horizonte predikcie[9].

Záver

„Sezónne očistenie je nekončiaci proces hľadania toho správneho, ale nikto nevie čo je vlastne to správne!!“
autori

V predchádzajúcich riadkoch sme načrtli hlavné časti procesu sezónneho očistenia časových radov. Na základe uvedených informácií a príkladov by sezónne očistenie už nemuselo byť len čiernou skrinkou, ktorá nejakým spôsobom poskytne vyhladenejší časový rad, ale pomôže pri ľahšom pochopení neočakávaných výsledkov a pri získaní z hľadiska ekonomickej interpretácie dôveryhodnejšieho časového radu. Treba však pripomenúť, že uvedené spôsoby a príklady nie sú univerzálne a nemožno ich automaticky použiť na všetky typy dát.

Aplikácia – Sezónne očisťovanie HICP : *Projekt sezónneho očisťovania cenových indexov slúžil hlavne na nájdenie najvhodnejších metód a postupov, ktoré by viedli k uspokojivým výsledkom zohľadňujúcim požiadavky, ako je ekonomická interpretovateľnosť, jednoduchosť z hľadiska pravidelných zverejnení a v neposlednom rade stabilita, čiže nízka volatilita medzi aktualizáciami v čase. Na základe výsledkov by sa mali sezónne očisťovať časové rady celkovej inflácie meranej HICP, spracovaných a nespracovaných potravín, priemyselných tovarov bez energií a služieb.* ***Ceny energií*** *vzhľadom na ich reguláciu nie je vhodné sezónne očisťovať.*

Sezónne očisťovanie prebiehalo súčasne metódami X12-ARIMA a Tramo/Seats. Výsledky ukázali len malé rozdiely. Uprednostnená by mala byť metóda X12-ARIMA, pri ktorej je možnosť širšieho nastavenia vplyvu veľkonočných sviatkov.

Pre získanie čo najlepších výsledkov boli všetky časové rady sezónne očistené priamo aj nepriamo. Zo všetkých 180 časových radov sa procesom postupnej selekcie vybralo 56, ku ktorým sa pristupovalo individuálne a podrobne.

Rozhodujúcim kritériom pri voľbe priameho či nepriameho očistenia cenového indexu bola volatilita (hladkosť) výsledného časového radu s ohľadom na stabilitu v čase. V prípade malých rozdielov v prospech nepriameho očistenia sa uprednostnilo priame očistenie z dôvodu jednoduchšej pravidelnej aktualizácie. Takáto situácia nastala len v prípade ***nespracovaných potravín****. Pri všetkých ostatných časových radoch (spracované potraviny, priemyselné tovary bez energií, služby) sa výsledky kritéria volatility priklonili k priamemu očisteniu*[10].

[9] V skutočnosti sa očisťuje iný (dlhší) časový rad.
[10] S vedomím nedodržanej aditivity.

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Zdroj: prepočty autorov

Keďže energie sú jednou z hlavných zložiek HICP s relatívne veľkou váhou a výraznou volatilitou, nedochádza pri nepriamom očistení celkového HICP k „želanému vyhladeniu“ a časový rad sa javí ako neočistený. Preto pri celkovom HICP by sme uprednostnili priame očistenie, ale vzhľadom na odporúčanie pracovnej skupín ECB pre sezónne očisťovanie HICP sa bude celkový index očisťovať nepriamo agregáciou sezónne očistených hlavných komponentov (t. j. cenových indexov pre nespracované potraviny, spracované potraviny, priemyselné tovary, služby) a sezónne neočistených energií.

Zdroj: prepočty autorov