
ОЦЕНА ТЕКУЋЕ ИНФЛАЦИЈЕ НА ОСНОВУ ЦЕНА СА ИНТЕРНЕТА

Мирко Ђукић, Ива Крмановић, Миодраг Петковић

© Народна банка Србије, март, 2023.

Доступно на www.nbs.rs

За ставове изнете у радовима у оквиру ове серије одговоран је аутор и ставови не представљају нужно званичан став Народне банке Србије.

Сектор за економска истраживања и статистику

НАРОДНА БАНКА СРБИЈЕ

Београд, Краља Петра 12

Тел.: (+381 11) 3027 100

Београд, Немањина 17

Тел.: (+381 11) 333 8000

www.nbs.rs

Оцена текуће инфлације на основу цена са интернета

Мирко Ђукић, Ива Крсмановић, Миодраг Петковић

Апстракт: Рад се бави приказом методологије коју Народна банка Србије примењује за оцену инфлације у реалном времену на основу цена са интернета, које скида аутоматски тзв. *web-scraping*-ом. Специфичност метода који користи Народна банка Србије огледа се у томе што се не базира само на ценама за онлајн куповину, већ на свим релевантним информацијама о ценама, укључујући оне које су на интернету приказане само информативно. Поред тога, Народна банка Србије настоји да изворима са интернета покрије што већи број производа из корпе индекса потрошачких цена (око 90% у тренутку писања овог рада), са идејом да тиме добије поузданију оцену централне тенденције инфлације. У првој години примене овог метода резултати су охрабрујући – оцењена инфлација се у просеку у посматраном периоду налазила на нивоу реализоване (оцене нису пристрасне), просечно апсолутно одступање износило је 0,20 процентних поена, а медијално 0,13 процентних поена, што није значајно с обзиром на то да је посматрани период карактерисала релативно висока и волатилна инфлација.

Кључне речи: прогнозирање инфлације, цене са интернета, *web-scraping*, *big data*

[JEL Code]: C53, E17, E58

Нетехнички резиме

Основни циљ Народне банке Србије је постизање и очување ценовне стабилности. Главну меру инфлације, стопу раста индекса потрошачких цена, Републички завод за статистику објављује 12-ог у месецу за претходни месец. Постоје бројни разлози зашто је креаторима монетарне политике важно да и пре тог датума имају индикацију у ком правцу се може очекивати кретање инфлације, односно да имају рани индикатор инфлације, а један од све популарнијих метода да се до њега дође је употреба цена са интернета, које се скидају аутоматски, тзв. web-scraping-ом. У прилог све веће употребе овог метода је доступност великог броја цена на интернету, што, уз растућу снагу компјутера, омогућава скидање велике количине информација са веба. У случају оцене индекса потрошачких цена, у питању је преузимање великог броја цена производа који су упоредиви са производима који се користе за израду ИПЦ-а.

Главна предност овог приступа огледа се у могућности праћења цена на недељном, или чак дневном нивоу, уз његову трошковну и временску ефикасност. Недостаци се односе на мањак гаранција да ће веб-сајтови или приватне компаније наставити да објављују цене у обиму, квалитету или форми на начин на који то сада раде. Тако добијене процене цена представљају комплемент званичној методологији израде ИПЦ-а и значајан инпут у предвиђању даљег кретања инфлације.

За потребе оцене текуће инфлације, у Народној банци Србије користимо све изворе информација који могу допринети бољој оцени, који укључују веб-сајтове малопродаваца, трговинских ланаца, информативне сајтове који садрже цене производа или се баве упоређивањем цена, као и сајтове државних органа који се баве регулисаним ценама. Користећи разне изворе, директно са интернета покрили смо око 90% корпе ИПЦ-а, што наш приступ знатно разликује од упоредивих радова, који се махом баве оцењивањем поткатогија у оквиру индекса цена, попут прехранбених производа. За цене које нисмо у могућности да преуземо са интернета користимо њихове оцене, добијене помоћу импутације.

Резултати примене овог приступа у Народној банци Србије у првој години су охрабрујући – оцењена инфлација у просеку се у посматраном периоду налазила на нивоу реализоване, а просечно апсолутно одступање износило је 0,2 процентна поена, што није значајно с обзиром на то да је инфлација у овом периоду била волатилна и релативно висока (1,2% месечно у просеку). Поред тога, корелисаност између остварених и оцењених доприноса инфлацији појединачних производа по месецима релативно је висока. У даљем раду, у жељи да побољшамо методологију, намеравамо да у процес укључимо нове изворе података са интернета, који ће додатно допринети већој прецизности оцене.

Садржај:

1. Увод	10
2. Инфлација и цене са интернета.....	13
3. Преузимање података са веба и њихово повезивање са индексом потрошачких цена	15
3.1. Избор репрезентативних производа.....	16
4. Процес оцене текуће инфлације	17
5. Остварења оцена текуће инфлације.....	20
6. Закључак.....	23
Додатак.....	24
Литература	25

1. Увод

Изобиље података о ценама разних производа и услуга на интернету може послужити као добар извор информација за оцену инфлаторних кретања у реалном времену. Предуслов за то је да се постави систем којим се та велика количина података преузима аутоматски, а затим и обрађује адекватним методама. Циљ овог рада је да представи начин на који се то ради у Народној банци Србије.

При изради средњорочних пројекција, текући податак о инфлацији представља значајну улазну информацију. Поред тога што је основ за анализу инфлаторних притисака, текући податак директно, обрачунски, улази у пројекцију међугодишње стопе за наредну годину дана, док на пројекцију утиче и индиректно, преко инерције у кретању месечне односно тромесечне стопе инфлације.

Проблем с којим се централне банке сусрећу у пракси, међутим, је тај што је званични податак о текућој инфлацији доступан са одређеним временским помаком, што онемогућава његово правовремено укључивање у пројекцију. Примера ради, званичан податак у Србији објављује се око 3 седмице након што је већина цена која улази у обрачун инфлације за тај месец већ снимљена од стране Републичког завода за статистику. С обзиром да је ово слично у многим земљама, развијен је велики број техника и модела – економетријски модели, анкете, истраживања јавног мњења и експериментална истраживања – који управо дају прве процене нивоа цена и инфлације, пре него што је званичан податак објављен (*nowcasting* – оцена текућег стања¹). Свакако званичан податак Републичког завода за статистику остаје најбоља и најпрецизнија мера инфлације управо због обухвата и процедура којима се ови подаци снимају.

Доступност великог броја цена на интернету, уз растућу снагу компјутера, условио је да се за оцене текућих инфлаторних кретања последњих година све више користи техника тзв. *web-scraping*-а, која заправо значи скидање података са веба. Овај метод је један од релативно трошковно и временски ефикаснијих. Поред цена, сет података може укључивати и информације о производу, попустима, доступности, а фреквенција може бити не само месечна већ и недељна или чак дневна (*Macias & Stelmasiak, 2019*). Релевантност прибављених цена овом методом веома је значајна, уз додатну предност што се њихова динамика може пратити у реалном времену (*Powell et al., 2018*). За разлику од анкетних метода, подаци су доступни одмах и могу се прибавити директно, без посредника. У зависности од потреба, цене прикупљене на овај начин могу бити засебна категорија цена на интернету, могу бити апроксимација само једне категорије (на пример цена хране у оквиру индекса потрошачких цена (ИПЦ)) или апроксимација целокупног ИПЦ-а.

С друге стране, неопходно је узети у обзир и недостатке оваквог приступа за оцену инфлације, који проистичу пре свега из саме природе податка који је непосредни инпут. Постоји и проблем с доступношћу података, при чему не постоји гаранција да ће веб-

¹ Енглески термин *nowcasting* буквално би се могао превести као „предвиђање садашњости“, при чему га ми за потребе овог рада преводимо као „оцена текућег стања (инфлације)“

-сајтови или приватне компаније наставити да објављују цене у обиму, квалитету или форми на начин на који то сада раде (*Kapetanios & Papailias*, 2018). Јасно је да је главна предност коришћења података добијених са интернета њихова висока фреквенција, али замка се скрива у претпоставци да ови подаци могу бити замена за традиционално прикупљање и анализу података, док би правилније било рећи да је у питању комплемент званичној методологији израде ИПЦ-а.

Један од првих аутора који се бавио оценом тренутне инфлације помоћу цена са интернета био је *Cavallo* (2013), који је у свом првом раду покушао да добије оцену у пет латиноамеричких држава користећи цене које се односе на храну, безалкохолна пића и производе коришћене у домаћинствима, а осим уобичајених разлога наводи да овакав индекс може бити коришћен за добијање алтернативних мера инфлације када је званична статистика изгубила свој кредибилитет. Добио је резултат да су онлајн индекси добро апроксимирани и нивое и динамику кретања званичне инфлације када су у питању Бразил, Чиле, Колумбија и Венецуела. С друге стране, када је у питању Аргентина, добијен је статистички значајан резултат који указује на несагласност између онлајн цена и званичне статистике, и то у периоду када је аргентинска влада критикована због сумњи у валидност званичних мера инфлације.

Значајан извор информација за ову врсту истраживања пружа „Пројекат милијарду цена”, који је био академска иницијатива на *MIT* и Харварду, у коме су у периоду 2008–2016. прикупљане цене од више стотина онлајн трговаца широм света на дневној бази, а чији је један од идејних креатора управо *Cavallo*. На његов рад својим истраживањем бразилске мере инфлације наставио се *Carvalho* (2020), који је методологију проширио динамичким факторским моделом за оцену тренутне инфлације. Моделски резултати показују статистичку значајност и унапређење пројекције коришћењем динамичких варијабли наспрам коришћења сирових података.

Macias и *Stelmasiak* (2019) из Народне банке Пољске у свом раду користе *web-scraping* као начин да унапреде пројекције прехранбене инфлације. Будући да је храна чинила око четвртине укупних издатака домаћинства у Пољској у тренутку када је рад објављен, као и то да постоји велика волатилност и сезоналност у ценама хране, унапређивање прецизности оцене ових цена показало се као значајно. Као потенцијалну препреку навели су чињеницу да се мали број куповине основних животних намирница обавља преко интернета, па се поставља питање упоредивости малопродајних и интернет цена. Цене добијене *web-scraping* методом агрегиране су у индекс цена хране, који се, упркос томе, показао као статистички значајан и с мањом грешком предвиђања будућих цена званичног индекса у односу на реперни модел *ARMA*. Перформансе модела у оцени прехранбене инфлације постају још боље када се укључе и доцње коришћених варијабли.

Централна банка Јерменије користи *web-scraping* као методолошки приступ чији су главни циљеви праћење динамике цена робе и услуга на тржишту (храна, непрехрамбена, услуга) у реалном времену и добијање брзих процена тренутне инфлације, али и праћење цена станова ради истраживања тржишта некретнина (*Aghajanyan et al.*, 2017). Поред тога, значајан број статистичких завода има своју меру тренутне инфлације базирану на онлајн ценама, укључујући и статистичке заводе у САД

(Horrigan, 2013), Уједињеном Краљевству (Breton et al., 2015), али и у Холандији, на Новом Зеланду и Норвешкој.

Jaworski (2021) бавио се оценом прехранбене инфлације у Пољској током кризе изазване пандемијом вируса корона, такође користећи онлајн цене. Посебно се истиче да је овај начин прикупљања цена изузетно значајан у периоду када није могуће физички прочитати цене као што је био случај за време затварања, као и када је број интернет куповина у великом порасту. Уз сличну методологију, Jaworski је показао да је поуздану процену прехранбене инфлације могуће добити месец дана пре званично објављених индекса. Сличан резултат добијен је и када је у питању прехранбена инфлација у Турској, где су Soybilgen et al. (2021) показали да онлајн индекси инфлације успешно оцењују цену хране.

Од радова који су се бавили оценом целокупне инфлације, Bertolotto et al. (2014) покушали су да унапреде пројекцију инфлације коришћењем модела VAR, који као варијабле узима реализоване индексе инфлације и онлајн цена у протеклих шест месеци. У индексу онлајн цена коришћене су цене које покривају већи део корпе када су у питању храна, безалкохолна пиће и транспорт, док се за остале подгрупе у оквиру ИПЦ-а које нису обухваћене овим категоријама (попут услуга) користе апроксимације на основу кретања групе. Показано је да су онлајн цене респонзивније од малопродајних цена за упоредиве производе, као и да се окидачи за промену цене на интернету спорије и постепеније инкорпорирају и у малопродајне цене.

У односу на поменуте радове, наш приступ има неколико карактеристика. Прва је да се наше оцене не базирају само на ценама за онлајн куповину, већ на свим релевантним информацијама о ценама, укључујући оне које су на интернету приказане само информативно. У ту сврху, поред веб-сајтова самих малопродаваца, користимо и оне који пореде цене различитих малопродаваца, као и сајтове државних органа и јавних предузећа са информацијама о ценама њихових услуга. Поред тога, док се у поменутим радовима аутори инфлацијом углавном баве парцијално (оцењујући поједине компоненте ИПЦ-а, пре свега цене хране), ми смо се определили да изворима са интернета покријемо што више производа из ИПЦ-а. У тренутку писања овог рада, око 90% корпе ИПЦ-а покривено је изворима са интернета, док недостајући део цена импутирамо неком од метода.

Наставак рада структуриран је на следећи начин. Најпре дајемо основне карактеристике метода снимања цена и обрачуна инфлације који примењује Републички завод за статистику (РЗС) и извора са интернета које користимо за снимање цена различитих категорија цена производа и услуга. Следе техничка објашњења начина на који скидамо податке и повезујемо их с корпом ИПЦ-а. Након тога следи део у коме дајемо детаљан приказ процеса оцене текуће инфлације на основу цена са веба који примењујемо у Народној банци Србије, да бисмо на крају приказали резултате досадашње примене овог метода.

2. Инфлација и цене са интернета

Као званична мера инфлације у Србији користи се стопа раста ИПЦ-а, коју РЗС објављује 12. у месецу за претходни месец. Корпа ИПЦ-а се, на најнижем нивоу агрегирања, састоји од 660 производа и услуга (према структури из 2023), чије цене РЗС прати у 15 српских градова, на разним продајним местима. За различите групе у индексу цене се снимају у различитим периодима у месецу:

- непрехрамбени производи – 1–10. у месецу;
- непољопривредни прехрамбени производи – 10–14. у месецу;
- пољопривредни производи – прва и трећа недеља у месецу;
- услуге – 14–23. у месецу;
- горива за аутомобиле – сваке среде у месецу.²

Може се приметити да се званични податак о инфлацији објављује са помаком од неколико седмица у односу на период снимања података. Тај временски помак пружа могућност оцењивања текуће инфлације пре званичног податка, што може бити веома корисно у процесу израде средњорочних пројекција инфлације.

На интернету се могу наћи цене великог броја производа из корпе ИПЦ-а. За потребе оцене текуће инфлације у Народној банци Србије користимо све изворе информација који могу допринети бољој оцени. То могу бити веб-сајтови малопродаваца, с приказаним ценама за онлајн куповину или једноставно као информација за купце; веб-сајтови са упоредним приказима цена различитих малопродаваца; веб-сајтови државних органа са информацијама о ценама њихових услуга. Сајтови могу садржати податак за само једну цену из ИПЦ-а, док неки могу садржати и преко стотину.

При оцењивању текуће инфлације подражавамо методологију РЗС-а, колико је могуће: цене појединих категорија производа прикупљамо у периодима када и РЗС то ради – углавном на неки од дана у каснијем делу распореда РЗС-а, најчешће на последњи радни дан у периоду; у складу с методологијом РЗС-а, из обрачуна искључујемо цене на привременим снижењима; добијене стопе раста са интернета за појединачне производе пондеришемо користећи пондере РЗС-а.

С друге стране, за разлику од РЗС-а, цене углавном не прикупљамо по градовима због тога што: најчешће није ни могуће наћи цене по градовима (малопродавци објављују јединствене цене за земљу); има смисла претпоставити да су промене цена по градовима сличне; укључивање превише извора чини цео процес тежим за праћење и одржавање.

Да бисмо процес што више поједноставили, преферирамо веб-сајтове који прикупљају цене од више малопродаваца (уместо њихових појединачних сајтова). Тако, на пример, користимо један сајт који на дневном нивоу пореди цене више од 6.000 производа (углавном прехрамбених) седам великих малопродајних ланаца; сајт Министарства трговине (СТИПС), који прикупља и објављује цене воћа, поврћа и

² https://data.stat.gov.rs/Metadata/03_Cene/Html/030103_ESMS_G0_2018_2.html

других прехранбених артикала с пијаца из 20 већих српских градова; веб-сајт који пореди цене горива за аутомобиле шест великих малопродаваца горива. Прва два извора покривају највећи део цена прехранбених производа (пољопривредних и непољопривредних), што је веома практично за оцену текуће инфлације.

Када су у питању цене непрехрамбених производа и услуга, ствари су нешто компликованије. У овом сегменту лакши део представљају регулисане цене, које имају значајан удео у домаћем ИПЦ-у (око петине). Ове цене могу се наћи на бројним сајтовима државних органа, као што су државне агенције, локалне самоуправе, јавна предузећа, или у државним билтенима. На пример, цена електричне енергије објављује се на сајту Електропривреде Србије, цене цигарета у „Службеном гласнику РС”, регулисане цене лекова на сајту Републичког фонда здравственог осигурања, цене локалних услуга (јавни транспорт, комуналије, вртићи и др.) на сајтовима локалних самоуправа итд. У случају локалних јавних услуга, док их РЗС прати у 15 већих градова, ми их снимамо у три највећа (Београд, Нови Сад, Ниш), који покривају око 55% узорка РЗС-а.

Слика 1. Део табеле с веб-сајта СТИПС с ценама поврћа

R.Br.	Proizvod	Jed.mere	Beograd(Kalenić)	Beograd(Skadarlija)	Čačak	Kragujevac
1	Blitva (sve sorte)v:srednja p:standardno	veza	40.00	40.00	50.00	
2	Brokola (sve sorte)v:srednja p:standardno	kg	400.00	450.00	350.00	
3	Celer (sve sorte)v:srednja p:standardno	kg	300.00	300.00	250.00	220.00
4	Cvekla (sve sorte)v:srednja p:standardno	kg	120.00		60.00	70.00
5	Dinja (sve sorte)v:srednja p:standardno	kg	500.00			
6	Karfiol (sve sorte)v:srednja p:standardno	kg	350.00	350.00	250.00	
7	Kelj (sve sorte)v:srednja p:standardno	kg	150.00	150.00		
8	Kelj pupčar (sve sorte)v:srednja p:standardno	kg	400.00	300.00		
9	Krastavac (Kornišon)v:srednja p:standardno	kg				
10	Krastavac (salatar)v:srednja p:standardno	kg	350.00	300.00	300.00	230.00
11	Krompir (beli)v:srednja p:standardno	kg	150.00	130.00	80.00	100.00
12	Krompir (crveni)v:srednja p:standardno	kg	150.00	130.00	80.00	100.00
13	Kupus (mladi)v:srednja p:standardno	kg	200.00			
14	Kupus (sve sorte)v:srednja p:standardno	kg	70.00	70.00	30.00	60.00
15	Luk beli (mladi)v:srednja p:standardno	veza	40.00			
16	Luk beli (sve sorte)v:srednja p:standardno	kg	900.00	700.00	500.00	450.00
17	Luk crni (mladi)v:srednja p:standardno	veza	40.00	50.00		40.00
18	Luk crni (sve sorte)v:srednja p:standardno	kg			70.00	80.00
19	Paprika (Babura)v:srednja p:standardno	kg	500.00	450.00		
20	Paprika (ljuta)v:srednja p:posebno	kg				
21	Paprika (ljuta)v:srednja p:standardno	kg				
22	Paprika (ostala)v:srednja p:standardno	kg	500.00	500.00	350.00	
23	Paprika (šilja)v:srednja p:standardno	kg	450.00	450.00	350.00	
24	Paradajz (chery)v:srednja p:posebno	kg	600.00	600.00		
25	Paradajz (sve sorte)v:srednja p:standardno	kg	400.00	350.00	300.00	260.00
26	Pasulj (beli gradištanac)v:srednja p:standardno	kg	400.00	350.00		
27	Pasulj (beli tetovac)v:srednja p:standardno	kg	500.00	400.00		
28	Pasulj (beli)v:srednja p:standardno	kg			350.00	320.00

Извор: РЗС и аутор.

Најизазовнији задатак у овој фази био је налажење извора са интернета за нерегулисани део цена непрехрамбених производа и услуга. (У наставку рада ову групу називаћемо само „не-храна“.) Ова група се састоји од 427 најразличитијих производа и услуга, као што су: обућа, одећа, алкохолна пића, лекови, намештај, посуђе, производи за личну хигијену, разни уређаји, аутомобили (нови и половни), поправка аутомобила, осигурање аутомобила, станарине, зубарске услуге, здравствене услуге, теретана, биоскоп итд. Покривање ове хетерогене групе захтевало је велики број извора. У почетку смо тражили сајтове који покривају већи број непрехрамбених производа, да

бисмо у каснијој фази ишли више у детаље, налазећи изворе за производ по производ, уз давање приоритета онима с већим пондером. Ако су били расположиви, за појединачне производе настојали смо да укључимо више извора.

Користећи разне изворе, директно са интернета покрили смо око 90% корпе ИПЦ-а. Ако се не рачуна група гориво за аутомобиле, која садржи само два производа (бензин и дизел), најбоље покривена група из ИПЦ-а јесте група воће и поврће (98%), затим храна без воћа и поврћа (94%), регулисане цене (91%) и не-храна (85%).

3. Преузимање података са вебa и њихово повезивање са индексом потрошачких цена

За прикупљање података о ценама производа и услуга користимо софтвер *R*. С обзиром на то да је асортиман производа на интернет продавницама променљив и да се позиције одређених производа на којима се заснива обрачун ИПЦ-а мењају унутар саме интернет продавнице, аутоматизовано прикупљање података је засновано на преузимању целокупног асортимана производа из одређене интернет продавнице.

У ту сврху смо креирали програме који аутоматски приступају страници интернет продавнице и у првом кораку прикупљају све веб-адресе путем којих се приступа свим производима у продавници. У наредном кораку прикупљају се цене и карактеристике производа с дефинисане веб-адресе, при чему се, у зависности од техничке карактеристике сајтова интернет продавница, користе пакети за преузимање података *Rvest* и *Rselenium*.

Пакет *Rvest* је био први избор за прикупљање података, пошто омогућава изузетно брзо скидање података у *HTML* формату, без директног приступа сајту интернет продавнице путем веб-прегледача. Ипак, већина модерних интернет продавница креирана је у динамичном окружењу које захтева интеракцију с корисником, тако да смо, ради аутоматизације комплетног процеса прикупљања цена, креирали и програме засноване на пакету *Rselenium*. Овај пакет омогућава да се путем веб-прегледача директно управља различитим захтевима који су неопходни за приказ производа у интернет продавницама. Између осталог, омогућава се одабир ставки из падајућег менија и аутоматско скривање прозора ради читавања свих производа који су доступни на једној веб-адреси. Треба истаћи да су процеси базирани на пакету *Rselenium* знатно спорији, пошто је за преузимање података неопходно да се комплетно учита садржај странице у веб-прегледачу.

Промена у структури веб-сајтова или у њиховом *HTML* коду захтева корекцију програма за скидање цена и њихово прилагођавање новој поставци веб-сајта. Због тога настојимо да избегнемо коришћење прекомерног броја извора, јер би у супротном одржавање функционалности процеса постало изузетно захтеван посао, с могућношћу честих грешака.

У целом процесу прикупљања података поштујемо политику сајтова у погледу расположивости њихових података за употребу. Пре преузимања података, проверавамо да ли интернет продавница (веб-сајт) експлицитно забрањује скидање и

употребу његових података за друге сврхе осим за онлајн куповину. Такође, да не бисмо оптеретили сајтове, настојимо да податке не скидамо чешће него што је минимално потребно за оцену текуће инфлације.

Након преузимања сирових података с веб-адреса, аутоматски се спроводи иницијална обрада података како би се цене исказале у нумеричком формату, с тачком као децималним сепаратором, и без икаквих текстуалних карактера. На пример, ако је цена оригинално изражена као „1.099,99 РСД/кг”, потребно ју је изразити као „1099.99”. Поред тога, неопходно је да имена производа у бази буду јединствена, барем оних изабраних који репрезентују производе из ИПЦ-а.

Конечно, пре него што податке искористимо за даљу обраду и оцену ИПЦ-а, они морају бити конвертовани у одговарајући формат. Једна база података са веба (са једног или више сајтова) налази се у *data-frame* форми, где се у првој колони налазе имена производа/услуга, а у осталим њихове цене за одређени дан. У појединим случајевима, база садржи више колона за цене, од различитих малопродаваца, с различитих продајних места или из различитих градова.

Базе се чувају одвојено по месецима. У последњој оцени текуће инфлације пре закључења овог рада користили смо укупно 28 база с подацима са 130 различитих сајтова, с ценама за тридесетак хиљада производа. Ове бројке су променљиве с обзиром на то да се од једне до друге оцене извори могу додавати или одузимати.

Потребно је напоменути да се подаци које преузимамо могу налазити у различитим формама. Најчешће се налазе у форми која омогућава аутоматско преузимање са именима производа и услуга и њиховим ценама, у идеалном случају директно на веб-страници. Ипак, у појединим случајевима подаци се могу налазити у *PDF* фајлу, што скидање података чини посебно изазовним, а понекад се цена за коју смо заинтересовани налази у оквиру текста.

Дешава се да поједине податке није могуће преузети програмски, најчешће када се он налази у *PDF* фајлу, који се не може конвертовати у текст или табелу или када сâм веб-сајт ограничава могућност скидања података. У тим случајевима прибегавамо ручном скидању, тако што једноставно укуцамо податке у нашу базу података. То радимо само изузетно, у случајевима када се цена ретко мења (једном, двапут годишње) и када је извор податка апсолутно поуздан, као што је случај с појединим регулисаним ценама на сајтовима државних органа.

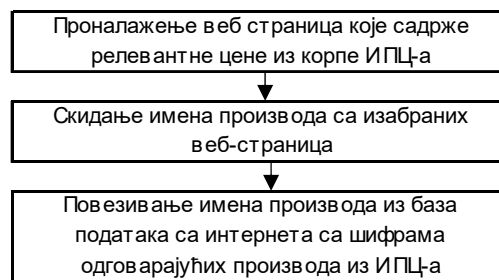
3.1. Избор репрезентативних производа

Када се нова база података са интернета укључује у систем, неопходно је одредити који производи из те базе репрезентују производе из ИПЦ-а. Технички гледано, ово повезивање врши се преко шифара производа које РЗС користи при својим обрачуницима. Када се подаци с једног веб-сајта први пут скину и трансформишу у одговарајућу базу података, неопходно је проћи кроз њу, производ по производ, и оним од њих за које закључимо да репрезентују производ из ИПЦ-а придружити његову шифру.

Већина производа из ИПЦ-а има више репрезентативних производа у базама података с веб-сајтова. Тако је, на пример, производ „сунцокретоно уље” из ИПЦ-а репрезентован у једној од база података уљима марки „Дијамант”, „Искон” и „Витал”, па њима придружујемо шифру за тај производ.

Уколико за неки производ из ИПЦ-а нисмо пронашли идентичан производ на интернету, као репрезентативни производ можемо изабрати неки који је најсличнији оном из ИПЦ-а. На пример, телевизор од 40 инча можемо репрезентовати телевизором од 43 инча, претпостављајући да се цене тако сличних производа мењају сличном динамиком.

Слика 2. Процес повезивања цена са интернета с производима из корпе ИПЦ-а



Извор: РЗС и аутор.

Поједине базе података садрже и више хиљада производа, па овај посао може бити временски изузетно захтеван. Ипак, он је углавном једнократан, тј. обавља се само када се база података први пут скине. Уколико се пак производи на веб-сајтовима који улазе у базу података често мењају, неопходно је с времена на време урадити ревизију шифара. Да бисмо минимизирали овај део посла, настојимо да бирамо сајтове који не мењају често производе.

Крајњи резултат овог дела посла је шифарник база, са именима производа из база у првој колони и придруженим шифрама РЗС-а у другој, који се у процесу оцене користи да производе са интернета повеже са одговарајућим производима из ИПЦ-а.

4. Процес оцене текуће инфлације

Оцена текуће инфлације почиње скидањем података са изабраних веб-страница.

Пратећи динамику РЗС-а, цене скидамо по следећем распореду:

- цене непрехрамбених производа – 10. у месецу,
- цене хране без воћа и поврћа – 14. у месецу,
- цене воћа и поврћа – крајем треће недеље у месецу,
- цене услуга – 17. у месецу,
- цене горива – сваке среде у месецу.

Ако неки од ових датума пада за викенд, цене скидамо у петак пре тог викенда.

На овај начин добијамо базе података које могу садржати податке с једног или више сајтова. У тренутку писања овог рада број база које користимо за оцену текуће инфлације износи 35, што укључује и базе из једног извора снимљене на више датума.

На пример, СТИПС снимамо двапут, једном за оцену цена хране без воћа и поврћа, а други пут за оцену цена воћа и поврћа.

Калкулација оцене инфлације састоји се из: 1) оцене раста цена појединачних производа за које постоје репрезентативни производи на интернету, 2) импутације недостајућих цена и 3) рачунања пондерисаног раста цена (снимљених и импутираних) на основу пондера ИПЦ-а који користи РЗС.

Раст цена појединачних производа за које постоје репрезентативни производи на интернету оцењује се у неколико корака:

- За сваки појединачни производ из сваке базе података рачуна се стопа раста цена у односу на претходни месец.
- Стопа раста цена производа из ИПЦ-а из једне базе рачуна се преко геометријског просека раста цена појединачних репрезентативних производа из те базе. На пример, раст цена сунцокретовог уља (из ИПЦ-а) рачуна се на основу раста цена уља „Дијамант”, „Искон” и „Витал” (из базе података). ИПЦ и базе се у овој фази повезују на основу шифарника, описаног у претходном делу.
- Уколико база садржи више продајних места (градови, пијаце, малопродајни ланци), онда се стопе по производима рачунају одвојено за продајна места (на горе описан начин), а затим се агрегатна стопа за производ за базу рачуна као пондерисани просек стопа по продајним местима. На пример, раст цена парадајза из базе СТИПС рачуна се као пондерисани просек раста цена парадајза по пијацама.
- Коначно, уколико неки производ из ИПЦ-а има репрезентативне производе у више база, финална оцена раста цена производа рачуна се као просек раста цена из више база (добијених у претходном кораку).

Овде треба додати да, када се неки производ снима више пута у току месеца, као што је случај с ценама горива, његова цена за дати месец рачуна се као просек цена снимљених у различитим периодима, а затим се на тим просецима примењује претходно описани поступак за рачунање стопе раста.

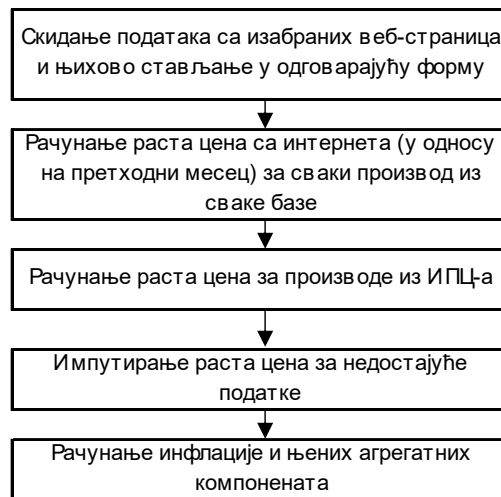
Цене производа за које, привремено или трајно, немамо одговарајуће репрезентативне производе на интернету импутирамо на одређене начине. Овде треба разликовати импутацију на основу методологије РЗС-а од оне коју ми примењујемо на производе за које немамо извор.

Најпре, у случају сезонског воћа и поврћа примењујемо метод импутације који РЗС спроводи у месецима у којима ових производа (сезонски) нема на тржишту у већим количинама. Цене ових производа импутирају се што на основу кретања цена остатка одговарајуће групе у том месецу, што на основу кретања цена самог производа из претходних неколико месеци. Примера ради, раст цена паприке у периоду од јануара до маја импутира се растом цена свежег поврћа у одговарајућем месецу, док се у првом месецу импутације, децембру, цена паприке рачуна као пондерисани просек њене цене из претходних месеци када се налазила на тржишту.

У овој фази, производи чије смо цене обухватили изворима са интернета и они које смо импутирала методологијом РЗС-а чине 90,5% ИПЦ-а. Преостали део цена апроксимирамо следећим методама:

- Импутација сродним производима – где претпостављамо да се цена производа за коју немамо извор мења као цена сличног производа за коју га имамо. Ову врсту импутације примењујемо на десетак производа из ИПЦ-а;
- Импутацију групом – примењујемо код хране без воћа поврћа, где за недостајуће цене претпостављамо да се крећу као остатак групе;
- Импутацију претходним растом – примењујемо код непрехрамбених производа, с обзиром на то да их карактерише перзистентније кретање од осталих група;
- Импутацију нултим растом – примењујемо код недостајућих регулисаних цена, с обзиром на то да се оне ретко коригују (углавном једном годишње).

Слика 3. Процес оцене текуће инфлације на основу цена са веба



Извор: РЗС и аутор.

Импутација служи томе да се при оценама текуће инфлације задрже непромењена учешћа различитих група у ИПЦ-у. У супротном, компоненте инфлације које имају већу покривеност ценама са веба, као што је воће и поврће, давале би већи допринос инфлацији, са свим последицама на поузданост оцене.

У последњем кораку рачунамо инфлацију и остале агрегатне индексе, као пондерисане просеке појединачних (снимљених и импутираних) стопа раста, примењујући пондере које РЗС користи у својим обрачунима.

Да бисмо дали финалну оцену текуће инфлације на бази цена свих производа који улазе у обрачун ИПЦ-а, морали бисмо да чекамо крај месеца (гориво се снима до задње среде у месецу). Из практичних разлога, међутим, по правилу је неопходно оценити инфлацију раније у току месеца. Тиме се у одређеној мери губе информације релевантне за оцену, али то не мора знатније да утиче на њену поузданост.

Тако, ако се прелиминарна оцена ради око 20. у месецу, практично једини недостајући податак јесте цена горива за једну или две последње среде у месецу. У том случају претпостављамо да ће се њихова цена задржати на нивоу од последње среде за

коју смо снимили цену. Како се месечне цене горива рачунају као просек недељних података у току месеца, евентуално одступање у задњој недељи не може знатније да промени финалну оцену.

Ако се прелиминарна оцена ради средином месеца, период снимања је добрим делом покривен, осим делимично горива (две среде), једне недеље код цена воћа и поврћа и неколико дана код услуга. Тиме и даље имамо релативно поуздану оцену већине категорија цена, с тим да су у току пољопривредне сезоне могућа знатнија одступања код воћа и поврћа у току једне недеље, што се узима у обзир при процени ризика остварења ових оцена.

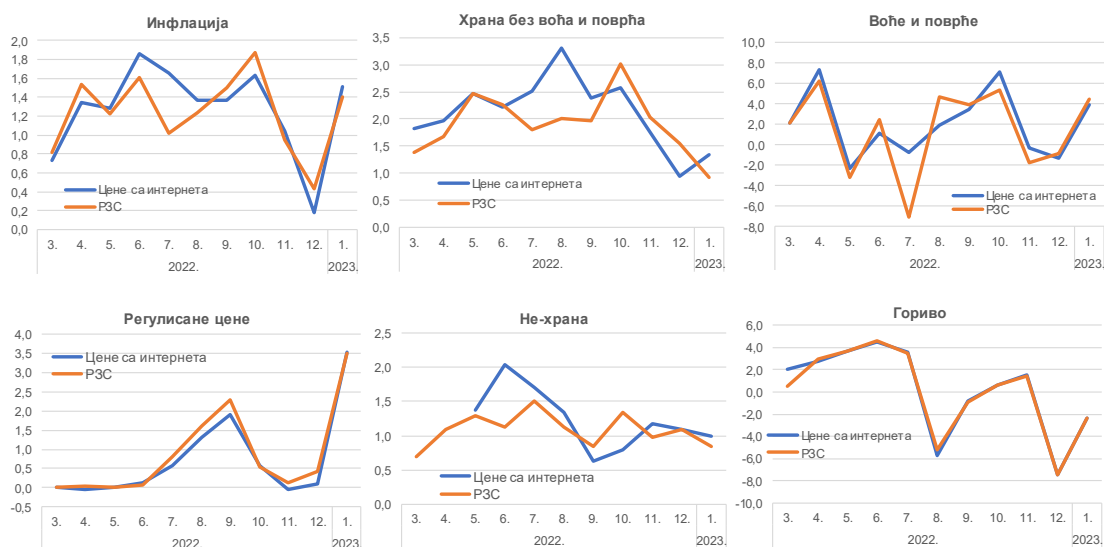
На крају додајмо и да процес оцене текуће инфлације укључује логичку контролу добијених стопа раста. Превелика или прениска стопа раста неког производа може сугерисати грешку уноса цене на веб-страници (погрешан број или децимални сепаратор) и у том случају се податак коригује или елиминише из обрачуна.

5. Остварења оцена текуће инфлације

У Народној банци Србије описани свеобухватни метод оцене текуће инфлације примењујемо од марта 2022. године (парцијалне пилот-оцене радили смо и раније, у току 2021), с тим да смо у прва два месеца посматраног периода за непрехрамбене цене импутирали експертску краткорочну пројекцију уместо оцена са интернета, с обзиром на то да тада ову групу још увек нисмо имали довољно покривену.

Мада период од непуне године дана није довољан да се донесе коначан суд о резултатима примене овог метода, можемо рећи да су досадашњи резултати охрабрујући. Од марта 2022. до јануара 2023. године наше оцене инфлације у просеку су биле веома блиске званичним подацима РЗС-а, што показује да не постоји систематска пристрасност оцена.

Графикон 1. Оцењене и остварене стопе инфлације и њених компонента



Извор: РЗС и аутор.

Просечно апсолутно одступање оцењене инфлације од званичне у посматраном периоду износило је 0,20 процентних поена, а медијално 0,13 процентних поена. Од 11 месеци колико смо користили описани метод, апсолутно одступање оцене од остварења износило је 0,1 процентни поен (заокружено на једну децималу) у шест случајева, 0,2 процентна поена у три случаја, 0,3 процентна поена у једном случају, док је једино веће одступање забележено у јулу 2022, када је инфлација на бази цена са веба била за 0,6 процентних поена виша од званичног податка РЗС-а услед снажног пада цена воћа и поврћа који је регистровао РЗС.

При томе, треба имати у виду да су ове оцене рађене у инфлаторно најтурбулентнијем периоду у последњој деценији, те да је у посматраном периоду месечна инфлација у просеку износила 1,2%, што генерално чини оцене текуће инфлације посебно изазовним, без обзира на метод.

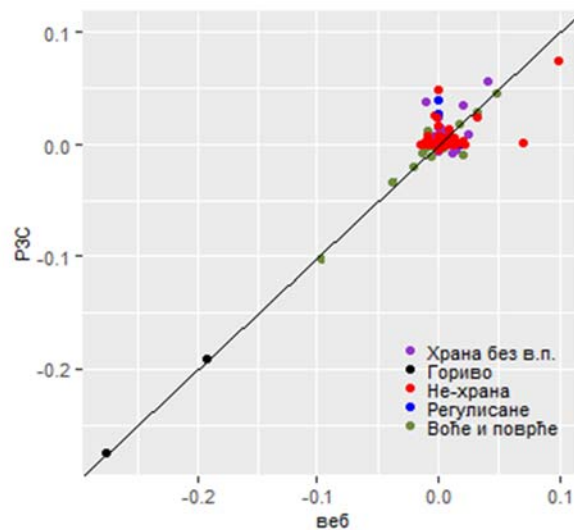
Табела 1. Статистике одступања веб-инфлације од података РЗС-а

	Просечно одступање	Просечно апсолутно одступање (у п.п.)	Просечно апсолутно одступање (у ст. Дев.)	Медијално апсолутно одступање	Коефицијент корелације појединачних производа
Инфлација	0,03	0,20	0,37	0,13	0,77
Гориво	0,08	0,23	0,05	0,05	1,00
Храна без в. и п.	0,20	0,45	0,56	0,43	0,56
Воће и поврће	0,52	1,57	0,33	1,04	0,82
Регулисане цене	-0,13	0,15	0,27	0,09	0,65
Не-храна	-0,07	0,39	0,87	0,21	0,61

Извор: РЗС и аутор.

Посматрано по групама, може се рећи да оцене углавном нису пристрасне, осим у случају воћа и поврћа, где просечно одступање износи 0,5 процентних поена, с тим да је то углавном вођено поменутих великим одступањем у јулу 2022. Из истог разлога, највеће просечно апсолутно одступање забележено је и код цена воћа и поврћа. То је донекле очекивано, с обзиром на то да ову групу карактеришу највеће осцилације цена, чак и на недељном нивоу, па релативно мало одступање од периода снимања РЗС-а може водити видљивој разлици између цена са интернета и званичних података.

Графикон 2. Оцењени и остварени доприноси инфлацији појединачних производа из корпе ИПЦ-а за децембар 2022. (у п.п.)



Извор: РЗС и аутор.

Док је на агрегатном нивоу одступање од података РЗС-а релативно мало, на нивоу појединачних производа могуће су значајне разлике. На Графикону 2. приказано је поређење доприноса појединачних производа на бази интернета (x-оса) са оствареним растом према РЗС-у (y-оса), на примеру децембра 2022. године, где одступање од линије од 45 степени представља грешку оцене доприноса појединачног производа. Одговарајући графикони за све месеце у посматраном периоду приказани су у Додатку.

Корелација између оцењених и остварених стопа (без импутираног дела) по месецима у посматраном периоду кретала се од 0,5 до 0,9. Ако се не рачуна гориво, које садржи само два производа, посматрано по групама, највећу корелисаност показују цене воћа и поврћа (око 0,8), док је код осталих група приближно исто, на око 0,6. То је, на први поглед, у супротности с налазом да управо ова група показује највеће одступање агрегатне оцене од остварења. Објашњење лежи у чињеници да ова група показује највеће варијације унутар групе (стопе некад иду од -50% до $+150\%$). Мада извори које користимо успевају добро да обухвате релативне промене између цена у групи (што води високој корелисаности), апсолутне разлике (оцене од остварења) ипак често буду значајне.

Када се при поређењу одступања узму у обзир и варијације група, добија се конзистентнија слика (с корелацијама). Тако, када се посматра одступање у односу на стандардну девијацију, група с највећим одступањем постаје она која има најстабилније кретање месечних стопа, не-храна, док је код воћа и поврћа и регулисаних цена овај показатељ знатно нижи.

За неке сложеније методе мерења перформанси описаног метода, као што је поређење с ванузорачким грешкама модела инфлације, неопходно је сачекати да прође још времена да би период за поређење био довољно дуг.

6. Закључак

Циљ овог рада био је приказ методологије оцене текуће инфлације на основу цена са интернета, која се у Народној банци Србије користи од марта 2022. године.

Наша методологија се, за разлику од већине оних које су описане у радовима на ову тему, не базира само на ценама за онлајн куповину, већ на свим релевантним информацијама о ценама, укључујући оне које су на вебу приказане само информативно. У ту сврху, поред веб-сајтова самих малопродаваца, користимо и оне који пореде цене различитих малопродаваца, као и сајтове државних органа и јавних предузећа са информацијама о ценама њихових услуга.

Још једна карактеристика нашег метода јесте свеобухватни приступ оцени. Док се друге методе оцене инфлацијом углавном баве парцијално (оцењујући поједине компоненте ИПЦ-а, пре свега цене хране), ми смо се определили да изворима са интернета покријемо што више производа из ИПЦ-а (око 90% у тренутку писања овог рада), што захтева велики број извора са интернета (око 130 веб-сајтова). То радимо због тога што нам је за потребе пројекција инфлације на дужи рок потребна оцена више компонената инфлације, као и због тога што би укључивањем већег броја производа у узорак требало да добијемо поузданију меру централне тенденције инфлације.

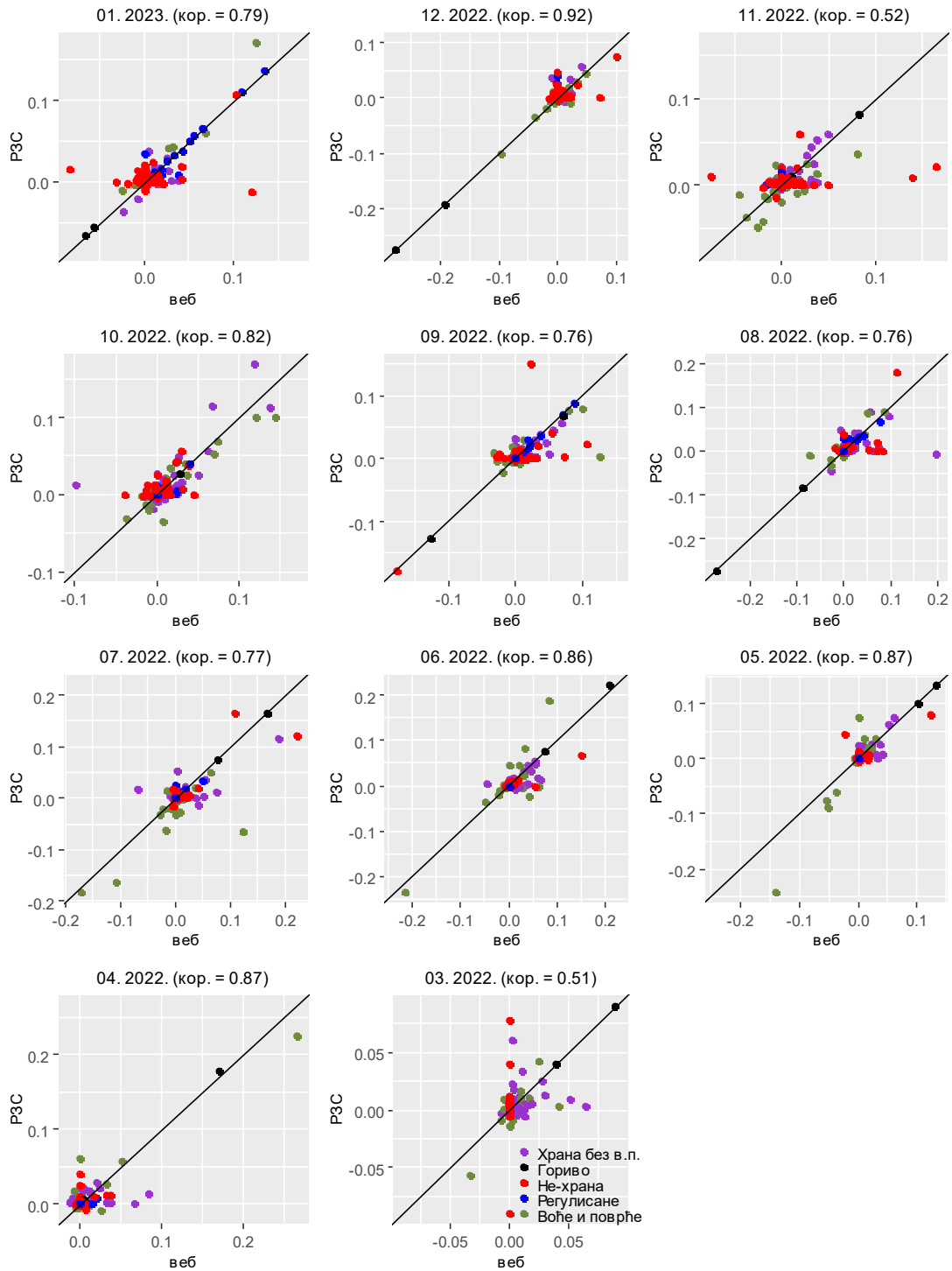
Недостајући део цена покривамо разним методима импутације – растом цена сродних производа, нулом, растом цена групе или претходним растом цена – у зависности од карактеристика кретања цена конкретних категорија производа. Тиме избегавамо могућност да компоненте инфлације које имају већу покривеност ценама са интернета дају већи допринос инфлацији, са свим последицама на поузданост оцене.

Досадашњи резултати примене овог метода су охрабрујући. Оцењена инфлација у просеку се у посматраном периоду налазила на нивоу реализоване (оцене нису пристрасне), док је просечно апсолутно одступање износило 0,20 процентних поена, а медијално 0,13 процентних поена, што није значајно с обзиром на то да је просечна месечна инфлација у том периоду износила 1,2%. На нивоу појединачних производа приметна су значајна одступања, али је корелисаност између остварених и оцењених доприноса инфлацији по месецима ипак релативно висока (око 0,8 у просеку).

Након непуне године дана примене овог метода још увек је рано да се донесе коначан суд о његовим резултатима. Неко будуће истраживање на ову тему, када период примене овог метода буде довољно дуг, засниваће се на поређењу одступања оцена на бази интернета, с ванузорачким грешкама прогнозе на основу неког од модела инфлације. У међувремену ћемо, као и до сада, у процес укључивати нове изворе података са веба и искључивати постојеће у зависности од њихове поузданости, ажурности и доступности, а све у настојању да тиме побољшамо прецизност оцена.

Додатак

Графикон А1. Оцењени (веб) и остварени (РЗС) доприноси инфлацији појединачних производа за период март 2022 – јануар 2023. (корелације у заградама)



Извор: РЗС и аутор.

Литература

- Aghajanyan, G., Baghdasaryan, T., & Lazyan, G. (2017). The use of Big Data in Central Bank of Armenia.
- Bertolotto, M., Cavallo, A., & Rigobon, R. (2014). Using Online Prices to Anticipate Official CPI Inflation. UTokyo Price Project Working Paper Series.
- Breton, R., Clews, G., Metcalfe, L., Milliken, N., Payne, C., Winton, J., & Woods, A. (2015). Research indices using web scraped data. Office for National Statistics (ONS).
- Carvalho, I. (2020). Nowcasting CPI using online retail prices: Forecasting combination of dynamic factor models.
- Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2).
- Horrigan, M. W. (2013). Big data: A perspective from the BLS. <http://magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/>
- Jaworski, K. (2021). Measuring food inflation during the COVID-19 pandemic in real time using online data: A case study of Poland. *British Food Journal*.
- Kapetanios, G., & Papailias, F. (2018). Big Data & Macroeconomic Nowcasting Methodological Review.
- Macias, P., & Stelmasiak, D. (2019). Food inflation nowcasting with web scraped data. 302.
- Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J., & Winton, J. (2018). Tracking and modelling prices using web-scraped price microdata: Towards automated daily consumer price index forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Soybilgen, B. ş, Kaya, H., & Yazgan, M. E. (2021). Nowcasting Turkish Food Inflation Using Daily Online Prices.
- Tissot, B. (2019). The Role of Big Data and Surveys in Measuring and Predicting Inflation. https://data.stat.gov.rs/Metadata/03_Cene/Html/030103_ESMS_G0_2018_2.html