
СТРУЧНИ РАДОВИ

ПРОЦЕНА ИНФЛАТОРНИХ ПРИТИСАКА ПУТЕМ АНАЛИЗЕ НОВИНСКИХ ТЕКСТОВА

Мирко Ђукић

© Народна банка Србије, септембар 2022.

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

Процена инфлаторних притисака путем анализе новинских текстова

Мирко Ђукић

Апстракт: У овом раду представљамо резултате речничке текстуалне анализе новинских чланака из економских рубрика четири српске дневне новине, како бисмо оценили да ли ти чланци садрже корисне информације за процену инфлаторних притисака. Укупно смо анализирали 117.113 економских чланака за период 2007–2022. године, пребројавајући изразе у вези с растом и падом цена, односно текстове који садрже те изразе. Мере новинског инфлационог сентимента, које смо добили на овај начин, показале су се високо корелисаним са инфлацијом, с тим да је то вођено периодима с високим осцилацијама инфлације. Током периода са стабилном инфлацијом корелисаност је знатно нижа. Узрочна веза иде од сентимента ка инфлацији, а модели инфлације с новинским инфлационим сентиментом показују боље резултате у прогнозирању изван узорка него реперни ауторегресиони модели инфлације. На основу анализе закључује се да се новински инфлациони сентимент може користити као показатељ инфлаторних притисака, пре свега у периодима високих осцилација инфлације.

Кључне речи: прогнозирање инфлације, текстуална анализа

[JEL Code]: C13, C55, E31, E37, E52

Нетехнички резиме

Новине у Србији често пишу о ценама и инфлацији. У овом раду користимо ове текстуалне податке ради мерења новинског инфлационог сентимента, а све како бисмо оценили да ли се он може користити као показатељ инфлаторних притисака.

Анализирали смо текстове из економских рубрика четири српске дневне новине за период од почетка 2007. до јуна 2022. године. Узорак се састоји од 117.113 чланака, или 650 месечно у просеку.

Текстове смо квантификовали тзв. речничким методом да бисмо добили мере инфлационог сентимента. Најпре смо дефинисали листе израза, једну која се односи на изразе у вези с растом цена и другу која се односи на изразе у вези с падом цена. Затим смо пребројали те изразе по месецима, а потом те бројке ставили у однос са укупним бројем речи у одговарајућим месецима. Користили смо и тзв. Булову претрагу, у којој смо по месецима пребројавали текстове који садрже бар један од израза с поменутих листа (у односу на укупан број текстова). Док број израза мери интензитет, број текстова мери учесталост помињања промена цена.

Резултати анализе показују да су мере инфлационог сентимента на целом узорку високо корелисане са самом инфлацијом. То је вођено пре свега заједничким кретањем серија у епизодама раста инфлације на двоцифрене нивое, коме су претходили порасти мера новинског инфлационог сентимента. Тако је у сва четири таква циклуса број израза у вези с растом инфлације на посматраном узорку растао изнад 2,5% укупног броја речи, док би се током периода ниске инфлације он углавном кретао испод 1%. Међутим, током периода ниске и стабилне инфлације (средина 2013. до средине 2021) корелација између инфлације и инфлационог сентимента знатно је нижа.

Узрочност између инфлације и новинског инфлационог сентимента, у принципу, може ићи у оба смера, с обзиром на то да новине могу писати и о претходној инфлацији, и о променама цена које ће тек бити укључене у инфлацију. Грејнцеров тест узрочности показује да у нашем случају, у свим комбинацијама сентимента и инфлације, узрочност иде од првог ка последњем, с тим да у неким случајевима узрочност иде у оба правца.

Поред тога, утврдили смо да модели с новинским инфлационим сентиментом показују боље перформансе него прост ауторегресиони модел инфлације када је у питању прогнозирање изван узорка.

Узимајући све у обзир, закључили смо да се новински инфлациони сентимент може користити као показатељ инфлаторних притисака, пре свега у периодима високих осцилација инфлације, док у периодима стабилне инфлације овај показатељ треба узети с резервом

Садржај:

1. Увод	46
2. База новинских текстова.....	49
3. Анализа текста – мерење новинског инфлационог сентимента.....	50
4. Веза између новинског инфлационог сентимента и инфлације	54
5. Закључак	60
Додатак	62
Литература.....	66

1. Увод

У тренутку писања овог рада, глобална инфлација достиже нивое какви нису виђени деценијама, изненађујући креаторе монетарне политике широм света из месеца у месец. Да ли би централне банке спремније дочекале шокове да су већи значај придавале писању новина? Према налазима представљеним у овом раду, у случају Србије, новински текстови заиста садрже информације које се могу искористити за процену инфлаторних притисака, нарочито у нестабилним временима.

Велики број текстова расположивих на интернету, попут новинских чланака, изјава, блогова, објава на друштвеним медијима итд., разматрају економска кретања у некој форми. Економски истраживачи све више су заинтересовани да ово обиље текстуалних података употребе за разне врсте анализа, као што је прогнозирање или креирање мера појединих економских концепата који се иначе не мере на адекватан начин.

Текстуална анализа углавном укључује трансформацију текстова као квалитативних података у нумеричку меру употребом неког метода – од једноставног бројања речи до софистицираних модела машинског учења, као и оцену њихове везе са економским варијаблама. Пошто су текстови високофреквентни подаци, корелација са економским варијаблама може се користити за процену економских трендова пре него што је податак за ту варијаблу објављен.

Најранији пример употребе текстуалне анализе потиче из 1933. године, када је *Cowles* класификовао едиторијале „Волстрит журнала” као оптимистичне, песимистичне или недефинисане, да би закључио да они не садрже корисне информације за прогнозу цена акција. С друге стране, *Tetlock* у свом познатом раду из 2007. користи Харвардски психолошки речник да би чланцима „Волстрит журнала” дао оцене сентимента, за које је установио да добро предвиђају кретања на финансијским тржиштима за следећи дан. Поједини истраживачи (*Lucca and Trebbi* (2009), *Born et al.* (2014), *Hansen et al.* (2018)) на разне начине анализирали су комуникацију централних банака и установили да њен сентимент утиче на финансијска тржишта. Још један познат пример текстуалне анализе дали су *Baker et al.* (2016), који су бројали чланке који садрже барем по једну реч из три категорије („економија”, „политика”, „неизвесност”) и установили да таква мера економске неизвесности добро предвиђа инвестиције, запосленост и производњу.

Ближе нашој теми су бројни радови који истражују везу између текстова и инфлационих очекивања. *Carroll* (2003) је употребом простог епидемиолошког модела утврдио да се током периода у коме је тема инфлације била интензивно покривена у вестима очекивања становништва брже прилагођавају очекивањима стручњака него током периода слабе покривености. Поједини радови баве се асиметријом у овом сегменту. *Lamla et al.* (2012) установили су да садржај вести понекад изазива пристрасност пренаглашавањем негативних вести, што резултира претераном реакцијом инфлационих очекивања, док *Drager* (2015) налази да само медијски извештаји с негативним вестима утичу на инфлациона очекивања. Према *Larsen et al.* (2021), медијска покривеност појединих тема, од којих су неке на први поглед неповезане са инфлацијом, може у знатној мери предвидети инфлациона очекивања

потрошача. Поједини истраживачи, као *Angelico et. al.* (2021), анализирају објаве с друштвених мрежа. Они су креирали меру инфлационих очекивања на основу Твитера (комбинујући *Latent Dirichlet Allocation (LDA)* с речничким приступом) и установили њену високу корелисаност са уобичајеним мерама инфлационих очекивања.

Изненађујуће мали број истраживача покушао је да установи директну везу између текстова и инфлације, што је тема нашег рада. Један од примера је *Rambaccussing et. al.* (2020), који нису нашли доказе да економске теме у медијима у Уједињеном Краљевству побољшавају краткорочну прогнозу инфлације (супротно прогнозама незапослености и производње). Насупрот томе, *Kalamara et. al.* (2021), комбинујући бројање израза с надгледаним машинским учењем, налазе да новински чланци побољшавају прогнозу инфлације у Уједињеном Краљевству, уз прогнозе производње и незапослености.

У нашем раду усредсредили смо се на оцену везе између новинских текстова и инфлације у Србији. При томе, не анализирамо везу са инфлационим очекивањима, зато што се у Србији она процењују на основу анкета, чији су резултати најчешће под утицајем текуће инфлације, а не обратно, о чему ће бити речи у раду. Поред тога, сматрамо да је из практичних разлога подршке монетарној политици корисније имати модел који директно прогнозира инфлацију, уместо индиректно, преко очекивања.

Веза између новинских текстова и инфлације проистиче из једноставне чињенице да у периодима већих и учесталијих промена цена новине више пишу о томе, тј. што је битно за нашу анализу, више користе речи које се односе на промене цена. То може бити помињање промена цена које су се већ десиле, али и оних које тек треба да се десе. Цене су честа тема у српским новинама. У нашем узорку, 22% чланака из економских рубрика помиње инфлацију или промене цена барем једном, а као што ћемо видети касније, учесталост писања о овој теми расте у периодима високе инфлације.

Ми смо анализирали чланке из четири српске дневне новине за период од почетка 2007. до јуна 2022. године, употребом речничког приступа и Булове претраге. Конкретно, најпре смо дефинисали листе израза, једну која се односи на изразе у вези с растом цена, а другу у вези с падом цена. Затим смо пребројали те изразе по месецима и ставили те бројке у однос са укупним бројем речи у одговарајућим месецима. Булова претрага се у нашем случају састоји из бројања чланака који садрже барем један израз с поменутих листа (према укупном броју чланака у одговарајућем месецу). Док се бројањем чланака мери колико новине често пишу о ценама, бројање израза мери интензитет писања о истој теми.

На тај начин смо креирали месечне серије разних мера новинског инфлационог сентимента: изрази о расту цена, разлика израза о расту и паду цена, чланци о расту цена, разлика чланака о расту и паду цена, као и филтриране варијанте прве две серије.

То што речнички приступ и Булова претрага узимају унапред дефинисану листу израза за претрагу чини их једноставним за употребу, али се то понекад узима и као њихова мана, јер у потпуности игноришу друге речи¹ (за разлику од сложенијих метода

¹ *Bholat et. al.* (2015).

машинског учења). Међутим, сматрамо да се изрази који се односе на промене цена у оба смера лако могу препознати, нарочито у српском језику, што чини речнички приступ погодним за ову врсту анализе. Поред тога, речнички приступ је погоднији када хоћемо да измеримо интензитет сентимента, а не само да тематски класификујемо текстове, за шта су погоднији методи машинског учења. Тако, на пример, *Angelico et. al.* (2021) најпре користе *LDA* да смање шумове, тј. да издвоје твитове о инфлацији, а потом користе речнички приступ да изброје изразе у вези с растом, односно падом цена. У нашем случају, шумове смо смањили тако што смо у анализу укључили само економске текстове, па није било потребе за машинским учењем у ту сврху.

У последњој фази анализирали смо везу између шест различитих мера новинског инфлационог сентимента и инфлације (међугодишња и месечна). Ове две класе серија показале су се високо корелисаном на целом узорку (2007–2022), али не и на периоду стабилне инфлације (средина 2013 – средина 2021). Висока корелација на целом узорку одређена је периодима снажних пораста инфлације, који су били праћени или су им претходили порасту инфлационог сентимента у новинама.

Према Грејнцеровом тесту узрочности, у свим комбинацијама мера сентимента и инфлације узрочност иде од првог ка последњем, с тим да у неким случајевима иде у оба смера. Ово је било важно испитати, с обзиром на то да новине некад пишу о прошлој инфлацији, а некад о текућим и будућим променама цена које ће тек бити укључене у инфлацију.

Када је у питању прогнозирање изван узорка, већина модела с новинским инфлационим сентиментом има боље перформансе од реперног ауторегресионог модела инфлације. Најмању грешку прогнозе у веома кратком року показали су модел с месечном инфлацијом и разликом броја изрза о поскупљењима и појефтињењима, за текући месец, односно модел с филтрираном верзијом исте серије, за један месец унапред. Поред тога, укључивање сентимента у постојећи модел за инфлацију даје статистички значајан коефицијент уз ту варијаблу, што показује да сентимент пружа додатне информације чак и када се узму у обзир релевантни фактори инфлације.

Треба нагласити да број изрза у највећем броју случајева показује боље резултате и у корелисаности са инфлацијом, и у њеном прогнозирању, што сугерише на то да је интензитет помињања цена релевантнији од његове учесталости.

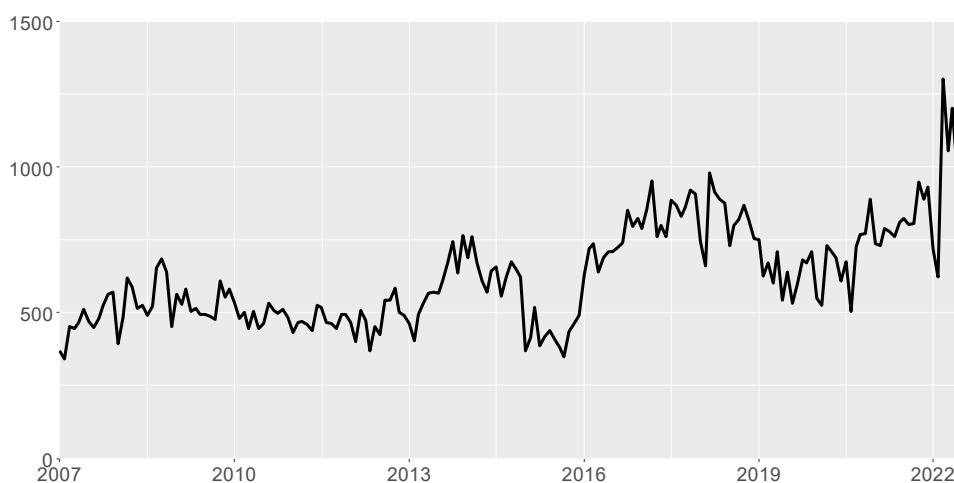
Узимајући све у обзир, закључили смо да се новински инфлациони сентимент може користити као показатељ инфлаторних притисака, али пре свега у периодима високих осцилација инфлације, док у периодима стабилне инфлације овај показатељ треба узети с резервом.

Рад је структуриран на следећи начин. Најпре, представљамо базу новинских текстова која је анализирана у раду. Након тога, објашњавамо методе текстуалне анализе које смо користили да креирамо серије новинског инфлационог сентимента. На крају, представљамо резултате оцене веза између мера сентимента и инфлације.

2. База новинских текстова

Први, најзахтевнији и временски најдужи корак у нашој анализи био је формирање базе економских текстова за анализу. У базу смо укључили четири српске дневне новине с довољно дугом историјом за ову врсту анализе (за детаље видети Додатак). Чланци су преузети употребом пакета *Rvest* у оквиру софтвера *R*, који смо користили и у каснијој фази за текстуалну анализу. Узорак почиње од јануара 2007. године, са изузетком једне новине, која има расположиве чланке од марта 2008. Чланке смо преузимали с веб-страница самих новина, или посредно с веб-страница које прикупљају и објављују вести из разних извора. Коначна база коју смо анализирали састојала се од 117.113 чланака (650 месечно у просеку) из економских рубрика поменутих новина.

Графикон 1. Број економских чланака у узорку, по месецима



Анализу смо ограничили на чланке из економских рубрика како су их саме новине класификовале из неколико разлога. Анализа на свим чланцима могла би створити велике шумове у серији, с обзиром на то да би поједини велики догађаји, као што је пандемија вируса корона, могли засенити друге теме, укључујући оне о инфлацији. Поред тога, поједине речи, као што су, на пример, криза или депресија, могу имати економско и некономско значење. У том смислу, ограничавање анализе само на економске текстове смањило би могуће грешке у том погледу, мада са изразима које користимо у нашој конкретној анализи то не би био толики проблем.

У целом узорку има преко 600.000 различитих речи, од којих смо заинтересовани за веома мали део. Такозвани „облак речи” (*word cloud*), на Слици 1, често је коришћен начин да се графички представе најчешће речи у тексту (већа фреквентност, већи фонт).

Табела 1. Листа израза у вези с поскупљењем и појефтињењем за претрагу

Изрази о поскупљењима	Изрази о појефтињењима
поскуп	појефт
скупљ	јефтени
раст/ раста/ расту /растом цен	пад /пада /паду цен
повећањ/ повећање/ повећању цен	снижење /снижења /снижењу цен
виша/више/вишој цен	нижа /ниже /нижој /нижу цен
висока/високе/високим/високој цен	ниска /ниске /ниским /ниској цен
скок/ скока/ скоку/ скокови цен	дефлаци
дигао/дигли/дичу/диче цен	пад инфлаци
инфлаци	

Напоменимо да претрага коју смо спровели не прави разлику између малих и великих слова, тако да у претходној фази није било неопходно велика слова трансформисати у мала, што је чест поступак у оваквој врсти анализе.

Као пример, у кратком чланку из 2008. године, на основу унапред дефинисаних израза из Табеле 1, избројано је десет израза који се односе на поскупљења (жуто) и два која се односе на појефтињења (плаво):

Литар бензина 101,6 динара

Уколико у наредна два дана не дође до **пада цене** сирове нафте у свету, гориво у Србији би већ од петка могло да **поскупи** у просеку за око четири динара за литар, сазнаје „Блиц”. Према овим проценама, литар безоловног бензина продавао би се за 101,6 динара.

Прогнозе да ће цена нафте расти уколико пређе магичну цифру од 100 долара за барел све више се остварују. Наиме, ових дана цена сирове нафте котирала се нешто мање од историјског максимума од 103,05 долара за барел (159 литара). Тако **висока цена** главне енергетске сировине највише је последица слабости долара, али и очекивања да чланице Организације земаља извозница нафте (ОПЕК) неће повећати ниво дневне производње. Сходно томе, и у Србији се већ крајем недеље може очекивати ново **повећање цене** деривата. Како „Блиц” сазнаје, гориво би могло да **поскупи** око четири динара, али постоји опасност да, уколико се настави **раст цене** нафте, ова сума буде и већа. Последње **поскупљење** десило се 20. фебруара, а Министарству енергетике су нам потврдили да ће се пресек цена радити у четвртак, тако да би возаче већ у петак ујутру могле да сачекају нови ценовници.

„Након овога **поскупљења**, мислим да се у наредних месец дана неће дешавати већи **скокови цене** горива код нас. С обзиром на то да је цена сирове нафте близу свог историјског рекорда, реално је очекивати да би у неком наредном периоду могла да падне. Међутим, извесно је да ће нафта током године **поскупљивати** – каже Небојша Атанацковић из „Нафте” а.д.

На **нижу цену** горива утицало би смањење акцизе на деривате, које су повећане приликом последњег **поскупљења**. Ове акцизе усклађује Министарство финансија – али, према ранијим изјава, тренутно се само посматрају дешавања на тржишту Србије и нема најављено њиховог смањења.

Свесни смо да овакво пребројавање није савршен одраз инфлационог сентимента. У нашем примеру укључена су два израза која се односе на пад цена, иако су у потенцијалу („Уколико у наредна два дана не дође до **пада цене** сирове нафте у свету...”), док у исто време није укључен израз који се односи на раст цена у трећој реченици („Прогнозе да **ће цена нафте расти**....”). Упркос томе, може се рећи да пребројани изрази у доброј мери одражавају инфлациони сентимент овог чланка, те да, с обзиром на велики број текстова у узорку (650 месечно у просеку), овакви повремени пропусти не могу искривити слику о тренду овог показатеља.

Још један приступ који смо користили јесте Булова претрага, која се састоји из пребројавања чланака који задовољавају одређени логички услов, у нашем случају који садрже било који израз из Табеле 1. Овај приступ је мање подложен екстремним вредностима, али му је мана што не узима у обзир интензитет употребе израза. Примећујемо да би применом овог метода поједини чланци могли истовремено бити категорисани тако и да се односе на поскупљења и на појефтињења, као што је случај с чланком у нашем примеру.

На крају, с обзиром на то да су овако добијене серије прилично волатилне, користили смо и њихове филтриране верзије, употребом Ходрик–Прескотоваг (HP) филтера с најмањим коефицијентом углачаности ($\lambda = 1$), са идејом да елиминишемо најволатилнији део серија, задржавајући истовремено краткорочне флукуације у тренду.

Укупно смо дефинисали **шест мера новинског инфлационог сентимента**:

- **Изрази о поскупљењима (PRT),²** као однос броја израза о поскупљењима и укупног броја речи у одговарајућем месецу (m), у промилима:

$$PRT_m = \frac{PRTC_m}{w_m} \cdot 1000$$

- **Изрази о променама цена (PCT),** као разлика између броја израза о поскупљењима и појефтињењима (PFTC) у односу на укупан број речи у одговарајућем месецу (m), у промилима:

$$PCT_m = \frac{PRTC_m - PFTC_m}{w_m} \cdot 1000$$

- **Текстови о поскупљењима (PRA),** као однос броја чланака о поскупљењима (PRAC) и укупног броја чланака (a) у одговарајућем месецу (m), у процентима:

$$PRA_m = \frac{PRAC_m}{a_m} \cdot 100$$

² У раду користимо скраћенице на основу енглеских назива: *Price-rise terms (count) – PRT(C)*, *price-change articles – PCA*, *newspaper inflationary sentiment – NIS*.

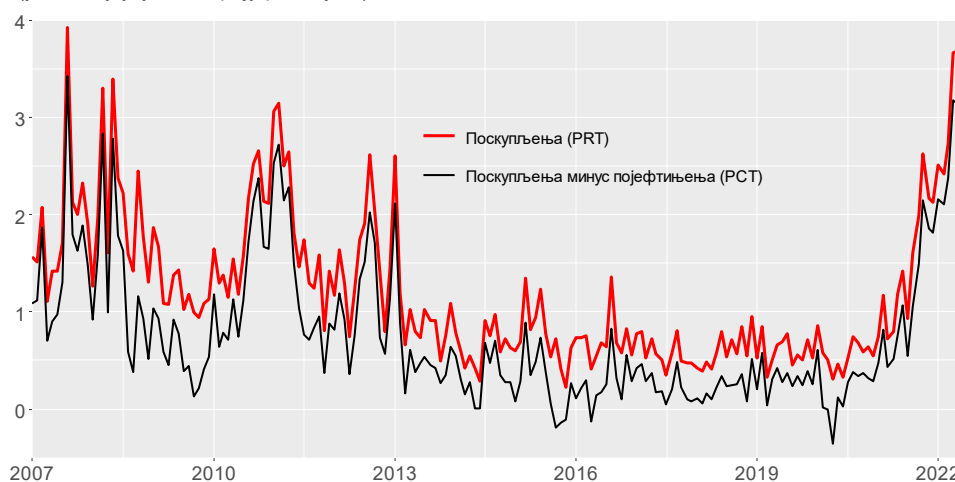
- **Текстови о променама цена (PCA)**, као разлика између броја текстова о поскупљењима ($PRAC$) и појефтињењима ($PFAC$) у односу на укупан број текстова у одговарајућем месецу (m), у процентима:

$$PCA_m = \frac{PRAC_m - PFAC_m}{a_m} \cdot 100$$

- **Изрази о поскупљењима – филтрирано ($PRT-HP$)**, као HP филтер од PRT -а, с параметром $lambda = 1$
- **Изрази о променама цена – филтрирано ($PCT-HP$)**, као HP филтер од PCT -а, с параметром $lambda = 1$.

Графикон 2 показује инфлациони сентимент мерен бројем одговарајућих израза у укупном броју речи, PRT и PCT . Уколико број израза о појефтињењима премашује број израза о поскупљењима, PCT може узети негативне вредности.

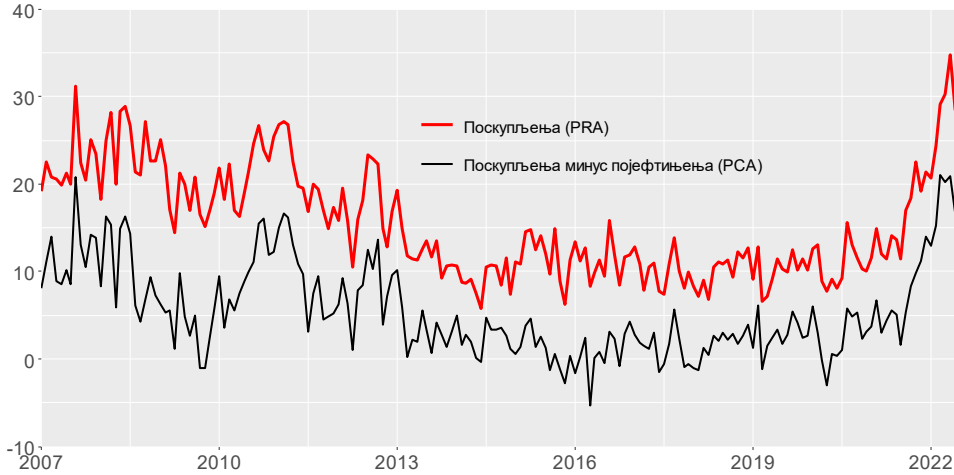
Графикон 2. **Мере новинског инфлационог сентимента на основу бројања израза** (учешће у укупном броју речи, у %)



Мада су ове серије веома волатилне, на анализираном периоду могу се јасно уочити четири велика циклуса у инфлационом сентименту. Тако обе серије достижу максимуме 2008, 2010, 2012. и 2022. године, на преко 2,5% у случају PRT -а, односно 2% у случају PCT -а. У периоду 2013–2021. серије су биле релативно стабилне, углавном у интервалу од 0% до 1%.

Када је у питању број чланака, такође се могу запазити четири велика циклуса, мада мање изражена него у случају броја израза. Учешће чланака с барем једним поменом поскупљења (PRA) варира између 10% у стабилним временима (2013–2021), до преко 30% у одређеним периодима. Разлика између броја текстова о поскупљењима и појефтињењима (PCA) варира између 0–5% у стабилним временима, па до око 20%.

Графикон 3. **Мере новинског инфлационог сентимента на основу бројања текстова**
(учешће у укупном броју текстова, у %)

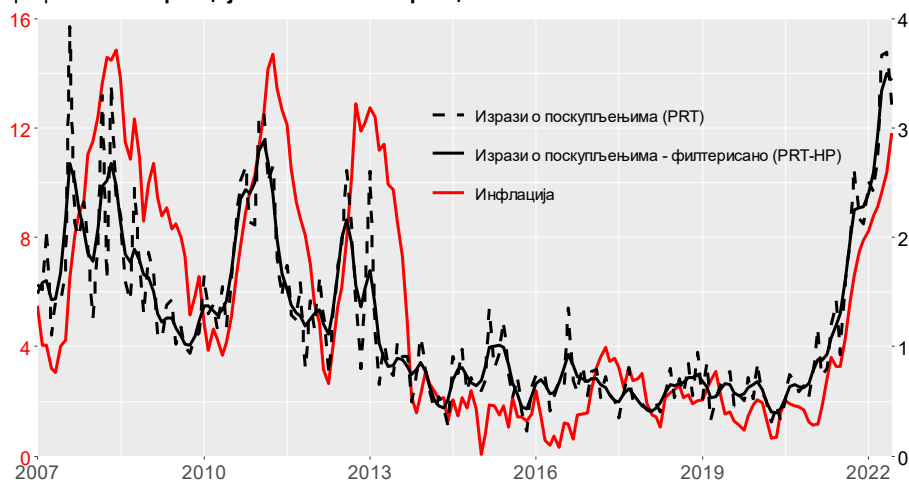


4. Веза између новинског инфлационог сентимента и инфлације

Више је начина на који изрази и текстови о ценама могу бити повезани са самом инфлацијом. На пример, након што се објави податак о порасту инфлације, он може изазвати пораст интересовања медија за ту тему, као и за појединачне промене цена које су довеле до тог раста. С друге стране, новине могу писати о текућим и будућим променама цена које ће тек бити укључене у инфлацију. Још једна могућност је да је инфлациони сентимент у новинама одраз кретања инфлационих очекивања, што је један од фактора који утичу на саму инфлацију.

Графикон 4 показује да је серија изрази о поскупљењима, иако веома волатилна, очигледно повезана са инфлацијом. (За графички приказ веза осталих мера сентимента с мерама инфлације видети Додатак.)

Графикон 4. **Инфлација и новински инфлациони сентимент**



Извор: РЗС и аутор.

Инфлација у посматраном периоду бележи четири значајна скока, с двоцифреним максимумима у 2008, 2011, 2012. и 2022. години. Свим овим циклусима претходио је или их је пратио снажан раст употребе израза о поскупљењима у новинама, на преко 2,5‰ укупног броја речи.

Може се приметити и да је током осмогодишњег периода стабилности инфлације (средина 2013 – средина 2021) *PRT* остао релативно низак, никад не премашивши 1,5‰. Кретање серија сентимента и инфлације у овом периоду често није било синхронизовано, као, на пример, у 2015, када повећано коришћење израза о поскупљењима није било праћено растом инфлације, или почетком 2017, када новине нису предвиделе привремени пораст инфлације на 4%.

Поред међугодишње, проверавали смо повезаност мера сентимента и са десезонираним месечним стопама инфлације. У овом случају анализирани серије су конзистентније – месечни пораст цена и месечни број израза/текстова.

Овде је битно нагласити и да ниједна од серија сентимента нема сезонско кретање, па стога није било неопходно њихово десезонирање. Поред тога, ове серије, као и две серије инфлације (месечна и међугодишња), стационарне су према Дики–Фулеровом тесту.

Табела 2 показује високу корелисаност између две групе серија на целом узорку. Корелација је најјача између међугодишње инфлације и филтриране серије израза о поскупљењима (*PRT-HP*). Корелисаност мера новинског инфлационог сентимента са месечном инфлацијом нешто су слабије.

Табела 2. Максималне корелације мера сентимента и инфлације за период 2007–2022. (доцње у заградама)

	Изрази		Изрази филтрирано		Чланци	
	<i>PRT</i>	<i>PCT</i>	<i>PRT-HP</i>	<i>PCT-HP</i>	<i>PRA</i>	<i>PCA</i>
Мг. инфлација	0,75 (2)	0,71 (3)	0,86 (4)	0,83 (4)	0,76 (3)	0,72 (4)
Мес. инфлација, дсз.	0,64 (0)	0,62 (0)	0,62 (0)	0,62 (0)	0,58 (0)	0,55 (0)

У складу с нашим ранијим опажањима, на периоду стабилне инфлације (средина 2013 – средина 2021) корелације су знатно слабије (Табела 3). У овом случају, месечне стопе инфлације показују нешто јачу корелацију с мерама сентимента од међугодишње стопе.

Табела 3. Максималне корелације мера сентимента и инфлације за период јун 2013 – јун 2021. (доцње у заградама)

	Изрази		Изрази филтрирано		Чланци	
	<i>PRT</i>	<i>PCT</i>	<i>PRT-HP</i>	<i>PCT-HP</i>	<i>PRA</i>	<i>PCA</i>
Мг. инфлација	0,22 (0)	0,26 (0)	0,23 (0)	0,26 (0)	0,14 (0)	0,23 (3)
Мес. инфлација, дсз.	0,30 (0)	0,37 (0)	0,25 (0)	0,31 (0)	0,22 (5)	0,17 (0)

Овде је важно нагласити да се подаци о инфлацији у Србији објављују 12. у месецу за претходни месец, тако да, и ако су корелације истовремене (без доцњи), може се рећи

да мере инфлационог сентимента претходе званичном податку о инфлацији, што је битно у контексту могућности коришћења овог показатеља у прогнозама инфлације.

Из истог разлога битно је утврдити правац узрочности. На основу Грејнцеровог теста узрочности, мере сентимента узрокују инфлацију у свим комбинацијама, при чему узрочност у неким случајевима иде у оба смера (Табела 4). Видети Додатак за детаљне резултате.

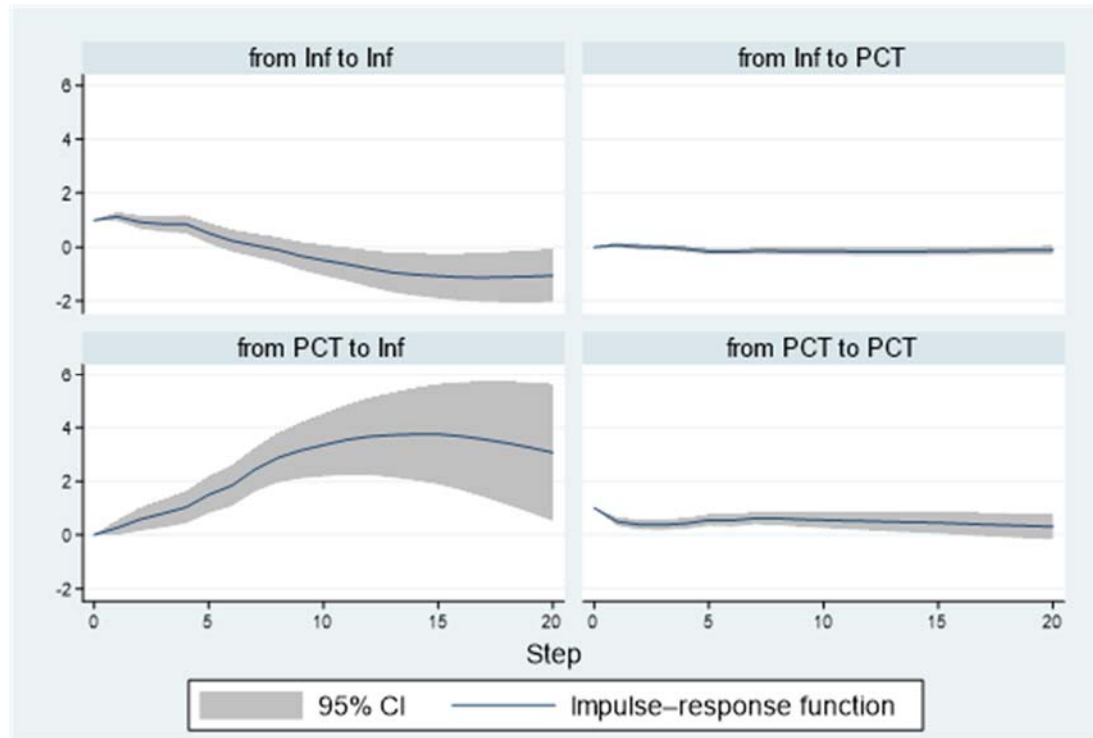
Табела 4. Смер Грејнцерове узрочности између мера сентимента и инфлације* (цео узорак)

	Изрази		Изрази филтрирано		Чланци	
	<i>PRT</i>	<i>PCT</i>	<i>PRT-HP</i>	<i>PCT-HP</i>	<i>PRA</i>	<i>PCA</i>
Мг. инфлација	↙↗	↙	↙↗	↙↗	↙	↙↗
Мес. инфлација, дсз.	↙	↙	↙↗	↙↗	↙	↙

* ↙ значи да узрочност иде од сентимента ка инфлацији, а ↗ у супротном смеру.

Функција одговора на импулсе из *VAR(7)* модела са инфлацијом и *PCT*-ом добра је илустрација ове везе. На графикону 5 јасно се може видети да инфлација реагује на инфлациони сентимент у новинама. Најјачи одговор инфлације на раст *PCT*-а за један промил је 4%, што је приближно у складу с нашим запажањима с Графикона 3, где се може видети да је раст *PCT*-а са 0–0,5‰ на 2–3‰ уобичајено био праћен растом инфлације од 8% до 10%.

Графикон 5. Функције одговора на имплусе из *VAR* модела с мг. инфлацијом и *PCT*

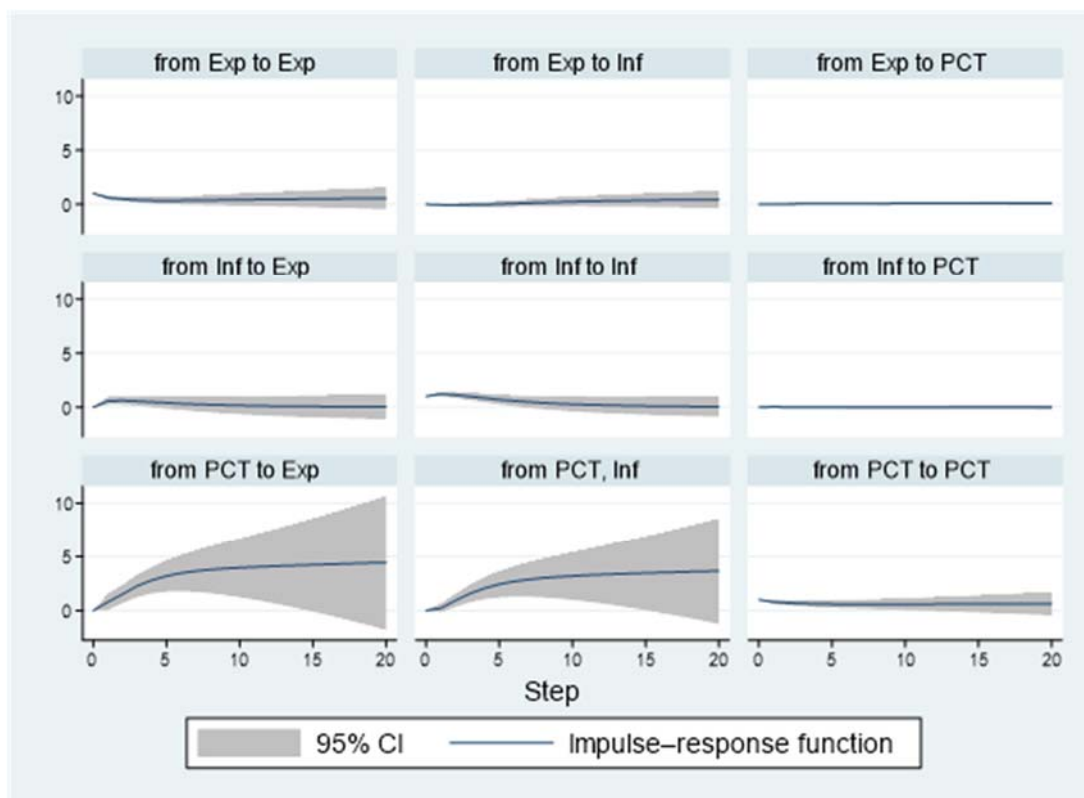


На бази претходно реченог, могли бисмо закључити да је инфлациони сентимент водећи показатељ инфлације. Шта би могао бити узрок те повезаности? Да ли то што новине пишу о текућим и будућим променама цена које тек треба да буду регистроване у инфлацији или су ти текстови одраз инфлационих очекивања, која су један од фактора инфлације?

Ако је ово последње случај, требало би да инфлациона очекивања узрокују инфлацију. Међутим, Грејндеров тест узрочности говори управо супротно, показујући да су инфлациона очекивања узрокована инфлацијом, а не обратно (видети Додатак).

Функције одговоре на импулсе из $VAR(2)$ модела са инфлацијом, PCT -ом и инфлационим очекивањима (Exp) показују да инфлациона очекивања реагују и на PCT , и на саму инфлацију. Такође, видимо да и по укључивању очекивања у модел стоји закључак да PCT утиче на инфлацију, а да обратно не важи (Графикон 6).

Графикон 6. Функције одговора на импулсе из VAR модела с међугодишњом инфлацијом, PCT -ом и инфлационим очекивањима



Да бисмо установили употребљивост овог односа за прогнозирање, спровели смо уобичајени приступ поређења ванузорачких грешака прогнозе модела инфлације с мерама сентимента с реперним моделом, у нашем случају ауторегресионим моделом инфлације, међугодишње и месечне. У анализу смо укључили $VAR(2)$ са инфлацијом и сентиментом, као и моделе с две једначине, од којих једна повезује сентимент и инфлацију, а друга је $AR(2)$ једначина за новински инфлациони сентимент (NIS):

$$Inf_m = \alpha_1 \cdot Inf_{m-1} + \alpha_2 \cdot NIS_{m-1} + \varepsilon_m$$

$$NIS_m = \beta_1 \cdot NIS_{m-1} + \beta_2 \cdot NIS_{m-2} + \gamma_m$$

Упоредивости ради, у случају међугодишње инфлације посматрамо грешку прогнозе l периода унапред, док у случају месечне инфлације посматрамо кумулативну грешку од првог до l -тог периода унапред (видети Додатак за детаље).

Мере грешке прогнозе, $RMSE$, различитих модела приказане су у Табели 5.

Табела 5. Ванузорачки $RMSE$ за различите моделе инфлације са мерама NIS -а (временски помаци у колонама)

	0	1	2	3	4	5	6
Модели с међугодишњом инфлацијом							
AR(1) – репер	0,613	0,613	0,929	1,172	1,384	1,556	1,739
PRT модел	0,567	0,549	0,770	0,958	1,145	1,307	1,516
PCT модел	0,558	0,548	0,770	0,968	1,160	1,303	1,499
PRA модел	0,582	0,569	0,843	1,062	1,253	1,411	1,574
PCA модел	0,601	0,563	0,824	1,054	1,249	1,414	1,609
PRT VAR	NA	0,662	1,007	1,286	1,583	1,893	2,194
PCT VAR	NA	0,658	0,994	1,275	1,576	1,880	2,189
PRT-HP модел	0,598	0,654	0,963	1,204	1,447	1,583	1,742
PCT-HP модел	0,587	0,641	0,956	1,240	1,520	1,652	1,804
Модели с месечном инфлацијом							
AR(1) – репер	0,396	0,396	0,699	1,016	1,345	1,649	1,969
PRT модел	0,324	0,338	0,522	0,704	0,890	1,048	1,258
PCT модел	0,322	0,342	0,540	0,748	0,960	1,141	1,372
PRA модел	0,334	0,338	0,527	0,712	0,879	1,004	1,188
PCA модел	0,347	0,342	0,546	0,766	0,974	1,156	1,400
PRT VAR	NA	0,350	0,552	0,764	0,986	1,176	1,413
PCT VAR	NA	0,346	0,536	0,728	0,932	1,111	1,336
PRT-HP модел	0,329	0,262	0,413	0,582	0,763	0,891	1,057
PCT-HP модел	0,326	0,256	0,422	0,629	0,830	0,969	1,153

Можемо видети да модели са NIS -ом у готово свим случајевима смањују грешку прогнозе у односу на реперни модел у моделима с међугодишњом инфлацијом, а у свим случајевима у моделима с месечном инфлацијом. Иако је месечна инфлација показала мању корелисаност са NIS -ом у односу на међугодишњу (Табела 2), модели с месечном инфлацијом имају боље перформансе када је у питању прогнозирање од модела с међугодишњом. Најнижу $RMSE$ постигли су модели с месечном инфлацијом и неком од мера базираних на бројању израза: PCT за текући месец, $PCT-HP$ за један месец унапред, модел са $PRT-HP$ од два месеца унапред, па надаље.

Разлог што смо ауторегресионе моделе користили као реперне јесте тај што се краткорочне прогнозе инфлације у Народној банци Србије раде на дезагрегираним

подацима, као комбинација моделског приступа (за неке компоненте), стручне процене и екстерних информација о променама појединих цена. Структурни модел с факторима инфлације у Народној банци Србије користи се за средњорочну пројекцију инфлације (Ђукић, Момчиловић и Трајчев (2010)), али како сентимент има сврхе користити само у кратком року, овај модел нисмо користили као реперни.

Последње што смо проверили јесте да ли новински инфлациони сентимент може побољшати објашњеност модела у који су укључени релевантни фактори инфлације. У ту сврху смо користи *ARDL* модел укупне инфлације из рада Ивковић, Јаковљевић и Милетић (2022), који као објашњавајуће варијабле укључује инфлацију у зони евра, цене пољопривредних производа, плате и девизни курс. У овај модел смо укључили *PCT* као додатну варијаблу и добили да је оцењен коефицијент уз њу статистички значајан и у кратком, и у дугом року (Табела 6). Из тога се може закључити да сентимент може побољшати објашњеност инфлације чак и када су у модел укључене најважније варијабле. Како је ово модел с једном једначином, не може се користити за прогнозу, због чега га нисмо користили у поређењу грешака прогнозе.

Табела 6. Оцењени резултати из *ARDL* модела инфлације

Conditional Error Correction Regression				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.632643	0.375823	-1.683354	0.0991
LOG(PRICES_SA(-1))*	-0.082318	0.027850	-2.955764	0.0049
LOG(EA_PRICES_SA)**	0.199081	0.135368	1.470666	0.1482
LOG(AGRIC_PRICES)**	-0.003742	0.008988	-0.416369	0.6791
LOG(WAGES_SA(-1))	0.018141	0.020716	0.875701	0.3857
LOG(NEER)**	-0.011635	0.021442	-0.542639	0.5900
PCT**	0.016931	0.002373	7.133902	0.0000
DLOG(WAGES_SA)	-0.193639	0.062357	-3.105336	0.0033
VES2012Q4	0.035339	0.005320	6.642341	0.0000

* p-value incompatible with t-Bounds distribution.
 ** Variable interpreted as $Z = Z(-1) + D(Z)$.

Levels Equation Case 2: Restricted Constant and No Trend				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(EA_PRICES_SA)	2.418436	1.214257	1.991700	0.0524
LOG(AGRIC_PRICES)	-0.045461	0.137099	-0.331593	0.7417
LOG(WAGES_SA)	0.220374	0.257840	0.854692	0.3972
LOG(NEER)	-0.141345	0.321418	-0.439756	0.6622
PCT	0.205681	0.097416	2.111374	0.0402
C	-7.685328	2.595410	-2.961123	0.0048

$$EC = LOG(PRICES_SA) - (2.4184 * LOG(EA_PRICES_SA) - 0.0455 * LOG(AGRIC_PRICES) + 0.2204 * LOG(WAGES_SA) - 0.1413 * LOG(NEER) + 0.2057 * PCT - 7.6853)$$

F-Bounds Test				
Test Statistic	Value	Signif.	I(0)	I(1)
F-statistic	37.18888	10%	2.08	3
		5%	2.39	3.38
		2.5%	2.7	3.73
		1%	3.06	4.15
Actual Sample Size	55	10%	2.226	3.241
		5%	2.617	3.743
		2.5%	3.06	4.15
		1%	3.543	4.839

5. Закључак

Сврха анализе коју смо представили у овом раду била је да се испита да ли новински текстови у Србији садрже корисне информације које би могле бити коришћене као показатељ инфлаторних притисака. Резултати показују да то јесте случај.

Текстуалну анализу смо спровели применом речничког приступа и Булове претраге на четири српске дневне новине, за период од почетка 2007. до јуна 2022. године. Најпре смо дефинисали листе израза за претрагу – једну у вези с поскупљењем и другу у вези с појефтињењем, а затим смо по месецима пребројали те изразе, као и чланке који садрже барем један од израза. Користећи те податке, креирали смо неколико временских серија мера новинског инфлационог сентимента: изрази о поскупљењима, изрази о променама цена, чланци о поскупљењима, чланци о променама цена, *HP*-филтер израза о поскупљењима и *HP*-филтер израза о променама цена.

Ове мере показале су се високо корелисаним с мерама инфлације на целом узорку. Корелација је снажнија с међугодишњом инфлацијом са одређеним временским помаком (око 0,8) него са месечном инфлацијом (око 0,6). Веза је нарочито упадљива у периодима већих осцилација инфлације. Тако су у сва четири циклуса када је инфлација премашивала 10% изрази о поскупљењима расли преко 2,5% укупног броја речи, док у периодима ниске инфлације овај показатељ углавном није прелазио 1%. С друге стране, у периодима релативно стабилне инфлације ова веза је знатно слабија. Ако се посматра само период ниске и стабилне инфлације (средина 2013 – средина 2021), веза између сентимента и инфлације пада на 0,2–0,3, што је знатно ниже него за цео узорак.

Овде је било битно установити правац узрочности, који би у принципу могао ићи у оба смера, с обзиром на то да новине пишу и о прошлој инфлацији, и о променама цена које ће тек бити укључене у инфлацију. Грејнцеров тест узрочности показао је да у свим комбинацијама узрочност иде од сентимента ка инфлацији, с тим да у понеким случајевима иде у оба смера. Функција одговоре на импулсе из *VAR* модела са изразима о променама цена и инфлацијом сугерише да постоји једносмерна веза између ове две варијабле, где једнопромили раст сентимента води расту инфлације од 4%.

Бројни радови третирају текстуалне информације као одраз инфлационих очекивања, која су даље битан фактор саме инфлације. Стога смо испитали могућност да веза између сентимента и инфлације иде преко инфлационих очекивања. Грејнцеров тест узрочности, те анализе функције одговора на импулсе из *VAR* модела са сентиментом, инфлацијом и очекивањима показали су да то у Србији није случај. За инфлациона очекивања становништва показало се да следе инфлацију (а да јој не претходе), при чему се веза између сентимента и инфлације не мења значајно укључивањем очекивања у модел. Закључили смо да је главни канал узрочности сентимента на инфлацију информациони садржај у новинама о променама цена које ће бити укључене у инфлацију.

Следеће смо утврдили да разне комбинације мера сентимента и инфлације у већини случајева имају боље перформансе када је у питању прогнозирање изван узорка у односу на реперни ауторегресиони модел инфлације. Мада су везе биле јаче код међугодишње инфлације, модели с месечном инфлацијом показали су се бољим за

прогнозирање. У веома кратком року најбоље су се показали модел с месечном инфлацијом и разликом броја израза о поскупљењима и појефтињењима (за текући месец), и модел с филтрираном верзијом исте серије (за један месец унапред).

Важно је нагласити и да су, по поменутиим критеријумима у анализи, мере сентимента базиране на броју израза показале боље резултате од оних базираних на броју текстова, што значи да је интензитет помињања цена релевантнији од његове учесталости.

На крају смо у модел укупне инфлације из рада Ивковић, Јаковљевић и Милетић (2022) укључили изразе о поскупљењима и добили статистички значајан коефицијент уз ту варијаблу, чак и када су најрелевантнији фактори инфлације укључени.

Узимајући све у обзир, закључили смо да су мере новинског инфлационог сентимента добар показатељ инфлаторних притисака, уз ограду да то не мора бити случај у стабилним временима. Овај показатељ се стога пре свега може користити као својеврстан аларм упозорења на потенцијални снажан раст инфлације, као онај коме смо сведочили од краја 2021. године.

Додатак

Прогнозе изван узорка

Овај метод се користи да би се проверило како се модели прогнозирају на основу историјских података. У нашем раду смо пуштали прогнозе с малим моделима инфлације с различитим мерама сентимента (*PRT*, *PCT*, *PRA*, *PCA*, *PRT-HP* и *PCT-HP*) и ауторегресионим једначинама за *NIS*:

$$Inf_m = \alpha_1 \cdot Inf_{m-1} + \alpha_2 \cdot NIS_{m-1} + \varepsilon_m$$

$$NIS_m = \beta_1 \cdot NIS_{m-1} + \beta_2 \cdot NIS_{m-2} + \gamma_m$$

У ову „трку” смо укључили и *VAR(2)* моделе са инфлацијом и *NIS*-ом, док смо као реперни модел користили *AR(2)* модел за инфлацију.

Цео процес смо спровели на уобичајени начин: прво смо оценили моделе на узорку који завршава у одређеном тренутку у прошлости (почињемо с јуном 2014), затим смо прогнозирали инфлацију за нула до шест периода унапред, а потом смо рачунали одступање прогнозиране од остварене инфлације за сваки временски помак. На крају смо рачунали грешку прогнозе (*RMSE* – *root-mean-squared error*) као корен просечне суме квадрата одступања за све узорке, од нула до шест месеци унапред. *RMSE* за све временске помаке *l* рачунали смо на следећи начин:

$$RMSE_l = \sqrt{\frac{\sum_{m=t+1}^{endhist} (Inf_{m+l}^{fcast} - Inf_{m+l})^2}{endhist - m}}$$

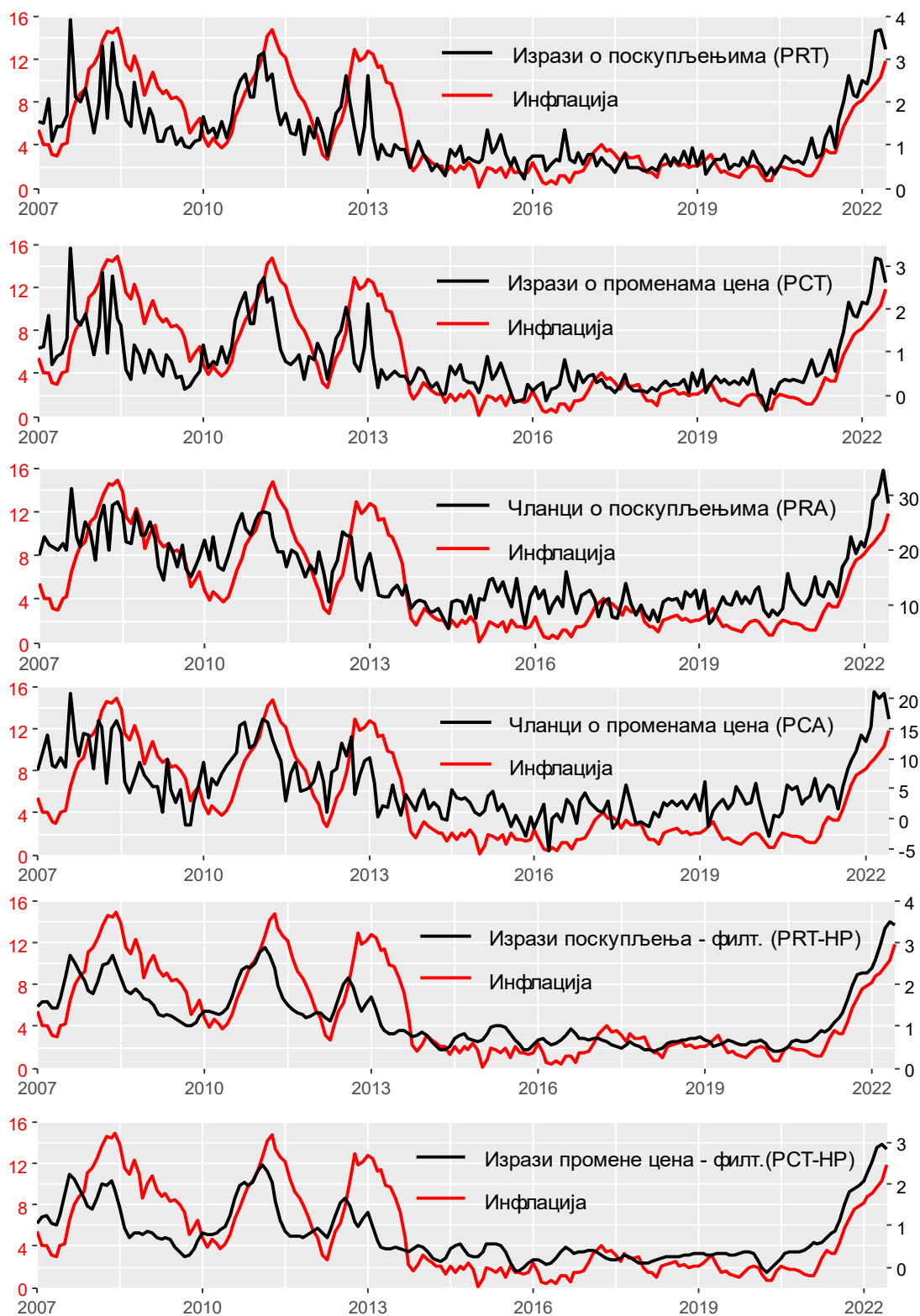
где је *t* последњи месец узорка за оцену, а *endhist* последњи месец целог узорка (јун 2022.).

Када су у питању модели с месечном инфлацијом, грешке смо кумулирали (од 1 до *l* периода унапред) и на тај начин их учинили упоредивим с грешкама прогнозе међугодишње инфлације:

$$RMSE_l = \sqrt{\frac{\sum_{m=t+1}^{endhist} (\sum_{i=1}^l (Inf_{m+i}^{fcast} - Inf_{m+i}))^2}{endhist - m}}$$

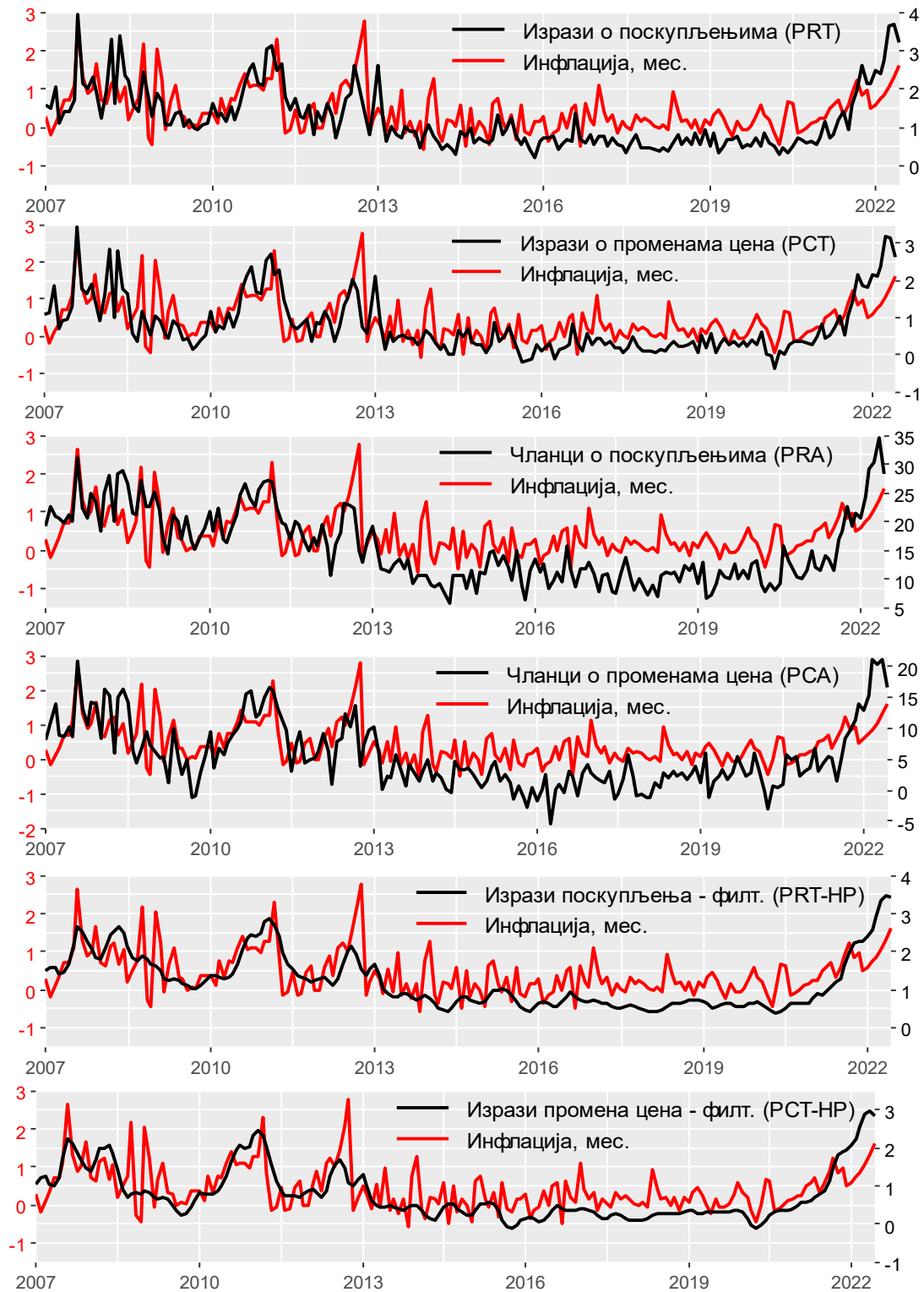
Овде је важно објаснити на који начин смо рачунали грешку прогнозе за текући месец (0 помака), с обзиром на то да, дословно речено, то не би била ванузорачка грешка. То смо урадили тако што смо прогнозирали инфлацију за један период унапред, али са оствареним *NIS*-ом за *t + 1* (што ову прогнозу чини ванузорачком). Идеја овог приступа је да симулира ситуацију на крају месеца када имамо расположиву остварену меру сентимента, али још увек немамо податак за инфлацију за тај месец. Ово се разликује од ванузорачке прогнозе за један месец унапред, где пуштамо прогнозу не са оствареним сентиментом, већ с његовом моделском прогнозом.

Графикон А1. Инфлација и показатељи новинског инфлационог сентимента



Извор: РЗС и аутор.

Графикон А2. Месечна инфлација и показатељи новинског инфлационог сентимента



Извор: РЗС и аутор.

Табела А1. Грејнџерова узрочност између мера НИС и инфлације** (цео узорак)

	Изрази		Изрази филтрирано		Чланци	
	<i>PRT</i>	<i>PCT</i>	<i>PRT-HP</i>	<i>PCT-HP</i>	<i>PRA</i>	<i>PCA</i>
Мг. инфлација	5.95017*	11.4523*	18.5593*	19.2880*	5.10133*	6.73472*
	3.04455*	0.98914	5.84182*	4.70734*	1.73623	2.56426*
Мес. инфлација, дсз.	10.8278*	2.31504*	11.8014*	12.1729*	10.2410*	11.2649*
	2.31504	0.80398	5.08440*	3.90030*	2.36838	1.95995

* Означава статистичку значајност на нивоу од 5%, тј. постоји узрочност

** Први број у пољу је F-статистика за узрочност од *NIS* ка инфлацији, а доњи број у супротном смеру

Табела А2. Грејнџерова узрочност између инфлације и инфлационих очекивања

Од очекивања ка инфлацији: $F = 0.35442$ ($p=0.7023$)
Од инфлације ка очекивањима: $F = 15.1907^*$ ($p=0.0000$)

База економских текстова

У базу су укључени економски текстови из четири српске дневне новине: „Политика”, „Вечерње новости”, „Данас” и „Блиц”. Три новине („Политика”, „Данас” и „Блиц”) имају расположиве текстове од почетка 2007, док за Вечерње новости узорак почиње од марта 2008. године. Док неке новине („Политика” и „Данас”) имају јавно расположиве архиве на својим веб-страницама, друге немају („Вечерње новости” и „Блиц”), па смо њихове чланке преузимали с других веб-сајтова који прикупљају и објављују вести из разних извора. Коришћени су само текстови из економских рубрика.

Литература

- Angelico, C., Marcucci, J., Miccoli, M., & Quarta F. (2021). „Can we measure inflation expectations using Twitter?“, *Temi di discussione (Economic working papers)* 1318, Bank of Italy.
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4), 1593–1636.
- Bholat D., Hans, S., Santos, P., & Schonhardt-Bailey, C. (2015). *Text mining for central banks, Handbooks*, Centre for Central Banking Studies, Bank of England, number 33.
- Carroll, C. D. (2003). „Macroeconomic expectations of households and professional forecasters”. *The Quarterly Journal of economics*, 118(1), 269–298.
- Cowles, A. (1933). „Can Stock Market Forecasters Forecast?”. *Econometrica*, 1(3), 309–324.
- Dräger, L. (2015). „Inflation perceptions and expectations in Sweden—Are media reports the missing link?”. *Oxford Bulletin of Economics and Statistics*, 77(5), 681–700.
- Dräger, L., & Lamla, M. J. (2012). „Updating inflation expectations: Evidence from micro-data”. *Economics Letters*, 117(3), 807–810.
- Djukic, M., Momcilovic, J., & Trajcevic, Lj. (2010). „Medium-term projection model of the National Bank of Serbia”, *Working papers 17*, National Bank of Serbia.
- Hansen, S, McMahon, M., & Tong, M. (2018). „The long-run information effect of central bank narrative”. *Working paper*, Oxford University and Bank of England.
- Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). „Making text count: economic forecasting using newspaper text.” *Staff Working Paper No. 865.*, Bank of England.
- Ivković, A., Jakovljević, S., & Miletić, M. (2022). „Estimation of the impact of global and domestic factors on inflation in Serbia”. *Working Papers Bulletin II – March 2022*, National bank of Serbia.
- Larsen, V. H., Thorsrud, L. A., & Zhulanova, J. (2021). „News-driven inflation expectations and information rigidities”. *Journal of Monetary Economics*, 117, 507–520.
- Lucca, D., & Trebbi, F. (2009). „Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements”. *NBER Working Papers 15367*, National Bureau of Economic Research, Inc.
- Rambaccussing, D., & Kwiatkowski, A. (2020). „Forecasting with news sentiment: Evidence with UK newspapers”. *International Journal of Forecasting*, 36(4), 1501–1516.
- Tetlock, P. (2007). „Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. *The Journal of Finance*, 62(3), 1139–1168.