



ҚАЗАҚСТАН ҰЛТТЫҚ БАНКІ

Galymzhan жүйесі: Қазақстандағы тұтынушылық инфляцияны online-бағалау

Ақша-кредит саясаты департаменті
№2021 экономикалық зерттеу-2

О. Төлеуов,
И. Ержан,
А. Сейдахметов

Қазақстан Республикасы Ұлттық Банкінің (бұдан әрі – ҚРҰБ) экономикалық зерттеулері мен талдамалық жазбалары ҚРҰБ зерттеулерінің, сондай-ақ ҚРҰБ қызметкерлерінің басқа да ғылыми-зерттеу жұмыстарының нәтижелерін таратуға арналған. Экономикалық зерттеулер пікірталастарды ынталандыру үшін қолданылады. Құжатта айтылған пікірлер автордың жеке ұстанымын білдіреді және ҚРҰБ ресми ұстанымымен сәйкес келмеуі мүмкін.

Galymzhan жүйесі: Қазақстандағы тұтынушылық инфляцияны online-бағалау.
NBRK – WP – 2021 – 2

© Қазақстан Республикасының Ұлттық Банкі

Ұсынылған материалдардың кез келген жариялануына авторлардың рұқсатымен ғана рұқсат беріледі

Galymzhan жүйесі: Қазақстандағы тұтынушылық инфляцияны online-бағалау

Олжас Төлеуов ¹
Ислам Ержан ²
Аңсар Сейдахметов ³

Аннотация

Инфляциялық таргеттеу саясатын ұстанатын орталық банктер үшін ағымдағы инфляциялық үрдістерді бақылау және талдау өте маңызды. Бағаның өзгеруі туралы жедел ақпараттың болуы әртүрлі күйзелістерге тиімді жауап беруге және қажет болған жағдайда ақша-кредит саясатын уақтылы түзетуге мүмкіндік береді.

Қазіргі уақытта тұтыну қоржынының тауарлары бағасының өзгеруін жедел талдаудың тежеуші факторы ресми баға статистикасы жарияланымдарының шектеулі жиілігі болып табылады. Осыған байланысты инфляцияның жанама жоғары жиілікті прокси-индикаторларын құру мәселесі ерекше маңызға ие.

Бұл жұмыста бағдарламалық алгоритмнің көмегімен іске асырылған веб-беттерден деректерді алу жолымен оларды автоматты түрде алуды білдіретін веб-скрепинг технологиясын пайдалану арқылы Ұлттық Банкте Қазақстан инфляциясының жоғары жиілікті прокси-көрсеткішін құру әдістемесі мен нәтижелері сипатталған.

Түйінді сөздер: инфляция, веб-скрепинг, Galymzhan, онлайн дүкен, бағалар, big data, ТБИ, парсинг, тұтыну тауарлары.

JEL-сыныптау: E31, E37, E39.

¹ Олжас Төлеуов – Қазақстан Республикасының Ұлттық Банкі Төрағасының кеңесшісі. E-mail: Olzhas.Tuleuov@nationalbank.kz

² Ислам Ержан – Қазақстан Республикасы Ұлттық Банкінің Ақша-кредит саясаты департаменті макроэкономикалық зерттеулер және болжау басқармасының бас маман-талдаушысы. E-mail: Islam.Yerzhan@nationalbank.kz

³ Аңсар Сейдахметов – Қазақстан Республикасы Ұлттық Банкінің Ақша-кредит саясаты департаменті макроэкономикалық зерттеулер және болжау басқармасының жетекші маман-талдаушысы. E-mail: Ansar.Seidakhmetov@nationalbank.kz

Мазмұны

1. Кіріспе	5
2. Әдебиетке және халықаралық тәжірибеге шолу	6
3. Пайдаланылатын деректер мен әдістеме.....	9
3.1 Galymzhan 1.0	10
3.2 Galymzhan 2.0	11
4. Нәтижелерді талқылау	13
5. Бұдан былайғы зерттеулерге арналған қорытындылар мен ұсынымдар	16
Әдебиеттер тізімі	17

1. Кіріспе

Инфляция-тұтыну қоржынының тауарлары бағасының орташа өлшемді өсу көрсеткіші, инфляциялық таргеттеу саясатын жүзеге асыратын орталық банктер үшін шешімдер қабылдау кезінде негіз қалаушы индикатор болып табылады. Инфляцияның тұрақты төмен қарқыны жоспарлау көкжиегінің кеңеюіне және инвестициялардың өсуіне ықпал етеді, бұл экономикалық белсенділіктің кеңеюіне және халықтың әл-ауқатының жақсаруына әкеледі.

Тұтыну қоржынының тауарлары бағасының тұрақты төмен өсу қарқынының сақталуына дәйекті ақша-кредит саясатын жүргізу ықпал етеді.

Түрлі дағдарыстар кезеңінде инфляциялық процестерді тежеу мақсатында монетарлық реттеуші болып жатқан және күтілетін макропроцестерге пәрменді ден қою, сондай-ақ қажет болған кезде ақша-кредит саясатын уақтылы түзету қажет. Тұтыну бағаларының өзгеруі туралы жедел ақпараттың болуы барынша сапалы шешімдер қабылдауға мүмкіндік береді.

Қазіргі уақытта Қазақстанда тұтыну қоржыны тауарларына бағалардың өзгеруін жедел талдаудың тежеуші факторы ресми статистика жарияланымдарының шектеулі жиілігі болып табылады. Осыған байланысты инфляцияның жанама жоғары жиілікті прокси-индикаторларын құру мәселесі ерекше маңызға ие.

Соңғы жылдары көптеген елдердің статистикалық органдары мен орталық банктерінің арасында жоғары жиілікті ақпаратты үлкен көлемде алуға мүмкіндік беретін бағаларды «сканерлеуге» және интернет ресурстардың веб-скрепингіне негізделген бағалық өзгерістерді жинау жүйелері ерекше танымалдылыққа ие болуда.

Бағаларды «сканерлеу» көбінесе бар-кодтар бойынша сәйкестендірілетін тауарлардың бағасы мен айналымы туралы ақпаратты қамтитын транзакциялар туралы ретейлерлердің деректерін пайдалануды білдіреді. Бұл технологияны қолдану осы ақпаратты ұсынатын дүкендермен тығыз ынтымақтастықты білдіреді (G. Chessa, 2016).

Әр түрлі сайттарды веб-скрепинг технологиясы тұтыну бағаларының өзгеруі туралы құнды жоғары жиілікті ақпаратты одан әрі өңдеу және алу үшін құрылымдық түрде баға деректерін жинауға мүмкіндік береді. Технология интернет-сатып алулардың танымал бола бастауына байланысты дамыды және көптеген сатушылар өздерінің ресми сайттарын барлық тауар өнімдерінің бағасымен ашып, онлайн-сатылымға көшуде.

Баға өзгерістерін жинаудың бұл түрі мемлекеттік мекемелерде (статистикалық органдар, орталық банктер), сондай-ақ уақытты, адам ресурстарын үнемдеуге және шығындарды азайтуға байланысты әртүрлі елдердің жеке ұйымдарында үлкен танымалдылыққа ие. Веб-скрепинг арқылы баға деңгейінің өзгеруі туралы мәліметтерді алу әсіресе 2020 жылы, COVID-19 коронавирустық пандемиясының аясында көптеген сауда нүктелері карантиндік шектеулер жағдайында өз қызметін тоқтата тұруға мәжбүр болған кезде өзекті болды, бұл баға өзгерістерін классикалық әдіспен жинауды қиындатты.

Қазақстан Ұлттық Банкі Galymzhan деп аталатын тұтыну бағаларының серпінін жинау және талдау жүйесі әзірленген 2018 жылы интернет ресурстардан деректерді жинау мәселесін зерттеуге кірісті.

Техникалық тұрғыдан алғанда, Galymzhan-бұл R бағдарламалау тілінде жазылған алгоритм, бұл кодты Windows жүйесіне біріктіру осы бағдарламаның автономиясына қол жеткізуге мүмкіндік берді. Осылайша, күн сайын, мереке және демалыс күндерін қоса алғанда, жүйе деректерді жинау және өңдеу алгоритмін іске қосады. Жүйе жұмысының нәтижесі инфляцияның күнделікті прокси-индикаторын құру болып табылады.

Бұл жұмыс бірнеше бөліктен тұрады. Бірінші бөлімде басқа авторлардың ұқсас жұмыстары қарастырылған әдебиеттерге шолу жасалады. Екінші бөлімде баға өзгерістерін жинау және соңғы индексті құру әдістемесі сипатталады. Төменде нәтижелерді талқылау бөлімі берілген, онда авторлар алынған индексті ресми статистикалық мәліметтермен салыстырады. Осы зерттеудің нәтижелері қорытынды бөлім болып табылады.

2. Әдебиетке және халықаралық тәжірибеге шолу

Құрылымдық түрде ашық көздерден қажетті ақпаратты алуға мүмкіндік беретін технологиялардың дамуымен бірге әртүрлі интернет-ресурстар санының артуы өмірдің көптеген салаларының дамуына күшті серпін берді. Бұрын қолжетімсіз ақпараттың болуы бизнес-процестерді автоматтандыруға және оңтайландыруға мүмкіндік берді.

Веб-скрепинг технологиясын қолдана отырып, статистикалық ақпаратты, атап айтқанда инфляцияның прокси-көрсеткішін алу процесі жақында танымал бола бастады. Бұл әдіс ресми статистикалық органдар жүргізетін бағаларды классикалық зерттеумен салыстырғанда, жоғары жиіліктегі баға өзгерістерін тіркеуге және адами ресурстарды босатуға мүмкіндік береді, бұл статистикалық органдарға жарияланған деректердің сапасы мен жиілігін бір уақытта арттыруға мүмкіндік береді.

Қазіргі уақытта Италия, Австралия, Бельгия, Германия, Нидерланды, Венгрия, Ұлыбритания, Канада сияқты елдердің статистикалық органдары өздерінің операциялық қызметінде веб-скрепинг жүйелерін игере бастады (Boshoff, 2020). Мәселен, Италияның Ұлттық статистика институты бағаларды автоматты түрде белгілеу жүйесін енгізу бойынша белсенді жұмыс жүргізуде (Polidoro, 2015). Атап айтқанда, статистиктер мен IT-мамандар тобы тауарлардың екі тобына: Тұрмыстық техника мен авиабилеттерге назар аудара отырып, тұтыну бағаларын тексеруде веб-беттерден деректерді алу әдістерін пайдалану мүмкіндігіне тестілеу жүргізді. Тестілеу нәтижелері статистикалық деректердің тиімділігі мен сапасын арттыру тұрғысынан интернетте автоматты түрде баға жинауды пайдаланудың үлкен әлеуетін растады.

Ең алдымен, инфляцияны өлшеу үшін деректерді алу құралы ретінде парсинг әдістерін пайдалану үлкен деректердің үш құрауышының бірімен (Big

Data), атап айтқанда, жылдамдықпен тікелей байланысты (Velocity)⁴. Статистика институтының қызметкерлері бұрын да веб-ресурстардан алынған бағаларды пайдаланғанын, бірақ бұған дейін бағаларды қызметкерлер қолмен тіркегенін («copy and paste» режимінде) айта кету керек.

Веб-ресурстардан бағаның өзгеруі туралы деректерді алу технологиясын қолданатын тағы бір статистикалық орган – Австралия статистика бюросы. 2017 жылдың ортасынан бастап Статистика бюросы тұтыну бағаларының индексі (ТБИ) есептеу кезінде пайдаланылатын деректер базасына веб-скрепинг⁵ нәтижесінде алынған көрсеткіштерді біртіндеп енгізеді. Бұған дейін Интернеттен алынған бағалар олардың сәйкес келетініне және баға серпінін өлшеуге жарамды екеніне көз жеткізу үшін соңғы 12 айда классикалық әдіспен жиналған бағалармен салыстырылды. Осыдан кейін қолмен жиналған белгілі бір өнімдердің бағалары интернеттен алынған деректермен алмастырылып, осы өнімдердің бағасын қолмен жинау тоқтатылды. Осылайша жиналған деректер негізінен алкоголь бағасының, саны аздаған киім-кешек тауарлары мен автомобиль бөлшектерінің индексі құру үшін қолданылады, бұл ТБИ-дің шамамен 5%-ын құрайды. Бұл тәсілдің бірқатар артықшылықтары бар. Аптасына екі рет жиналған деректерге негізделген орташа баға айына бір рет жиналған бағадан гөрі көбірек көрінеді. Сонымен қатар бұл ТБИ себетіне көбірек өнімдерді қосуға және жинау шығындарын азайтуға мүмкіндік береді. Қазіргі уақытта апта сайын 500 000 өкіл тауарға баға жинау жүргізілуде, ал дәстүрлі жинау тәсілдерімен барлығы шамамен 1000 позиция⁶ жиналатын еді.

Германияның Федералды статистикалық басқармасы 2017 жылы онлайн-баға жинау жұмысын жүргізе бастады. Тәжірибе кезеңінде қызмет көрсету бағасын қоспағанда, тұтыну себетінің 242 тауары бойынша жинау жүргізілді. Жалпы, Германияның тұтыну себетінде 600-ге жуық тауарлар мен қызметтер бар. Веб-скрепинг технологиясы сағатына 3050 бағаны жинауға мүмкіндік берді. Мұндай жоғары жиілік бұл жұмыстың бастапқы кезеңде қолданбалы сипатқа қосымша зерттеу бағыты болғандығымен түсіндіріледі (Blaudow and Burg, 2018). Жоғары жиілікті баға жинау нәтижесінде авторлар тәулік ішінде интернет-дүкендердегі баға серпінінің толық көрінісін анықтай алды. Мәселен, онлайн-дүкендердің басым бөлігі тәулік басында (00:00) бағаға түзетулер енгізеді.

Германияның статистикалық басқару бағасын автоматты түрде жинау жүйесі Java бағдарламалау тілінде жазылған. Веб-парақтар парсингі браузердегі кодта алдын-ала көрсетілген әрекеттерді көрсету арқылы веб-браузерді басқаратын Selenium құралын қолдану арқылы жүзеге асырылды. Интернет

⁴ Үлкен деректерді мына үш V көмегімен түсіндіруге болады: Velocity (жылдамдық), Variety (әртүрлілік) және Volume (көлем). 2001 жылғы зерттеу есебінде META Group (қазіргі Gartner) талдаушысы Дуг Лейни үлкен деректерді үш өлшемді деп анықтады, яғни өсіп келе жатқан көлем (деректер көлемі), жылдамдық (деректерді қабылдау және беру жылдамдығы) және әртүрлілік (деректер түрлері мен көздер ауқымы). Кейінірек, 2012 жылы Gartner үлкен көлемді, жоғары жылдамдықты және үлкен әртүрлілік ретінде үлкен деректердің анықтамасын жанартты.

⁵ Австралия статистика бюросының ресми сайты. Қолжетімділік режимі: <https://www.abs.gov.au/articles/web-scraping-australian-cpi>.

⁶ Australian Bureau of Statistics (ABS), Maximizing the use of Web Scraped Data in the Australian CPI.

ресурстарының бұғаттауын болдырмау үшін жобада серпінді IP мекенжайлар пайдаланылады.

Қазіргі уақытта инфляцияны есептеу кезінде азық-түлік, алкогольсіз сусындар, алкогольді ішімдіктер және темекі өнімдері сияқты топтар бойынша 10 000 өкіл тауарларға бағаның өзгеруі туралы онлайн деректер пайдаланылады (Boshoff, 2020).

Бельгия статистика офисі соңғы жылдары интернет-дүкендерден деректер жинауды жүргізуде. Жүйе R бағдарламалау тілінде жазылған және rvest⁷ және Rselenium⁸ сияқты топтамаларды қолданады. Алынған мәліметтер халықаралық сапарлар мен видео ойындарға инфляцияны есептеу кезінде пайдаланылады.

Сонымен қатар күніне немесе аптасына бірнеше рет 70-ке жуық парсер киім, аяқ киім, әуе билеттері, қонақ үй бөлмелері, пайдаланылған автомобильдер, дәрі-дәрмектер, кітаптар мен тұрмыстық техника және т. б. бағаларын жинайды. Бельгияның статистикалық басқармасы жақын арада веб-скрепинг арқылы алынған жоғарыда аталған тауарлар бойынша деректерді ресми жарияланатын тұтыну бағаларының индекcін есептеуге қосуды жоспарлап отыр (Loon, Roels, 2018).

Нидерланды статистика басқармасы қазіргі уақытта күн сайын 100 000 өкіл тауарға бағаның өзгеруін тіркей отырып үш ірі ретейлердің интернет-ресурстарынан баға жинауды жүргізуде. Бұл ретте баға өзгерістерін тіркеу тек киім заттары үшін ғана жүргізіледі. Бұл екі негізгі себепке байланысты. Біріншіден, 2010 жылдан бастап Нидерланды статистикалық басқармасы бағаны сканерлеу технологиясы арқылы азық-түлік, азық-түлікке жатпайтын тауарлардың бір бөлігіне бағаның өзгеруін тіркейді, қазіргі уақытта офлайн дүкендерде деректерді жинаудың көп бөлігі киім-кешектерге тиесілі (Grient, Naan, 2010). Екіншіден, киім – бұл процесті автоматтандыруды жеңілдететін, тауарлардың сынамасын жиі түзетуді қажет етпейтін тауар. Қазіргі уақытта тест режимінде 16 робот 19 бөлшек сауда дүкендерінде деректерді жинайды. Тексеруден кейін бұл деректер негізгі үлгіге қосылады, бұл оны күнделікті 500 000 бақылауға дейін арттыруға мүмкіндік береді (Griffioen, Bosch, 2016).

Ресми статистикалық органдардан басқа, прокси-ТБИ құру үшін веб-скрепинг технологиясын қолдана отырып алынған деректерді пайдалану мемлекеттік емес мекемелер арасында да кең таралды. Ресурстар мен тиісті дағдылар болған кезде жанама ТБИ құру жеке зерттеушілер үшін мүмкін болатын міндетке айналады. Polidoro өз жұмысында болашақта құралдар мен ақпараттың кең қол жетімділігі аясында статистикалық институттар бәсекеге қабілеттілігін жоғалтуы мүмкін және олар қазір олардың статистикалық ақпаратты ресми деңгейде жинау мен таратудың ерекше құқығы қолдап отырған өздерінің монополиясын жоғалтуы мүмкін екенін жазады (Polidoro, 2015).

Ресми емес органдардың баға статистикасын тарату мысалдары – «Billion Prices Project» және «PriceStats» жобалары. «The billion Prices Project» жобасы – америкалық экономистер Альберто Кавалло мен Роберто Ригобонның

⁷ Веб-парақтардан ақпарат жинауға мүмкіндік беретін кітапхана.

⁸ Веб-браузерді автоматтандыру үшін R-де «Selenium 2.0 WebDriver» пайдалану мүмкіндігін беретін топтама.

академиялық бастамасы, оның шеңберінде 2008 жылдан бастап бүкіл әлем бойынша жүздеген интернет-дүкендерден күнделікті баға жинау жүргізіліп, одан әрі нақты уақыт режимінде инфляцияны бағалау беріледі. Аргентинада 2007 жылдан 2015 жылға дейінгі инфляция статистикасына айла-шарғы жасау осы жүйенің құрылуына түрткі болды. Осы кезеңде Аргентинаның Ұлттық статистикалық басқармасы есептейтін инфляцияның ресми деңгейі бағаның нақты өзгеруін көрсетпейтіні көпшілікке белгілі болды. Статистика институтының ресми деректері бойынша инфляцияның орташа жылдық деңгейі 2007 жылдан бастап 2015 жылға дейін 8% болды, ал онлайн-деректерге сәйкес орташа мән 20%-дан астам болды, бұл жекелеген өңірлердегі бағалауларға және үй шаруашылықтарының инфляциялық күтулеріне сәйкес келді. 2011 жылы Аргентина үкіметі айыппұл салып, жергілікті сарапшыларды деректерді жинауды тоқтатуға мәжбүр ете бастаған кезде, елден тыс жерде онлайн-баға жинау мүмкіндігі аса пайдалы болды. Үкімет ауысқаннан кейін 2015 жылдың соңында ресми статистикамен айла-шарғы жасау тоқтатылды.

Қазіргі кезде «The Billion Prices Project» күн сайын 50-ден астам елде 300 онлайн-дүкендерде сатылатын 5 миллионға жуық тауарлардың бағасын тіркейді (Cavallo, 2016).

«The Billion Prices Project» негізінде 2011 жылы «PriceStats» коммерциялық жобасы жасалды, ол деректерді жинайды және оларды қолдана отырып, орталық банктер мен қаржы секторы үшін жоғары жиілікті индекстерді құрады. «PriceStats»-да пайдаланылатын деректердің саны мен сапасы едәуір арттырылды. Қазіргі уақытта компания 23 елде күнделікті инфляция индексін және 10 елде өндірушілердің баға индексін одан әрі құру үшін 1200-ден астам бөлшек сатушылардан шамамен 15 миллион тауардың бағасын тіркейді. Барлығы 60 ел үшін баға жиналады.

Тұтыну бағаларының жанама индексін құру үшін веб-скрепинг технологиясын пайдаланудың халықаралық тәжірибесін талдау мемлекеттік мекемелер мен жеке сектор арасында да баға жинаудың осы әдісінің қолайлы жұмыс істеп тұруы мен оған деген өсіп келе жатқан қызығушылықты көрсетті. Интернет-дүкендердің санын арттыру және қолжетімді ақпарат алу құралдарын, атап айтқанда онлайн-дүкендердегі бағаларды үнемі жетілдіру жақын арада статистикалық ақпараттың негізгі көзі бола алады.

3. Пайдаланылатын деректер мен әдістеме

Қазақстан Ұлттық Банкінде инфляцияның прокси-индикаторын құру арқылы тұтыну бағалары нарығындағы ағымдағы жағдайды талдау үшін қолданылатын құрал ретінде Galymzhan жүйесін дамыту және оны қалыптастыру процесін екі кезеңге бөлуге болады. 2018 жылдың басынан 2020 жылғы қазан аралығында Galymzhan 1.0. жүйесінің бірінші нұсқасы пайдаланылды. 2020 жылғы қазанда жүйе айтарлықтай жетілдірілді. Мәселен, екінші нұсқада бақыланатын тауарлар саны артып, географиялық көрінісі кеңейтілді және жүйенің дербестігі артты.

Катализаторы COVID -19 коронавирус пандемиясы болған Қазақстанның ірі қалаларында ғана емес, өңірлік орталықтарында да көптеген жаңа интернет-алаңдардың ашылуы, сондай-ақ Galymzhan жүйесімен жұмыс істейтін талдаушылардың деректерін жинау және өңдеу дағдыларын жақсарту және жинақтау қызметтің осы бағытын елеулі дамытуға мүмкіндік берді.

Бағдарламалау дағдыларының белгілі бір тізімін қажет ететін веб-скрепинг техникалық тұрғыдан өте күрделі процесс болуымен қатар, индексті есептеу тәсілін таңдау да тривиалды емес мәселе болып табылады. Қазақстанның Стратегиялық жоспарлау және реформалар агенттігінің Ұлттық статистика бюросы (бұдан әрі – ҰСА) жүзеге асыратын баға өзгерісін дәстүрлі жинау мақсатты іріктемеге негізделген, ол тұтыну сегменттері бойынша алдын ала таңдалған тиісті тауарлар мен қызметтер тізімінің болуын болжайды. Бұл тізімге үй шаруашылықтары тұтынатын тауарлар мен қызметтер кіреді, олардың салмағы осы тауарлар мен қызметтерді сатып алу шығындарының үлесі бойынша анықталады. Бұдан әрі, бағаларды зерттеуді инспекторлар оқшауланған жерде әр түрлі сауда орындарында жүзеге асырады. ҰСБ⁹ мен Galymzhan жүйесі пайдаланатын әдіснама бірінші кезекте тауарлар орталықтандырылған түрде анықталатын, содан кейін баға жинау процесінің өзі жүретін ұлғаюға қарай тәсілге негізделген.

Galymzhan жүйесінің мақсаты баға серпіні бойынша барлық ақпаратты жинау емес, инфляцияның прокси-индикаторын құру екенін атап өткен жөн. ҰБС-мен ұқсас тәсілді пайдалану алынған индекстердің түсіндірілуін сақтауға және оларды Қазақстандағы тұтынушылық инфляцияны прокси-бағалау ретінде пайдалануға мүмкіндік береді.

3.1 Galymzhan 1.0

2018 жылдың басында іске қосылған Galymzhan 1.0. ол R бағдарламалау тілінде жазылған арнайы алгоритмді білдіреді. Бұл кодты Windows жүйесімен ықпалдастыру мереке және демалыс күндерін қоса алғанда, жүйе деректерді жинау және өңдеу алгоритмін іске қосады, кейіннен күнделікті инфляцияның прокси-көрсеткішін алуға мүмкіндік берді.

Жүйе алдын ала көрсетілген HTML-сілтемелері бойынша 13 дүкенде тұтыну себеті тауарының әрбір санатына бір атау бойынша азық-түлік және азық-түлікке жатпайтын тауарлардың 395 түріне күнделікті баға жинайды. Баға өзгерістерін тіркеу екі ірі қалада: Нұр-Сұлтан мен Алматыда жүргізіледі. Жүйенің бірінші нұсқасының тар географиялық көрінісі Қазақстанның ірі қалаларындағы интернет-алаңдардың шектеулі санымен және олардың өңірлерде болмауымен байланысты болды. Осыған байланысты, жиынтық индексті құру кезінде екі қаланың бірінде белгілі бір тауар бағасының өзгеруін жүйе тауардың осы санатына бағаның жалпы республикалық өзгерісі ретінде

⁹Тұтынушылық бағалар индексін құру әдіснамасы. <https://stat.gov.kz/api/getFile/?docId=ESTAT109961>

қабылдайды, яғни тауардың барлық бас жиынтығы үшін тиісті болып табылады деген болжам жасалады.

«Жалпы» жинау нәтижесінде алынған тауарлар бағасы бойынша деректер индекстерге түрлендіріледі және әрбір тауардың ҰСБ салмағына сәйкес жиынтық жуықталған ТБИ-ға топтастырылады. ТБИ күнделікті прокси-индикаторын жинақтау Galymzhan 1.0. жұмысының соңғы нәтижесі болады.

3.2. Galymzhan 2.0

Қазіргі кезде Galymzhan 2.0 тестік режимде жұмыс істейді және баға өзгерістерін жинау 19 азық-түлік санатынан тұратын әлеуметтік маңызы бар азық-түлік тауарларының тобы (бұдан әрі – ӘМАТ) бойынша жүргізіледі¹⁰. Galymzhan 2.0. Қазақстанның алты қаласында¹¹ орналасқан түрлі сауда маркаларының 16 ресми интернет-ресурстарын өңдейді және 5000-ға жуық жалпы-азық-түлік тауарларының бағалары бойынша күнделікті ақпарат жинайды. Орташа алғанда, тұтыну қоржыны тауарларының бір санатына шамамен 263 жалпы тауар тиесілі, ал жүйенің бірінші нұсқасында тұтыну қоржыны тауарларының бір санатына бір жалпы тауар тиесілі болды. Жалпы тауарлардың көбірек санын пайдалану іріктеудің өзгешелігін арттыруға және сонымен қоса индекстің құбылмалылығын төмендетуге мүмкіндік берді.

ҰСБ әдіснамасын толық сақтау жыл ішінде алдын-ала таңдалған тауарлардың бағаларын жинауды көздейді. Алайда, физикалық тексеруден айырмашылығы, интернетте баға жинаудың бірқатар ерекшеліктері бар. Осы ерекшеліктердің бірі тауарларға сілтемелердің жиі өзгеруін жатқызуға болады. Осыған байланысты, бақыланатын тауарлар мен оларға сілтемелерді тіркеу жиілігін арттыру туралы шешім қабылданды. Мәселен, екі апта сайын жүйе барлық өнімдерді және олардың қолда бар метадеректерін жинап, дүкендерді толық «сканирлейді». Бұл кезеңде тауарларға барлық сілтемелер, олардың санаттарға жататындығы, дүкен атауы және оның орналасқан жері жазылады. Жалпы 140 000 тауар бойынша ақпарат жиналады. Содан кейін жүйе ҰСБ-ның санаттарға бөлуін қолдана отырып, өнімдерді сыныптайды және қалыпты тілді өңдеудің қарапайым алгоритмдерін қолдану арқылы қажетті тауарларды іріктейді. Іріктелген тауарлардың сипаттамалары жеке базаға жазылады.

Бұдан әрі келесі екі апта ішінде өнімдерді іріктеу өзгеріссіз қалады және осы іріктемеге сүйене отырып, инфляцияның прокси-көрсеткіші есептеледі. Екі аптадан кейін процесс қайталанады. Яғни, біз қолданатын рәсім уақыт шығынын азайта отырып, қажетті өкіл тауарларды және олардың сипаттамаларын анықтау кезеңін автоматтандырады. Бұл тәсіл сонымен қатар, пайдаланылатын компьютердің өңдеу қуатына жүктемені азайтуға және күнделікті жинақ жүргізуге мүмкіндік береді.

¹⁰ Қазақстан Республикасы Үкіметінің 2017 жылғы 27 наурыздағы № 137 қаулысына сәйкес: ұн, нан, кеспе, қарақұмық, күріш, картоп, сәбіз, жасыл пияз, орамжапырақ, қант, сұйық май, сиыр еті, тауық еті, сүт, айран, сары май, жұмыртқа, тұз, ірімшік. <http://adilet.zan.kz/rus/docs/P1700000137#2>

¹¹ Нұр-Сұлтан, Алматы, Шымкент, Қарағанды, Павлодар, Өскемен

Веб-скрепинг процесі бүкіл html-бетін бағдарламалау ортасына¹² жүктеуден басталады. Алайда, бұл үшін ақпарат жиналатын сайттың түрін анықтап алу қажет. Сондықтан, осы кезеңде сайттың түрі анықталады. Сайттың екі негізгі түрі бар – динамикалық және статикалық. Бірінші, динамикалық түрде сайт деректеріне тек сұрату бойынша қол жеткізуге болады және олар пайдаланушының әрекеттеріне байланысты динамикалық түрде жылжып отырады. Мұндай сайттағы сұрату сайттың бастапқы жүктелуі, веб-бетті «айналдыру», веб-беттің белгілі бір элементтерін курсормен басу және т.б. болуы мүмкін. Динамикалық веб-сайттарды сканерлеу үшін пайдаланушының іс-әрекеттерін жасау қажет болғандықтан, мұнда html-файлын сұратуы бойынша қайтарылатын қарапайым кітапханалар жарамайды. Осындай беттермен жұмыс істеу үшін браузер қолданатын және пайдаланушының іс-іәрекеттерін «ойната» алатын Selenium (RSelenium) сияқты үшінші тарап қосалқы жүйелері қажет. Екінші жағдайда, егер сайт түрі статикалық болса, элементтер пайдаланушының әрекеттеріне қарамастан алдын-ала белгіленген жерлерде болады. Бұл түрдегі беттер веб-скрепингке оңай түсетіні анық, өйткені олар қолданушының с-әрекетін салуды қажет етпейді. Мұндай сайттардың көмегімен rvest сияқты веб-бетімен жұмыс істеу үшін кіріктірілген R кітапханаларын пайдалануға болады, оның артықшылығы жұмыс жылдамдығы және баптау мен пайдаланудағы қарапайымдылық. Әрбір веб-сайт бірегей болып табылады. Сондықтан Galymzhan 2.0 жүйесіндегі әрбір жеке сайтқа деректерді алу үшін әртүрлі сценарийлер бағдарламаланған.

Веб-сайттан деректерді алу үшін XPath-жолдары деп аталатын бет элементтері тобының бірегей сәйкестендіргіштері қолданылады. Осындай «жолдардың» көмегімен бағаларға, атауларға және т.б. қатысты барлық элементтерді бөліп көрсетуге болады.

Алынған мәліметтерді өңдеу барысы барлық өкіл тауарларға күнделікті бағаның өзгеруін анықтаудан басталады. Бұдан әрі әрбір өңірдегі тұтыну себетінің бір тауарына жататын өкіл тауарларға бағаның өзгеруі жалпы орташа геометриялық көрсеткішті есептеу арқылы топтастырылады. Содан кейін алынған баға өзгерістері өңірлер мен республика бойынша ӘМАТ инфляциясына, сондай-ақ барлық 19 тауарға бағалардың жалпы республикалық өзгеруіне біріктіріледі.

Күнделікті прокси-инфляциядан басқа, жүйе қосымша екі көрсеткішті – апта сайынғы инфляцияны және жинақталған айлық инфляцияны есептейді. Апта сайын ӘМАТ-та жарияланған инфляциямен салыстыру үшін күнделікті прокси-инфляция көрсеткіші алдыңғы күндізгі өзгерістерді жинақтау арқылы апталық жиілік көрсеткішіне айналады. Жинақталған айлық көрсеткіш қосымша талдау және айдың қорытындысы бойынша ресми деректермен салыстыру үшін пайдаланылады. Алынған индекстерді талдау инфляцияның ағымдағы болжамдарын түзетуге, сондай-ақ Қазақстандағы инфляциялық процестерді тереңірек зерттеуге және талдауға мүмкіндік береді.

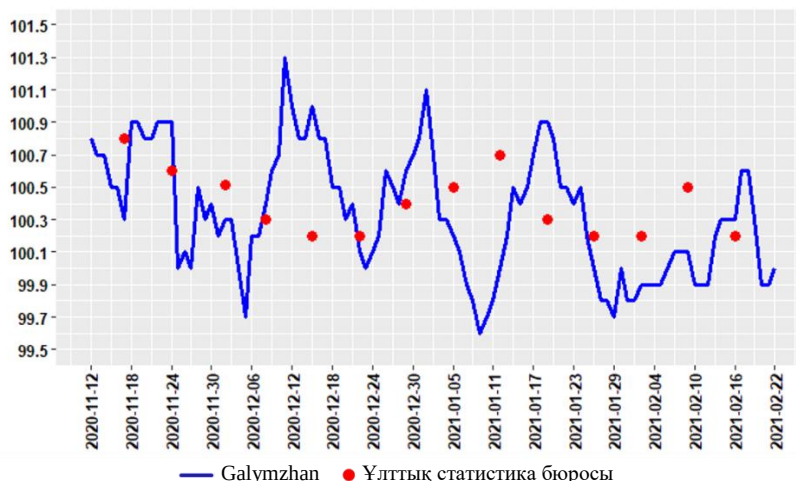
¹² Бұл жағдайда R бағдарламалау және статистикалық есептеу ортасы

Алынған нәтижелер Tableau – барлық аймақтардағы тауарлардың 19 санатының әрқайсысының динамикасын мүмкіндігінше тез және ыңғайлы талдауға мүмкіндік беретін интерактивті талдау жүйесінде дашбордтар түрінде көрініс табады.

4. Нәтижелерді талқылау

Galymzhan жүйесі қызметінің соңғы нәтижесі инфляцияның біріктірілген прокси-индикаторын, аталған жағдайда ЭМАТ инфляциясын құру болып табылады. Galymzhan жүйесінің ағымдағы нәтижелерін көрсету үшін 1-суретте БНС жариялайтын 19 әлеуметтік маңызы бар тауарлардың ресми апталық баға динамикасы және веб-скрепинг технологиясы алған деректер негізінде есептелген апталық прокси-инфляция көрсетілген.

1-сурет. Апта алдындағы тиісті күнге ЭМАТ-қа баға индексінің және прокси-инфляцияның %-бен серпіні



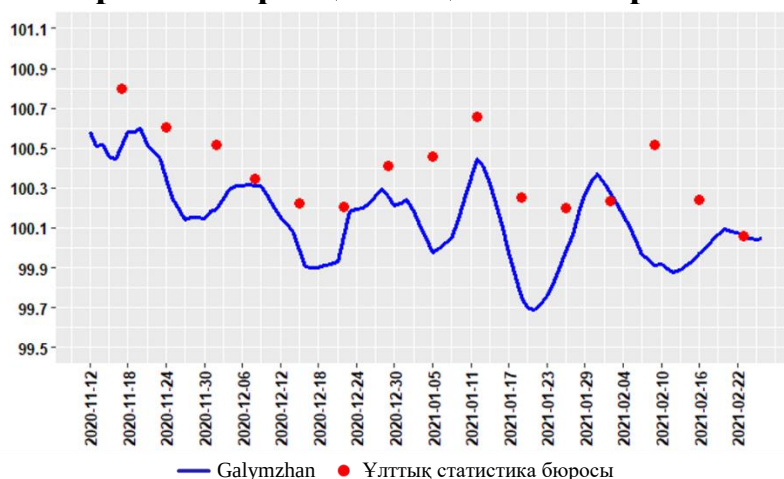
Дереккөзі: ҰСБ, автордың есептеулері

Жалпы, суреттен көрініп тұрғандай, прокси-инфляцияның есептік мәндерінің жоғары көлеміне қарамастан, ол көп жағдайда әлеуметтік маңызы бар тауарлардың нақты апталық инфляциясына сай келеді.

Талдау барысында онлайн-дүкендердегі деректер классикалық зерттеу әдісімен белгіленген бағалармен салыстырғанда тезірек өзгеруге бейім екендігі анықталды, бұл тауарлардың жоғары айналымымен және баға белгілерін жаңартуға нақты шығындардың болмауымен байланысты болуы мүмкін. Макроэкономикадағы ұқсас құбылыс «мәзір эффектісі» деп аталады. Бұл ретте, онлайн-бағаның өсу қарқыны көп жағдайда офлайн-бағаның өсу қарқынынан төмен болады. Бұл факті меншікті жеке онлайн-дүкеннің болуы негізінен ірі азық-түлік желілерінің құзыреті болып табылатынына байланысты болуы мүмкін, мұнда түпкілікті өнімдердің құны ауқым әсерінің есебінен жекелеген сауда нүктелерімен салыстырғанда төмен болуы мүмкін. Онлайн-бағалар мен классикалық зерттеу әдісімен белгіленетін бағалар арасында уақытша артта қалудың болуы Galymzhan 2.0 жүйесімен есептелген көрсеткішті әлеуметтік

маңызы бар өнімдерге бағалардың өзгеруі туралы ресми деректердің озыңқы индикаторы ретінде пайдалануға негіз береді. 2-суретте 10 күндік кешуілдеумен алынған апта сайынғы прокси-инфляция және әлеуметтік маңызы бар азық-түлік тауарларына баға индексі көрсетілген. Сонымен қатар, визуалды шуды азайту үшін онлайн индексі жеті күндік жылжымалы орташа деңгеймен тегістелді.

2-сурет. Апта алдындағы тиісті күнге ӘМАТ-қа баға индексінің және прокси-инфляцияның* %-бен серпіні



* Galymzhan жүйесінің деректері 7 күндік МА-мен тегістеліп, көкөністердің бағасын ескерместен 10 күндік лагпен алынады
Дереккөзі: ҰСБ, автордың есептеулері

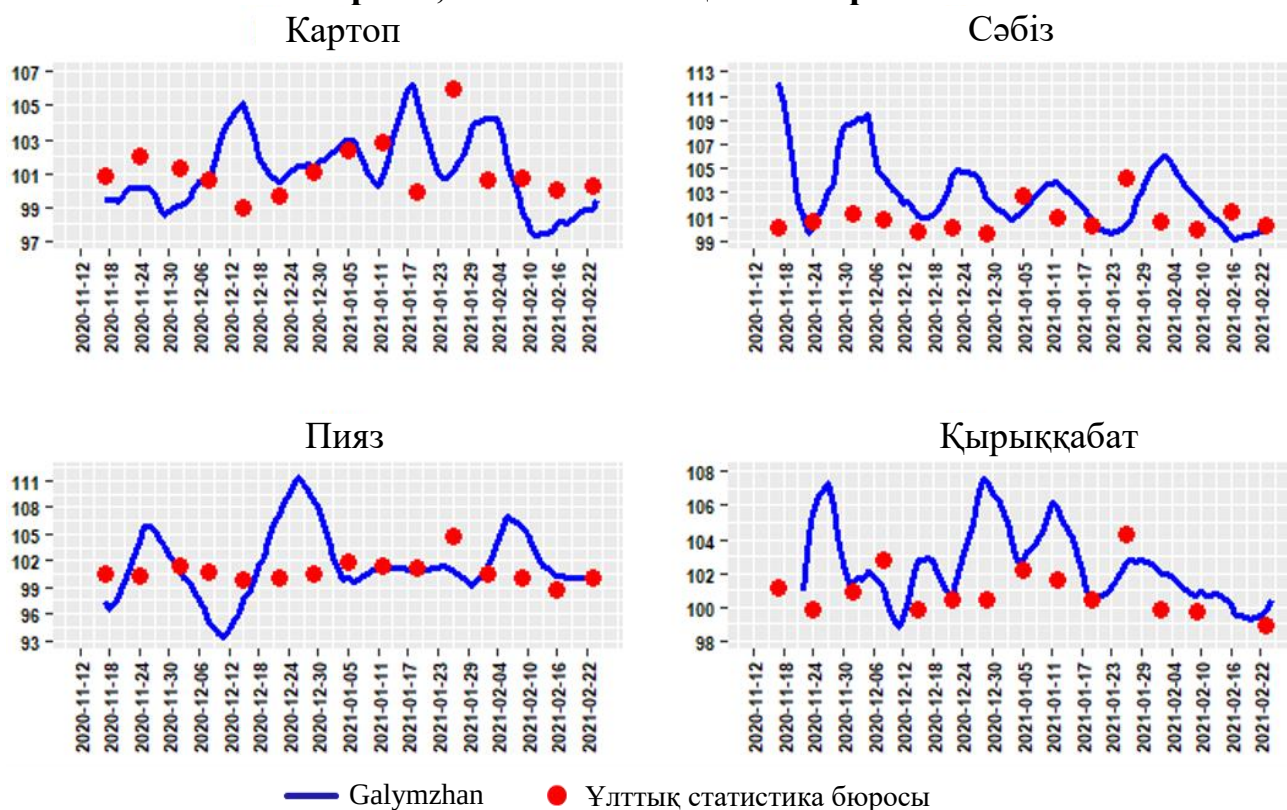
Сонымен қатар, бірқатар тауарлар бойынша, атап айтқанда көкөніс өнімдері бойынша ресми статистикадан біршама сәйкессіздік байқалады. Бұл факт онлайн-дүкендерде бірегей өкіл тауарлардың аз ұсынылуымен байланысты (3-сурет).

Әрі қарай, жүйе кеңейген сайын (жаңа дүкендер қосылып, іріктеу жоғарылайды) ҰБС деректері мен Galymzhan 2.0 арасындағы дифференциал төмендейді деп күтілуде.

Өңірлік бөліністің мысалы ретінде онлайн-дүкендер саны ең көп қала ретінде Нұрсұлтан және Алматы қалалары таңдалды (4 және 8 онлайн-алаңдар, тиісінше). Бұл ретте, Нұрсұлтан қаласында Galymzhan жүйесінің деректері мен ҰБС ресми көрсеткіштері арасында қатты тәуелділік байқалады (4-сурет).

Нұрсұлтан қаласына қарағанда, Алматы қаласында жоғары құбылмалылық байқалады. Мәселен, жаңа жыл қарсаңында (2020 жылғы 26-27 желтоқсан) көптеген азық-түлік өнімдеріне бағаның күрт өсуі тіркелді. Бұл бағаның өсуінің түсіндірмесі жаңа жылдық мерекелерге байланысты сұраныстың артуы болып табылады. Бұл ретте, бағаның өсуі ірі сауда желілерінде байқалады, ал жергілікті нарықтардағы бағалар іс жүзінде өзгерген жоқ. Осылайша, бұл жүйенің артықшылығы нарықтар құбылып тұрған кезеңде баға серпініне терең талдау жүргізу мүмкіндігі болып табылады. Сонымен, қазіргі уақытта Galymzhan түрлі сауда нүктелерінің сөрелеріндегі белгілі бір тауар түрлерінің баға серпінін бақылауға мүмкіндік береді.

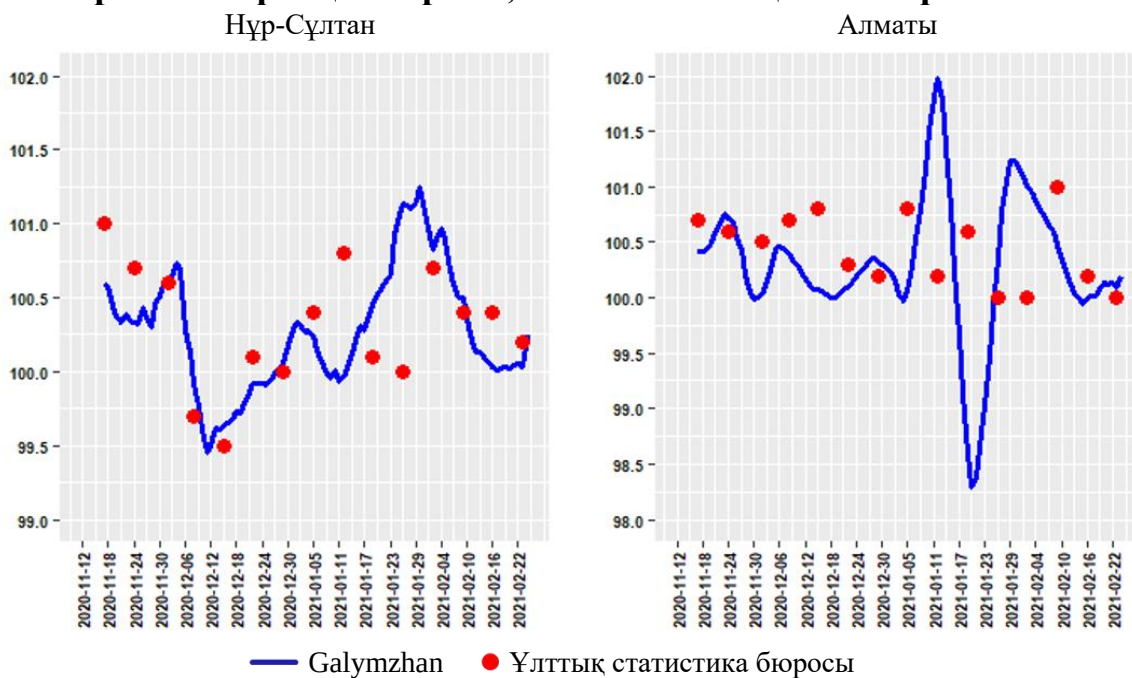
3-сурет. ӘМАТ баға индексінің және көкөністерге прокси-инфляция серпіні, өткен аптаның тиісті күніне %



Galymzhan жүйесінің деректері 7 күндік МА-мен деңгейлестірілді және 10 күндік лагпен алынды

Дереккөзі: БНС, автордың есептеулері

4-сурет. ӘМАТ баға индексінің және Нұр-Сұлтандағы және Алматыдағы прокси-инфляция серпіні, өткен аптаның тиісті күніне %



Galymzhan жүйесінің деректері 7 күндік МА-мен деңгейлестірілді және 10 күндік лагпен алынды

Дереккөзі: ҰБС, автордың есептеулері

Болашақта жүйе жақсарған сайын, атап айтқанда географиялық көріністің кеңеюі, жаңа тауар түрлерін қосу және тұтыну себеті тауарлары санаттарының жалпы санын ұлғайту кезінде Galymzhan жүйесі есептейтін индекс пен ресми деректер арасындағы конвергенция төмендейді деп күтілуде, ал кеңейтілген индекс тауар және өңірлік көрініс бойынша тұтыну бағаларының индексіне барынша жақын болады және қысқа мерзімді болжамдарды құру кезінде, ақша-кредит саясаты бойынша талдау және шешімдер қабылдау мақсаттары үшін пайдаланылатын болады.

5. Бұдан былайғы зерттеуге арналған қорытындылар мен ұсынымдар

Ағымдағы зерттеу шеңберінде ресми статистикалық органдарда да, сондай-ақ жеке ұйымдарда да инфляцияны құру кезінде веб-скрепинг технологиясын пайдаланудың әлемдік тәжірибесі зерделенді. Бұдан басқа, Galymzhan 2.0 жүйесі Қазақстанда әлеуметтік маңызы бар азық-түлік тауарларына инфляцияның прокси-көрсеткішін жинау және одан әрі құру үшін пайдаланатын әдіснаманың сипаттамасы келтірілген.

Алынған нәтижелер ресми деректер мен құрылған индекс арасындағы корреляцияның болуы туралы көрсетеді. Бұл ретте, онлайн-дүкендердегі бағалар офлайн сауда нүктелеріне қарағанда тезірек өзгереді және нақты дүкендердегі бағаның өсу қарқыны біршама жоғары екендігі айтылды.

Айта кету керек, Galymzhan 2.0 жүйесі қазіргі уақытта тестілік режимде жұмыс істейді. Жүйені одан әрі дамыту мақсатында авторлар географиялық қамтуды кеңейту және көптеген дүкендерді, тұтыну себеті тауарларының санаттарын және тауарлардың түрлерін қосу бойынша жұмыстар жүргізуді жоспарлап отыр.

Визуализация процесінің одан әрі даму әлеуеті тек аймақтар мен дүкендерде ғана емес, сонымен қатар белгілі бір дүкендердің интернет-сөрелеріндегі тауар түрлерінің нақты өнімдерін бақылау мүмкіндігін бақылау дашбордына қосумен байланысты.

Жүйені одан әрі кеңейту географиялық және тауарлық қамту бойынша тұтыну бағаларының индексіне барынша жақындауға мүмкіндік береді, бұл оны толық көлемде көшірмелеуге, сондай-ақ Galymzhan 2.0 жүйесінің нәтижелерін қысқа мерзімді болжамдарды құру және монетарлық саясатты талдау кезінде құрал ретінде пайдалануға мүмкіндік береді.

Әдебиеттер тізімі

1. A Guide to Data Integration for Official Statistics, *UNECE High Level Group for the Modernisation of Official Statistics Data Integration Project*, January 2017;
2. Alan Bentley, Frances Krsinich, Towards a big data CPI for New Zealand, *Ottawa Group*, 2017;
3. Alberto Cavallo, Roberto Rigobon, The Billion Prices Project: Using Online Prices for Measurement and Research, *Journal of Economic Perspectives*, Spring 2016;
4. Alberto Cavallo, Scraped Data and Sticky Prices, *MIT Sloan School of Management*, April 2012;
5. Antonio G. Chessa, A new methodology for processing scanner data in the Dutch CPI, EURONA — *Eurostat Review on National Accounts and Macroeconomic Indicators*, January 2016;
6. Antonio G. Chessa, Processing scanner data in the Dutch CPI: A new methodology and first experiences, *Statistics Netherlands*;
7. Antonio G. Chessa, Robert Griffioen, Comparing Price Indices of Clothing and Footwear for Scanner Data and Web Scraped Data, *Economics and Statistics*, April 2019;
8. Australian Bureau of Statistics (ABS), Maximizing the use of Web Scraped Data in the Australian CPI;
9. Ben Powell, Guy Nason, Duncan Elliott, Matthew Mayhew, Jennifer Davies, Joe Winton, Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting, *Journal of the Royal Statistical Society*, 2017;
10. Christian Blaudow, Florian Burg, Dynamic pricing as a challenge for consumer price statistics, EURONA — *Eurostat Review on National Accounts and Macroeconomic Indicators*, March 2018;
11. Federal Statistical Office of Germany, *Your Benefit. Our Mission.*, 2018;
12. Federico Polidoro, Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation, *Statistical Journal of the IAOS* 31 (2015) 165–176;
13. Heymerik van der Grient, Jan de Haan, The use of supermarket scanner data in the Dutch CPI, *Statistics Netherlands*, July 2010;
14. Janine Boshoff, Xuxin Mao, Garry Young, Outlier detection methodologies for alternative data sources, *ESCoE Technical Report 07*, July 2020;
15. Karola Brunner, Automated price collection via the internet, *Wirtschaft und Statistik*, April 2014;
16. Ken Van Loon, Dorien Roels, Integrating big data in the Belgian CPI, *Meeting of the Group of Experts on Consumer Price Indices Geneva, Switzerland*, May 2018;
17. Ken Van Loon, Dorien Roels, Web scraping and online data collection and processing for the consumer price index, *Belgian statistical office*, February 2018;

18. Lincoln T. da Silva, Ingrid L. de Oliveira, Tiago M. Dantas, Vladimir G. Miranda, *Studies of new data sources and techniques to improve CPI compilation in Brazil*, Brazilian Institute of Geography and Statistics;
19. Matthew Mayhew, Comparison of index number methodology used on UK web scraped price data, *Office for National Statistics*;
20. Robert Breton, Tanya Flower, Matthew Mayhew, Elizabeth Metcalfe, Natasha Milliken, Christopher Payne, Thomas Smith, Joe Winton, Ainslie Woods, Research indices using web scraped data, *Office for National Statistics*, May 2016;
21. Robert Griffioen, Olav ten Bosch, On the use of internet data for the Dutch CPI, *Meeting of the Group of Experts on Consumer Price Indices Geneva, Switzerland*, May 2016;
22. Австралияның статистика бюросының ресми сайты, қол жеткізу режимі: <https://www.abs.gov.au/articles/web-scraping-australian-cpi>.