NATIONAL BANK OF KAZAKHSTAN

Galymzhan System: Online Assessment of Consumer Inflation in Kazakhstan

Monetary Policy Department
Working Paper No.2021-2

Tuleuov O.,
Yerzhan I.,
Seidakhmetov A.

The Working Paper and Analytical Note series of the National Bank of the Republic of Kazakhstan ("the NBRK") are intended to disseminate the deliverables of the National Bank's research work as well as other studies of the NBRK staff. The results of research activity are disseminated with a view to encourage discussions. Opinions delivered in the Paper express the author's personal attitude and may not coincide with the NBRK's official position.

Galymzhan System: Online Assessment of Consumer Inflation in Kazakhstan.
NBRK – WP – 2021 – 2

Galymzhan System: Online Assessment of Consumer Inflation in Kazakhstan

Tuleuov Olzhas[1]
Yerzhan Islam[2]
Seidakhmetov Ansar[3]

Abstract

It is critical for the central banks that adhere to the inflation targeting policy to monitor and analyze the existing inflationary trends. The presence of up-to-date information about the price changes enables to effectively respond to various shocks and, where necessary, to adjust the monetary policy in a timely fashion.

At present, a limited frequency of publications of the official pricing statistics serves as a constraining factor for the real-time analysis of change in prices of goods in the consumer basket. Hence, the designing of indirect high-frequency proxy indicators of inflation comes to the fore.

This Paper describes the methodology and outcomes of designing a high-frequency inflation proxy in Kazakhstan at the National Bank by using the web scrapping technology, which implies the automated data generation by deriving the data from web pages implemented with a help of a software algorithm.

Key words: inflation, web scrapping, Galymzhan, online store, prices, big data, CPI, parsing, consumer goods.

JEL classification: E31, E37, E39.

[1] Tuleuov Olzhas is the Advisor to the Governor of the National Bank of the Republic of Kazakhstan.
E-mail: Olzhas.Tuleuov@nationalbank.kz
[2] Yerzhan Islam is the chief specialist-analyst, Division of Macroeconomic Research and Forecasting, Monetary Policy Department of the National Bank of the Republic of Kazakhstan.
E-mail: Islam.Yerzhan@nationalbank.kz
[3] Seidakhmetov Ansar is the leading specialist-analyst, Division of Macroeconomic Research and Forecasting, Monetary Policy Department of the National Bank of the Republic of Kazakhstan.
E-mail: Ansar.Seidakhmetov@nationalbank.kz

Contents

1. Preamble

Inflation as an indicator of the weighted average growth in prices of goods in the consumer basket is a fundamental indicator in the decision-making by the central banks pursuing the inflation targeting policy. Steadily low inflation rates promote the widening of planning horizon and the growth of investments thus expanding the economic activity and improving the people's welfare.

Implementation of a consistent monetary policy helps maintaining the steadily low rates of growth in prices for goods in the consumer basket.

In the periods of various crises, the monetary regulator, in order to curb inflationary processes, needs to effectively react to the existing and expected macro-processes as well as to adjust the monetary policy in a timely fashion. The existence of up-to-date information enables to make better quality decisions.

At present, a limited frequency of official statistics publications serves as a constraining factor for the real-time analysis of change in prices of goods in the consumer basket in Kazakhstan. Hence, the designing of indirect high-frequency proxy indicators of inflation comes to the fore.

In the recent years, the systems of collection of information about price changes based on price scanning and web scrapping of the Internet resources that allow obtaining high-frequency information in bulk become especially popular among statistics authorities and central banks of many countries.

In broad terms, the price scanning implies the use of retailer data about transactions where information regarding the price and turnover of goods identified based on bar codes is contained. The use of this technology implies close collaboration with stores that provide such information (G. Chessa, 2016).

The web scrapping technology of various web sites allows collecting the data on pricing in the structured form for their subsequent processing and generation of high-frequency information about the change in consumer prices. The technology gained its momentum due to the fact that online purchases become increasingly popular and many vendors turn to online sales opening their official web sites with prices for all traded products.

This type of data collection about the price changes is gaining an increasing momentum both at the government authorities (statistics authorities, central banks) and in private companies of different countries, due to the economy of time, human resource saving and minimization of expenses. Generation of data about the change in price levels by way of the web scrapping became especially popular in 2020, when, against the backdrop of COVID-19 pandemic, many sales outlets had to suspend their operation because of quarantine restrictions that complicated the collection of data regarding the price changes with the use of a classical method.

The National Bank of Kazakhstan proceeded with the study of data collection from the Internet resources in 2018, when the system of data collection and analysis of the consumer price behavior named Galymzhan was developed.

From a technical standpoint, the Galymzhan system represents an algorithm written in the R programing language. Integration of this code into the Windows system allowed attaining autonomy of the program. Thus, every day, including holidays and

weekends, the system contemporaneously launches the algorithm of data collection and processing. Generation of the daily figures of inflation proxy is an outcome of the system's operation.

This Working Paper consists of several parts. The first section presents an overview of references where similar papers of other authors are reviewed. The second section describes the methodology for collection of price changes and generation of the final index. Further, there is a section with the discussion of outcomes where the authors compare the resulting index with the official statistical data. The findings of this study are put together in the final section.

2. Overview of References and International Experience

The increasing number of various Internet resources coupled with the development of technologies that enable to derive the required information from open sources in a structured form gave a strong impetus to the progress in many spheres of life. The availability of information that was previously inaccessible allowed automating and optimizing the business processes.

The process of obtaining statistical information, particularly inflation proxy, by using the web scrapping technology has lately become rather popular. This method, as compared to a classical price survey conducted by official statistics authorities enables to capture price changes with a higher frequency and to release the human resource, thus providing the statistics authorities with a possibility to increase the quality of published data and their frequency.

Nowadays, the statistics authorities in such countries as Italy, Australia, Belgium, Germany, the Netherlands, Hungary, the UK and Canada started to adopt and implement the systems of web scrapping in their operations (Boshoff, 2020). So, the Italian National Institute of Statistics is making an active effort to implement the system of automated price monitoring (Polidoro, 2015). Specifically, the group of statisticians and IT-specialists tested the possibility of using the data derivation methods from web pages in the consumer price surveys; they focused on two groups of goods: household appliances and air tickets. The testing results proved the existence of a huge potential for using the automated price data collection in the Internet in terms of increasing effectiveness and quality of statistical data.

First of all, the use of parsing methods as a tool of data generation for measuring inflation is directly connected with one of the three Big Data components, namely with Velocity[4]. It is worth mentioning that the staff of the Institute of Statistics have already been using prices derived from web resources; however, such prices were captured by the staff manually (in the "copy and paste" mode).

[4] It is generally believed that big data can be explained with the help of three V: Velocity, Variety and Volume. In his research report of 2001, a META Group (now – Gartner) analyst Doug Laney described big data as three-dimensional, that is the volume of data is more and more massive, velocity (velocity of data collection and processing) and variety (the range of data types and their sources). Later, in 2012, Gartner updated the definition of big data as the data of massive volume, high velocity and a great variety.

Another statistical authority that uses the derivation technologies for the data on price changes from web-resources is the Australian Bureau of Statistics. Since the middle of 2017, the Bureau of Statistics has been gradually posting into the database used in calculation of the consumer price index (CPI) the indicators generated as a result of the web scrapping[5]. Before doing so, the prices derived from the Internet had been compared with the prices collected via a classical method during the last 12 months in order to be certain that they coincide and are fit for measuring the price dynamics. After that, prices for certain goods that were collected manually had been replaced with the data derived from the Internet, and the manual data collection for such products was terminated. The data collected in such a way are mainly used to build the alcohol price index, prices for a small quantity of clothes and car spare parts accounting for about 5% of CPI. This approach has a number of advantages. The average price on the basis of data collected twice a week is more representative than the price collected once a month. This also allows including more products into the CPI basket and reducing the costs of collection. At present, prices of 500 000 representative goods are being collected on a weekly basis whereas under traditional methods of collection only about 1000 positions had been collected[6].

The Federal Statistical Office of Germany in 2017 started the online price data collection. At the pilot stage, collection was conducted among 242 goods in the consumer basket, excluding prices of services. There are about 600 goods and services in the consumer basket of Germany. The web scrapping technology allowed collecting 3050 prices every hour. Such a high frequency is explained by the fact that at the initial stage this effort, in addition to an applicable nature, had also a research focus (Blaudow and Burg, 2018). As a result of a high-frequency collection of prices, the authors managed to reveal a detailed picture of the price dynamics in online stores within 24 hours. So, a large portion of online stores make price adjustments at the beginning of the day (00:00).

The system of automatic price collection of the Statistical Office of Germany is written in the Java programming language. The web pages were parsed with the help of Selenium tool, which operates the web browser by reproducing actions in the browser that were specified in the code. In order to prevent from being blocked by the Internet resources, the project uses dynamic IP addresses.

At present, when calculating the inflation, the online data about the change in prices of 10 000 representative goods are used on such groups as foodstuffs, non-alcohol beverages and tobacco products (Boshoff, 2020).

The Belgian Statistical Office has been also collecting data from online stores over the recent years. The system is written in the R programming language and uses such packages as rvest[7] and Rselenium[8]. The obtained data has already been used in calculating the inflation of international travel and video games.

---

[5] Official web-site of the Australian Bureau of Statistics, available at: 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.

[7] The library which allows the information gathering from web pages.

[8] A package which enables to use Selenium 2.0 WebDriver in R for a web-browser automation.

In addition, about 70 parsers collect data on prices of clothes, footwear, air tickets, used cars, medications, books and household appliances, etc. on a daily basis or several times a week. In the near future, the Belgian Statistical Office is planning to include the data on the above goods obtained with the help of web scrapping into the calculation of the officially published consumer price index (Loon, Roels, 2018).

Statistics Netherlands is currently collecting the data on prices from the Internet resources of three large retailers by capturing changes in prices of 100 000 representative goods on a daily basis. In doing so, they capture price changes only in respect of garments. This is explained by two fundamental reasons. First, since 2010, Statistics Netherlands have been capturing the change in prices of a part of food and non-food products via the price scanning technology and currently garments account for a large portion of data collection in offline stores (Grient, Haan, 2010). Second, garments are the kind of goods that do not require a frequent adjustment of the sample of goods thus simplifying the process automation. Now, 16 robots are collecting data at 19 retail stores in a test mode. After their validation, these data will be included into the main sample, which allows expanding it to 500 000 daily observations (Griffioen, Bosch, 2016).

In addition to official statistics authorities, the use of data generated with the help of the web scrapping technology for constructing a CPI proxy has gone mainstream also among non-government institutions. Provided there are resources and relevant skills, the building of an indirect CPI becomes a realistic objective also for individual researchers. Polidoro in his study wrote that in future, given a wide access to tools and information, statistics institutions may lose their competitiveness and they may probably lapse their monopoly, which is now supported by their exclusive right for collection and dissemination of statistical information at the official level (Polidoro, 2015).

The examples of dissemination of price statistics by unofficial authorities include such projects as “The Billion Prices Project” and “PriceStats”. “The Billion Prices Project” is the academic initiative of American economists Alberto Cavallo and Roberto Rigobon; as part of the initiative, since 2008 prices from hundreds of retail stores all over the world have been collected on a daily basis followed up by the real time presentation of inflation assessment. An impetus to creation of the system had been the manipulation of inflation statistics in Argentina from 2007 to 2015. During this period, it became apparent for many people that the official inflation rate that is calculated by the National Statistical Office of Argentina does not reflect the real price changes. According to the official data of the Institute of Statistics, the average annual inflation rate from 2007 to 2015 accounted for 8%, whereas the average figure based on the online data exceeded 20%, which was in line with assessments in some regions and inflation expectations among households. A possibility to collect price data online beyond the country appeared to be especially useful in 2011, when the Government of Argentina started to impose penalties and to force local experts to stop the data collection. At the end of 2015, after the change of government, manipulations of the official statistics ceased.

Nowadays, “The Billion Prices Project” every day is capturing prices for about 5 million goods sold in 300 online stores, in more than 50 countries (Cavallo, 2016).

In 2011, a commercial project entitled “PriceStats” was created on the basis of “The Billion Prices Project”, which collects the data and, by using such data, constructs high-frequency indices for central banks and the financial sector. In the “PriceStats”, the quantity and quality of used data was increased significantly. Now, the company captures prices of about 15 million goods from more than 1200 retailers with a view to further build a daily inflation index in 23 countries and the producer price index in 10 countries. In total, the data for 60 countries are collected.

The review of international experience in the use of the web scrapping technology for constructing an indirect consumer price index has demonstrated viability and the increasing interest in this method of price data collection both among government institutions and the private sector. The growing number of online stores and an ongoing sophistication of available tools for derivation of information, prices from online stores in particular, may become the main source of statistical information in the near future.

3. The Data and the Methodology Used

The process of development and buildup at the National Bank of Kazakhstan of the Galymzhan system as a tool used to analyze the current situation in the consumer price market by constructing an inflation proxy may be divided into two periods. From the beginning of 2018 through October 2020, the first version of the system – Galymzhan 1.0 – was used. In October 2020, the system had been fine-tuned significantly. So, in the second version, the quantity of monitored goods increased, the geographic coverage expanded and the system’s autonomy strengthened.

A great number of newly-opened Internet sites not only in large cities but also in the regional centers of Kazakhstan that were triggered by the COVID-19 pandemic as well as the improvement and accumulation of data collection and processing skills by the analysts working on the Galymzhan system enabled to ensure a significant development in this area.

In addition to the fact that the web scrapping is quite a complex process in technical terms requiring a certain set of programming skills, a selection of approach to calculation of the index is also a non-trivial problem. A traditional collection of data regarding the price changes that is conducted by the Bureau of National Statistics with the Agency for Strategic Planning and Reforms of Kazakhstan (“the BNS”) is based on the target population, which implies the presence of a list of representative goods and services that was earlier selected based on consumption segments. This list includes goods and services consumed by households where weights are determined based on the percentage of their expenditures for purchasing such goods and services. Further, the price survey is conducted by inspectors locally in various sales outlets. The methodology applied by the BNS[9] and the Galymzhan system is based on the bottom-

[9]The Methodology for designing the consumer price index. https://stat.gov.kz/api/getFile/?docId=ESTAT109961

up approach, under which the first step is to identify goods in a centralized fashion and then the price data collection process itself takes place.

It should be noted that the purpose of the Galymzhan system is to build an inflation proxy and not to collect all available information regarding the price dynamics. The use of approach similar to that of the BNS helps retain the interpretability of generated indices and use them as a proxy assessment of the consumer inflation in Kazakhstan.

3.1 Galymzhan 1.0

Galymzhan 1.0., which was launched at the beginning of 2018, is a special algorithm written in the R programming language. Integration of the code into the Windows system allowed attaining that the system launches the algorithm of data collection and processing followed by generation of a daily inflation proxy every day, including weekends and holidays.

Every day, the system collects the data about prices of 395 types of food and non-food products, one name per each category of product in the consumer basket at 13 stores based on earlier indicated HTML-links. Price changes are monitored in two large cities: Nur-Sultan and Almaty. A narrow geographic coverage of the first version of the system was related to a limited number of Internet sites in large cities of Kazakhstan and their absence in the regions. To this end, when building an aggregate index an assumption is made that a change in the price of a certain product in one of the two cities will be perceived by the system as the nation-wide change in the price for this product category, which is representative for the entire population of this product.

The data obtained as a result of a "field" collection of product prices are transformed into indices and are grouped into the approximate aggregate CPI in line with the BNS weights of each product. The concluding result of running the Galymzhan 1.0. system is accumulation of the daily CPI proxy.

3.2 Galymzhan 2.0

At present, Galymzhan 2.0 functions in the test mode and the price changes are collected on the group of socially important foodstuffs ("CIFs), consisting of 19 food categories of products[10]. Galymzhan 2.0. processes 16 official Internet resources of various trademarks located in six towns of Kazakhstan[11] and collects daily information on prices of about 5000 representative foodstuffs. On average, there are 263 representative goods per one product category in the consumer basket, whereas in the

[10] According to the Government Decree of the Republic of Kazakhstan as dated March 27, 2017 No. 137: flour, bread, elbows, buckwheat, rice, potatoes, carrots, onions, cabbage, sugar, vegetable oil, beef, chicken, milk, kefir, butter, eggs, salt, cottage cheese. http://adilet.zan.kz/rus/docs/P1700000137#2

[11] Nur-Sultan, Almaty, Shymkent, Karaganda, Pavlodar, Ust-Kamenogorsk.

first version of the system there had been one representative product per one product category in the consumer basket. The use of a larger quantity of representative goods enabled to increase the sample's representation and to reduce the index volatility at the same time.

A full compliance with the BNS methodology implies the collection of prices for goods that were earlier selected during the year. However, as opposed to a physical survey, the price collection in the Internet has a number of specific features. A frequent change in links to products can be referred to as one of such specific features. In this connection, the decision was made to increase the frequency of capturing the monitored goods and their links. So, every two weeks, the system conducts a full "scanning" of stores with collection of all products and their meta data that are available. At this stage, all links to goods, their rating, the name of store and its location are recorded. In total, information about 140 000 goods is collected. Then the system conducts classification of products using the BNS categorization and selects necessary products by applying the simplest algorithms of natural language interpretation. Characteristics of selected goods are recorded in the separate database.

Further, within two weeks to follow, the sample of products remains unchanged and, basing on that sample, an inflation proxy is calculated. After two weeks, the procedure is repeated. That is, the procedure we use is automating the stage where the required representative goods and their characteristics are determined, thus reducing the time costs. Such approach also allows reducing the burden on the processing capacity of the computer used and collecting the data on a daily basis.

The web scrapping process itself begins with the uploading of the entire html-page to the programming environment[12]. However, a web site from which information will be gathered needs to be identified. Therefore, the type of site is identified at this stage. There are two main types of web sites – the dynamic and the static one. In the first, the dynamic type, the web site data are accessible only upon inquiry and are dynamically moving depending on actions taken by the user. An inquiry in such a web site could be a primary loading of the web site, scrolling a web page, clicking a cursor on certain elements of the web page, etc. Since in order to scan the dynamic web sites one needs to simulate the user actions, simple libraries that bring back an html-file upon inquiry are not appropriate. Running such pages requires external subsystems such as Selenium (RSelenium), which use a browser and are capable of "reproducing" the user actions. In the second case, however, if the web site is of the static type, elements are in the pre-determined places irrespective of the user actions. Such pages are apparently web scrapped easier since they do not require simulation of the user actions. With such sites, one can use the built-in R libraries to handle web pages such as rvest, where the speed of operation and simplicity in setting and application are advantages. Each web site is unique. Therefore, different scenarios of data derivations are programmed virtually for each individual web site in the Galymzhan 2.0. system.

In order to derive data from a web site, unique identifiers of the group of page elements are used that are called XPath. With the help of such "paths", all elements related to prices, names, etc. can be obtained.

[12] In this case, it's the R environment for programming support and statistical computing

The processing of obtained data starts with determination of a daily price change for all representative goods. Further, the change in prices of representative goods that belong to the same product in the consumer basket in each region are grouped by calculating the geometric mean. After that, the generated price changes are aggregated into inflation of SIFs by regions and in total for the country, including a nation-wide change in prices for the entire 19 goods.

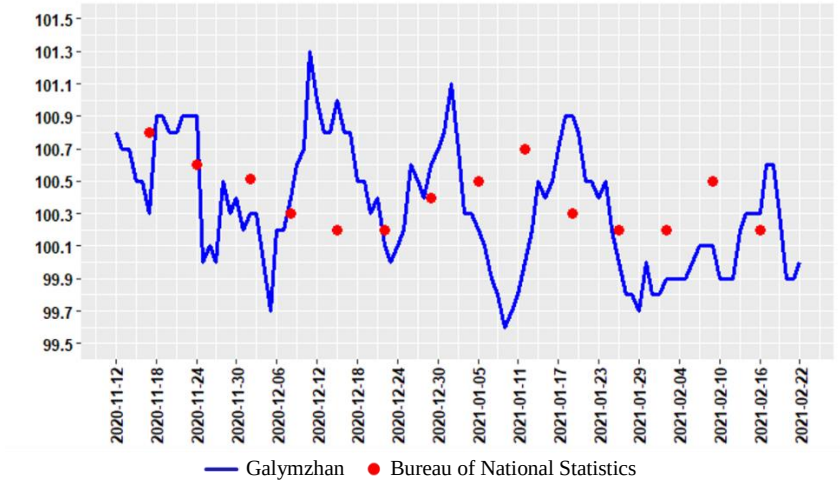
Apart from a daily inflation proxy, the system calculates two additional indicators – weekly inflation and accumulated monthly inflation. With a view to compare a weekly SIFs inflation published, a daily inflation proxy is transformed into a weekly frequency indicator through accumulation of previous daily changes. The accumulated monthly indicator is used for an additional analysis and comparison with the official data at the end of the month. The review of generated indices allows adjusting the current inflation forecasts and also to have a better insight into inflationary processes in Kazakhstan.

The obtained outcomes are visualized in the form of dashboards in Tableau – the system of interactive analytics, which enables to analyze the dynamics of each of 19 categories of goods in all regions as fast and convenient as possible.

4. Discussion of Outcomes

The concluding result of running the Galymzhan system is the building of an aggregate inflation proxy, SIFs in this case. To illustrate the current outcomes of the Galymzhan system, Figure 1 presents the official weekly dynamics of prices for 19 socially important goods published by the BNS, and a weekly inflation proxy calculated based on the data obtained through the web scrapping technology.

Figure 1. Dynamics of Price Indices for SIFs and Inflation Proxy, as % to the Corresponding Day of the Previous Week

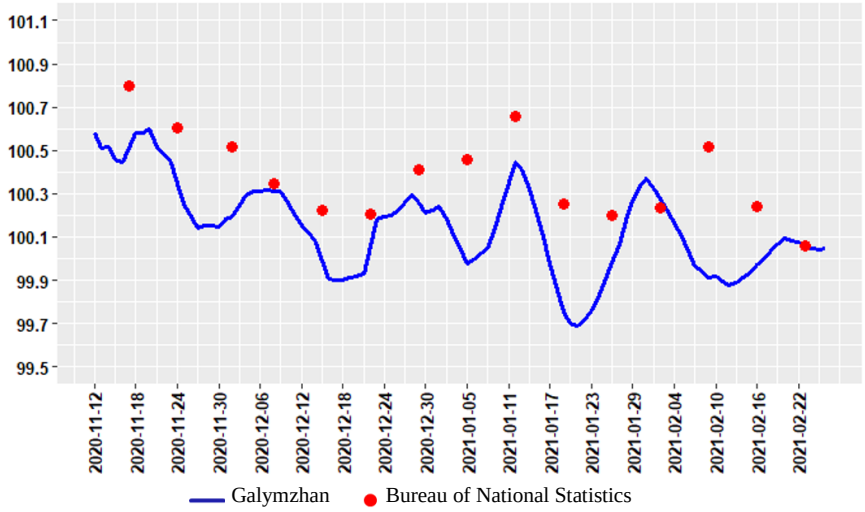


Source: BNS, derivations by the authors

As can be generally seen from the Figure, despite a wide range of computed values of inflation proxy, in most cases it converges with the actual weekly в of socially important goods.

In the course of the analysis, it was determined that data from online stores, in comparison with prices captured by the classical survey method, tend to change more rapidly, which is possibly due to the high turnover of goods and the absence of obvious costs for updating the price tag. In macroeconomics, this phenomenon is called the "menu effect". At the same time, the growth rate of online prices in most cases is lower than the growth rate of offline prices. This is probably due to the fact that having your own online store is mainly the prerogative of large food chains, where the cost of final products may be lower in comparison with single outlets due to economies of scale. The presence of a time lag between online prices and prices captured by the classical survey method provides grounds to use the indicator calculated by the Galymzhan 2.0 system as a leading indicator of official data on price changes for socially important products. Figure 2 shows the weekly inflation proxy taken with a 10-day lag and the price index for socially important foodstuffs. Additionally, to reduce visual noise, the online index was smoothed with a 7-day moving average.

Figure 2. Dynamics of Price Indices for SIFs and Inflation Proxy, *, as % to the Corresponding Day of the Previous Week



* Data from Galymzhan System are smoothed by a 7-day MA and are taken with a 10-day lag excluding the prices of vegetables
Source: BNS, derivations by the authors

At the same time, there is some discrepancy with the official statistics for a number of goods, in particular for vegetable production. This fact is associated with the low representation of unique representative products in online stores (Figure 3).

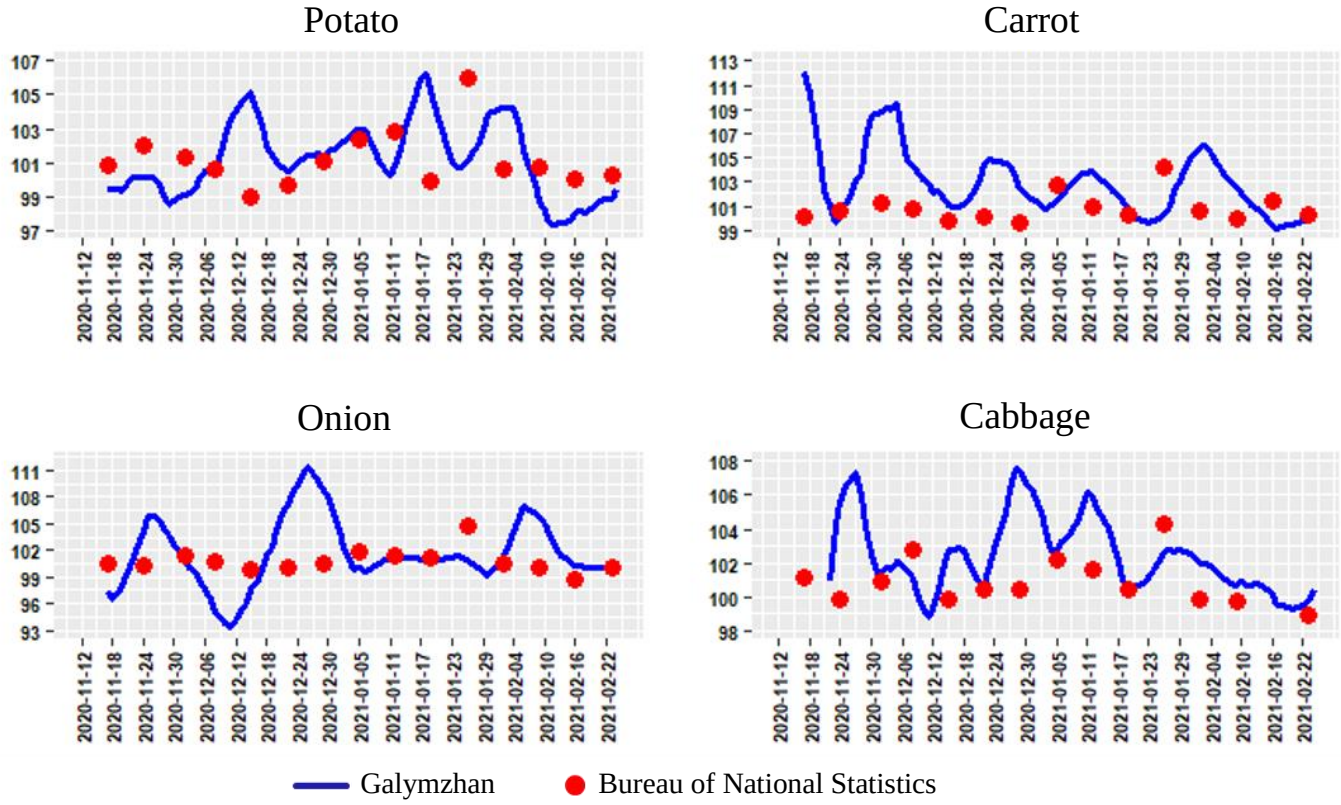
In future, as the system expands (adding new stores and increasing the sample), it is expected that the differential between the BNS and Galymzhan 2.0 data will narrow.

Nur-Sultan and Almaty cities were selected as examples of regional breakdown as cities with the largest number of online stores (4 and 8 online sites, respectively). At the same time, in the city of Nur-Sultan, there is a strong correlation between the data of the Galymzhan system and the official indicators of the BNS (Figure 4).

Contrary to Nur-Sultan, there was heightened volatility in Almaty. So, during the period preceding the New Year (December 26-27, 2020), a sharp rise in prices for

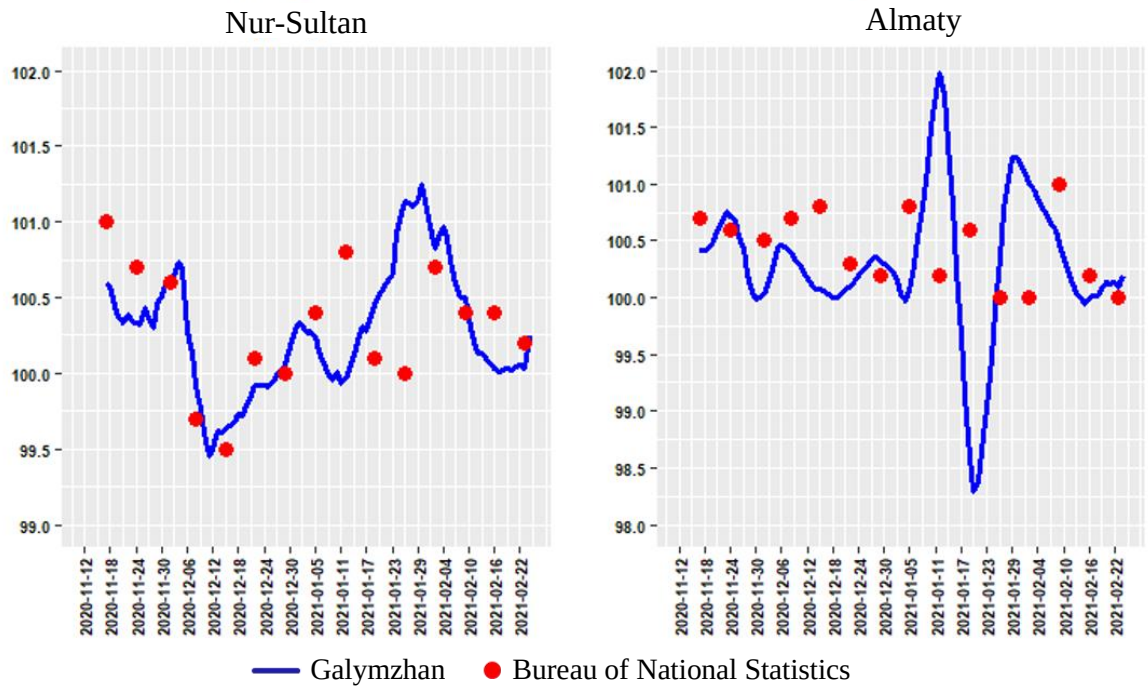
many foodstuffs was observed. A possible explanation for such growth in prices is the increased demand due to the New Year holidays. At the same time, the rise in prices is observed in large retail chains, while prices in local markets have practically not changed. Thus, the advantage of this system is also the ability to conduct an in-depth analysis of price dynamics during periods of abnormal market behavior. So, at present, Galymzhan allows you to track the dynamics of prices of certain representative goods on the shelves of various sales outlets.

Figure 3. Dynamics of Price Indices for SIFs and Inflation Proxy for Vegetables*, as % to the Corresponding Day of the Previous Week



*Data from Galymzhan System are smoothed by a 7-day MA and are taken with a 10-day lag
Source: BNS, derivations by the authors

Figure 4. Dynamics of Price Indices for SIFs and Inflation Proxy* in Nur-Sultan and Almaty, as % to the Corresponding Day of the Previous Week



* Data from Galymzhan System are smoothed by a 7-day MA and are taken with a 10-day lag
Source: BNS, derivations by the authors

With time, as the system improves, namely as the geographical representation is expanding, new representative goods are added and the total number of categories of goods in the consumer basket increases, it is expected that the convergence between the index calculated by the Galymzhan system and the official data will shrink and the extended index will be as close as possible to the consumer price index in terms of the product and regional representation and will be used building the short-term forecasts, for the purposes of analysis and decision-making on the monetary policy.

5. Findings and Recommendations for Future Research

As part of the current research, the world experience of using the web scrapping technology in building inflation both in official statistical bodies and in private organizations was studied. In addition, a description of the methodology used by the Galymzhan 2.0 system for collection and further construction of the inflation proxy for socially important foodstuffs in Kazakhstan is provided.

The results obtained indicate the presence of a correlation between the official data and the constructed index. At the same time, it was noted that prices in online stores change faster than in offline retail outlets, and the growth rate of prices in real stores is slightly higher.

It should be noted that the Galymzhan 2.0 system is currently run in the test mode. In order to further develop the system, the authors plan to expand the geographical coverage and add more stores, categories of goods in the consumer basket and representative goods.

The potential for further development of the visualization process is connected with the inclusion into the dashboard of the capability to track price adjustments not only in regions and stores, but also of specific representative goods on the online shelves of certain stores.

Further extension of the system will enable to get as close as possible in geographical and product coverage to the consumer price index, which will make it possible to replicate it in full, as well as to use the results of the Galymzhan 2.0 system as a tool for building short-term forecasts and analyzing monetary policy.

References

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, Wirts*chaft 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. Official web-site of the Australian Bureau of Statistics, Available at: https://www.abs.gov.au/articles/web-scraping-australian-cpi.