

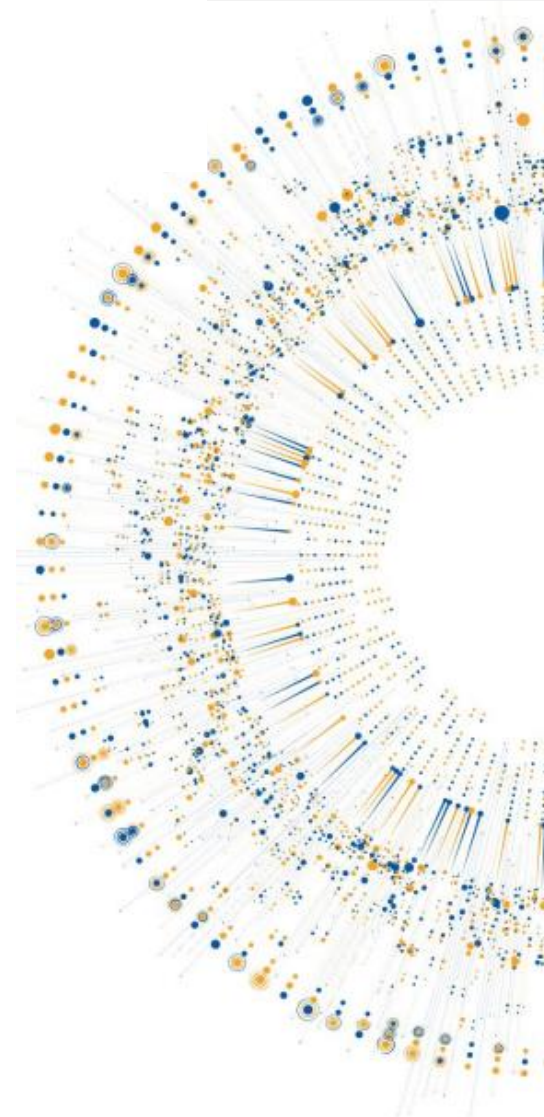


Agency for Strategic planning  
and reforms of the Republic of Kazakhstan  
Bureau of National statistics

## Price monitoring based on fiscal receipt data

Experimental statistics  
Working article No.2024-02

Astana 2024



## About experimental statistics

Experimental statistics are indicators and tools that are at the **testing stage** and are **not included in official statistics**. Indicators and tools are based on innovative methods and/or new data sources.

Taking into account the analysis of the results of experimental calculations, the recommendations of international experts, as well as receiving feedback from users, methodological approaches, respectively, and **the results may change** in the process of forming the final methodology.

Thus, experimental statistics are **intended to test new approaches and cannot be used** as official statistical information.

Within the framework of experimental statistics, **the confidentiality of the primary statistical data used** is maintained, in accordance with paragraphs 1-7 of Article 8 "Ensuring guarantees of confidentiality and protection of submitted data" of the Law of the Republic of Kazakhstan "On State Statistics" dated March 19, 2010 No. 257-IV.

## 1. Key points

- This article is published to inform users about ongoing research on the modernization of **inflation calculation** in the Republic of Kazakhstan, as well as as an explanation for the developed interactive dashboard.

- The Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan (hereinafter referred to as the Bureau), within the framework of the Portfolio of Development of State Statistics and the national Data ecosystem for 2023-2025, is implementing a project **to switch to a new format for monitoring inflation** in a pilot mode based on alternative data (hereinafter referred to as the Project). **Data from fiscal data operators** (hereinafter referred to as FDO) are considered as alternative data for calculating inflation.

- In order to implement the Project under a joint order, the Bureau receives data on fiscal receipts from the Ministry of Finance. An average of **130-145** million records are received weekly.

- Experimental calculations were made for the period **from January 2023 to June 2024**.

- **The operator of fiscal data** is a legal entity that ensures the transfer of information on monetary settlements online to the tax authorities via public telecommunications networks, determined by the authorized body in coordination with the authorized body in the field of informatization.

- **A cash register** is an electronic device with a fiscal memory unit without a data transfer function, a hardware and software complex with (without) the function(s) of fixing and (or) transmitting data, an electronic device with the function of fixing and (or) transmitting data, providing registration and display of information about monetary settlements carried out when selling goods, performing works, and providing services.

- **The receipt of the cash register** is the primary accounting document of the cash register confirming the fact of a monetary settlement between the seller (supplier of goods, work, services) and the buyer (client), issued in paper form or in electronic form.

- **The results of experimental price indices** for individual goods according to the FDO data for the previous month, by December of the previous year, by the corresponding month of the previous year are presented in the dashboard.

## 2. Introduction

**The use of big data** in statistical production, including price statistics, is one of the priority areas of development.

European countries such as Belgium, Sweden, and Switzerland are actively using alternative data – scanning data to calculate inflation in addition to traditional ways of registering prices by bypassing surveillance facilities.

**Scanning data** is data on the sale of goods obtained from scanned barcodes of goods from electronic terminals of retail outlets.

In order to improve the efficiency and quality of statistical data, the Bureau, within the framework of the Project, began using fiscal data actually paid for by the buyer to build inflation. Work has been carried out to study the data of the FDO, a draft methodology has been developed, and the results of experimental calculations have been obtained.

The data of the FDO contains the **following information**: the business identification number (hereinafter referred to as the BIN) of the trading enterprise, the address of the place of use of the cash register, the quantity of the purchased goods, the total amount of the sale of the goods, the date and time of purchase.

One of the key aspects of working with big data is **its quality**. However, the data entering the system from the receipts of cash registers have a number of limitations that make it difficult to use them for price analysis. Difficulties occurred in identifying and comparing specific product items, since the data is **not structured** and does not contain information on the characteristics of goods that affect pricing.

The resulting data array is **cleaned and marked up** using machine learning tools. Using a machine learning model allows you to recognize certain types of product names.

No matter how large the volumes of FDO data are, they cover only a limited part of the objects, which may not coincide with the general population we are interested in. In this regard, based on **the volume of retail turnover** by product type (data from domestic trade statistics: statistical form 1-VT "Report on the sale of goods and services"), work has been carried out **to determine the representativeness** of the FDO data. According to the results of which, the most representative goods (**20 items**) were identified in **the cities of Astana and Almaty, Karagandy region**. The experimental calculation was carried out for **13** food items, **7** non-food items.

### 3. Methodology for calculating the individual price index

Work on innovations in the methodology for calculating inflation based on FDO data is carried out by the Bureau as part of the provision of technical assistance by **international experts from the IMF**.

The following approaches to the use of big data in calculating inflation are used in international practice: static, dynamic and multi-sided.

**The static approach is the closest to the** traditional sampling method. With the static approach, a sample of the best-selling products is determined at the beginning of the year and a price survey is conducted for them during the reporting year. In this approach, new types of goods that appeared on the market during the reporting year are not taken into account in the calculation of inflation.

With a **dynamic approach**, the basket of consumer price index products is updated monthly, ensuring the relevance of the selection of the assortment type of goods. Every month, the products sold are selected and available for two consecutive months. This method is used in most European countries (Luxembourg, Australia, Holland, Belgium, etc.), which assumes an index approach, without forming average prices.

**The multi-pronged approach** takes into account all products available in at least two-time window periods. When using a multi-party index, the time interval during which the index is calculated is updated every month. Sliding time windows are most often used: each month, the time window is shifted forward by 1 month. The longer the time window, the more data from the previous one will affect the compilations of the current month.

In accordance with the recommendations of the international expert of the IMF, the Bureau chose to use a dynamic approach in Kazakhstan to calculate experimental inflation according to the FDO data.

The goods of the consumer basket in the republic are classified according to the Directory of commodity items to the Classification of Individual Consumption by Purpose (hereinafter referred to as the COICOP). At the initial stage, the data of the FDO are interconnected with the classification of the elements of the COICOP.

Next, several stages of **cleaning up big data** are performed, which include the exclusion of irrelevant information, errors, outliers beyond the scope of price observation and **data normalization**.

Taking into account the influence of various factors on price changes, such as the region, the type of base object, it is important to determine the optimal **stratum level**. The stratum serves as the smallest aggregate for calculating the price index, and subsequently aggregating it by weighing.

The generated stratum represents grouped **FDO data by product, region, and type of base object**.

According to the classification and coding of the dimension of legal entities, branches and representative offices, as well as individual business entities by the number of employees (hereinafter referred to as DCE), the basic objects are divided into three types of stratification: **large, medium and small**.

**Weights** by types of basic objects in the context of goods are determined by correlating the annual **volume of retail turnover** of the corresponding type of basic object to the total volume of retail turnover of the region.

The calculation of the monthly price index takes into account FDO data for **the first 3 weeks of the reporting month**.

**The price of a subspecies of goods (product name)** by region according to BIN is based on the formula:

$$P = \frac{\sum Q}{\sum K}$$

where,

$P$  – the price of the product subspecies (product name);

$Q$  – the total amount of the sale of a subspecies of goods;

$K$  – the number of purchased subspecies of goods.

**The calculation of the individual price index** of the reporting period to the previous one (month) is carried out in several stages:

1) **The price index for a subspecies of goods (product name)** by BIN by region to the previous month is based on the formula:

$$I_{jr} = \frac{\overline{P_{jr(t)}}}{\overline{P_{jr(t-1)}}} \times 100$$

where,

$I_{jr}$  – price index by product type (product name)  $j$  by region  $r$  in period  $t$  to the previous period  $(t-1)$  by BIN;

$P_{jr(t)}$  – the price of the product subspecies (product name)  $j$  of the reporting period  $t$  of the region  $r$  by BIN;

$P_{jr(t-1)}$  – the price of the subspecies of the product (product name)  $j$  of the previous period  $t-1$  of the region  $r$  according to BIN.

2) **Price index by type of product (product category)  $j$**  for the DCE BO  $k$  for the region  $r$  is calculated as the geometric mean using the formula:

$$\overline{I_{jkr}} = \sqrt[n]{I_{j1kr} \times I_{j2kr} \times \dots \times I_{jnkr}}$$

where,

$\overline{I_{jkr}}$  – price index of the type of product (product category)  $j$  by DCE BO  $k$  by region  $r$ ;

$i_{j1kr}, i_{j2kr}, \dots, i_{jnkr}$  – the price index of subspecies of goods (product name)  $j$  (1, 2..n) by DCE BO  $k$  by region  $r$ ;

$n$  – the number of subspecies of goods (names of goods) in the form of goods (categories of goods) about in this DCE BO in the  $k$  region.

3) **The price index of the type of product (product category) by region** is based on the formula:

$$\bar{I}_{jr} = \frac{(I_{jk1r} \times W_{jk1r}) + (I_{jk2r} \times W_{jk2r}) + \dots + (I_{jknr} \times W_{jknr})}{W_{jk1r} + W_{jk2r} + \dots + W_{jknr}}$$

where,

$\bar{I}_{jr}$  – the price index of the type of product (product category)  $j$  by region  $r$ ;

$I_{jk1r}, I_{jk2r}, \dots, I_{jknr}$  – the price index of the type of product (product category)  $j$  by DCE BO  $k$  (1, 2..n) by region  $r$ ;

$W_{jk1r}, W_{jk2r}, \dots, W_{jknr}$  – the weight values to the DCE BO  $k$  (1, 2..n) by type of product  $j$  by region  $r$ .

## 4. Results

As a result of **expert calculations**, price indices for 20 goods in the cities of Astana, Almaty and Karaganda region for the period from February 2023 to June 2024 were formed **in comparison with the official data of the Bureau** (the traditional approach).

A **similar trend** in prices for some goods was noted over **the month**: for cabbage, carrots, onions, cucumbers, tomatoes in all three regions, for apples – in Astana and Karaganda region, for potatoes – in Astana.

Individual examples by product and region are shown in Figures 1 and 2.

Figure 1

### Consumer price index for carrots in the Karagandy region

as a percentage, compared to the previous month, the increase

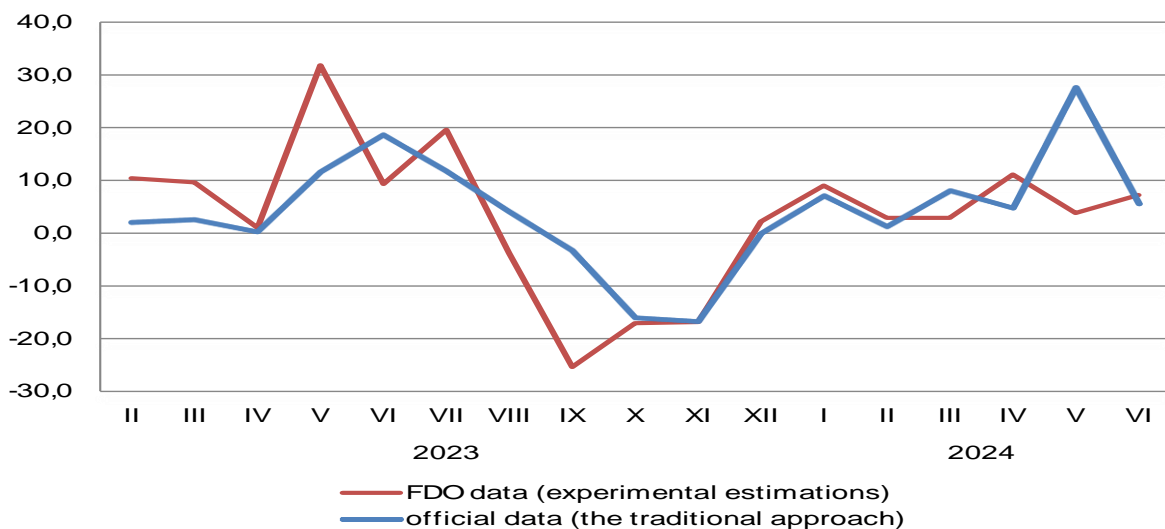
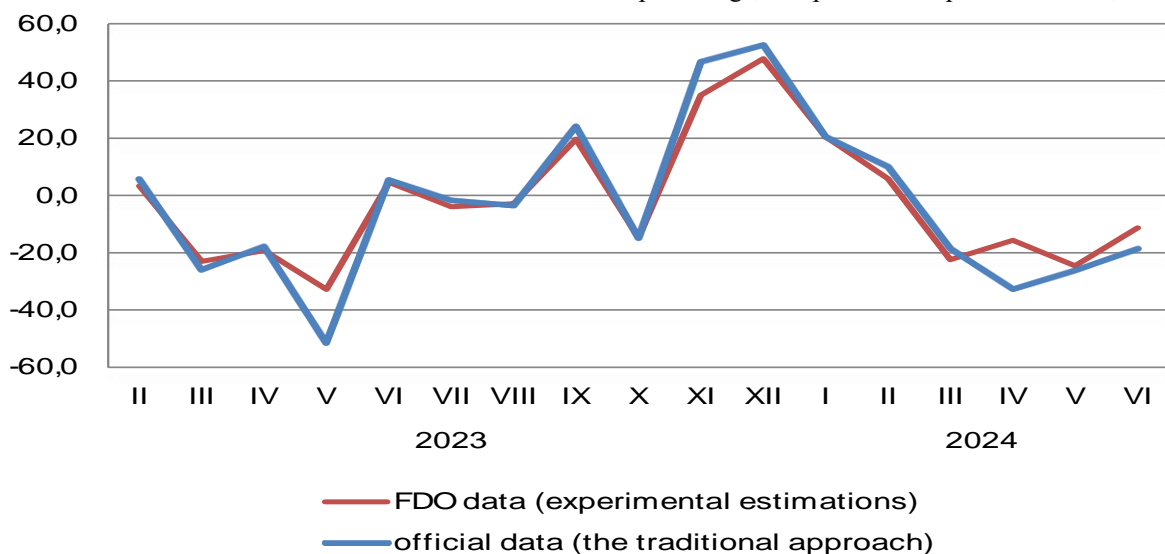


Figure 2

### Consumer price index for cucumbers in Astana

as a percentage, compared to the previous month, the increase



At the same time, in certain months there are **significant discrepancies** in the price indices for individual regions. Table 1 shows a list of products with similar dynamics, where price indices have formed above or below **15** percentage points in comparison with data from the FDO and the Bureau.

Table 1

**Price indices with similar dynamics**

to the previous month, as a percentage

Product name	Region	Period	FDO data	Bureau data
A	Б	В	1	2
White cabbage	Astana city	April 2023	109,5	134,2
		May 2023	103,5	118,7
		September 2023	71,9	93,8
	Almaty city	July 2023	155,2	124,7
		May 2024	93,0	76,1
	Karagandy region	February 2023	133,8	110,9
		June 2023	64,9	81,5
		July 2023	160,9	119,7
		June 2024	71,6	92,0
Potato	Astana city	September 2023	77,6	94,3
Bulb onion	Astana city	June 2023	82,8	98,8
		May 2024	110,5	130,3
	Almaty city	May 2024	108,9	143,9
Carrot	Astana city	September 2023	75,4	97,1
	Almaty city	May 2023	120,6	102,2
		September 2023	72,8	91,9
	Karagandy region	May 2023	131,8	111,5
		September 2023	74,6	96,7
Cucumbers	Astana city	May 2023	67,3	48,5
		April 2024	84,4	67,2
	Almaty city	May 2023	68,3	52,3
		July 2023	83,8	68,0
		August 2023	121,4	100,9
		December 2023	127,1	182,3
	Karagandy region	December 2023	156,4	135,9
		February 2024	103,4	135,2
		Июнь 2024	93,8	75,4
	Tomatoes	Astana city	February 2023	109,9
June 2023			69,9	88,0
Almaty city		July 2023	97,5	66,6
		December 2023	102,6	133,8
Television	Astana city	April 2024	106,4	126,9
	Almaty city	April 2024	101,0	126,9
	Karagandy region	April 2024	101,7	126,9
Gasoline AI-92	Karagandy region	May 2023	115,3	100,1

**The largest difference** in values was recorded: for cucumbers and tomatoes – in Almaty in December 2023 at 55.2 percentage points and 31.2 percentage points, respectively; for cabbage – in Almaty and Karaganda region in July 2023 at 30.5 percentage points and 41.2 percentage points, respectively.

In January-June 2024, 70-85% of goods (17 items out of 20) have **similar price dynamics, for the year** (June 2024 by June 2023) – 55-80% of goods (Table 2).

Table 2

**The number of products with similar price index dynamics according to the experimental and traditional approach**

	January-June 2024r.	June 2024r. to June 2023r.	units
Astana city	17		16
Almaty city	17		14
Karagandy region	14		11

A **similar dynamics** of price changes based on the results of the experiment and the traditional approach in the regional context has developed as follows:

1) *For the period January-June of the current year:*

- in Astana in 17 positions (85%): buckwheat, butter and sunflower oil, sour cream, potatoes, bulb onions, carrots, cabbage, cucumbers, tomatoes, apples, toilet soap, laundry soap, vinyl wallpaper, vacuum cleaner, TV, gasoline AI-95/96;

- in Almaty, 17 items (85%): rice, buckwheat, butter and sunflower oil, sour cream, salt, potatoes, carrots, cabbage, cucumbers, tomatoes, apples, toilet soap, laundry soap, vinyl wallpaper, vacuum cleaner, TV;

- in the Karagandy region in 14 positions (70%): rice, buckwheat, butter and sunflower oil, potatoes, carrots, cabbage, cucumbers, tomatoes, apples, laundry soap, vinyl wallpaper, vacuum cleaner, gasoline AI-95/96.

2) *For the year (June 2024. by June 2023.):*

- in Astana by 16 items (80%): rice, buckwheat, butter and sunflower oil, sour cream, salt, potatoes, bulb onions, carrots, cabbage, apples, toilet soap, laundry soap, vacuum cleaner, TV, gasoline AI-95/96;

- in Almaty by 14 items (70%): rice, buckwheat, butter and sunflower oil, sour cream, salt, bulb onions, cucumbers, tomatoes, apples, toilet soap, laundry soap, vacuum cleaner, gasoline AI-95/96;

- in the Karagandy region in 11 positions (55%): buckwheat, sunflower oil, sour cream, salt, onion, cabbage, tomatoes, apples, vacuum cleaner, TV, gasoline AI-95/96.

At the same time, there was a **multidirectional change in prices** for individual goods.

1) *For the period January-June of this year:*

Astana: according to experimental calculations, an increase in salt prices by 3.6% was recorded (the traditional approach is a decrease by 10.8%);

Almaty: according to experimental calculations, there is a decrease in prices for onions by 1.9% (the traditional approach is an increase of 4.6%);

Karagandy region: according to experimental calculations, prices for onions decreased by 33.5% (the traditional approach is an increase of 32.9%), TV – by 2.1% (by 33.2%), gasoline AI-95/96 and toilet soap – by 0.3% (by 0.7% and 1.4%). Also, according to the traditional approach, sour cream fell in price by 4.5% (experimental calculations – increased in price by 1.9%), salt – by 0.1% (by 1.1%).

2) *For the year (June 2024. by June 2023):*

- Astana: according to experimental calculations, prices for vinyl wallpaper decreased by 10.2% (the traditional approach – an increase of 6.5%), tomatoes also fell by 4.9% according to the traditional approach (experimental calculations – increased in price by 12.4%), cucumbers – by 3.2% (by 7.3%);

- Almaty: according to experimental calculations, there is a decrease in potato prices by 21.4% (the traditional approach is an increase of 10.2%), carrots – by 16% (by 16.1%), TV – by 9.1% (by 56.3%), vinyl wallpaper – by 0.2% (by 8.9%). Also, according to the traditional approach, cabbage fell in price by 27.1% (experimental calculations – increased in price by 1.9%);

- Karagandy region: according to the experimental calculation, there was a decrease in prices for carrots by 13.8% (the traditional approach is an increase of 29.3%), vinyl wallpaper – by 13.3% (by 0.8%), household soap – by 11.7% (by 2%), potatoes – by 11.1% (by 0.6%), toilet soap – by 3.7% (by 3.5%), butter – by 1.6% (by 28.6%). Also, according to the traditional approach, cucumbers fell in price by 17.8% (experimental calculations – increased in price by 23.8%), AI-92 gasoline – by 0.5% (by 2.8%).

## 5. Conclusions

Users of the interactive dashboard have the opportunity to assess the results of experimental price index calculations based on the following criteria (region, product, period). The presented experimental indices based on FDO are **preliminary**, and the methodology for experimental calculation may be revised.

As part of the project, **validation** of the consumer price index calculation methodology using big data will be sought from the IMF to ensure alignment with international standards.

However, transitioning to a progressive methodology for processing large price statistics data, as recommended by the IMF, requires **detailed information** from cash register receipts.

One of the challenges in processing FDO is its lack of structure. This necessitates **the development of additional algorithms** for automatic recognition and classification of products on an ongoing basis.

To ensure product identification, the Ministry of Trade and Integration is working on the development and implementation of a **National Commodity Catalog** (NCC).

At the same time, the dashboard **does not include** data on average prices. The absence of detailed information (unit identifiers, packaging) in fiscal data has affected the determination of average prices per standard unit of measurement. International experts do not recommend calculating average prices based on big data due to their heterogeneity.

Despite these current challenges, using FDO in calculating the consumer price index will **improve the quality** of primary information, **increase the coverage** of surveyed products, and **enhance the timeliness** of statistical data

[www.stat.gov.kz](http://www.stat.gov.kz) / Experimental statistics and tools

---

<b>Responsible for the release:</b> Department of Price Statistics	<b>Executor:</b> A.Makizhanova Tel. +7 7172 749265	<b>Director of the Department:</b> K. Zhakypbekuly Tel. +7 7172 749057	<b>Address:</b> 010000, Astana city Mangilik el avenue, 8 House of Ministries, entrance 4
---	--	--	--

---