

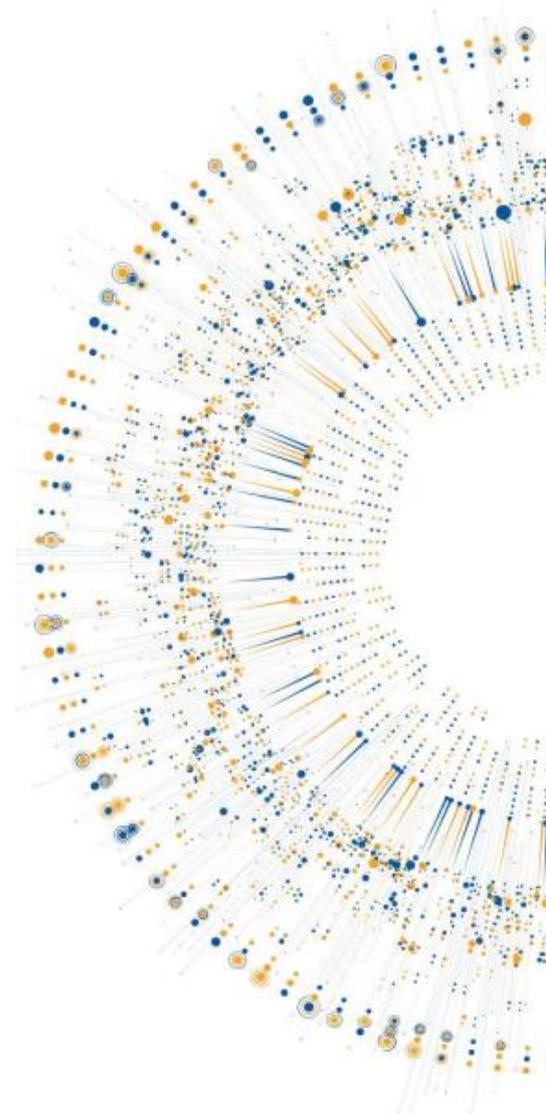


Бюро национальной статистики
Агентства по стратегическому планированию и реформам
Республики Казахстан

Мониторинг цен на основе данных фискальных чеков

Экспериментальная статистика
Рабочая статья №2024-02

Астана 2024



Об экспериментальной статистике

Экспериментальная статистика – это показатели и инструменты, которые находятся на **этапе тестирования и не входят в официальную статистику**. Показатели и инструменты основаны на инновационных методах и/или новых источниках данных.

С учетом анализа результатов экспериментальных расчетов, рекомендаций международных экспертов, а также получения обратной связи от пользователей, методологические подходы, соответственно, и **результаты могут меняться** в процессе формирования окончательной методологии.

Таким образом, экспериментальная статистика **предназначена для тестирования новых подходов и не может быть использована** в качестве официальной статистической информации.

В рамках экспериментальной статистики сохраняется **конфиденциальность используемых первичных статистических данных**, в соответствии с пунктами 1-7 статьи 8 «Обеспечение гарантий конфиденциальности и защиты представляемых данных» Закона Республики Казахстан «О государственной статистике» от 19 марта 2010 года № 257-IV.

1. Ключевые моменты

- Данная статья публикуется для информирования пользователей о проводимых исследованиях по модернизации **расчета инфляции** в Республике Казахстан, а также в качестве пояснения к разработанному интерактивному дашборду.

- Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан (далее – Бюро) в рамках Портфеля развития государственной статистики и национальной экосистемы данных на 2023-2025 годы реализует проект **по переходу на новый формат мониторинга инфляции** в пилотном режиме на основе альтернативных данных (далее – Проект). В качестве альтернативных данных для расчета инфляции рассматриваются **данные операторов фискальных данных** (далее – ОФД).

- Для реализации Проекта в рамках совместного приказа Бюро получает данные о фискальных чеках от Министерства финансов. Еженедельно в среднем поступает **130-145** млн. записей.

- Экспериментальные расчеты произведены за период **с января 2023г. по июнь 2024г.**

- **Оператор фискальных данных** – юридическое лицо, обеспечивающее передачу сведений о денежных расчетах в оперативном режиме в налоговые органы по сетям телекоммуникаций общего пользования, определенное уполномоченным органом по согласованию с уполномоченным органом в сфере информатизации.

- **Контрольно-кассовая машина** – электронное устройство с блоком фискальной памяти без функции передачи данных, аппаратно-программный комплекс с (без) функцией (-и) фиксации и (или) передачи данных, электронное устройство с функцией фиксации и (или) передачи данных, обеспечивающие регистрацию и отображение информации о денежных расчетах, осуществляемых при реализации товаров, выполнении работ, оказании услуг.

- **Чек контрольно-кассовой машины** – первичный учетный документ контрольно-кассовой машины, подтверждающий факт осуществления между продавцом (поставщиком товара, работы, услуги) и покупателем (клиентом) денежного расчета, выданный в бумажном виде либо в электронном виде.

- **Результаты экспериментальных индексов цен** на отдельные товары по данным ОФД к предыдущему месяцу, к декабрю предыдущего года, к соответствующему месяцу предыдущего года представлены в дашборде.

2. Введение

Использование больших данных в статистическом производстве, в том числе статистики цен, является одним из приоритетных направлений развития.

Такие европейские страны, как Бельгия, Швеция, Швейцария, активно используют альтернативные данные – данные сканирования для расчета инфляции в дополнении к традиционным способам регистрации цен путем обхода объектов наблюдения.

Данные сканирования — это данные о продаже товара, полученные по сканированным штрих-кодам товаров из электронных терминалов торговых точек.

В целях повышения оперативности и качества статистических данных Бюро в рамках Проекта приступило к использованию фискальных данных, фактически оплаченных покупателем для построения инфляции. Проведена работа по изучению данных ОФД, разработан проект методики, получены результаты экспериментальных расчетов.

Данные ОФД содержат **следующую информацию**: бизнес-идентификационный номер (далее – БИН) торгового предприятия, адрес места использования контрольно-кассовой машины, количество приобретаемого товара, общая сумма продажи товара, дата и время покупки.

Одним из ключевых аспектов работы с большими данными является **их качество**. Однако, данные, поступающие в систему из чеков контрольно-кассовых машин, имеют ряд ограничений, которые затрудняют их использование для анализа цен. Сложности произошли в идентификации и сравнении конкретных товарных позиций, так как данные **не структурированы** и не содержат информацию по характеристике товаров, влияющих на ценообразование.

Получаемый массив данных **подвергается очистке и разметке** с использованием инструментов машинного обучения. Использование модели машинного обучения позволяет распознавать некоторые виды наименований товаров.

Какими бы большими ни были объемы данных ОФД, они охватывают только ограниченную часть объектов, которая может не совпадать с интересующей нас генеральной совокупностью. В связи с этим, исходя из **объема розничного товарооборота** по видам продукции (данные статистики внутренней торговли: статистическая форма 1-ВТ «Отчет о реализации товаров и услуг») осуществлена работа по **определению репрезентативности** данных ОФД. По результатам которой, определены наиболее репрезентативные товары (**20 наименований**) по городам Астана и Алматы, Карагандинской области. Экспериментальный расчет был проведен по **13** позициям продуктов питания, **7** - непродовольственным товарам.

3. Методология расчета индивидуального индекса цен

Работа по нововведению в методике расчета инфляции на основе данных ОФД осуществляется Бюро в рамках оказания технической помощи **международными экспертами МВФ**.

В международной практике применяются следующие подходы к использованию больших данных в расчете инфляции: статический, динамический и многосторонний.

Наиболее приближенный к традиционному методу формирования выборки является **статический подход**. При статическом подходе в начале года определяется выборка наиболее продаваемых товаров и в течение отчетного года по ним проводится ценовое обследование. В таком подходе новые виды товаров, появившиеся на рынке, в течение отчетного года не учитываются в расчете инфляции.

При **динамическом подходе** корзина товаров индекса потребительских цен обновляется ежемесячно, обеспечивая актуальность выборки ассортиментного вида товара. Каждый месяц выбираются продаваемые товары, доступные в течение двух месяцев подряд. Данный метод используется в большинстве европейских стран (Люксембург, Австралия, Голландия, Бельгия и др.), который предполагает индексный подход, без формирования средних цен.

Многосторонний подход учитывает все товары, доступные как минимум в двух периодах временного окна. При использовании многостороннего индекса временной интервал, в течение которого рассчитывается индекс, обновляется каждый месяц. Чаще всего используются скользящие временные окна: каждый месяц временное окно сдвигается вперед на 1 месяц. Чем длиннее временное окно, тем больше данных из предыдущего повлияют на компиляции текущего месяца.

В соответствии с рекомендациями международного эксперта МВФ для расчета экспериментальной инфляции по данным ОФД, Бюро выбрало использование **динамического подхода** в Казахстане.

Товары потребительской корзины в республике классифицируются согласно Справочнику товарных позиций к Классификатору индивидуального потребления по целям (далее – НК ИПЦ). На первоначальном этапе осуществляется **взаимоувязка** данных ОФД с **классификацией** элементов НК ИПЦ.

Далее производится несколько этапов **очистки больших данных**, которые предусматривают исключение нерелевантной информации, ошибок, выбросов, выходящих за рамки ценового наблюдения и **нормализацию данных**.

Принимая во внимание влияние разных факторов на изменение цены, таких как регион, тип базового объекта, важно определить оптимальный **уровень страты**. Страта служит наименьшим агрегатом для расчета индекса цен, и в последующем его агрегации путем взвешивания.

Сформированная страта представляет сгруппированные данные ОФД по товару, региону и типу базового объекта.

Согласно классификации и кодирования размерности юридических лиц, филиалов и представительств, а также субъектов индивидуального предпринимательства по численности работников (далее – КРП) базовые объекты подразделены на три типа стратификации: **крупные, средние и малые.**

Вес по типам базовых объектов в разрезе товара определяются путем соотнесения годового **объема розничного товарооборота** соответствующего типа базового объекта к общему объему розничного товарооборота региона.

В расчете месячного индекса цен учитываются данные ОФД за **первые 3 недели отчетного месяца.**

Цена подвида товара (наименование товара) по региону по БИНу находится по формуле:

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

где,

P – цена подвида товара (наименование товара);

Q – общая сумма продажи подвида товара;

K – количество приобретенного подвида товара.

Расчет индивидуального индекса цен отчетного периода к предыдущему (месяц) осуществляется в несколько этапов:

1) **Индекс цен по подвиду товара (наименование товара)** по БИНу по региону к предыдущему месяцу находится по формуле:

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

где,

I_{jr} – индекс цен по подвиду товара (наименование товара) j по региону r в период t к предыдущему периоду $(t-1)$ по БИНу;

$P_{jr(t)}$ – цена подвида товара (наименование товара) j отчетного периода t региона r по БИНу;

$P_{jr(t-1)}$ – цена подвида товара (наименование товара) j предыдущего периода $t-1$ региона r по БИНу.

2) **Индекс цен по виду товара (категория товара) j** по КРП БО k по региону r рассчитывается как среднегеометрическая по формуле:

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

где,

I_{jkr} – индекс цен вида товара (категория товара) j по КРП БО k по региону r ;

$i_{j1kr}, I_{j2kr}, \dots, I_{jnkr}$ – индекс цен подвидов товара (наименование товара) j (1,2.. n) по КРП БО k по региону r ;

n – количество подвидов товара (наименований товаров) в виде товаров (категории товаров) j в данном КРП БО k в регионе r .

3) **Индекс цен вида товара (категория товара) по региону** находится по формуле:

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

где,

\bar{I}_{jr} – индекс цен вида товара (категория товара) j по региону r ;

$I_{jk1r}, I_{jk2r}, \dots, I_{jknr}$ – индекс цен вида товара (категория товара) j по КРП БО k (1,2.. n) по региону r ;

$W_{jk1r}, W_{jk2r}, \dots, W_{jknr}$ – весовые значения объема розничного товарооборота по КРП БО k (1,2.. n) по виду товара j по региону r .

4. Результаты

В результате экспериментальных расчетов сформированы индексы цен по 20 товарам по городам Астана, Алматы и Карагандинской области за период с февраля 2023г. по июнь 2024г. в сравнении с официальными данными Бюро (традиционным подходом).

На некоторые товары отмечена аналогичная тенденция изменения цен за месяц: на капусту, морковь, лук репчатый, огурцы, помидоры – во всех трех регионах; на яблоки – в г.Астана и Карагандинской области; на картофель – в г.Астана.

Отдельные примеры по товарам и регионам приведены на рисунках 1 и 2.

Рисунок 1

Индекс потребительских цен на морковь в Карагандинской области

в процентах, к предыдущему месяцу, прирост

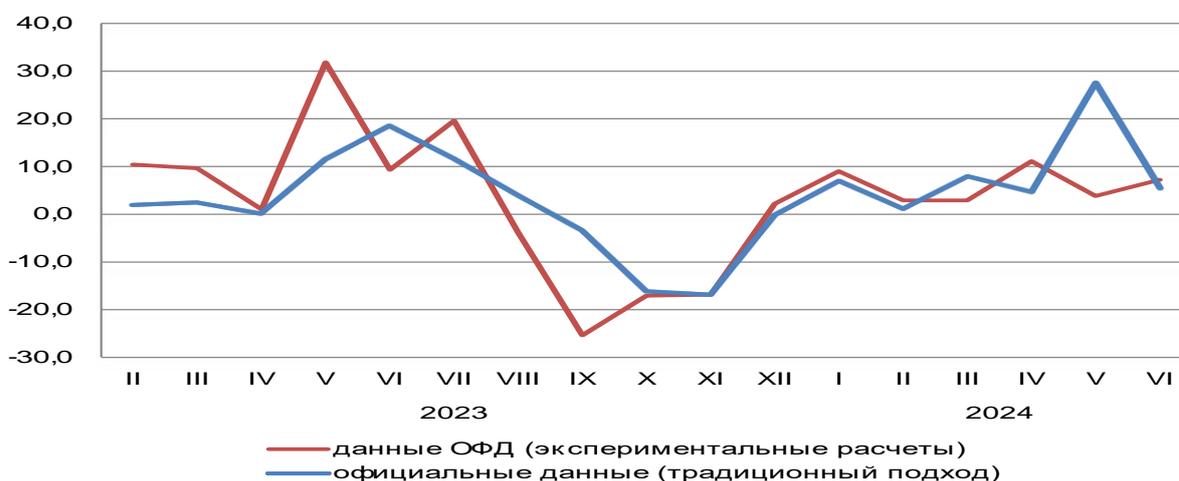


Рисунок 2

Индекс потребительских цен на огурцы в г.Астана

в процентах, к предыдущему месяцу, прирост



В то же время в определенные месяцы наблюдаются **значительные расхождения** в индексах цен по отдельным регионам. В таблице 1 представлен список товаров с аналогичной динамикой, где индексы цен сложились выше или ниже **15** процентных пунктов в сравнении данных ОФД и Бюро.

Таблица 1

Индексы цен с аналогичной динамикой

к предыдущему месяцу, в процентах

Наименование товара	Регион	Период	Данные ОФД	Данные Бюро
А	Б	В	1	2
Капуста белокочанная	г.Астана	Апрель 2023	109,5	134,2
		Май 2023	103,5	118,7
		Сентябрь 2023	71,9	93,8
	г.Алматы	Июль 2023	155,2	124,7
		Май 2024	93,0	76,1
	Карагандинская область	Февраль 2023	133,8	110,9
		Июнь 2023	64,9	81,5
Июль 2023		160,9	119,7	
Июнь 2024		71,6	92,0	
Картофель	г.Астана	Сентябрь 2023	77,6	94,3
Лук репчатый	г.Астана	Июнь 2023	82,8	98,8
		Май 2024	110,5	130,3
	г.Алматы	Май 2024	108,9	143,9
Морковь	г.Астана	Сентябрь 2023	75,4	97,1
	г.Алматы	Май 2023	120,6	102,2
		Сентябрь 2023	72,8	91,9
	Карагандинская область	Май 2023	131,8	111,5
		Сентябрь 2023	74,6	96,7
Май 2024	103,8	127,5		
Огурцы	г.Астана	Май 2023	67,3	48,5
		Апрель 2024	84,4	67,2
	г.Алматы	Май 2023	68,3	52,3
		Июль 2023	83,8	68,0
		Август 2023	121,4	100,9
		Декабрь 2023	127,1	182,3
	Карагандинская область	Декабрь 2023	156,4	135,9
Февраль 2024		103,4	135,2	
Июнь 2024	93,8	75,4		
Помидоры	г.Астана	Февраль 2023	109,9	137,5
		Июнь 2023	69,9	88,0
	г.Алматы	Июль 2023	97,5	66,6
		Декабрь 2023	102,6	133,8
Телевизор	г.Астана	Апрель 2024	106,4	126,9
	г.Алматы	Апрель 2024	101,0	126,9
	Карагандинская область	Апрель 2024	101,7	126,9
Бензин АИ-92	Карагандинская область	Май 2023	115,3	100,1

Наибольшее различие значений зафиксировано: на огурцы и помидоры – в г.Алматы в декабре 2023г. в 55,2 п.п. и 31,2 процентных пункта, соответственно; на капусту – в г.Алматы и Карагандинской области в июле 2023г. в 30,5 п.п. и 41,2 процентных пункта, соответственно.

За январь-июнь 2024 года 70-85% товаров (17 позиций из 20) имеют схожую динамику изменения цен, за год (июнь 2024г. к июню 2023г.) – 55-80% товаров (таблица 2).

Таблица 2

Количество товаров с аналогичной динамикой индексов цен по экспериментальным расчетам и традиционному подходу

	Январь-июнь 2024г.	Июнь 2024г. к июню 2023г.	единиц
г. Астана	17		16
г. Алматы	17		14
Карагандинская область	14		11

Аналогичная динамика изменения цен по результатам эксперимента и традиционного подхода в региональном разрезе сложилась следующим образом:

1) За период январь-июнь текущего года:

- в г.Астана по 17 позициям (85%): на крупу гречневую, масло сливочное и подсолнечное, сметану, картофель, лук репчатый, морковь, капусту, огурцы, помидоры, яблоки, мыло туалетное, мыло хозяйственное, обои виниловые, пылесос, телевизор, бензин марки АИ-95/96;

- в г.Алматы по 17 позициям (85%): на рис шлифованный, крупу гречневую, масло сливочное и подсолнечное, сметану, соль, картофель, морковь, капусту, огурцы, помидоры, яблоки, мыло туалетное, мыло хозяйственное, обои виниловые, пылесос, телевизор;

- в Карагандинской области по 14 позициям (70%): на рис шлифованный, крупу гречневую, масло сливочное и подсолнечное, картофель, морковь, капусту, огурцы, помидоры, яблоки, мыло хозяйственное, обои виниловые, пылесос, бензин марки АИ-95/96.

2) За год (июнь 2024г. к июню 2023г.):

- в г.Астана по 16 позициям (80%): на рис шлифованный, крупу гречневую, масло сливочное и подсолнечное, сметану, соль, картофель, лук репчатый, морковь, капусту, яблоки, мыло туалетное, мыло хозяйственное, пылесос, телевизор, бензин марки АИ-95/96;

- в г.Алматы по 14 позициям (70%): на рис шлифованный, крупу гречневую, масло сливочное и подсолнечное, сметану, соль, лук репчатый, огурцы, помидоры, яблоки, мыло туалетное, мыло хозяйственное, пылесос, бензин марки АИ-95/96;

- в Карагандинской области по 11 позициям (55%): на крупу гречневую, масло подсолнечное, сметану, соль, лук репчатый, капусту, помидоры, яблоки, пылесос, телевизор, бензин марки АИ-95/96.

Вместе с тем, отмечено **разнонаправленное изменение цен** на отдельные товары.

1) За период январь-июнь текущего года:

- г.Астана: по экспериментальным расчетам зафиксировано повышение цен на соль на 3,6% (традиционный подход – снижение на 10,8%);

- г.Алматы: по экспериментальным расчетам наблюдается снижение цен на лук репчатый на 1,9% (традиционный подход – рост на 4,6%);

- Карагандинская область: по экспериментальным расчетам отмечено снижение цен на лук репчатый на 33,5% (традиционный подход – рост на 32,9%), телевизор – на 2,1% (на 33,2%), бензин марки АИ-95/96 и мыло туалетное – по 0,3% (на 0,7% и 1,4%). Также по традиционному подходу сметана подешевела на 4,5% (экспериментальные расчеты – подорожала на 1,9%), соль – на 0,1% (на 1,1%).

2) За год (июнь 2024г. к июню 2023г.):

- г.Астана: по экспериментальным расчетам зафиксировано снижение цен на обои виниловые на 10,2% (традиционный подход – рост на 6,5%), Также по традиционному подходу помидоры подешевели на 4,9% (экспериментальные расчеты – подорожали на 12,4%), огурцы – на 3,2% (на 7,3%);

- г.Алматы: по экспериментальным расчетам наблюдается снижение цен на картофель на 21,4% (традиционный подход – рост на 10,2%), морковь – на 16% (на 16,1%), телевизор – на 9,1% (на 56,3%), обои виниловые – на 0,2% (на 8,9%). Также по традиционному подходу капуста подешевела на 27,1% (экспериментальные расчеты – подорожала на 1,9%);

- Карагандинская область: по данным экспериментального расчета отмечено снижение цен на морковь на 13,8% (традиционный подход – рост на 29,3%), обои виниловые – на 13,3% (на 0,8%), мыло хозяйственное – на 11,7% (на 2%), картофель – на 11,1% (на 0,6%), мыло туалетное – на 3,7% (на 3,5%), масло сливочное – на 1,6% (на 28,6%). Также по традиционному подходу огурцы подешевели на 17,8% (экспериментальные расчеты – подорожали на 23,8%), бензин марки АИ-92 – на 0,5% (на 2,8%).

5. Выводы

Пользователи интерактивного дашборда имеют возможность оценить результаты экспериментальных расчетов индексов цен по следующим критериям (регион, товар, период).

Представленные экспериментальные индексы по данным ОФД **неокончательные**, методика по экспериментальному расчету может быть пересмотрена.

В ходе работы над Проектом планируется получить со стороны МВФ **валидацию** методологии расчета индекса потребительских цен с использованием больших данных на соответствие международным стандартам.

Вместе с тем, для перехода к прогрессивной методологии обработки больших данных по статистике цен, рекомендуемой МВФ, необходимо использовать **подробную информацию** по чекам контрольно-кассовых машин.

Одним из сложностей обработки данных ОФД является их не структурированность. Это **требует разработки дополнительных алгоритмов** для автоматического распознавания и классификации товаров на постоянной основе.

В целях обеспечения идентификации товаров Министерством торговли и интеграции ведется работа по разработке внедрения **Национального каталога товаров (НКТ)**.

В то же время, дашборд **не содержит** данные об уровне **средних цен**. Отсутствие подробной информации (идентификаторы единицы измерения, расфасовка) в фискальных данных повлияло на определение средней цены за стандартную единицу измерения. Международные эксперты рекомендуют не рассчитывать средние цены на основе больших данных из-за их неоднородности.

Несмотря на текущие проблемы, использование данных ОФД в расчете индекса потребительских цен **позволит повысить качество** первичной информации, **увеличить охват** обследуемых товаров и **оперативность получения** статистических данных.

www.stat.gov.kz / Экспериментальная статистика и инструменты

Ответственные за выпуск: Департамент статистики цен	Исполнитель: А.Макижанова Тел. +7 7172 749265	Директор департамента: Қ. Жақыпбекұлы Тел. +7 7172 749057	Адрес: 010000, г. Астана Проспект Мәңгілік ел, 8 Дом министерств, 4 подъезд
---	--	--	---