



М. В. Казакова

ВЗАИМОСВЯЗЬ ЦЕНОВОЙ ДИНАМИКИ В ОНЛАЙН-СЕКМЕНТЕ И СТАНДАРТНЫХ ЦЕНОВЫХ ПОКАЗАТЕЛЕЙ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО
ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ
ФЕДЕРАЦИИ»
(РАНХиГС)

**ВЗАИМОСВЯЗЬ ЦЕНОВОЙ ДИНАМИКИ В ОНЛАЙН-СЕКМЕНТЕ
И СТАНДАРТНЫХ ЦЕНОВЫХ ПОКАЗАТЕЛЕЙ**

Казакова М.В., Центр изучения проблем центральных банков Института при-
кладных экономических исследований (ИПЭИ) Российской академии народного хо-
зяйства и государственной службы при Президенте РФ (РАНХиГС), с.н.с., к.э.н.,
ORCID: 0000-0002-7396-4666, kazakova@ranepa.ru

Москва

2022

FEDERAL STATE BUDGETARY EDUCATIONAL INSTITUTION OF HIGHER EDUCATION
«RUSSIAN PRESIDENTIAL ACADEMY OF NATIONAL ECONOMY
AND PUBLIC ADMINISTRATION» (RANEPA)

**THE RELATIONSHIP BETWEEN PRICE DYNAMICS
IN THE ONLINE SEGMENT AND STANDARD PRICE INDICATORS**

M.V. Kazakova, Center for the Study of Central Banking Problems, Institute of Applied Economic Research, Russian Presidential Academy of National Economy and Public Administration (RANEPA), senior researcher, Cand. Sci. (Econ.), ORCID ID: 0000-0002-7396-4666, kazakova@ranepa.ru

Moscow

2022

Аннотация

Развитие интернет-торговли в условиях цифровизации экономики способствует внедрению и расширению возможностей применения новых макро- и микро-экономических индикаторов, основанных на больших массивах данных, для изучения инфляционных процессов. **Актуальность** настоящей работы определяется тем, что в силу своей доступности и высокой частотности онлайн-цены мультиканальных ритейлеров могут быть более репрезентативным источником информации для измерения и прогнозирования уровней инфляции, чем традиционные данные. В частности, ценовые индексы, основанные на онлайн-данных, позволили отслеживать динамику цен в период пандемии COVID-19 в режиме реального времени. Основная **цель** исследования заключается в систематизации результатов эмпирических работ по использованию данных онлайн-торговли для анализа особенностей инфляционных процессов (**предмет** исследования). Кроме того, работа направлена на изучение возможностей применения онлайн-данных и онлайн-ценовых индексов для прогнозирования офлайн-цен. Достижению заявленной цели способствует обзор академической литературы (**источник** информации), посвященной использованию данных об онлайн-ценах для построения высокочастотных онлайн-индикаторов, измерения и прогнозирования уровня инфляции, анализа ценовой динамики в онлайн-сегменте и сравнения жесткости онлайн-цен и цен в традиционной розничной торговле (**задачи** исследования). Исследование проводилось с использованием таких **методов**, как дескриптивный, статистический и графический анализ, анализ данных, системный подход, сравнительный анализ. По **результатам** исследования эмпирического опыта применения данных об онлайн-ценах можно прийти к **выводу**, что в настоящее время данные интернет-торговли интенсивно используются зарубежными статистическими службами для построения высокочастотных индексов цен и могут служить репрезентативным источником информации об уровне инфляции. Тем не менее проведенный обзор выявил, что ни в одном из существующих исследований официальный ИПЦ не был реплицирован в полном объеме ввиду высокой сложности сбора данных и поддержания базы данных в рабочем состоянии (**научная новизна** работы). В этой связи **перспективы** дальнейшего исследования данной проблемы представляются в максимально возможном устранении этого недостатка с опорой на предшествовавший мировой опыт применения данных об онлайн-ценах и высокочастотных онлайн-индикаторов.

В свою очередь, результаты проведенного обзора могут быть **использованы** в интересах монетарных властей РФ для построения прогнозных моделей инфляции, учитывающих высокочастотные онлайн-данные по ценам.

Ключевые слова: инфляция, интернет-торговля, факторы инфляции, модели инфляции, цифровизация, индекс потребительских цен, онлайн-данные, индекс онлайн-цен, парсинг данных

JEL: E17 E31, E52, C81

Abstract

The development of e-commerce in the context of digitalization of the economy contributes to the introduction and expansion of opportunities for the use of new macro- and microeconomic indicators based on big data to study inflationary processes. The **relevance** of this work is determined by the fact that due to their availability and high frequency online prices of multi-channel retailers can be a more representative source of information for measuring and forecasting inflation levels than traditional data. Price indices based on online data made it possible to track price dynamics during the COVID-19 pandemic in real time. The main purpose of the study is to systematize the results of empirical work on the use of online trading data to analyze the features of inflationary processes (**subject** of the study). In addition, the work is **aimed** at exploring the possibilities of using online data and online price indices to predict offline prices. The achievement of the stated goal is facilitated by a review of academic literature devoted to the use of online price data for constructing high-frequency online indicators, measuring and forecasting the inflation rate, analyzing price dynamics in the online segment and comparing the rigidity of online prices and prices in traditional retail (**tasks** of research). The study was conducted using relevant academic literature and as the major **source** of information and **methods** such as descriptive, statistical, graphical analysis, a systematic approach, and comparative analysis. Based on the **results** of the study of the empirical experience of using online price data, it can be **concluded** that at present, online trade data are intensively used by foreign statistical agencies to build high-frequency price indices and can serve as a representative source of information about the level of inflation. Nevertheless, the review revealed that the official CPI was not replicated completely in any of the existing studies due to the high complexity of data collection and maintaining the database in working order (**scientific novelty** of the work). In this regard, the **prospects** for further research of the problem are presented in the maximum possible elimination of this

shortcoming based on the previous world experience in the use of online price data and high-frequency online indicators. The results of the review can be **used** in the interests of the monetary authorities of the Russian Federation to build forecast models of inflation, considering high-frequency online data on prices.

Key words: inflation, e-commerce, inflation factors, inflation models, digitalization, consumer price index, online data, online price index, data parsing

JEL: E17 E31, E52, C81

Содержание

| | |
|--|----|
| Введение | 7 |
| 1. Мировой опыт построения и использования онлайн-индикаторов цен..... | 9 |
| 2. Анализ эмпирических исследований связи онлайн- и офлайн-цен..... | 23 |
| Заключение..... | 34 |
| Благодарности..... | 36 |
| Список использованных источников..... | 37 |

Введение

По мере развития интернет-торговли расширяются возможности использования онлайн-данных для построения своевременных высокочастотных индексов цен. Процесс сбора (парсинга) данных об онлайн-ценах, определенный Дж. Хиллен как автоматизированный процесс доступа к веб-документам и загрузки информации, ее преобразования и сохранения в структурированном формате [1], собирает и копирует данные из Интернета в электронную таблицу для последующего анализа. При низкой стоимости каждого наблюдения парсинг позволяет отслеживать цены ежедневно, избавляя исследователей от необходимости использования усредненных временных показателей. Цены регистрируются с первого дня предложения продукта потребителям до дня его снятия с продажи, что гарантирует отсутствие пробелов в собранных данных.

Онлайн-данные быстрее реагируют на шоки и затем превосходят их в официальном уровне инфляции. Это можно объяснить более низкими затратами на корректировку, более интенсивной конкуренцией среди онлайн-продавцов и высокой чувствительностью онлайн-потребителей по отношению к изменениям цен [2].

Методологию, используемую для построения индексов онлайн-цен, можно свести к двум этапам: сбор данных и их агрегирование. На первом этапе программное обеспечение для парсинга веб-страниц анализирует общедоступный HTML-код ритейлера, который включает в себя информацию о каждом продукте, такую как цена, описание, бренд и размер. Ритейлер присваивает каждому товару уникальный идентификатор (ID), который зачастую не меняется во времени. Этот идентификатор позволяет алгоритму каждый день создавать панельный набор данных по ценам на уровне различных продуктов. На втором этапе изменения цен от нескольких розничных продавцов объединяются в уникальный индекс цен с использованием весов ИПЦ. Методология агрегирования аналогична методологии, используемой национальными статистическими службами.

В последнее десятилетие национальные статистические управления развитых стран, среди которых Великобритания, США, Нидерланды, Норвегия и другие страны, начали интенсивно применять данные об онлайн-ценах в микро- и макроэкономических исследованиях, а также инициировали процесс, предполагающий возможность использования онлайн-данных для расчета официального ИПЦ. Целью и

задачами данного обзора литературы представляется систематизация эмпирических исследований, посвященных использованию данных об онлайн-ценах для построения высокочастотных онлайн-индикаторов, измерения и прогнозирования уровня инфляции, анализа ценовой динамики в онлайн-сегменте и сравнения жесткости онлайн-цен и цен в традиционной розничной торговле.

1. Мировой опыт построения и использования онлайн-индикаторов цен

Несмотря на тот факт, что объем литературы, посвященной использованию онлайн-цен для измерения и прогнозирования инфляции, быстро растет, число таких работ пока все равно остается небольшим. Одно из первых исследований было проведено А. Кавальо в рамках проекта Массачусетского технологического института «Billion prices project» (BPP), запущенного в 2008 г. [3]. Для латиноамериканских стран — Аргентины, Бразилии, Чили, Колумбии и Венесуэлы — Кавальо показал, что онлайн-цены могут эффективно использоваться в качестве альтернативного источника информации для построения высокочастотных ценовых индексов.

В период с октября 2007 по март 2011 гг. на ежедневной основе производился парсинг данных шести крупнейших супермаркетов, являющихся лидерами рынка в каждой стране. В случае Аргентины парсинг онлайн-цен осуществлялся для двух разных супермаркетов. Собранные данные классифицировались по категориям, предусмотренным официальным ИПЦ каждой страны. Основной объем выборки составил 37% весов официального ИПЦ в Чили, 48,5% — в Аргентине, 27,9% — в Бразилии, 28,5% — в Колумбии. Другими словами, данные официальных ИПЦ рассматриваемых стран не были реплицированы в полном объеме, поскольку товары, представленные в супермаркетах, не покрывают корзину товаров и услуг ИПЦ.

Для построения высокочастотного индекса онлайн-цен автором используется стандартная методология построения ИПЦ. В первую очередь рассчитывается невзвешенное геометрическое среднее изменение цен на уровне товара категории j за каждый день t (формула (1)):

$$R_{t,t-1}^j = \prod_i \left(\frac{P_t^i}{P_{t-1}^i} \right)^{\frac{1}{n_{j,t}}}, \quad (1)$$

где P_t^i — это цена товара i в момент t ;

$n_{j,t}$ — это число товаров в составе категории j , доступных для покупки в момент t .

Далее по формуле (2) осуществляется усреднение каждого индекса в момент t :

$$I_t^j = R_{1,0}^j \cdot R_{2,1}^j \cdots R_{t,t-1}^j. \quad (2)$$

Индивидуальный для супермаркета индекс в момент времени t — это взвешенное среднее арифметическое значение индексов всех категорий, которое строится по формуле (3):

$$S_t = \sum_j \frac{w^j}{W} I_t^j, \quad (3)$$

где w^j — это веса товарных категорий, предусмотренных в официальном ИПЦ для каждой страны;

W — это сумма всех весов категорий, включенных в выборку.

Согласно вышеизложенной методологии, Кавальо рассчитывает онлайн-индексы для каждой из стран и далее приводит их к аннуализированному виду.

Отдельным направлением исследований статистических служб разных стран становится разработка надежных автоматизированных процедур для оценки индексов цен на основе онлайн-данных с поправкой на изменение качества. Цель независимых исследований статистических служб при построении индексов онлайн-цен заключается в создании более своевременного и удобного способа определения динамики инфляции, на основе которого они в дальнейшем могут проводить другие макроэкономические исследования. Статистические службы Великобритании, Норвегии, Новой Зеландии, Нидерландов, США и других стран изучают онлайн-данные для разработки более дешевых и эффективных методов сбора данных для составления официальных индексов потребительских цен с использованием автоматических инструментов на основе машинного обучения и новых методологий, способных обрабатывать большие данные. Следовательно, статистические службы склонны формировать такую выборку данных об онлайн-ценах, которая позволяет покрывать большую часть потребительских расходов. Использование данных, собранных в интернете, в основном относится к определенным категориям потребительских товаров, например, бытовой электронике, товарам личной гигиены, продуктам питания.

По мнению Кавальо, традиционные подходы к построению индексов онлайн-цен оказываются неэффективными в связи с тем, что в онлайн-данных отсутствует

информация о количестве проданных товаров [4]. Отсутствие данных о количестве товаров и информации о весе на уровне их позиций в классификации продуктов вынуждают статистические службы использовать невзвешенные индексы. В связи с этим в выборку попадают те продукты, которые считаются популярными и/или репрезентативными, что вызывает неявный перевес в сторону товаров с более высокими расходами на них. Выборка для каждого продукта узко определенной товарной позиции остается фиксированной, а индексы строятся на сопоставлении цен товаров, принадлежащих к этой позиции. Основная проблема построения индексов онлайн-цен возникает, когда в выборку вводятся новые товары для замены исчезающих товаров и требуется проводить корректировку на изменение качества для измерения среднего изменения цены товаров определенной позиции.

Собираемые онлайн-данные, как правило, не содержат информацию о характеристиках продуктов, необходимую для использования методов поправки на изменение качества в явном виде, с помощью которых оценивается стоимость различий в качестве между старым и новым продуктами, попадающими в корзину. В целях нивелирования отсутствия информации о качестве товаров в собираемых данных, статистические службы начали разрабатывать модели построения индексов онлайн-цен, в которых гедонистические ценовые эффекты заменены фиксированными ценовыми эффектами для конкретных товаров, а также используются альтернативные методы корректировки на изменение качества.

В исследовании [5] аналитики Статистического управления Нидерландов Я. де Хаан и Р. Хендрик рассматривают и сопоставляют между собой три наиболее распространенные на момент исследования модели высокочастотных онлайн-индикаторов для измерения ИПЦ, перечисленные ниже.

1. Индекс онлайн-цен TPD (time-product dummy method) — это логлинейная гедоническая модель с фиксированными ценовыми эффектами для каждого товара. Модель TPD направлена на разложение цены продукта на основную цену продукта и часть цены, обусловленную периодом времени, в течение которого наблюдался продукт. Таким образом, модель TPD отражает статистическую взаимосвязь между ценами, продуктами и временем для оценки разложения цены.

2. Индекс онлайн-цен GEKS-Jevons — это среднее геометрическое всех возможных парных двусторонних индексов Джевонса (которые рассчитываются по формуле (4), представленной ниже) от базового до текущего периода. Модель GEKS-

Jevons позволяет автоматически сравнивать разницу цен во времени, что делает его более подходящим для анализа фрагментированных данных. За счет разделения длинного временного ряда на более короткие интервалы в два периода использование индекса позволяет избежать сокращения выборки в случае замены продукта.

3. Двусторонний (цепной) индекс Джевонса, используемый компанией Pricestats для построения дневных индексов онлайн-цен. Индекс онлайн-цен конструируется по формуле (4):

$$P_{CJ}^{0,t} = \prod_{i=1}^t \left(\prod_{j \in S^{i-1,i}} \frac{p_j^i}{p_j^{i-1}} \right)^{\frac{1}{n^{i-1,i}}}, \quad ((4))$$

где p_j^i — это цена товара j в период i ;

$S^{i-1,i}$ — это группа товаров, наблюдаемых в периоды i и $i-1$;

$n^{i-1,i}$ — это количество товаров в группе $S^{i-1,i}$.

Модели высокочастотных онлайн-индикаторов сравниваются с целью проверки на способность вывести с их помощью индексы цен с поправкой на качество и выяснения, полезны ли данные методы для оценки ИПЦ на основе данных об онлайн-ценах товаров, изменение качества которых не представляется серьезной методологической проблемой.

Для модели TPD предполагается, что фиксированные эффекты могут рассматриваться как гедонистические, что позволяет построить скорректированный на качество индекс онлайн-цен. Авторы рассматриваемого исследования утверждают обратное: если в гедонистической модели контролируется изменение характеристик товаров, то модель TPD не подразумевает такой возможности вследствие отсутствия вспомогательной информации о характеристиках. При этом фиктивные коэффициенты модели, в сущности, не измеряют изменение цены с поправкой на качество, а демонстрируют определенный тип изменения цены индекса Джевонса.

Метод GEKS использует цену двустороннего индекса Джевонса — в данном случае двустороннего индекса Jevons, который удовлетворяет критериям отсутствия количественных данных — и автоматически сравнивает изменение цен всех наблюдаемых товаров. При разложении формул невзвешенного индекса TPD, GEKS-Jevons и

двустороннего индекса Джевонса, авторы обнаруживают, что эти модели имеют одинаковую конструкцию. Это объясняется тем, что методы используют единственный доступный источник данных (более чем однократное изменение цен всех наблюдаемых товаров), хоть и по-разному обрабатывают эти данные. Авторы утверждают, что «вменение» отсутствующих цен, которое предполагает метод TPD, не может сильно изменить динамику тренда по сравнению с индексом GEKS-Jevons, но может способствовать снижению его волатильности.

Авторы сравнивают все три модели, используя данные об онлайн-ценах, взятые с сайта голландского интернет-магазина одежды Dutch, для трех товаров: женские футболки, мужские часы и кухонная техника. Эмпирические результаты по этим товарам подтверждают, что дневные индексы онлайн-цен могут быть очень волатильными, поскольку ни один из этих методов не позволяет получить более точные индексы онлайн-цен товаров, для которых важно изменение качества или у которых часто меняются веб-идентификаторы. Однако модель TPD легче оценить при условии, что статистический пакет или компьютерная система могут обрабатывать большие объемы данных и проводить регрессии, включающие множество временных и фиктивных индексов.

Альтернативный метод корректировки изменения качества товаров представлен в работе Статистического управления Новой Зеландии [6]. Автор предлагает использовать индекс онлайн-цен FEWS (fixed-effects with a window-splice) — это регрессионная модель, которая автоматически измеряет ценовые эффекты наблюдаемых товарных групп и контролирует наборы характеристик, соответствующих каждому продукту.

В работе предполагается, что индекс онлайн-цен FEWS будет наиболее эффективным в статистическом смысле для группы товаров (таких как бытовая электроника), имеющих более короткий жизненный цикл на рынке в связи с их стремительным технологическим развитием. Применение традиционного подхода к ценообразованию репрезентативной фиксированной корзины для таких продуктов оказывается методологически сложным по тем же причинам, поскольку для измерения изменения онлайн-цен с более высокой точностью и частотой необходим индекс с автоматизированным расчетом изменений цен при быстрой смене продукта.

Обращаясь к исследованию [5], автор конструирует индекс FEWS как модель TPD (формула (5)):

$$P_{FE}^{0,t} = \frac{\prod_{j \in S^t} (p_j^t)^{\frac{1}{n^t}}}{\prod_{j \in S^0} (p_j^0)^{\frac{1}{n^0}}} \exp(\bar{\hat{\gamma}}^0 - \bar{\hat{\gamma}}^t), \quad (5)$$

где p_j^t — это цена товара i в период t ;

S^t — это категория товаров, наблюдаемых в период t ;

n^t — это количество товаров в категории S^t ;

$\bar{\hat{\gamma}}^t = \sum_{j \in S^0} \hat{\gamma}_j / n^t$ — это выборочное среднее значение оцененных фиксированных эффектов в момент времени t . В свою очередь, фиксированный эффект представляет собой сумму параметров основных ценовых эффектов и всех комбинаций характеристик продукта. Оценка ценовых эффектов соответствует оцененному коэффициенту регрессии.

$\hat{\gamma}^0$ — это среднее значение оцененного коэффициента регрессии с фиксированными эффектами в момент времени 0 .

Для пересмотра движения онлайн-цен за предыдущий квартал требуется автоматизированная система обновления оценок с фиксированными эффектами. Это позволит учесть изменения цен на новые продукты и включить их в индекс с фиксированными эффектами. Для включения неявных изменений наряду с движением за последний период используется два наблюдения и «оконное соединение» (window splice), позволяющее обновлять оценки с фиксированными эффектами для всех других продуктов, наблюдаемых в окне оценки.

Для обновления оценок с фиксированными эффектами для всех товаров автор применяет к модели TPD «оконное соединение» по формуле (6):

$$P_{WS}^{0,d+t} = P_{[0,d]}^{0,1} \times P_{[t,d+1]}^{1,2} \times \dots \times P_{[t,d+1]}^{t,d+1}, \quad (6)$$

где $P_{[0,d]}^{t,d+1}$ — это индекс цен от периода t до периода $d+1$ с использованием окна оценки $[t, d + 1]$.

Таким образом, индекс FEWS позволяет наблюдать движение цен за период конкретного окна оценки, а не только за последний период. Комбинация индекса с

фиксированными эффектами и «оконного соединения» позволяет создавать непере-сматриваемые индексы цен с поправкой на качество для онлайн-данных, характеристики которых недоступны. Соответственно, когда в индекс онлайн-цен включается новый продукт, модель с фиксированными эффектами определяет, какая часть ценового разрыва между новым товаром и товаром, снятым с производства, обусловлена различиями в качестве продуктов, а какая отражает инфляцию. Преимущество этого подхода заключается в том, что в сравнении с индексом TPD модель FEWS благодаря «оконному соединению» позволяет учитывать ненаблюдаемые изменения цен на новые товары с запаздыванием. В долгосрочной перспективе индекс FEWS устраняет любое систематическое смещение, вызванное отсутствием методов поправки на корректировку качества в неявном виде, для изменения цен на новые и исчезающие товары.

Эмпирические результаты этой работы, как и в исследовании [5], демонстрируют возможность высокой волатильности индексов онлайн-цен из-за отсутствия количественных данных и характеристик, необходимых для идентификации товаров. Несмотря на это, модель FEWS позволяет получить полностью скорректированный на изменение качества индекс цен при условии наличия информации о количестве новых или исчезающих товаров по сравнению с предыдущим периодом на уровне подробной спецификации товара. Автор также отмечает, что индекс FEWS может не работать в случае продуктов, даты начала и окончания жизненного цикла которых сильно зависят от сезонности (таких как женская одежда). В данном случае для успешной оценки индекса FEWS должно соблюдаться условие, чтобы спецификация товаров по категориям совпадала между двумя последовательными периодами оценки.

По мнению автора, использование онлайн-данных может быть перспективным с точки зрения официальных статистических агентств. Чтобы избежать высокой волатильности индексов онлайн-цен вследствие отсутствия характеристик и количественных данных товаров, автором также предполагается разработка гибридного подхода, при котором высокочастотные индексы, полученные на основе онлайн-данных, могут быть откалиброваны совместно с менее частотными официальными показателями, полученными на основе более полной информации (например, данных контрольно-кассовой техники и данных, собранных на местах).

Резюмируя выводы работ [5] и [6] можно заключить, что основная проблема методологии построения индексов онлайн-цен заключается в том, что модели не позволяют идентифицировать качественно схожие товары, которые автоматически предполагаются разными. Последнее приводит к смещению индекса цен в сторону понижения. Это связано с тем, что индексы онлайн-цен фиксируют видимое снижение цен в течение срока службы каждого продукта, но не могут определить ненаблюдаемый рост цен, с которым сталкиваются потребители, когда на рынке появляются новые модели товаров. Дополнительный метод расчета индексов онлайн-цен, позволяющий компенсировать описанную выше погрешность при составлении выборки, представлен в работе [7].

Используемая в исследовании Бертолотто база онлайн-данных по ценам на одежду предоставлена частной компанией PriceStats. Эта компания производила ежедневный парсинг данных крупных ритейлеров товаров одежды Испании, США, Великобритании, Нидерландов и Германии. Онлайн-данные по ценам собирались одновременно с подробной информацией о размещенных на сайтах компаний товарах (название, марка, модельные характеристики, веб-идентификатор). Преимущество этой базы состоит в том, что товары, которых нет в наличии, из нее исключаются, поэтому онлайн-индексы включают только товары, которые потребители могут купить в каждый момент времени. Для уравнивания компонентов цены каждого товара онлайн-цены включают налог на добавленную стоимость, но не включают транспортные расходы.

Метод наиболее близкого соответствия (*the closest match method*), предлагаемый Бертолотто, имитирует процесс принятия решений специалистом, который рассматривает товары, подлежащие принудительной замене. Метод сопоставляет каждый вновь представленный товар с его ближайшим из существующих альтернативным товаром в соответствующей базе данных. Затем он сравнивает первую цену каждого нового товара с последней ценой его ближайшего альтернативного товара и записывает это как первое изменение цены нового товара. Выбор наиболее подходящего альтернативного товара выполняется в два этапа.

Цель первого этапа — идентифицировать товары из предыдущего сезона и убедиться, что они регулярно присутствуют в товарообороте, избегая конкретных продуктов, которые присутствуют на рынке только в течение очень короткого периода времени. Товары отфильтровываются по следующим критериям (*рисунок 1*):

- альтернативный товар присутствовал в товарообороте более 90 дней ($t-90$);
- альтернативный товар присутствовал в товарообороте менее 365 дней ($t-365$);
- альтернативный товар присутствовал в выборке не менее 10 дней.

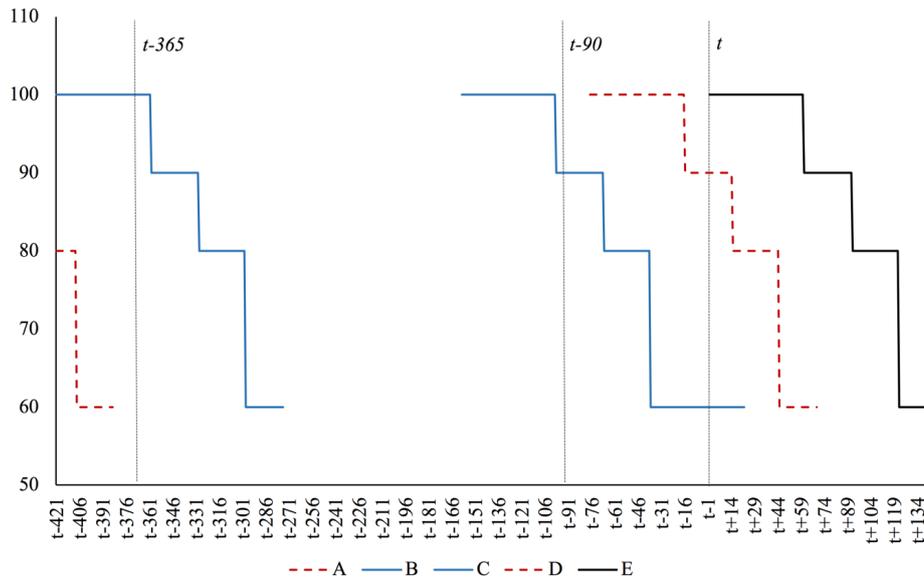


Рисунок 1. Идентификация альтернативных товаров с наиболее близким соответствием

Источник: [7].

Примечания:

- по оси абсцисс отложены дни присутствия товаров в базе данных, по оси ординат — цена каждого товара;
- рисунок демонстрирует, какие товары соответствуют критериям, описанным выше. Товар Е вводится на рынок в момент t , поэтому фильтр ищет альтернативные товары, которые были в базе до этой даты. Товары В и С удовлетворяют критериям первого этапа.

Таким образом, на первом этапе отфильтровывается большинство продуктов, присутствующих в базе данных, а оставшиеся распределяются в группу товаров-кандидатов.

На втором этапе с помощью формулы (8) вычисляется оценка характеристики каждого возможного альтернативного товара-кандидата на наличие схожести с вы-

бывшим из корзины товаром (автор использует поисковую систему Elasticsearch, которая автоматически сопоставляет между собой результаты оценок характеристик товаров):

$$S(\mathbf{q}, \mathbf{d}) = r(\mathbf{q}, \mathbf{d}) * \sum_{w=1}^N wf(w \text{ in } \mathbf{d}) \cdot idf(w)^2 \cdot fln(w, \mathbf{d}), \quad (7)$$

где $r(q, d)$ — коэффициент релевантности описания товара q и альтернативного товара d . Этот коэффициент рассчитывается как количество совпадающих слов в описании товаров-кандидатов, разделенное на общее количество слов в поисковом запросе. Соответственно, чем больше совпадающих слов в описании товаров-кандидатов, тем выше его релевантность;

$f(w \text{ in } d)$ представляет частоту слова w в описании товара. Этот фактор придает более высокую значимость описаниям, в которых одно и то же слово повторяется дважды или более. Автор рекомендует установить это значение равным единице;

$idf(w)$ — обратная частота описания слова w : это логарифм количества описаний товаров в наборе продуктов, деленный на количество описаний, содержащих это слово. Чем чаще слово встречается в описаниях в наборе товаров, тем меньше вес этого слова;

$fln(w, d)$ — это обратный квадратный корень из количества слов в описании продукта: чем ниже показатель этого коэффициента, тем выше его вес.

Наилучший альтернативный товар, рассчитанный с помощью формулы (8), включается в индекс, когда старый продукт снимается с производства. Таким образом, Бертолотто предлагает методологию, позволяющую наиболее оперативно формировать релевантную выборку потребительской корзины для построения индексов онлайн-цен. Метод наиболее близкого соответствия автоматически ищет лучший альтернативный товар, когда он появляется на рынке, в то время как традиционный подход предполагает ручной поиск замены выбывшего продукта и не позволяет своевременно включать в выборку новые товары. Автор также отмечает, что в случае, когда старый товар снимается с производства, а новый товар, его заменяющий, еще не успел появиться на рынке, метод позволяет подобрать товар, наиболее подходящий для включения в индекс.

Индекс онлайн-цен, основанный на методе наиболее близкого соответствия (closest-match index), рассчитывается Бертолотто как невзвешенное среднее геометрическое всех ценовых соотношений товаров (для этого рассчитывается разница между первой ценой за единицу каждого нового товара и последней ценой его ближайшего альтернативного товара). Для построения индекса онлайн-цен Бертолотто применяет простую методологию, однако метод наиболее близкого соответствия может масштабироваться до неопределенного количества товаров без ручного вмешательства. Если традиционные методы ищут замену товара, снятого с производства, то метод наиболее близкого соответствия ищет альтернативный товар каждый раз, когда на рынке появляется новый товар.

Поскольку индексы онлайн-цен содержат репрезентативные данные о розничных транзакциях и аналогично традиционному ИПЦ отслеживают потребительские цены в основном из одних и тех же секторов экономики, это позволяет им повторять динамику официального ИПЦ, несмотря на их высокую волатильность. Вследствие того, что традиционные методы расчета ИПЦ могут задерживать запись определенных ценовых изменений, для отражения их в официальной статистике требуется больше времени. Данные об онлайн-ценах, напротив, моментально отражают ценовые изменения и могут предвосхищать их в ИПЦ. В качестве одного из направлений исследований данных об онлайн-ценах выступает разработка инструментов прогнозирования уровня ИПЦ на основе индексов онлайн-цен. Высокая частота онлайн-цен позволяет улучшить точность прогнозов моделей, в которые они включены.

Исследование о прогностических свойствах индексов онлайн-цен, проведенное в 2017 г., было представлено в работе [8]. Авторы показывают, что в экономиках Австралии, Канады, Франции, Германии, Греции, Ирландии, Италии, Нидерландов, Великобритании и США вневыборочные прогнозы инфляции с использованием онлайн-данных статистически эффективнее многих альтернативных моделей прогнозирования.

Авторы используют онлайн-индексы, разработанные компанией PriceStats для измерения совокупного уровня инфляции в каждой стране. Их ежедневная частота (данные собирались ежедневно с июля 2008 г. по сентябрь 2016 г.) служит преимуществом по сравнению с другими показателями, используемыми для прогнозирования инфляции, поскольку онлайн-индексы позволяют фиксировать изменения инфляции уже до наступления конца месяца.

Онлайн-модель прогнозирования уровня инфляции ИПЦ на месячном горизонте предлагается авторами в уравнении (9):

$$\mathbf{E}_{t-1}\mathbf{p}_t = \hat{\alpha} + \sum_{i=1}^p \hat{\beta}_{t-i}\mathbf{p}_{t-i} + \sum_{i=0}^p \hat{\theta}_{t-i}\mathbf{f}_{t-i} + \sum_{i=0}^p \hat{\gamma}_{t-i}\mathbf{o}_{t-i} + \sum_{i=0}^p \hat{\eta}_{t-i}\mathbf{o}\mathbf{f}_{t-i}, \quad (8)$$

где \mathbf{p}_t — это официальный ИПЦ;

\mathbf{f}_t — предиктор инфляции, основанный на офлайн-ценах на бензин и дизельное топливо;

\mathbf{o}_t и $\mathbf{o}\mathbf{f}_t$ — индекс онлайн-цен на бензин и дизельное топливо и темпы инфляции онлайн-цен на бензин и дизельное топливо за период t соответственно.

Авторы используют данные об онлайн-ценах на бензин и дизельное топливо, исходя из предположения, что они представляют собой наиболее точный предиктор инфляции.

Для прогнозирования уровня инфляции в период t авторы используют ежедневные данные об онлайн-ценах, опубликованные до 15 числа месяца $t-1$ (то есть за месяц до публикации официального ИПЦ). Индексы корректируются с учетом сезонных колебаний при помощи дамми-переменных.

Далее авторы рассчитывают все возможные случаи оценки уравнения (9) и выделяют 100 наиболее эффективных моделей на основе их RMSE. Для любого заданного месяца авторы усредняют прогнозы этих моделей. Результирующий прогноз сравнивается с ИПЦ аналогично сравнению ИПЦ с любой отдельной моделью, оцененной в уравнении (9).

Результаты расчета кумулятивного распределения абсолютных ошибок прогноза отдельных моделей показывают, что использование онлайн-цен снижает вероятность больших ошибок прогнозов для Австралии и США. Аналогичный результат справедлив и для других стран. Прогноз на 1 месяц вперед в среднем на 14% точнее моделей, использующих офлайн-данные по ценам, и более чем в три раза превышает точность прогнозов Bloomberg.

Кроме того, с помощью модели (9) авторы рассчитывают квартальный прогноз инфляции в США, который строится на основе прогнозов на один, два и три месяца вперед из заданного квартала базовой спецификации, а затем пересчитывается на год.

При сравнении результатов онлайн-модели с результатами средних и медианных показателей прогностической модели «Опрос профессиональных прогнозистов» (SPF) Федерального резервного банка Филадельфии и показателями AR(p), кривой Филлипса и RW (Random walk) авторы констатируют, что качество прогнозов, полученных на основе онлайн-индексов, существенно превосходит качество прогнозов, полученных при помощи альтернативных методов.

Результаты использования ежедневных онлайн-данных, демонстрирующие повышение точности краткосрочных прогнозов инфляции, представлены в исследовании Центрального банка Швеции (Риксбанка) [9]. Авторы разработали систему краткосрочного прогнозирования инфляции путем сбора данных об онлайн-ценах для товаров на фрукты и овощи крупных онлайн-ритейлеров, расположенных в Швеции. Система была установлена на сервере Linux, для которого было разработано приложение, собирающее данные для 4 онлайн-магазинов. Для создания категорий продуктов была создана база данных на основе классификации по кратким описаниям товаров. Такая классификация позволила сопоставить категории, предусмотренные в официальном ИПЦ, с критериями, по которым производился парсинг данных по онлайн-ценам.

Авторы рассчитывают онлайн-индекс цен на фрукты и овощи с использованием весов, предусмотренных официальным ИПЦ. Несмотря на тот факт, что основной объем выборки для онлайн-индекса покрыл всего 3% корзины ИПЦ, авторы показали, что онлайн-индекс способен предвосхитить динамику волатильности основного индекса для фруктов и овощей. При сравнении с существующими альтернативными моделями прогнозирования Риксбанка взвешенный ежемесячный онлайн-индекс демонстрирует наименьшее среднеквадратическое отклонение (1,6 против 2,0 RMSE).

Как показал приведенный выше обзор литературы, статистические ведомства разных стран начинают интенсивно использовать данные интернет-торговли для построения высокочастотных индексов цен. Методология построения индексов онлайн-цен подразумевает, как правило, два этапа: парсинг данных на площадках крупных ритейлеров и агрегирование онлайн-цен по товарным категориям с учетом весов официального ИПЦ [3], [5], [6]. Ввиду отсутствия в собранных онлайн-данных идентификаторов товаров и информации об их количестве также предполагается агрегирование

данных об онлайн-ценах, используя классификатор, построенный на основе текстовых описаний для выявления качественно схожих товаров [7].

Разработка высокочастотных онлайн-индексов перспективна с точки зрения отражения фактических краткосрочных колебаний цен, на основании которых можно построить ежемесячную/квартальную оценку текущей динамики ИПЦ, а также для улучшения прогноза инфляции [8], [9]. В частности, в исследовании [8] показано, что качество прогнозов официальной инфляции, основанных на данных об онлайн-ценах, существенно возрастает по сравнению с традиционными методами прогнозирования.

2. Анализ эмпирических исследований связи онлайн- и офлайн-цен

В некоторых рассмотренных выше эмпирических работах приводится сравнение динамики высокочастотных индексов онлайн-цен с эквивалентными официальными показателями.

Результаты исследования [3] Кавальо демонстрируют, что в Бразилии, Чили, Колумбии и Венесуэле онлайн-индексы способны отразить как средний уровень, так и динамику официальной инфляции. Автор наблюдает наилучшее совпадение в Чили, где средняя годовая инфляция в рассматриваемом интервале составляет 3,00% (онлайн-индекс) и 3,19% (официальная инфляция), а корреляция между рядами онлайн-индекса и общего тренда официальной инфляции равна 0,97. Для Колумбии онлайн-индекс повторяет снижение инфляции с 8% в конце 2008 г. до примерно 3% в конце 2009 г. В Бразилии и Венесуэле онлайн-индексы хоть и соответствуют динамике инфляции, однако ряды менее синхронизированы во времени. Кавальо объясняет это меньшей репрезентативностью данных: если товары, представленные в супермаркетах в Чили, занимают 27% в индексе онлайн-цен, то доля товаров бразильского онлайн-ритейлера составляет всего 15%.

В Аргентине в период с октября 2007 г. по март 2011 г. онлайн-индекс вырос более чем на 100%, в то время как официальный индекс цен вырос всего на 35%. Аналогично инфляция, рассчитанная на основе онлайн-данных, постоянно превышала официальную инфляцию в 2-3 раза (в среднем 20,14% против 8,38% соответственно). Отклонение индекса онлайн-цен от официальной инфляции наблюдается на длительном периоде. Автор не смог объяснить столь значительные расхождения онлайн-данных с официальными. Необъяснимое расхождение между данными об онлайн-ценах и официальными данными свидетельствует о вероятном манипулировании официальной статистикой. Таким образом, в случае недостаточной надежности официальных данных высокочастотные индексы онлайн-цен могут использоваться в качестве альтернативного источника статистической информации.

В исследовании [10], проведенном в 2015 г., Кавальо снова использует данные об онлайн-ценах для изучения влияния погрешности измерений на три общих статистических параметра жесткости цен:

- продолжительность ценовых изменений;

- форма распределения величины ценовых изменений;
- форма функции риска изменения цены, которая демонстрирует условную вероятность изменения цен с точки зрения времени.

Данные об онлайн-ценах позволяют Кавальо доказать, что погрешности изменений могут быть связаны с характеристиками выборки данных официального ИПЦ и данных контрольно-кассовой техники. Подобная закономерность обнаруживается автором при использовании на данных формы распределения величины изменений цен.

Распределение изменений цен проводится с использованием онлайн-данных и данных контрольно-кассовой техники, полученных от одного из крупнейших американских продовольственных ритейлеров, с использованием одинакового адреса торговой точки и периода выборки. Распределение изменений цен для обоих типов данных генерируется фиксированными ценами на сырье, однако результаты распределения сильно отличаются. Если для онлайн-данных наблюдается бимодальное распределение изменения цен, сосредоточенное около нулевого процента, то данные контрольно-кассовой техники демонстрируют колоколообразное (унимодальное) распределение с относительно значимой долей изменения цен. Кавальо находит этому очевидное объяснение: данные контрольно-кассовой техники представлены в виде средних значений цен, наблюдавшихся в течение недели, в то время как частота данных об онлайн-ценах ежедневная.

На основе других эмпирических сравнений данных в исследовании делается вывод, что онлайн-цены имеют большую продолжительность наблюдений, характеризуются меньшими изменениями, близкими к нулю, а также функциями риска изменения цены, которые первоначально увеличиваются с течением времени. Автор утверждает, что расхождения данных об онлайн-ценах с альтернативными данными связаны с усреднением по времени и условными ценами на временно отсутствующие товары в данных контрольно-кассовой техники и данных ИПЦ.

В работе [11] Кавальо совместно с Б. Нейманом и Р. Ригобоном показывает, что онлайн-цены могут быть использованы для изучения уровней и динамики реального обменного курса. Онлайн-индикаторы относительных цен и реального обменного курса построены PriceStats на основе данных об онлайн-ценах более чем 250 товаров продуктов питания, электроники и топлива. В случае Аргентины авторы рассчитывают обменный курс на черном рынке.

Эмпирические результаты исследования указывают на то, что динамика относительных цен тесно связана с колебаниями номинального обменного курса. Так, в период с 2008 г. по 2011 г. относительные цены Австралии падали по мере укрепления национальной валюты, при этом в 2013 г. относительные цены выросли, когда австралийский доллар начал обесцениваться. Онлайн-индикаторы, представленные для Аргентины, демонстрируют похожую динамику. Рост относительных цен сопровождается постепенным снижением уровня обменного курса на черном рынке и скачкообразным — номинального обменного курса. Такое наблюдение подтверждается авторами на примере других стран, однако, как номинальный обменный курс, так и уровень относительных цен могут оказывать влияние друг на друга. В случае Австралии динамика розничных цен была подвержена шокам номинального курса, а на примере Аргентины корректировка номинального курса происходит в ответ на изменение розничных цен.

Кроме того, в течение длительного времени авторы наблюдают значимое расхождение динамики реального обменного курса Аргентины с номинальным. Эти отклонения были связаны с тем периодом, когда аргентинское правительство фальсифицировало официальную статистику [3]. При продолжительном росте цен официальный обменный курс оставался фиксированным, на что указывают динамика обменного курса на черном рынке и динамика реального обменного курса. Результаты также демонстрируют внезапные корректировки номинального курса, которые произошли в связи с девальвацией валюты в январе 2014 г. и декабре 2015 г. Скачки официального обменного курса выровняли его по отношению к уровню относительных цен (предполагаемому обменному курсу по ППС). Таким образом, онлайн-данные могут успешно использоваться для изучения жесткости цен и обеспечить более точные показатели степени отклонения реальных обменных курсов от номинальных в режиме реального времени.

Идейным продолжением представленных выше работ стало исследование [12], в котором Кавальо сравнивает онлайн- и офлайн-цены 56 крупных многоканальных ритейлеров в 10 странах: Аргентине, Австралии, Бразилии, Канаде, Китае, Германии, Японии, Южной Африке, Великобритании и США. Исследование показало, что временные ряды онлайн-цен были на 72% идентичны фактическим ценам. При этом изменения цен не синхронизированы, но имеют схожую частоту и средние масштабы.

Эмпирические данные о частоте и размере изменений онлайн- и офлайн-цен, свидетельствуют о том, что поведение онлайн-цен аналогично поведению традиционных цен, однако изменения цен обоих типов данных происходят не синхронизировано во времени. Последнее, по мнению авторов, связано с тем, что данные об онлайн-ценах быстрее реагируют на рыночные шоки и опережают ценовую динамику в офлайн-сегменте. Помимо этого, эмпирические результаты исследования также показали, что данные контрольно-кассовой техники, ежедневный сбор которых затруднителен, потенциально могут быть заменены онлайн-ценами. Авторы подкрепляют это суждение тем, что ценообразование в онлайн-сегменте, как правило, имеет более стабильную динамику.

Сравнение динамики индекса наибольшего соответствия (closest-match index), разработанного Бертолотто в [7], с официальными показателями ИПЦ исследуемых стран представлено на *рисунке 2* на примере Испании.

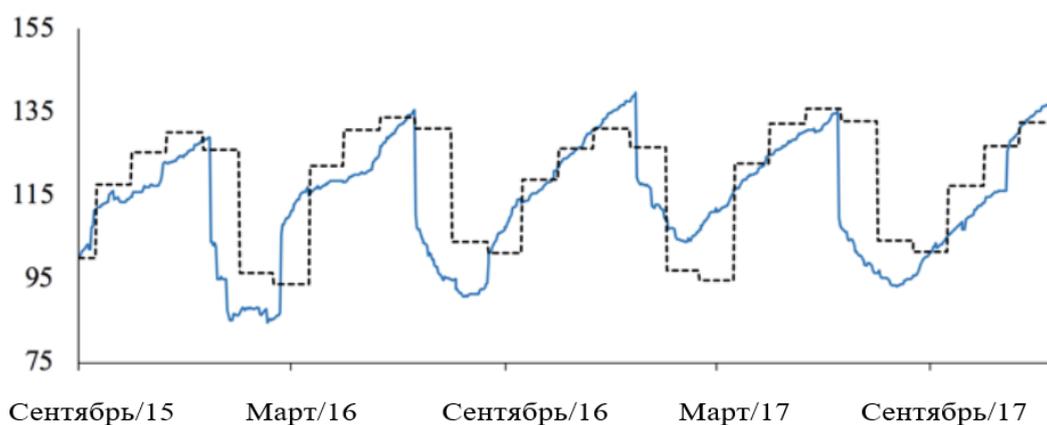


Рисунок 2. Индекс наибольшего соответствия в сравнении с официальным ИПЦ Испании

Источник: [7].

Примечание: на рисунке приведена динамика онлайн-индекса наибольшего соответствия (Closest Match) и официального ИПЦ (CPI) для категории «одежда».

Индекс онлайн-цен, рассчитанный для Испании, начал расти в сентябре 2015 г. По данным индекса онлайн-цен, спустя три месяца динамика изменения индекса составила 29,5%. Автор отмечает, что статистическое управление Испании также зарегистрировало кумулятивный рост официального ИПЦ в размере 30%. Индекс онлайн-цен начал падать в январе 2016 г., и ИПЦ продемонстрировал аналогичное изменение, однако официальные показатели инфляции были опубликованы только 15 февраля.

Главный вывод из этого рисунка заключается в том, что обе методологии дают очень похожие результаты. Вышеописанное наблюдение подтверждается на данных для США, Великобритании, Нидерландов и Германии. При этом Бертолотто отмечает, что небольшие отклонения между индексами онлайн-цен и официальными данными могут объясняться методологическими различиями и различиями в выборке. Во-первых, для построения официального ИПЦ используется более сложный подход (например, гедонистические модели). Во-вторых, в выборке исследования участвовало ограниченное количество розничных продавцов. В связи с этим показатели онлайн-индекса наибольшего соответствия были более волатильными по сравнению с оценками ИПЦ.

Тем не менее высокочастотные индексы онлайн-цен способны повторять динамику и уровень официального ИПЦ. Это объясняется тем, что онлайн-данные быстрее реагируют на шоки и затем превосходят их в официальном уровне инфляции. При этом онлайн-ценообразование принципиально отличается от ценообразования в традиционной розничной торговле, что может свидетельствовать о несходстве показателей индекса онлайн-цен и официального ИПЦ.

Вспышка COVID-19 и меры, предпринимаемые для ограничения его распространения (в частности, закрытие офлайн магазинов), привели к резким изменениям расходов по всем категориям товаров, а также не позволили статистическим агентствам вручную собирать данные о ценах в офлайн магазинах. Это сделало онлайн-данные и высокочастотные ценовые индексы полезным источником данных для мониторинга экономической активности и динамики цен в режиме реального времени.

Исследование о возможности использования онлайн-цен для измерения и прогнозирования инфляции было представлено в [13]. В работе предлагается полностью автоматизированная система сбора и анализа данных об онлайн-ценах для категории «продукты питания и безалкогольные напитки» в Польше. Анализ данных использовался для выявления механизма ценообразования в супермаркетах во время пандемии.

Ежедневный парсинг данных об онлайн-ценах производился с июля 2015 г. по август 2020 г. Расчет для 20 тыс. продуктов питания основан на списке репрезентативных позиций для каждой группы продуктов (то есть подкатегорий, находящихся на самом низком уровне агрегирования). Были отобраны отдельные товары, отражающие изменение цен в каждой из групп. Всего в исследование было

включено 205 отдельных продуктов, которые охватывают все 86 групп, представляющих цены на продукты питания.

Дневные индексы онлайн-цен по группам товаров на низших уровнях агрегирования (86 групп в «продуктах питания и безалкогольных напитках») рассчитываются как среднее геометрическое ежедневных цен на отдельные продукты (формула (10)):

$$P_{it}^d = N \sqrt{x_{1t}^1 \cdot x_{2t}^1 \cdot \dots \cdot x_{\tau t}^1 \cdot \dots \cdot x_{1t}^n \cdot x_{2t}^n \cdot \dots \cdot x_{\tau t}^n} \text{ для } \tau \leq d, \quad (9)$$

где x_{at}^b — это цена, зафиксированная в день a месяца t для продукта b , принадлежащего к группе i ;

n — это количество продуктов, принадлежащих к группе i ;

τ — это количество ежедневных ценовых предложений в месяце t ;

$N = n \cdot \tau$ — это общее количество индивидуальных цен, зарегистрированных для всех n продуктов в течение месяца t .

Агрегированный индекс онлайн-цен для каждой подкатегории продуктов питания рассчитывается по формуле Ласпейреса (11):

$$P_{kt}^d = \sum_i P_{it}^d \cdot w_{i,t-12}, \quad (10)$$

где P_{kt}^d — это индекс потребительских цен продовольственной подкатегории k ;

P_{it}^d — это индивидуальный индекс цен для группы i , относящейся к подкатегории k , в месяце t ;

$w_{i,t-12}$ — это веса, используемые при расчете официального ИПЦ.

На основе индексов онлайн-цен рассчитываются относительные изменения цен по сравнению с предыдущим месяцем для подкатегории k , то есть инфляция в месячном исчислении (формула (12)):

$$\pi_{kt}^M = \frac{P_{kt}^d}{P_{kt-1}} \cdot 100. \quad (11)$$

Аналогичным образом рассчитывается инфляция в годовом исчислении (формула (13)):

$$\pi_{kt}^Y = \frac{P_{kt}^d}{P_{kt-12}} \cdot 100. \quad (12)$$

Показатели ежемесячной и годовой инфляции на продукты питания оцениваются в режиме реального времени и обновляются ежедневно.

Рассчитывая RMSE показателей инфляции, оцениваемых с помощью онлайн-данных, и уровень их корреляции с официальной статистикой, авторы обнаруживают, что точные прогнозы инфляции цен на продукты питания предоставляются за несколько дней до окончания отчетного периода PCSO. Наблюдается тенденция к усилению корреляции и одновременному уменьшению RMSE по мере увеличения количества дней, необходимых для публикации официальных показателей инфляции. Как правило, точность и опережающая динамика показателей онлайн-инфляции сохраняются для всех 10 основных подкатегорий. Авторы констатируют, что методология позволяет получить месячный индекс онлайн-цен, аналогичный официальному ИПЦ, на продукты питания за две недели до окончания отчетного периода Центрального статистического управления Польши (Poland Central statistics office, PCSO) и примерно за 30 дней до публикации официальных данных.

Подход, используемый в этом исследовании, позволил достаточно точно отслеживать официальную продовольственную инфляцию в течение периода, охваченного пандемией COVID-19 (рисунок 3).

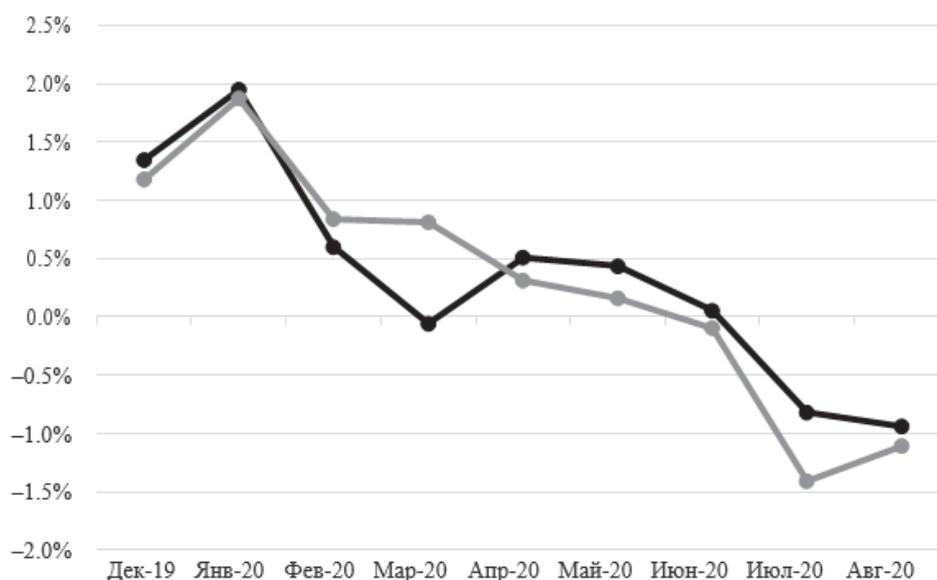


Рисунок 3. Сравнение динамики инфляции в Польше, оцененной на основе онлайн-данных, с аналогичными данными официальной статистики во время пандемии COVID-19

Источник: [13].

Примечания:

- на графике черная линия отражает показатель месячной инфляции на основе онлайн-данных, серая линия отображает динамику официального ИПЦ;
- по горизонтальной оси отложены месяцы, по вертикальной оси — величина инфляции в процентных пунктах.

По результатам исследования показатель месячной инфляции на основе онлайн-данных оказался более репрезентативным, чем официальная статистика. Например, наибольшее расхождение между двумя показателями можно наблюдать в марте 2020 г., когда Центральное статистическое управление Польши сообщило о трудностях со сбором информации о ценах в офлайн магазинах по причине введения в стране локдауна. После того, как PCSO скорректировало свои методы сбора цен, расхождение между оценками снизилось с апреля по июнь 2020 г. до 0,21 п. п. Таким образом, как показано в работе [13], онлайн-данные могут быть полезными для оперативной диагностики шоков спроса и предложения в экономике.

Используя онлайн-данные, авторы наблюдают, как изменялось ценообразование супермаркетов во время пандемии COVID-19. Во-первых, онлайн-продажи про-

дуктов питания почти удвоились в период с января по апрель 2020 г. из-за распространения пандемии. Во-вторых, хоть в целом цены на продукты питания не изменились, масштабы рекламных акций значительно уменьшились. Согласно расчетам, выполненным на базе данных этого исследования, количество товаров со скидкой уменьшилось на 31,3% в период с конца февраля по середину марта, когда в Польше был официально введен режим локдауна. Акции были частично возобновлены до конца месяца: количество товаров со скидкой оказалось на 20,3% больше, чем в середине марта.

Более детальное исследование, посвященное анализу динамики онлайн-цен в период локдауна, представлено в работе [14] для Швейцарии.

Ежедневный парсинг данных производился с мая 2019 по июнь 2020 гг. для товаров из категорий «электроника», «канцтовары», «одежда и обувь», «продукты питания» «топливо коммунально-бытового назначения», «мебель». Для построения потребительской корзины авторами используются веса для соответствующих категорий товаров, предусмотренные Федеральным статистическим управлением Швейцарии (Swiss Federal statistical office). Чтобы зафиксировать изменения потребительских предпочтений, связанные с введением режима локдауна в стране, с января 2020 г. в качестве весов в индексе авторы используют данные ежедневных расходов по дебетовым картам. База данных об объеме и назначении транзакций по дебетовым картам публикуется «Monitoring Consumption Initiative for Switzerland».

Авторы рассчитывают индекс онлайн-цен в соответствии с формулой для индекса Фишера. Четыре этапа расчета индекса описаны далее.

1. Для каждой товарной категории строится индекс Джевонса (формула (14)):

$$P_j^t = \prod_{i=1}^N \left(\frac{P_i^t}{P_i^0} \right)^{\frac{1}{n}}, \quad (13)$$

где $i = 1 \dots N$ — это совокупность товаров в каждой категории;

$j = 1 \dots J$ — это товарные категории;

P_i^0 — это цена товара i , взятая в базовый период. За базовый период взят период с 9 по 15 марта 2020 г., то есть это неделя, предшествующая введению режима локдауна в Швейцарии.

2. Далее по формуле (15) конструируется ежедневный индекс цен Ласпейреса:

$$P_t^{Las} = \sum_{i=1}^J \frac{P_j^t}{P_t^0} \cdot w_j^{0,CPI}, \quad (14)$$

где P_j^t — это индекс цен Джевонса (формула (14));

$w_j^{0,CPI}$ — это веса, используемые при расчете официального ИПЦ.

3. Ежедневный индекс цен Пааше строится по формуле (16):

$$P_t^{Paa} = \left[\sum_{i=1}^J \left(\frac{P_j^t}{P_t^0} \right)^{-1} \cdot w_j^{t,Covid19} \right]^{-1}, \quad (15)$$

где $w_j^{t,Covid19}$ — это веса, представляющие собой долю расходов по дебетовым картам на соответствующую категорию в общих расходах.

4. Индекс Фишера для онлайн-цен рассчитывается как среднее геометрическое индексов Пааше и Ласпейреса (формула (17)):

$$P_t^{Fis} = (P_t^{Las} \cdot P_t^{Paa})^{0,5}. \quad (16)$$

В целях снижения зашумленности данных к индексу Фишера применяется формула скользящего среднего с лагом в 7 дней. Сопоставляя результаты, авторы наблюдают, что в 2019 г. индекс онлайн-цен в значительной степени повторял динамику и уровень официального ИПЦ. После введения режима локдауна авторы зафиксировали мгновенное снижение индекса онлайн-цен примерно на 0,4 процентных пункта, в то время как официальный ИПЦ продемонстрировал соответствующее снижение только спустя 6 недель.

Результаты данного исследования подтверждают тезис [13] о том, что в случае внешнего шока онлайн-данные позволяют предсказать изменения официального ИПЦ. В ситуации отсутствия шоков авторы не обнаруживают значимого различия между онлайн- и офлайн-ценами. Авторы приходят к выводу, что данные об онлайн-ценах (примерно 25% всей выборки) могут использоваться в качестве исходных данных для расчета официального ИПЦ.

Резюмируя выводы [13] и [14], можно утверждать, что в результате высокой неопределенности, вызванной пандемией COVID-19, а также увеличения доли интернет-торговли онлайн-данные и высокочастотные индексы оказались репрезентативным источником информации для интерпретации инфляционного тренда и диагностирования внешних шоков. Как показывают работы [13] и [14], индексы онлайн-цен оказались способны более точно отображать ценовые тенденции во время введения режима локдауна.

Заключение

Как следует из исследования [12], данные об онлайн-ценах могут служить репрезентативным источником информации об уровне инфляции, поскольку их поведение несущественно отличается от поведения цен в традиционной розничной торговле. При этом онлайн-данные имеют ряд преимуществ по сравнению с данными контрольно-кассовой техники и офлайн-ценами, среди которых выделяются высокая частота их изменений [15], способность быстро реагировать на рыночные шоки и опережать их в ценовой динамике офлайн-сегмента [7]. Таким образом, онлайн-данные могут эффективно использоваться в качестве альтернативного источника статистической информации [3], для обнаружения доминирующих шоков в экономике [7], [14], изучения жесткости цен [11], [15], а также для выявления механизма ценообразования онлайн-ритейлеров [13].

На основании обзора международного опыта выявлены следующие направления использования данных об онлайн-ценах странами:

- построение стандартных высокочастотных индексов цен и получение альтернативного источника статистической информации [3], [5], [6], [16];
- разработка новых методов построения индексов цен, позволяющих компенсировать погрешности выборки [6], [7];
- выявление доминирующих шоков в экономике [7];
- повышение точности прогноза инфляции на коротком горизонте [8], [9];
- выявление особенностей механизма ценообразования онлайн-ритейлеров, в том числе во время пандемии COVID-19 [13], [14];
- корректировка оценки макроэкономических (реальных) показателей с опорой на показатели индекса онлайн-цен [3];
- оценка эффекта переноса валютного курса во внутренние цены, жесткости цен и др. [11], [12], [15].

Проведенный обзор выявил один существенный недостаток всех описанных исследований: ни в одном из них официальный ИПЦ не был реплицирован в полном объеме в связи с высокой сложностью сбора данных и поддержания базы данных в рабочем состоянии. В этой связи перспективы дальнейшего исследования проблемы представляются в максимально возможном устранении данного недостатка с опорой

на предшествовавший мировой опыт применения данных об онлайн-ценах и высокочастотных онлайн-индикаторов. В свою очередь, результаты проведенного обзора могут быть использованы в интересах монетарных властей РФ для построения прогнозных моделей инфляции, учитывающих высокочастотные онлайн-данные по ценам.

Благодарности

Материал подготовлен в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hillen J., "Web scraping for food price research," 2019.
2. Vicente H., "Forecasting inflation using online daily prices: a MADAS approach for Brazil," 2021. pp. 1-60.
3. Cavallo A., "Online and official price indexes: Measuring Argentina's inflation," *Journal of Monetary Economics*, No. 60, 2013. pp. 152-165.
4. Cavallo A., Rigobon R., "The Billion Price Project: using online prices for measurement and research," *Journal of economic perspectives*, Vol. 30, No. 2, 2016. pp. 151-178.
5. de Haan J., Hendriks R., "Online data, fixed effects and the construction of high-frequency price indexes," *Economic measurement group workshop*, 2013. pp. 28-29.
6. Krsinich F., "Price Indexes from online data using the fixed-effects window-splice (FEWS) index," *Ottawa Group*, Tokyo, Japan, 20-22, 2015. 20-22 pp.
7. Bertolotto M. I. 16th Meeting of the Ottawa Group on Price Indices // *Online Price Index with Product Replacement: The Closest-Match Approach*. Rio de Janeiro. 2019.
8. Aparicio D., Bertolotto M. I., "Forecasting inflation with online prices," *International Journal of Forecasting*, Vol. 36, No. 2, 2020. pp. 232-247.
9. Hull I. et al. Proc. IFC-Bank Indonesia Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference // *Price information collected online and short-term inflation forecasts*. 2017.
10. Cavallo A. *Scraped data and sticky prices // Review of Economics and Statistics*. 2018. Vol. 100. No. 1. pp. 105-119.
11. Cavallo A., Neiman B., Rigobon R. *Product introductions, currency unions, and the real exchange rate*. National Bureau of Economic Research, 2012.
12. Cavallo A., "Are online and offline prices similar? Evidence from large multi-channel retailers," *American Economic Review*, Vol. 107, No. 1, 2017. pp. 283-303.
13. Jaworski K., "Measuring food inflation during the COVID-19 pandemic in real time using online data: a case study of Poland," *British Food Journal*, 2021.
14. Alvarez S. E., Lein S. M., "Tracking inflation on a daily basis," *Swiss Journal of Economics and Statistics*, Vol. 156, No. 1, 2020. pp. 1-13.

15. Lünemann P., Wintr L. Price stickiness in the US and Europe revisited: evidence from internet prices // Oxford bulletin of economics and statistics. 2011. Vol. 73. No. 5. pp. 593-621.

16. Metcalfe E. et al., "Research Indices Using Web Scraped Data: Clustering Large Datasets into Price Indices (CLIP)," Office of National Statistics (ONS), London, 2016.

В СЕРИИ ПРЕПРИНТОВ
РАНХиГС РАССМАТРИВАЮТСЯ
ТЕОРЕТИЧЕСКИЕ
И ПРАКТИЧЕСКИЕ ПОДХОДЫ
К СОЗДАНИЮ, АКТИВНОМУ
ИСПОЛЬЗОВАНИЮ
ВОЗМОЖНОСТЕЙ
ИННОВАЦИЙ В РАЗЛИЧНЫХ
СФЕРАХ ЭКОНОМИКИ
КАК КЛЮЧЕВОГО УСЛОВИЯ
ЭФФЕКТИВНОГО УПРАВЛЕНИЯ



РАНХиГС

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ