Щербаков В.С. [1, 2], Харламова М.С. [3], Яковина М.Ю. [3]

[1] Омский государственный университет им. Ф.М. Достоевского, Омск, Россия
[2] Омский государственный педагогический университет, Омск, Россия
[3] Федеральное государственное бюджетное учреждение науки Омский научный центр Сибирского отделения Российской академии наук, Омск, Россия

Статистика поисковых запросов как прокси-показатель региональной ценовой динамики

ЦИТИРОВАТЬ СТАТЬЮ:

Щербаков В.С., Харламова М.С., Яковина М.Ю. Статистика поисковых запросов как прокси-показатель региональной ценовой динамики // Креативная экономика. – 2022. – Том 16. – № 11. – С. 4475–4490. doi: 10.18334/ce.16.11.116419

АННОТАЦИЯ:

На сегодняшний день апробирован широкий перечень моделей наукастинга ценовой динамики с использованием статистик пользовательской активности в Интернете, наиболее доступными из которых являются поисковые запросы. Многие из них демонстрируют позитивные результаты в улучшении качества прогнозов. Однако зачастую наилучшие предсказания достигаются за счет существенного увеличения числа таких статистик, что не вполне соответствует целям наукастинга, требующего оперативного и частого обновления прогнозов. Кроме того, использование распространенных статистик запросов может быть нерелевантным на региональном уровне. Анализу этих проблем и возможных решений посвящена данная работа.

ФИНАНСИРОВАНИЕ:

Исследование выполнено в рамках государственного задания Омского научного центра СО РАН (номер госрегистрации проекта 121022000112-2).

КЛЮЧЕВЫЕ СЛОВА: наукастинг, динамика цен, поисковые запросы, региональный рост цен

ОБ АВТОРАХ

Щербаков Василий Сергеевич, доцент, к.э.н. (chsherbakov.v@gmail.com)
Харламова Мария Сергеевна, ведущий инженер (m.kharlamova@alumni.nsu.ru)
Яковина Максим Юрьевич, инженер (maksim_yakovina@mail.ru)

издательство КРЕАТИВНАЯ ЭКОНОМИКА

creativeconomy.ru

ISSN 1994-6929

Shcherbakov V.S. [1, 2], Kharlamova M.S. [3], Yakovina M.Yu. [3]

[1] Dostoevsky Omsk State University, Russia
[2] Omsk State Pedagogical University, Russia
[3] Omsk Scientific Center of the Siberian Branch of the Russian Academy of Sciences, Russia

Search query statistics as a proxy indicator of regional price dynamics

CITE AS:

Scherbakov V.S., Kharlamova M.S., Yakovina M.Yu. (2022) Statistika poiskovyh zaprosov kak proksi-pokazatel regionalnoy tsenovoy dinamiki [Search query statistics as a proxy indicator of regional price dynamics]. *Kreativnaya ekonomika. 16*. (11). – 4475–4490. doi: 10.18334/ce.16.11.116419

ABSTRACT:

To date, a wide range of nowcasting models of price dynamics has been tested. The statistics of user activity on the Internet were used. Search queries are the most accessible ones. Many of them demonstrate positive results in improving the forecast quality. However, the best predictions are often achieved due to a significant increase in the number of such statistics. This does not fully meet the goals of nowcasting, which requires prompt and frequent updating of forecasts. In addition, the use of common query statistics may not be relevant at the regional level.

FUNDING:

The research was carried out within the framework of the state assignment of the Omsk Scientific Centre of the Siberian Branch of the Russian Academy of Sciences (state registration number 121022000112-2).

KEYWORDS: nowcasting, price dynamics, search queries, regional price growth

JEL Classification: E31, L86, O31

Received: 11.10.2022 / **Published:** 30.11.2022
© Author(s) / Publication: CREATIVE ECONOMY Publishers
For correspondence: Shcherbakov V.S. (chsherbakov.v@gmail.com)

Введение

В последние годы интерес к наукастингу общей динамики цен и инфляционных ожиданий существенно возрос. Это обусловлено как необходимостью повышения точности краткосрочного прогнозирования для принятия решений в условиях экономической нестабильности, так и развитием методов и информационного обеспечения для анализа широкого ряда доступных высокочастотных данных.

Для целей наукастинга в основном используются четыре типа данных: опросные; рыночные, в частности биржевые; новостные, в том числе комментарии и посты; и другие данные об активности пользователей в интернете, в особенности запросы к поисковым системам. Последние наиболее активно задействуются исследователями за счет их высокой доступности.

Улучшение прогнозов с использованием таких данных показано как в зарубежных, так и в отечественных работах. При этом опыт исследований для российской экономики существенно меньше, интерес возрастает лишь в последние годы. В связи с этим имеется множество перспектив углубления анализа и выработки решений существующих проблем, особенно в контексте объемов информации, необходимой для существенного улучшения прогнозов.

В региональном разрезе найти работы, использующие поисковые запросы для прогнозирования общей динамики цен, не удалось. При этом представляется, что на региональный рост цен, вероятно, оказывают влияние специфические факторы, которые могут быть отражены в поисковой активности.

Целью данной работы является оценка возможностей применения существующих подходов к наукастингу общей динамики цен с использованием поисковых запросов на региональном уровне.

В связи с этим необходимо, во-первых, рассмотреть имеющийся исследовательский опыт в этой области. Затем оценить проблемы экстраполяции существующих подходов на российский, в особенности региональный уровень. В заключение – предложить возможные решения выявленных проблем и дальнейшие перспективы исследования.

Исследовательский опыт наукастинга общей динамики цен с использованием поисковых запросов

В недавней работе [1] *(Yurevich, 2021)* приведен достаточно подробный обзор работ по наукастингу инфляции с использованием статистики поисковых запросов. Результат анализа исследований представлен в наглядной для сравнения табличной форме, где определены используемые авторами ключевые

creativeconomy.ru

ISSN 1994-6929

слова и модели. Обращает внимание разнообразие апробированных моделей, преимущественно эконометрических. Наряду с этим в последние годы возрастает распространенность и значимость методов машинного обучения. Также вариативен перечень рассматриваемых поисковых образов, где авторов может интересовать от одного до широкого ряда ключевых слов (так, например, в работе [2] *(Petrova, 2019)* отобрано 75 слов и сочетаний).

Стоит отметить, что интерес к теме развивается в последнее десятилетие. Число публикаций с каждым годом заметно возрастает, особенно с 2020 года, что может быть обусловлено высоким запросом на оперативные оценки важнейших экономических показателей в условиях экономического спада из-за пандемии. Обратим внимание на ряд интересных работ последних лет в дополнение к упомянутому выше обзору.

В последние годы разнообразие методов и моделей, а также география зарубежных исследований ценовой динамики с использованием поисковых запросов продолжает расширяться. В работе [3] *(Drachal, 2020)* для прогнозирования инфляции в США и Польше применяется метод динамического усреднения моделей (DMA), в том числе с построением спецификаций, включающих данные по поисковым запросам Google. В качестве ключевых слов используются такие формулировки, как инфляция, индекс цен, инфляция в США, цены в США, ИПЦ, индекс потребительских цен. Такие спецификации показывают относительно лучшие характеристики.

Данные поисковых запросов Google (по ключевому слову «инфляция») также были использованы при построении гибридной новокейнсианской кривой Филлипса для Индии [4] *(Jha, Sahu, 2021)*. Авторы приходят к выводу, что модель с включением данной переменной демонстрирует относительно лучшую прогностическую силу по сравнению со спецификацией, включающей данные по инфляционным ожиданиям на основе опросов домохозяйств.

Кроме того, Aromi D. и Llada M. показали состоятельность использования данных из Twitter для построения косвенного показателя для измерения инфляционных ожиданий [5] *(Aromi, Llada, 2020)*. Исследователи конструируют показатель внимания к инфляции на основе корпуса аргентинских твитов. Основная идея состоит в том, чтобы вычислить относительную частоту употребления словоформ существительного «инфляция» и прилагательного «инфляционный» в собранных постах. Далее построенный показатель используется в качестве одной из объясняющих переменных для инфляции наряду с ее темпами за предыдущие периоды и обменными курсами. Показано, что включение данного индикатора улучшило прогнозные свойства модели.

В свою очередь, Angelico C. и коллеги работают с корпусом итальянских твитов. На начальном этапе исследования определен словарь слов, которые, по мнению авторов, могут быть связаны с ценами и ценовой динамикой в целом. В словарь вошло 20 слов, в том числе: *цены, инфляция, дорого, дефляция, продажи, высокие цены на бензин* и другие. Для снижения уровня шума в данных авторы применяют трехэтапную процедуру, включающую тематический анализ с помощью метода скрытого распределения Дирихле (LDA). На основе отфильтрованных данных строятся направленные индикаторы ценовых ожиданий, которые предоставляют дополнительную информацию об инфляционных процессах в Италии [6] *(Angelico et al., 2021)*. По своей сути идея аналогична подходу Aromi D. и Llada M., но при этом основана на более широком словаре, который может раскрывать особенности данных процессов в рассматриваемой стране или регионе.

Объясняющая способность поисковых запросов для анализа ценовой динамики была доказана и для российской экономики, в последние годы интерес отечественных исследователей к данной теме возрос. Широкий перечень моделей (с главными компонентами, методы со штрафом, метод наименьших углов (LARS), случайный лес, градиентный бустинг) был апробирован в работе Д.А. Петровой [2] *(Petrova, 2019)*. Работа следует логике более раннего исследования [7] I. Baybuza *(Baybuza, 2018)*, дополняя анализ статистиками поисковых запросов Google, связанных с ценовыми ожиданиями. Наилучшие результаты показала модель с главными компонентами, позволившая извлечь наиболее важную информацию из большего числа коррелированных данных.

В упомянутой выше работе [1] *(Yurevich, 2021)* протестирован ряд предложенных автором эконометрических моделей прогнозирования индекса потребительских цен (ИПЦ) и его ожиданий. Индекс Google Trends по ключевому слову «инфляция» показал значимые результаты при прогнозировании месячного ИПЦ. Автор также отмечает потенциал использования показателей частоты поисковых запросов в качестве оценки инфляционных ожиданий, в частности для замещения оценок, получаемых с помощью социологических опросов.

Нахождение такого прокси-показателя представляется значимым результатом как с точки зрения экономии ресурсов на сбор данных и проведение социологических исследований, так и роста доступности, частоты и оперативности получения таких данных.

В результате статистика по поисковым запросам зарекомендовала себя как информативный источник данных, имеющий большой потенциал для на-

creativeconomy.ru

ISSN 1994-6929

укастинга ценовой динамики и ее ожиданий. За счет их включения множество исследований для различных стран пришло к положительному результату в улучшении показателей качества прогнозов. Однако, на взгляд авторов, можно выделить ряд содержательных и технических проблем, связанных с использованием этих данных.

Проблемы применения данных поисковых запросов для анализа и прогнозирования общей динамики цен

Как было отмечено, большинство исследований, посвященных наукастингу и прогнозированию общей динамики цен, подтверждают значимое улучшение моделей при включении статистики тематических поисковых запросов. Тем не менее прирост прогнозной способности моделей после включения статистики запросов не всегда выглядит существенным, а где-то и вовсе отсутствует.

В качестве возможной причины отрицательного результата отмечается относительно низкая личная заинтересованность населения в информации об измерителях общего уровня цен. Вне профессионального сообщества потребителей могут чаще интересовать, например, цены на определенные группы товаров. В частности, поэтому использование поисковых запросов получило большее распространение и демонстрирует лучшие результаты для прогнозирования безработицы по сравнению с прогнозом инфляции [8] *(Niesert et al., 2020)*.

Профессиональное сообщество, напротив, может производить запросы, связанные с различными показателями роста цен, влияющими на него факторами, а также основными источниками информации, где эти данные публикуются. Кроме того, запросы могут варьироваться в зависимости от целей поиска. Экономист, ищущий данные о росте цен, может обратить запрос непосредственно к известным ему релевантным источникам данных, где публикуется нужная информация, введя, например, «росстат», «витрина данных», «цб» и подобное. Представляется, что такие запросы, как, например, *рост цен, инфляция, индекс цен,* и похожие в большей степени служат целям отслеживания новостного фона и факторов. В связи с этим перечень потенциально информативных запросов расширяется.

Тем не менее расширение перечня статистик запросов (как и включение других частотных данных, например, часто используемых биржевых котировок и индексов) приводит к другой проблеме – росту объемов необходимых данных.

Кроме содержательной стороны, больших объемов информации также требуют для обучения многие из тестируемых исследователями алгоритмов. Демонстрируя сравнительно лучшие результаты, многие из таких моделей в полной мере не отвечают целям наукастинга, так как подготовка данных для их работы оказывается слишком трудоемкой по отношению к требуемой оперативности и частоте обновления. Даже при достаточной автоматизации сбора данных меняющаяся конъюнктура и необходимость постоянной коррекции прогнозов затрудняют применение таких алгоритмов в регулярной практике.

Отдельного внимания заслуживает проблема региональной специфики. Разные территории, будь то страны или регионы, могут характеризоваться уникальными для них запросами, которые обусловлены отраслевой специализацией, языковыми особенностями и прочими факторами неоднородности. Кроме того, между регионами отличаются уровень проникновения Интернета и популярность различных поисковых систем.

Как было показано выше, одним из устоявшихся стандартов в области использования статистики поисковых данных для целей наукастинга и прогнозирования экономических показателей является применение сервиса Google Trends. Начало использования данного подхода было заложено в работе сотрудников Google более 10 лет назад [9] (*Choi, Varian, 2009*).

Как известно, в сервисе Google Trends данные по запросам доступны с 2004 года. На *рисунке 1* приведено сопоставление уровней проникновения Интернета в России и ряде стран, исследования по которым приведены в обзоре.

Согласно представленным данным видно, что уровень проникновения Интернета в России приблизился к 85% к 2020 году, но по-прежнему уступает США и Великобритании. Наряду с этим он больше, чем в Китае и Индии. В связи с этим использование статистики по поисковым запросам для России, например, с 2010 года, вероятно, будет более релевантным.

При этом в разрезе субъектов Российской Федерации уровень проникновения Интернета остается дифференцированным. Если в г. Москве, по данным 2020 года, 87,6% населения пользовалось Интернетом почти каждый день, то в Омской области – только 77,6%, а например, в Ульяновской – 61,5%[1].

[1] Росстат. Регионы России. Социально-экономические показатели. Использование сети интернет населением. URL: https://rosstat.gov.ru/storage/mediabank/Region_Pokaz_2021.pdf (дата обращения: 01.08.2022).

creativeconomy.ru

ISSN 1994-6929

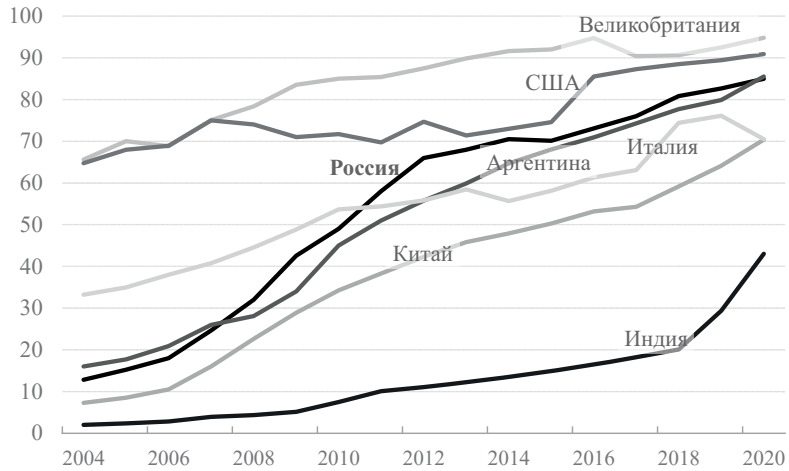


Рисунок 1. Удельная доля населения страны, использующего Интернет, %

Источник: Всемирный банк. Процент населения, использующего Интернет. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2020&locations=RU-US-GB-IN-CN-AR-IT&start=2004&view=chart (дата обращения: 08.02.2022).

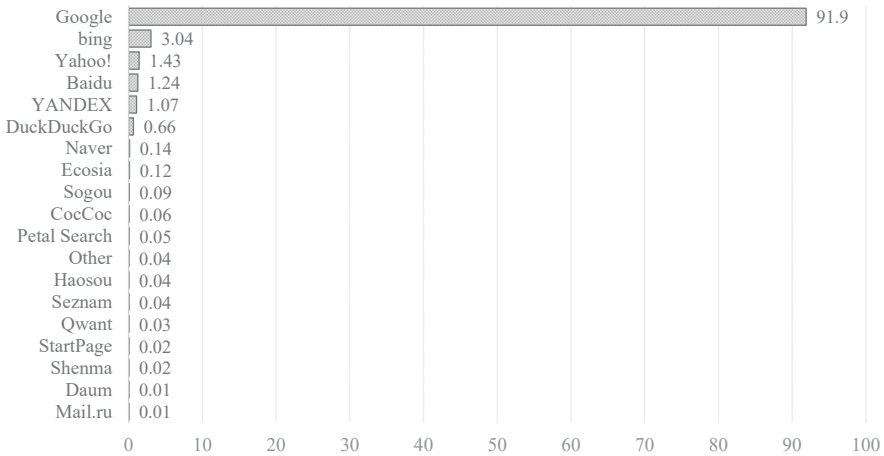


Рисунок 2. Средняя доля рынка поисковых систем на сентябрь 2022, %

Источник: Аналитическое агентство Statcounter. Доля рынка поисковых систем в мире. URL: https://gs.statcounter.com/search-engine-market-share (дата обращения: 03.10.2022).

Следующим важным вопросом является оценка уровня использования той или иной поисковой системы и, как следствие, репрезентативности ста-

тистики ее запросов для оценки определенных экономических процессов. Безусловным мировым лидером среди поисковых систем является компания Google. Превалирующая доля запросов населения стран мира приходит через данный ресурс. Поэтому использование данных специального сервиса Google Trends в рамках вышеупомянутых исследований является весьма оправданным *(рис. 2)*.

Существует также определенная страновая специфика, отражающая предпочтение пользователей относительно поисковых систем (рис. 3).

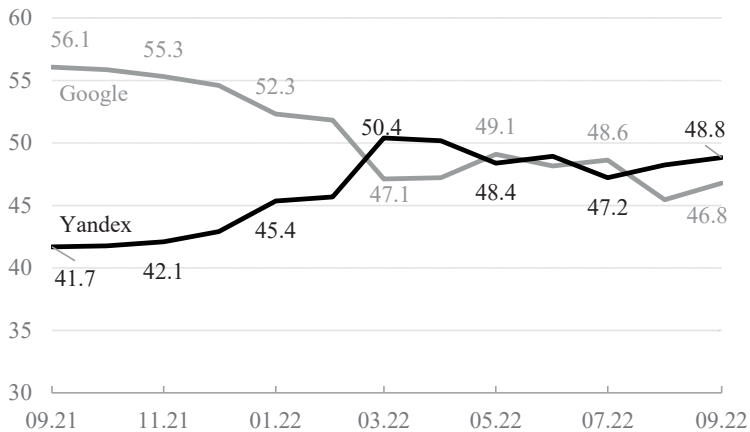


Рисунок 3. Наибольшая рыночная доля поисковых систем в России, %*
**доля остальных поисковых систем в среднем не превышает 3%.*

Источник: Аналитическое агентство Statcounter. Доля рынка поисковых систем в России. URL: https://gs.statcounter.com/search-engine-market-share (дата обращения: 03.10.2022).

Таким образом, в России наблюдается усиленная конкурентная борьба за внимание пользователей со стороны двух компаний – Google и «Яндекс». Тем не менее есть ряд причин, по которым использование данных Google Trends может быть затруднено.

Во-первых, сам формат предоставления данных Google. Сервис представляет информацию в нормализированном виде для удобства сравнения поисковых запросов. Данные нормализуются с учетом времени и места отправки запроса[2]. Такой подход, с одной стороны, обладает рядом преимуществ,

[2] Google. Как выполняется нормализация данных Google Trends. URL: https://support.google.com/trends/answer/4365533?hl=ru&ref_topic=6248052 (дата обращения: 05.03.2022).

creativeconomy.ru

ISSN 1994-6929

в частности, позволяет сопоставлять данные в географическом и временном разрезе.

С другой стороны, в регионах, где определенная тема пользуется примерно одинаковой популярностью, общее количество запросов может быть разным. Не учитывается и тот момент, что релевантность некоторых ключевых слов может меняться в разных регионах со временем, а относительная представленность данных не позволяет отследить этот нюанс [10] *(Simionescu, Zimmermann, 2017)*.

Яндекс предоставляет возможность получать статистические данные по поисковым запросам через свою систему с помощью сервиса «Яндекс. Подбор слов» (Wordstat). Значительным преимуществом этого сервиса по сравнению с конкурирующим и общепринятым в данной сфере Google Trends является предоставление данных не в индексной форме, а в абсолютном значении, то есть количестве запросов по тем или иным ключевым словам. Также доступны данные об относительном числе запросов, где базой сравнения выступает общее число запросов по выбранной территории за рассматриваемый период. Такой вид представления дает дополнительное преимущество, связанное с тем, что индексы Google пересчитываются относительно максимального значения за выбранный период, то есть при изменении периода осуществляется пересчет всего ряда.

Необходимо учитывать и некоторые ограничения в работе с открытыми данными Wordstat. История запросов по ключевым словам предоставляется за 24 последних месяца на день обращения для данных частотой в месяц и за 12 последних месяцев – для недельных. При моделировании это ограничение является существенным, так как число доступных наблюдений резко сокращается по сравнению с Google Trends.

Частично нивелировать эту проблему позволяет использование недельных данных. Так, по наиболее популярному запросу, который используется как прокси-показатель в приведенных выше исследованиях, для российских данных наблюдается схожая динамика статистики Yandex и Google (рис. 4).

Корреляция между статистиками Yandex и Google за рассматриваемый период составляет около 79%, в связи с чем использование обеих может быть информативным. Тогда как на региональных данных, например для Омской области, данные Google Trends нерепрезентативны, в отличие от Yandex (рис. 5).

В результате наиболее популярный для наукастинга ценовой динамики источник данных Google Trends не всегда может быть полезен на региональном уровне.

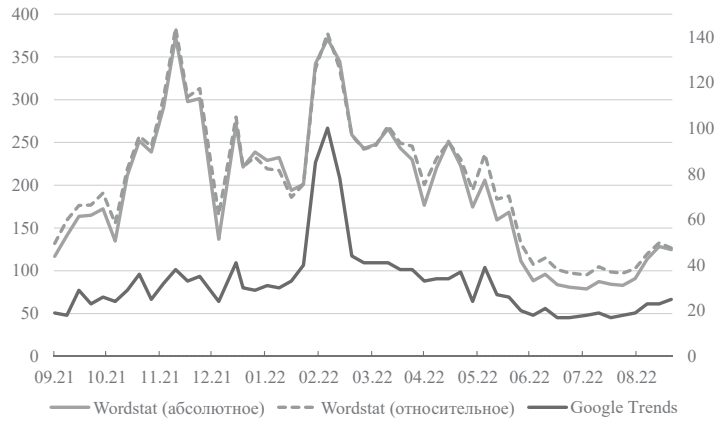


Рисунок 4. Динамика поисковых запросов по ключевому слову «инфляция» в России, недельные данные*

**левая шкала – абсолютное число запросов Wordstat в тысячах; правая шкала – Google Trends и умноженное на 1 млн относительное число запросов Wordstat.*

Источник: Wordstat Yandex. URL: https://wordstat.yandex.ru (дата обращения: 30.09.2022); Google Trends. URL: https://trends.google.ru/trends/ (дата обращения: 30.09.2022).

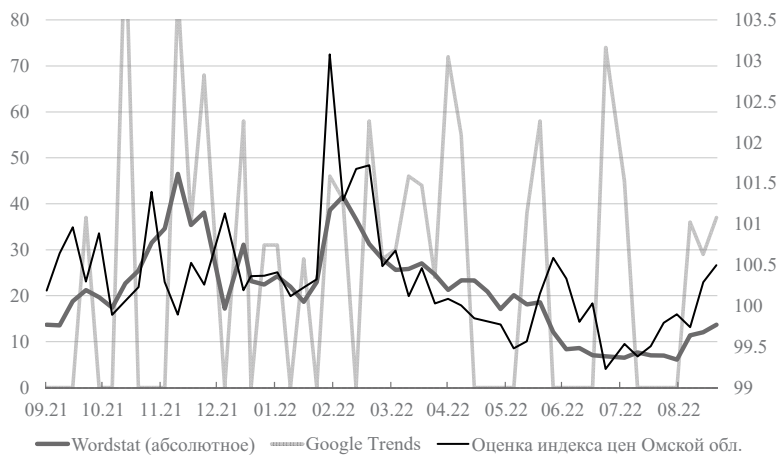


Рисунок 5. Динамика поисковых запросов по ключевому слову «инфляция» и оценка индекса цен по Омской области, недельные данные*

**левая шкала – индекс Google Trends и абсолютное число запросов Wordstat в сотнях; правая шкала – оценка недельного индекса цен на основе данных о средних потребительских ценах и их изменении*

Источник: расчеты авторов; Омскстат. URL: https://omsk.gks.ru/prices?ysclid=l8zplxmniy594530276 (дата обращения: 30.09.2022); Wordstat Yandex. URL: https://wordstat.yandex.ru (дата обращения: 30.09.2022); Google Trends. URL: https://trends.google.ru/trends/ (дата обращения: 30.09.2022).

creativeconomy.ru

ISSN 1994-6929

В отличие от Google, данные Yandex могут служить дополнительным источником информации для наукастинга ценовой динамики на региональном уровне за счет как формата предоставления данных, так и популярности поисковой системы. Статистика по поисковым запросам Wordstat Yandex значимо коррелирует с ростом цен в регионе. С использованием данного подхода открываются перспективы экстраполяции разработанных на страновом уровне методов и дальнейших исследований региональной ценовой динамики.

Перспективы дальнейших исследований

Как показывает проведенный анализ, большинство работ, которые используют данные поисковых запросов для анализа и прогнозирования общей динамики цен применяют экспертный подход для определения ключевых слов. Другими словами, авторы самостоятельно предопределяют перечень слов, изменение поисковой активности по которым может свидетельствовать о происходящих инфляционных процессах.

С одной стороны, данный подход является простым и интуитивно понятным, с другой стороны, в нем отсутствует объективность в обосновании выбора тех или иных ключевых слов. Велика вероятность того, что авторы не учли все значимые слова, так как на инфляционные ожидания той или иной страны, региона влияют специфические термины, которые не являются очевидными для оценки рассматриваемых взаимосвязей.

Необходимо отметить, что определение ключевых слов, поисковые запросы по которым могут выступать прокси-показателями для оценки инфляционных ожиданий, не является конечной целью подобных исследований. В любом случае основная задача заключается в анализе и прогнозировании роста цен. Если данные по поисковым запросам позволят улучшить показатели сопоставимых моделей, только в этом случае можно считать, что поставленная цель достигнута.

В связи с этим потенциальная механика дальнейших исследований в данной области схематично представлена на *рисунке 6*.

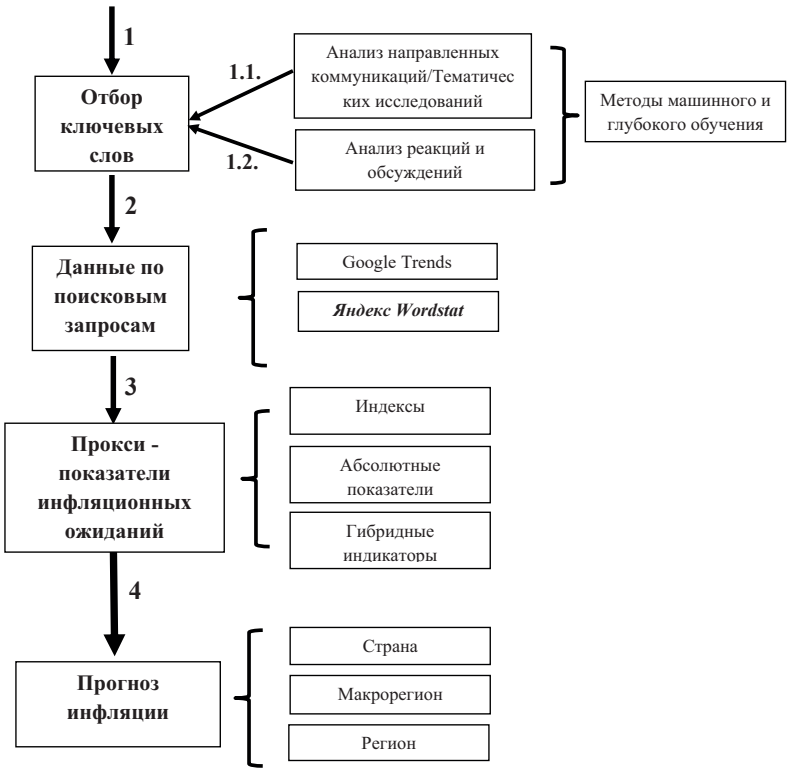


Рисунок 6. Процесс конструирования прокси-показателей инфляционных ожиданий на основе поисковых запросов в сети Интернет

Источник: составлено авторами.

Известно, что одним из ключевых факторов, влияющих на общую динамику цен, выступают инфляционные ожидания. В контексте режима инфляционного таргетирования речь идет о процессе заякоривания ожиданий на цели. Это приводит к тому, что экономические агенты в своих прогнозах не учитывают кратковременные факторы, а инфляция быстро возвращается к цели, будучи отклоненная от таргета под действием разного рода шоков.

Так, в России начиная с 2017 г. инфляционные ожидания аналитиков являются заякоренными. В свою очередь, ожидания населения оставались незаякоренными, они чувствительны к динамике цен на ряд товаров-маркеров, например медицинских средств и часто приобретаемых продуктов питания [11] *(Grishchenko et al., 2022)*.

На наш взгляд, ожидания вполне могут быть оценены на основе поисковых запросов. Для получения репрезентативных оценок необходим эконо-

creativeconomy.ru

ISSN 1994-6929

мически логичный и практико-обоснованный выбор ключевых слов, которые население использует для обращения к поисковым системам. В данном случае предлагается воспользоваться как минимум двумя подходами *(см. 1.1. и 1.2. на рис. 6)*.

Во-первых, это текстовый анализ большого массива тематических текстов, относящихся к анализу факторов роста цен в стране и регионах, в том числе целевых коммуникационных материалов регулятора, направленных на объяснение наблюдаемых тенденций. Оценка частотности применения тех или иных ключевых слов в подобных текстах может выступить существенной основой для определения слов-маркеров. Этот подход можно условно называть «сверху вниз».

Во-вторых, может быть применен подход, когда анализу подлежит большой массив данных, формируемый пользователями сети Интернет, в особенности социальных сетей, относительно наблюдаемых и ощущаемых инфляционных процессов. Речь идет о публикуемых постах, комментариях, реакциях в наиболее распространенных в России социальных сетях – «Вконтакте», возможно, «Одноклассники», а также в «Телеграм». В данном случае задача заключается в том, чтобы с использованием специальных алгоритмов глубокого обучения (deep learning), например архитектур таких нейронных сетей, как Seq2seq, BERT и т.д., получить список значимых для общей динамики цен ключевых слов без предопределения их значения на старте исследования. Данный подход, в свою очередь, может быть определен как «снизу вверх».

После того как списки ключевых слов будут получены, они могут быть квантифицированы для той или иной территории как раз с помощью статистики поисковых запросов. Далее на их основе предполагается построение разного рода индикаторов, которые будут использованы в моделировании.

Как было отмечено ранее, конечной целью запланированной работы выступает построение моделей прогнозов динамики цен с лучшими метриками качества относительно спецификаций, не включающих данные по поисковым запросам, или получение содержательных доказательств того, что они не приносят дополнительных знаний с точки зрения объяснения изучаемых процессов. Заметим, что в данном случае идет речь не только о макроэкономическом, но и мезоэкономическом уровне исследования.

Представляется, что предлагаемый нами подход, с одной стороны, может быть автоматизирован на базе современных библиотек текстового анализа, с другой стороны, позволит не допустить появления «проклятия размерности», о котором упоминалось ранее.

Заключение

Поисковые запросы зарекомендовали себя как информативный источник данных для наукастинга общей динамики цен. Значимые результаты были показаны в зарубежных, а также недавних российских исследованиях. Однако выявлен ряд проблем, нуждающихся в дополнительной проработке.

Во-первых, рост прогнозной силы моделей требует существенного расширения необходимых данных, что затрудняет апробацию таких алгоритмов для целей наукастинга, требующего оперативного мониторинга и переоценки. Во-вторых, при использовании поисковых запросов требует внимания репрезентативность отобранных ключевых слов, которые в основном определяются экспертно и, кроме того, могут быть недостаточны на мезоуровне.

В результате проведенного анализа предложены источники данных, релевантные на региональном уровне, а также алгоритмы выделения ключевых слов для обоснованного и более информативного конструирования прокси-показателей для наукастинга общей динамики цен.

creativeconomy.ru

ИСТОЧНИКИ:

1. Юревич М.А. Инфляционные ожидания и инфляция: наукастинг и прогнозирование // Journal of Economic Regulation. – 2021. – № 2. – с. 22–35. – doi: 10.17835/2078–5429.2021.12.2.022–035.
2. Петрова Д.А. Прогнозирование инфляции на основе Интернет-запросов // Экономическое развитие России. – 2019. – № 11. – с. 55–62.
3. Drachal K. Forecasting the Inflation Rate in Poland and U.S. Using Dynamic Model Averaging (DMA) and Google Queries // Ournal for Economic Forecasting. – 2020. – № 2. – p. 18–34.
4. Jha S., Sahu S. Forecasting inflation for India with the Phillips Curve: Evidence from internet search data // Economics Bulletin. – 2020. – № 3. – p. 2372–2379.
5. Aromi D., Llada M. Forecasting inflation with twitter. Asociacion Argentina de Economia Politica: Working Papers, 4308. – 2020
6. Angelico C., Marcucci J., Miccoli M., Quarta F. Can We Measure Inflation Expectations Using Twitter? Bank of Italy Temi di Discussione // Bank of Italy Temi di Discussione (Working Paper). – 2021. – doi: 10.2139/ssrn.3827489.
7. Baybuza I. Inflation forecasting using machine learning methods // Russian Journal of Money and Finance. – 2018. – № 4. – p. 42–59. – doi: 10.31477/rjmf.201804.42.
8. Niesert R., Oorschot J., Veldhuisen C., Brons K., Lange R-J. Can Google search data help predict macroeconomic series? // International Journal of Forecasting. – 2020. – № 3. – p. 1163–1172. – doi: 10.1016/j.ijforecast.2018.12.006.
9. Choi H., Varian H. Predicting initial claims for unemployment benefits. Static.googleusercontent.com. [Электронный ресурс]. URL: http://static.

ISSN 1994-6929

googleusercontent.com/media/research.google.com/en/us/archive/papers/initialclaimsUS.pdf.

10. Simionescu M., Zimmermann K. Big Data and Unemployment Analysis. GLO Discussion Paper. [Электронный ресурс]. URL: http://hdl.handle.net/10419/162198 (дата обращения: 05.03.2022).
11. Грищенко В., Кадрева О., Поршаков А., Чернядьев Д. Оценка инфляционных ожиданий для России. Аналитическая записка Банка России. Банк России. [Электронный ресурс]. URL: http://www.cbr.ru/content/document/file/139272/analytic_note_20220728_dip.pdf (дата обращения: 20.08.2022).

REFERENCES:

Angelico C., Marcucci J., Miccoli M., Quarta F. (2021). *Can We Measure Inflation Expectations Using Twitter? Bank of Italy Temi di Discussione Working Paper*. doi: 10.2139/ssrn.3827489.

Baybuza I. (2018). *Inflation forecasting using machine learning methods Russian Journal of Money and Finance*. *77* (4). 42–59. doi: 10.31477/rjmf.201804.42 .

Choi H., Varian H. Predicting initial claims for unemployment benefitsStatic.googleusercontent.com. Retrieved from http://static.googleusercontent.com/media/research.google.com/en/us/archive/papers/initialclaimsUS.pdf

Drachal K. (2020). *Forecasting the Inflation Rate in Poland and U.S. Using Dynamic Model Averaging (DMA) and Google Queries Ournal for Economic Forecasting*. *23* (2). 18–34.

Jha S., Sahu S. (2020). *Forecasting inflation for India with the Phillips Curve: Evidence from internet search data Economics Bulletin*. *40* (3). 2372–2379.

Niesert R., Oorschot J., Veldhuisen C., Brons K., Lange R-J. (2020). *Can Google search data help predict macroeconomic series? International Journal of Forecasting*. *36* (3). 1163–1172. doi: 10.1016/j.ijforecast.2018.12.006.

Petrova D.A. (2019). *Prognozirovanie inflyatsii na osnove Internet-zaprosov* [Inflation forecasting based on internet search queries]. *Russian Economic Developments*. *26* (11). 55–62. (in Russian).

Simionescu M., Zimmermann K. Big Data and Unemployment Analysis-GLO Discussion Paper. Retrieved March 05, 2022, from http://hdl.handle.net/10419/162198

Yurevich M.A. (2021). *Inflyatsionnye ozhidaniya i inflyatsiya: naukasting i prognozirovanie* [Inflation expectations and inflation: nowcasting and forecasting]. *Journal of economic regulation*. *12* (2). 22–35. (in Russian). doi: 10.17835/2078–5429.2021.12.2.022–035.