

## Возможности прогнозирования уровня безработицы на основе анализа статистики запросов (в поисковых системах)

Максим Андреевич Юревич

Финансовый университет при Правительстве РФ, Москва, Россия, e-mail: MAYurevich@fa.ru

Денис Рашидович Ахмадеев

Финансовый университет при Правительстве РФ, Москва, Россия, e-mail: DRAhmadeev@fa.ru

**Цитирование:** Юревич М.А., Ахмадеев Д.Р. (2021). Возможности прогнозирования уровня безработицы на основе анализа статистики запросов (в поисковых системах) // *Terra Economicus* 19(3): 53–64. DOI: 10.18522/2073-6606-2021-19-3-53-64

*Безработица является одним из ключевых макроэкономических показателей, играющих значительную роль при формировании и разработке экономической политики государства. Принимая во внимание серьезное влияние, которое оказал коронакризис в том числе и на рынок труда, прогнозирование уровня безработицы представляется весьма актуальной задачей. Авторы отмечают, что на данный момент существует значительное разнообразие методов прогнозирования ситуации на рынке труда и уровня безработицы, а также показывают эволюцию подходов к этому вопросу: от традиционных методов (связанных с использованием модели баланса трудовых ресурсов, изучением структурных сдвигов региональных рынков труда, использованием многофакторных моделей рынка труда, таких как ARIMA, TAR, ARNN, на основе анализа различных индексов) до новаторских. Прогнозирование ситуации на рынке труда и безработицы связано с цифровизацией и развитием интернет-технологий, предоставляющих исследователям новые инструменты анализа. Данные о поисковых запросах в системе Google и другие данные об интернет-активности помогают решить проблему с запаздыванием официальной статистики, а также добавляют релевантную информацию для анализа и прогнозирования безработицы по сравнению с традиционными методами, что приводит к улучшению качества прогнозов. На основе зарубежного и российского опыта использования статистики поисковых запросов при прогнозировании и наукастинге уровня безработицы авторами построено несколько моделей прогнозирования уровня безработицы в России. Результаты исследования показали, что добавление в авторегрессионную модель нескольких переменных частоты запросов способно повысить прогностическую точность модели. Содержательные причины превосходства гибридной модели над авторегрессионной вариацией состоят в способности первой реагировать на будущие шоки рынка труда.*

**Ключевые слова:** уровень безработицы; рынок труда; запросы; GoogleTrends; большие данные

**Благодарность:** Статья подготовлена по результатам исследований, выполненных за счет средств Научного фонда Финуниверситета, по теме «Прогнозирование макроэкономических индикаторов на основе анализа статистики запросов в поисковых системах».

# Predicting the unemployment rate: Analyzing statistics on search engine query

Maksim A. Yurevich

Financial University, Moscow, Russia, e-mail: MAYurevich@fa.ru

Denis R. Akhmadeev

Financial University, Moscow, Russia, e-mail: DRAkhmadeev@fa.ru

**Citation:** Yurevich M.A., Akhmadeev D.R. (2021). Predicting the unemployment rate: Analyzing statistics on search engine query. *Terra Economicus* 19(3): 53–64. DOI: 10.18522/2073-6606-2021-19-3-53-64

*Unemployment is one of the key macroeconomic indicators that play a significant role in the state economic policy. There is a significant variety of methods for forecasting the situation on the labor market and the unemployment rate. We show the evolution of relevant approaches, including traditional methods (labor balance model, structural changes in regional labor markets, multifactorial labor market models such as ARIMA, TAR, ARNN) and modern ones. The development and the possibility of applying innovative approaches to forecasting unemployment is associated with digitalization and the development of Internet technologies, which provide researchers with new analysis tools. When compared to traditional methods, Google search query data and other internet activity data result in better forecasts, helping to solve the problem of lagging data provided by the official statistics, and adding relevant information to analyze and predict unemployment. Based on the experience of using the statistics of search queries in predicting and nowcasting the unemployment rate, the authors developed several models for predicting the unemployment rate in Russia. The research findings show that adding multiple query variables to the autoregressive model is able to improve the predictive accuracy of the model. Superiority of the hybrid model over the autoregressive variation is due to its ability to respond to future labor market shocks.*

**Keywords:** unemployment rate; labor market; queries; Google Trends; Big Data

**Acknowledgement:** The research was carried out with the support of the Scientific Foundation of the Financial University, project “Forecasting of macroeconomic indicators based on the analysis of query statistics in search engines”.

**JEL codes:** J64, D8, C22, C55

## Введение

Безработица представляет собой сложное многоаспектное социально-экономическое явление, а также является одним из ключевых макроэкономических показателей. Текущий глобальный кризис, вызванный внешним фактором – пандемией коронавируса, привел к снижению экономической активности не только в сфере производства, но и в большей степени в сфере услуг. Это в значительной степени является особенной характеристикой текущего кризиса, ведь главная причина периодичности промышленных кризисов – обновление основных фондов производства. Хотя острой необходимости в этом не было. Эта ситуация, безусловно, повлияла на многие сферы, в том числе и на рынок труда. Уровень безработицы вырос как в мире в целом, так и в России. Если перед пандемией, по данным Росстата, в России он не превышал 5%, то в пиковый период составил 6,4%.

Учитывая текущий кризис, сопровождавшийся значительным спадом экономической активности во всем мире, безработица является таким макроэкономическим показателем, который представляет особый интерес и для рядовых граждан, и для академических исследователей, и для представителей государственных органов. Это поднимает вопрос об эффективности прогнозирования уровня безработицы.

### **Обзор литературы: традиционные и новаторские способы прогнозирования уровня безработицы**

Методы прогнозирования весьма разнообразны, они постоянно обогащаются и совершенствуются. На данный момент можно выделить более сотни методов прогнозирования, однако наиболее часто используются лишь несколько десятков из них.

Классифицировать методы прогнозирования можно по многим критериям. Если осуществлять классификацию по такому критерию, как степень формализации, то можно выделить экспертные методы и формализованные методы.

Экспертные методы, также иногда называемые эвристическими или интуитивными, неотделимы от делающего прогноз лица и не изложены в явной форме, так как при их проведении все основывается на интуиции, прошлом опыте и накопленных знаниях эксперта. Такие методы весьма удобны и наиболее эффективны в случае высокой сложности изучаемого объекта, а также при высоком уровне неопределенности имеющихся данных. Принято выделять индивидуальные (опрос, аналогия, метод гирлянд случайностей и ассоциаций, морфологический анализ) и коллективные (комиссия, разнообразные формы мозговых штурмов, метод «Дельфи») экспертные методы.

Формализованные методы предполагают использование фактической информации об объекте и его прошлых состояниях. Можно выделить следующие формализованные методы прогнозирования:

- 1) методы экстраполяции тренда (скользящая средняя, экспоненциальное сглаживание, наименьшие квадраты);
- 2) эконометрический метод;
- 3) математико-статистический метод;
- 4) метод исторических аналогий;
- 5) имитационный метод прогнозирования;
- 6) методы логического моделирования и др.

Следует сказать о наличии так называемых наивных методов прогнозирования, т.е. простейших методов, которые используются в случае нехватки информации об объекте. Они дают менее точные оценки, но весьма просты в реализации. К данной группе методов относят:

- 1) метод предыдущего фактического значения;
- 2) метод простых средних величин;
- 3) метод простой скользящей средней;
- 4) метод взвешенной скользящей средней;
- 5) метод двойной скользящей средней;
- 6) метод экспоненциальной скользящей средней.

При попытке прогнозирования уровня безработицы обычно используют многофакторную модель, в число объясняющих переменных которой часто входят: доля убыточных предприятий в общем количестве предприятий; доля занятых на МСП среди всех занятых; общая численность ЭАН; среднесписочная численность работников предприятий; степень текучести кадров; численность обратившихся в службу занятости и некоторые другие.

Также можно использовать модель баланса трудовых ресурсов, которая является ключевым элементом при регулировании рынка труда в целом и принятии решений относительно безработицы в частности. Сводный баланс отражает величину трудовых ресурсов и миграционные потоки, на его основе можно определить количество безработных (как в стране в целом, так и по отдельным регионам).

Дать прогнозное значение уровня безработицы можно с помощью индекса структурных сдвигов, который комплексно характеризует функционирование региональных рынков труда. Анализ взаимодействий региональных рынков труда предполагает изучение динамики, структуры и направлений перемещения населения, прежде всего трудоспособной его части, которая формирует потенциал региональных рынков труда. Для более детального анализа необходимо исследовать динамику межрегионального движения населения и трудовых ресурсов по конкретным позициям: занятые, безработные, экономически неактивное население. Для этого можно использовать модель движения населения и трудовых ресурсов, учитывающую неоднократные переходы людей (Коровкин, 2001; Коровкин и др., 2002).

Довольно часто уровень безработицы прогнозируется в рамках многофакторной модели рынка труда для краткосрочного периода. Проблема таких моделей является то, что они предназначены в большей степени для комплексного анализа рынка труда, нежели для его оперативного мониторинга, а также высокий уровень их специфичности. Универсальность использования таких моделей не бесспорна, зачастую они эффективно работают только в рамках той страны, для которой разрабатывались. В качестве примера можно привести модель оперативного мониторинга рынка труда в США. Она строится на анализе 19 показателей, и ее результат представлен в виде индекса Labor Market Conditions Index (LMCI). 19 показателей этой модели разделены на 9 блоков, в том числе присутствует блок «безработица и неполная занятость».

Также можно выделить ежемесячный комплексный индекс Chicago Fed National Activity Index (CFNAI). Модели, состоящие из сложных индексов для комплексной оценки рынка труда, разрабатывались разными научными коллективами (Zmitrowicz, Khan, 2014; Barnes et al., 2007). И, как заявляют международные организации, данные модели бесспорно хороши для оценки рынка труда в первую очередь со стороны занятости. При этом, как отмечалось выше, они теряют в универсальности и оперативности прогноза, так как зачастую требуют весьма значительный массив данных.

С точки зрения оперативности прогноза следует отметить индекс рынка труда Labor Index (LI), который рассчитывается Евростатом. Данный индекс включает четыре компонента, одним из которых является процент безработных<sup>1</sup>.

Вышеобозначенные индикаторы используются в моделях класса ARIMA (autoregressive integrated moving average – авторегрессионная интегрированная скользящая средняя) для вычисления прогнозного уровня безработицы, однако такого рода модели не обладают достаточным уровнем надежности. Так, классические линейные модели данного класса хорошо зарекомендовали себя при прогнозировании уровня безработицы в некоторых европейских странах (Edlund, Karlsson, 1993; Dumičić et al., 2015) и Канаде (Khan Jafur et al., 2017), но были не так эффективны в США, где лучше себя проявила модель пороговой авторегрессии (TAR), классическая нелинейная модель временных рядов (Montgomery et al., 1998). Повышению эффективности нелинейных моделей поспособствовал прогресс в области современной статистики и машинного обучения, который снабдил их новыми инструментами анализа. Модели данного типа весьма эффективны для прогнозирования уровня безработицы в долгосрочном периоде (Feuerriegel, Gordon, 2019).

Оба типа моделей не идеальны и имеют некоторые недостатки. Чтобы их преодолеть, была предложена модель авторегрессионной нейронной сети (ARNN) (Faraway, Chatfield, 1998), которая использует элементы линейных и нелинейных моделей и положила начало развитию моделей гибридного типа (Chakraborty et al., 2021).

Помимо традиционных методов, используемых в прогнозировании, сейчас набирают популярность методы прогнозирования различных ключевых макроэкономических показателей, в том числе и безработицы, основанные на поисковых запросах. Такой подход позволяет исследователям использовать массивы данных, альтернативные по сравнению с данными, публикуемыми национальными статистическими службами и соответствующими международными организациями либо получаемыми в ходе различных опросов. «Эти источники можно назвать традиционными, и им присущи следующие недостатки: 1) затратность получения, сложность организации системы сплошного статистического наблюдения и проведения опросов; 2) вре-

<sup>1</sup> <https://1prime.ru/files/pdf/methodology.pdf>

менной лаг получения сведений, который для национальной статистики может составить год и больше; 3) нерепрезентативность выборки в ходе проведения социологических исследований; 4) ограниченные возможности по дезагрегации данных для проведения микроэкономического анализа» (Юревич и др., 2020: 40). Доступность же онлайн-данных в режиме реального времени может помочь исследователям лучше объяснить и спрогнозировать различные экономические показатели, по которым официальные данные предоставляются с задержкой. Это, безусловно, относится и к такому показателю, как уровень безработицы, в связи с чем повышается актуальность оценки уровня безработицы в реальном времени. Чтобы решить проблему с запаздыванием официальной статистики и попытаться повысить точность прогнозирования различных экономических показателей, рядом исследователей были использованы данные об интенсивности поиска, полученные посредством инструмента GoogleTrends.

Так, ряд исследователей анализировали влияние поиска работы через сеть Интернет на возможность прогнозирования уровня безработицы.

Впервые высокий уровень корреляции между уровнем безработицы и конкретными поисковыми запросами в системе Google по ключевым словам был показан в работе (Askittas, Zimmermann, 2009). Исследователи использовали метод причинно-следственной связи Грейнджера для временных рядов при оценке ежемесячного уровня безработицы в Германии через изменения значимых переменных. Модели коррекции ошибок были построены с использованием сезонно нескорректированного уровня безработицы в период с января 2004-го по апрель 2009 года. Авторы использовали такие поисковые запросы, как «уровень безработицы», «бюро или агентство по трудоустройству» и «консультант по персоналу». Такой концептуальный подход к моделированию уровня безработицы с использованием данных об интернет-активности в дальнейшем начал использоваться и исследователями из других стран. Эмпирические результаты показывают, что данные о поисковых запросах в системе Google или другие данные об интернет-активности добавляют релевантную информацию для анализа безработицы по сравнению с показателями бизнес-цикла или традиционными моделями временных рядов.

В одной из работ был изучен уровень безработицы в различных частях Румынии, неоднородных с точки зрения экономического и социального развития, по панельным данным (Simionescu, 2020). Рассмотрены две модели прогнозирования безработицы: одна из них базируется только на использовании официальных макроэкономических показателей, полученных официальными статистическими структурами, другая включает как официальные макроэкономические показатели, так и данные, полученные с использованием инструмента Google Trends. В результате исследователь приходит к выводу, что точность прогноза уровня безработицы при использовании инструмента Google Trends совместно с официальными статистическими данными превосходит точность прогноза, основанного на использовании лишь официальных статистических данных.

К схожим выводам приходят и в (González-Fernández, González-Velasco, 2018). Исследователи проанализировали возможность использования информации, полученной с использованием инструмента Google Trends, при прогнозировании уровня безработицы в Испании. Исследователи отметили высокий уровень корреляции между поисковыми запросами в Google и уровнем безработицы в Испании, а также пришли к выводу об улучшении качества прогнозирования безработицы при использовании информации о соответствующих поисковых запросах в Google.

К выводам о возможности и целесообразности применения интернет-данных при прогнозировании безработицы, а также об улучшении точности прогноза по сравнению с традиционными моделями прогнозирования пришли также (Fondeur, Karamé, 2013) на основе анализа данных по Франции, (Choi, Varian, 2012; Shintaro Nagao et al., 2019) на основе данных по США, (McLaren N., Shanbhogue R., 2011) на основе данных по Великобритании, (Suhoy, 2009) на основе данных по Израилю, (D'Amuri, 2009; Naccarato, et al., 2015) на основе данных по Италии, (Anvik, Gjelsstad, 2010) на основе данных по Норвегии, (Chadwick, Sengül, 2015) на основе данных по Турции, (Lasso, Snijders, 2016) на основе данных по Бразилии. Согласно тесту причинно-следственной связи Грейнджера, индексы поисковых запросов, связанных с безработицей, могут улучшить прогнозы различных макроэкономических показателей также в Китае (Su, 2014).

До того как стали доступны данные о запросах в Google, некоторые экономисты использовали данные об использовании ключевых слов в поисковых системах из отчета Word Tracker's Top 500 Keyword Report, который еженедельно публикуется Rivergold Associates Ltd (Ettredge et al., 2005). Этот отчет охватывал крупнейшие поисковые системы Интернета. Авторы работали с наиболее употребляемыми выражениями, которые могут использовать люди, ищущие работу, и учитывали их для прогнозирования уровня безработицы в США.

Подводя итог, отметим, что доступность данных по запросам в интернете побудила исследователей использовать их для анализа и прогнозирования различные макроэкономические показатели. Рост популярности такого подхода к исследованиям связан с доступностью, большим массивом и достаточным уровнем достоверности данных, полученных с помощью таких новых инструментов анализа, как Google Trends.

Данные служб официальной статистики относительно уровня безработицы, так или иначе, имеют определенные ограничения относительно степени надежности (не говоря уже об оперативности), особенно если речь идет о развивающихся странах. Появление новых методов и инструментов анализа логично. Исследования относительно возможностей использования таких новых инструментов показывают их высокий потенциал. В большинстве исследований отмечается увеличение точности прогнозирования уровня инфляции при использовании интернет-данных. Однако точность прогноза зависит от ряда переменных: степени распространения интернета, возрастной структуры пользователей интернета и т.д.

### Модели прогнозирования уровня безработицы по статистике запросов

Основываясь на зарубежном и российском опыте использования статистики поисковых запросов при прогнозировании и наукастинге уровня безработицы, был составлен список поисковых терминов (табл. 1). Они отражают либо обеспокоенность населения состоянием рынка труда, либо заинтересованность в поиске нового места занятости, либо опасение увольнения. В силу разных предполагаемых мотивов ввода тех или иных слов и словосочетаний между поисковыми образами ожидаема низкая корреляция (пожалуй, единственное явное исключение – запросы «центр занятости» и «биржа труда»).

Из всех поисковых образов достаточно заметной зашумленностью обладает «работа». Судя по сопутствующим словам, с которыми часто ищут этот термин, немалая часть запросов приходится на учащих, интересующихся контрольными и классными работами. Устранение этой части запросов из поискового образа сильно затруднено, так как, например, запрос «работа 2» отражает и режим работы «два через два», и классные работы во втором классе. Все остальные термины, как показывает анализ связанных слов, являются специфичными к рассматриваемой области и не имеют посторонних для данного исследования смыслов.

**Таблица 1**

**Характеристика популярности поисковых терминов, связанных с безработицей**

Поисковый запрос	Количество запросов в месяц, тыс. ед.
«работа»	79 207
«вакансии»	25 570
«центр занятости»	2250
«увольнение»	1999
«безработица»	910
«биржа труда»	626

Источник: <https://wordstat.yandex.ru/> – Дата обращения: 15.04.2021.

В зарубежной практике в поисковые образы часто включаются названия порталов, агрегирующих информацию о доступных вакансиях, таких как Superjob или Indeed. Отечественные аналоги (Head Hunter, Работа Mail, Авито.работа и др.) также имеют высокую популяр-

ность в российском сегменте интернета. Но при рассмотрении динамики запросов с названиями этих сервисов видно, что их востребованность заметно увеличилась за последние 3–5 лет, т.е. для изучения зависимостей на длительном временном интервале они все же не подходят.

В качестве измерителя уровня безработицы используется традиционный показатель, рассчитываемый Росстатом, – «доля безработных в общей численности рабочей силы в возрасте 15 лет и старше»<sup>2</sup> (единица измерения – проценты). Показатели частоты запросов автоматически нормируются по отношению к максимальному значению за рассматриваемый период и относительно общего числа запросов за каждый временной период (диапазон изменения – от 0 до 100 ед.). И по динамике запросов, и по уровню безработицы данные доступны с глубокой ретроспективой. Однако ряд индексов запросов до 2011 года имеют излишне высокую волатильность, что может быть связано с относительно небольшим количеством запросов по данному термину и в целом по всем поисковым операциям в браузере Google. Поэтому временной период наблюдения выбран с января 2011 года по март 2021 года.

На рис. 1–3 продемонстрирована сравнительная динамика уровня безработицы и индексов частоты отобранных поисковых запросов. С целью более наглядной графической иллюстрации каждый из показателей был шкалирован от 0 до 1 по процедуре МИН-МАКС.

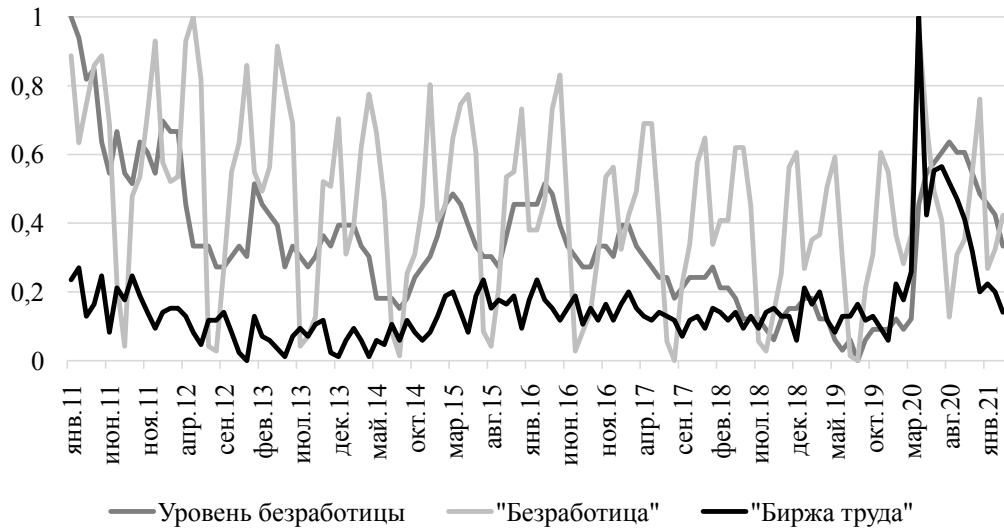


**Рис. 1.** Динамика уровня безработицы и запросов «Вакансии» и «Работа» в браузере Google  
 Источник: расчеты авторов по данным Росстат и GoogleTrends.

Большинство из переменных, сконструированных по статистике запросов, имеют достаточно высокую волатильность. При этом запросы «Безработица», «Биржа труда» и «Центр занятости» демонстрируют высокую чувствительность к будущим скачкам уровня безработицы. На всех графиках заметна ярко выраженная сезонность каждой из рассматриваемых переменных. Ее присутствие было подтверждено QS-тестом, являющимся аналогом теста Льюнга – Бокса, и F-тестом на сезонные фиктивные переменные. Устранение сезонной компоненты было выполнено при помощи алгоритма X-13ARIMA-SEATS<sup>3</sup>. После этой операции были построены корреляционные матрицы с целью обнаружения оптимальной величины лага зависимости между уровнем безработицы и переменными по статистике запросов (табл. 2), а также определения степени коррелированности последних между собой (табл. 3).

<sup>2</sup> Росстат. Уровень безработицы (по методологии Международной организации труда) среди населения в возрасте 15 лет и старше (<https://rosstat.gov.ru/opendata/7708234640-unemploymentrate2>).

<sup>3</sup> SEASONAL X-13ARIMA-SEATS (<http://www.seasonal.website/>).



**Рис. 2.** Динамика уровня безработицы и запросов «Безработица» и «Биржа труда» в браузере Google  
 Источник: расчеты авторов по данным Росстат и GoogleTrends.



**Рис. 3.** Динамика уровня безработицы и запросов «Центр занятости» и «Увольнение» в браузере Google  
 Источник: расчеты авторов по данным Росстат и GoogleTrends.

**Таблица 2**

**Корреляционный анализ уровня безработицы и переменных, отражающих интенсивность запросов**

	Величина лага, мес.						
	0	1	2	3	4	5	6
«Вакансии»	-0,45	-0,46	-0,46	-0,45	-0,45	-0,44	-0,43
«Работа»	0,02	0,02	0,00	0,02	0,03	0,03	0,05
«Безработица»	0,59	0,60	0,57	0,57	0,57	0,51	0,51
«Биржа труда»	0,42	0,40	0,41	0,39	0,39	0,34	0,28
«Центр занятости»	0,15	0,17	0,15	0,14	0,11	0,06	0,01
«Увольнение»	-0,16	-0,06	-0,02	0,06	0,16	0,23	0,29

Источник: расчеты авторов.



Наиболее плотная корреляционная связь уровня безработицы наблюдается с динамикой запросов «Вакансии», «Безработица» и «Биржа труда» с минимальными лагами. С остальными переменными эта связь не слишком заметна. Неожиданным следует признать отрицательный знак коэффициентов корреляции уровня безработицы с динамикой запросов «Вакансии». Можно предположить, что слово «вакансии» больше ассоциируется со сменой работы при наличии текущего места трудоустройства. Иными словами, в данном случае поиск нового места занятости происходит добровольно, а не вынужденно, когда увольнение уже произошло или ожидается в ближайшем будущем. Кроме того, важно отметить, что некоторые из переменных сохраняют или приобретают достаточно заметную корреляционную связь с уровнем безработицы также и с лагом в 6 месяцев, что может трактоваться в качестве предпосылки для построения прогнозных моделей с протяженным горизонтом. Что касается взаимозависимости переменных частоты запросов, то подтвердилось ранее сформулированное предположение о наличии сильной корреляции между запросами «Биржа труда» и «Центр занятости». Остальные переменные в достаточной мере независимы друг от друга.

Таблица 3

## Корреляционная матрица переменных, отражающих интенсивность запросов

	«Вакансии»	«Работа»	«Безработица»	«Биржа труда»	«Центр занятости»	«Увольнение»
«Вакансии»	1,00					
«Работа»	0,33	1,00				
«Безработица»	-0,39	0,03	1,00			
«Биржа труда»	0,12	-0,24	0,34	1,00		
«Центр занятости»	0,33	-0,38	0,02	0,82	1,00	
«Увольнение»	-0,09	-0,09	0,32	0,12	0,02	1,00

Источник: расчеты авторов.

Распространенным математическим приемом понижения размерности матрицы объясняющих переменных является метод главных компонент. При использовании динамики запросов, полученных по нескольким поисковым образам, этот подход получил особую популярность, в том числе было доказано его превосходство над остальными алгоритмами слияния переменных (Петрова, 2019; Kreiner, Duca, 2020). Для отобранных шести переменных статистики запросов, согласно анализу дисперсии, целесообразно использовать две главные компоненты (суммарно они объясняют более 70% дисперсии исходных данных):

$$PC1 = 0.66 * \text{"Вакансии"} + 0.06 * \text{"Работа"} - 0.71 * \text{"Безработица"} - 0.04 * \text{"Биржа труда"} + 0.18 * \text{"Центр занятости"} - 0.15 * \text{"Увольнение"}; \quad (1)$$

$$PC2 = 0.28 * \text{"Вакансии"} - 0.11 * \text{"Работа"} + 0.38 * \text{"Безработица"} + 0.51 * \text{"Биржа труда"} + 0.70 * \text{"Центр занятости"} + 0.11 * \text{"Увольнение"} . \quad (2)$$

Эталонная авторегрессионная модель (модель 1) имеет высокую долю объясненной дисперсии (нормированный R-квадрат) и достаточно низкую среднюю абсолютную ошибку (MAE) (табл. 4). Модели 2 и 3, в которых объясняющими переменными являются лагированные на один месяц частоты запросов, оказались менее точными в прогнозировании. При этом статистическая значимость была зафиксирована для всех переменных без исключения. При использовании главных компонент также была получена статистически значимая форма модели, но с высокой погрешностью прогнозов. Снижения MAE удалось достичь при совмещении лагированного уровня безработицы и трех переменных частот запросов (модель 4). Сочетание главных компонент с уровнем безработицы предыдущего периода не дало положительных результатов. Что касается

сравнения моделей по байесовскому информационному критерию (BIC), то модели 1 и 4 оказались почти эквивалентны, а остальные вариации заметно уступили в релевантности.

Таблица 4

**Модели прогнозирования уровня безработицы**

	<b>Модель 1</b>	<b>Модель 2</b>	<b>Модель 3</b>	<b>Модель 4</b>	<b>Модель 5</b>
Константа	0,003** (0,001)	0,018*** (0,005)	0,019*** (0,006)	0,006*** (0,002)	0,051*** (0,001)
Уровень безработицы (–1)	0,934*** (0,025)			0,824*** (0,041)	
«Вакансии» (–1)		–0,030*** (0,003)	–0,032*** (0,003)	–0,005*** (0,002)	
«Работа» (–1)		0,046*** (0,005)	0,050*** (0,006)	0,007** (0,003)	
«Безработица» (–1)		0,008*** (0,003)	0,012*** (0,003)		
«Биржа труда» (–1)		0,041*** (0,003)		0,009*** (0,003)	
«Центр занятости» (–1)			0,026*** (0,003)		
«Увольнение» (–1)		0,016** (0,005)	0,015** (0,006)		
PC1 (–1)					–0,023*** (0,002)
PC2 (–1)					0,011*** (0,002)
Количество наблюдений	123	123	123	123	123
Нормированный R-квадрат	0,92	0,72	0,64	0,92	0,46
MAE	0,104	0,236	0,276	0,094	0,356
BIC	–1213,7	–1046,3	–1015,4	–1211,3	–976,4

\* $p < 0,1$ ; \*\* $p < 0,05$ ; \*\*\* $p < 0,01$ , в скобках величина стандартной ошибки.

Источник: расчеты авторов.

Таким образом, добавление в авторегрессионную модель нескольких переменных частоты запросов оказалось способно повысить прогностическую точность модели. С одной стороны, разница средней абсолютной ошибки прогноза в 0,01% не выглядит очень существенной, а с другой – произошло увеличение точности прогнозов почти на 10%. К сожалению, в России не строятся месячные прогнозы уровня безработицы на регулярной основе с длинной историей. Пожалуй, единственным исключением является портал [investing.com](https://ru.investing.com)<sup>4</sup>, публикующий такую информацию. Однако о способе прогнозирования почти нет информации, лишь указано, что он экспертный. С июня 2011-го по март 2021 года средняя абсолютная ошибка прогноза согласно этому источнику составила 0,126%. Даже с поправкой на достоверность данных портала полученная гибридная модель характеризуется значительно более высокой точностью.

Содержательные причины превосходства гибридной модели над авторегрессионной вариацией состоят в способности первой реагировать на будущие шоки рынка труда. Иными словами, эта модель отражает массовые ожидания увольнения или фиксирует их до факта отражения в официальной статистике. Учитывая, что в гибридную модель вошла статистика запросов «Вакансии», «Работа» и «Биржа труда», по всей видимости, удалось охватить запросы и тех, кто предсказывает собственное

<sup>4</sup> <https://ru.investing.com/economic-calendar/russian-unemployment-rate-556>

увольнение, и тех, у кого оно уже случилось. В любом случае полученный результат свидетельствует в пользу ценности статистики запросов при прогнозировании уровня безработицы.

### Заключение

Результаты исследования показывают, что статистика запросов пользователей вполне может использоваться в качестве предиктора изменений макроэкономических индикаторов. Соответственно, основная цель проведенного исследования, которая состояла в доказательстве релевантности наукастинга по данным запросов как метода краткосрочного прогнозирования, была достигнута. Этот факт позволяет сделать предположение о грядущем расширении применения данного подхода как перспективного инструмента мониторинга настроений экономических агентов, функции которого доселе традиционно отводились социологическим исследованиям. Хотя в силу небольшой изученности объясняющих и предиктивных свойств статистики запросов могут возникать «подводные камни» (например, из-за присутствия различных смыслов в одних и тех же словах или влияния ненаблюдаемых событий). Тем не менее перспективы этого инструмента видятся достаточно ясными.

Практическая ценность полученных результатов заключается в расширении инструментария мониторинга и прогнозирования безработицы. Учитывая, что анализ активности интернет-пользователей обладает такими неоспоримыми преимуществами по сравнению с социологическими исследованиями, как оперативность получения информации, большой охват населения и, конечно же, меньшая затратность, применение этого источника данных выглядит вполне целесообразным для профильных органов власти или аналитических центров. Вдобавок, можно предположить, что изучение запросов способно давать более репрезентативные показания об ожиданиях населения относительно их возможного увольнения. В итоге конечная выгода состоит в получении более надежного инструмента прогнозирования уровня безработицы и шоков рынка труда.

Дальнейшее развитие разработанного инструментария видится в двух направлениях. Во-первых, прогностическая точность может быть повышена за счет расширения поискового образа, а также более скрупулезной валидации отдельных терминов. Во-вторых, все чаще в моделях на основе больших данных методы глубокого машинного обучения демонстрируют превосходство над классическими эконометрическими подходами, в частности моделями авторегрессии. Соответственно, внедрение нового алгоритма расчета также имеет большой потенциал повышения точности прогнозов уровня безработицы.

### Литература / References

- Коровкин А.Г. (2001). *Динамика занятости и рынка труда: вопросы макроэкономического анализа и прогнозирования*. М.: МАКС Пресс. [Korovkin A.G. (2001). *Dynamics of Employment and Labor Market: Issues of Macroeconomic Analysis and Forecasting*. Moscow: MAX Press (in Russian)].
- Коровкин А.Г., Полежаев А.В., Андрунин А.В. (2002). Структурные изменения и взаимодействия региональных рынков труда: методы и анализ // *Проблемы прогнозирования* (4): 134–138. [Korovkin A.G., Polezhaev A.V., Andryunin A.V. (2002). Structural changes and interactions of regional labor markets: Methods and analysis. *Problems of Forecasting* (4): 134–138 (in Russian)].
- Петрова Д.А. (2019). Прогнозирование инфляции на основе интернет-запросов // *Экономическое развитие России* 26(11): 55–61 [Petrova D.A. (2019). Inflation forecasting based on Internet requests. *Economic Development of Russia* 26(11): 55–61 (in Russian)].
- Юревич М.А., Екимова Н.А., Балацкий Е.В. (2020). Цифровая трансформация экономической науки // *Информационное общество* (2): 39–47. [Yurevich M.A., Ekimova N.A., Balatsky E.V. (2020). Digital transformation of economic science. *Information society* (2): 39–47 (in Russian)].
- Anvik S., Gjelstad K. (2010). Just Google it. Forecasting Norwegian unemployment figures with web queries. *Working Paper 11*. Oslo: Center for Research in Economics and Management.
- Askitas N., Zimmermann K.F. (2009). Google econometrics and unemployment forecasting. *Appl. Econ. Q.* 55(2): 107–120.

- Barnes M., Chahrour R., Olivei G., Tang G. (2007). A principal components approach to estimating labor market pressure and its implications for inflation. *Federal Reserve Bank of Boston Public Policy Brief 07-2*: 1–23.
- Chadwick M.G., Sengül G. (2015). Nowcasting the unemployment rate in Turkey: Let's ask google. *Central Bank Review* **15**(3): 15.
- Chakraborty T., Chakraborty A.K., Biswas M., Banerjee S., Bhattacharya S. (2021). Unemployment rate forecasting: A hybrid approach. *Computational Economics* **57**: 183–201.
- Choi H., Varian H. (2012). Predicting the present with google trends. *Econ. Rec.* **88**(s1): 2–9.
- D'Amuri F. (2009). *Predicting Unemployment in Short Samples with Internet Job Search Query Data*. Germany: University Library of Munich.
- Dumičić K., ČehČasni A., Žmuk B. (2015). Forecasting unemployment rate in selected European countries using smoothing methods. *World Academy of Science, Engineering and Technology: International Journal of Social, Education, Economics and Management Engineering* **9**(4): 867–872.
- Edlund P.O., Karlsson S. (1993). Forecasting the Swedish unemployment rate var vs. transfer function modeling. *International Journal of Forecasting* **9**(1): 61–76.
- Ettredge M., Gerdes J., Karuga G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM* **48**(11): 87–92.
- Faraway J., Chatfeld C. (1998). Time series forecasting with neural networks: A comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**(2): 231–250.
- Feuerriegel S., Gordon J. (2019). News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions. *European Journal of Operational Research* **272**(1): 162–175.
- Fondeur Y., Karamé F. (2013). Can google data help predict French youth unemployment? *Econ. Modell.* **30**: 117–125.
- González-Fernández M., González-Velasco C. (2018). Can Google econometrics predict unemployment? Evidence from Spain. *Economics Letters* **170**: 42–45.
- Khan Jafur Z.R., Sookia N.U.H., NunkooGonpot P., Seetanah B. (2017). Out-of-sample forecasting of the Canadian unemployment rates using univariate models. *Applied Economics Letters* **24**(15): 1097–1101.
- Kreiner A., Duca J. (2020). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters* **27**(17): 1434–1437.
- Lasso F., Snijders S. (2016). The power of Google search data; an alternative approach to the measurement of unemployment in Brazil. *Student Undergraduate Research E-journal!* (2).
- McLaren N., Shanbhogue R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* (Q2).
- Montgomery A.L., Zarnowitz V., Tsay R.S., Tiao G.C. (1998). Forecasting the us unemployment rate. *Journal of the American Statistical Association* **93**(442): 478–493.
- Naccarato A., Pierini A., Falorsi S. (2015). *Using Google trend data to predict the Italian unemployment rate*. Departmental Working Papers of Economics – University “Roma Tre” 0203. Department of Economics – University Roma Tre.
- Nagao S., Takeda F., Tanaka R. (2019). Nowcasting of the U.S. unemployment rate using Google Trends. *Finance Research Letters* (30): 103–109.
- Simionescu M. (2020). Improving unemployment rate forecasts at regional level in Romania using Google Trends. *Technological Forecasting & Social Change* **155**(C).
- Su Z. (2014). Chinese online unemployment-related searches and macroeconomic indicators. *Frontiers of Economics in China* **9**(4): 573–605.
- Suhoy T. (2009). *Query indices and a 2008 downturn: Israeli data*. Discussion Paper № 2009.06. Bank of Israel.
- Zmitrowicz K., Khan M. (2014). Beyond the unemployment rate: Assessing Canadian and U.S. labor markets since the Great Recession. *Bank of Canada Review* Spring: 42–53.