

Обзор успешных практик применения наукастинга в социально-экономическом прогнозировании

Д. В. Фирсов  , Т. К. Чернышева 

Финансовый университет при Правительстве Российской Федерации

г. Москва, Россия

 dvfirsov@fa.ru

Аннотация. Необходимой компетенцией в современных реалиях становится умение анализировать большие данные в экономике, в связи с чем одним из ключевых вопросов становится выбор инструментов для проведения подобного анализа. Одним из таких перспективных инструментов является наукастинг, который позволяет с высокой точностью определять экономические изменения на очень коротких временных отрезках. Актуальность исследования связана с тем, что наукастинг как инструмент позволяет, помимо прочего, прогнозировать поведение пользователей, выявлять ключевые макроэкономические тенденции и, таким образом, являться основной для принятия управленческих решений как в корпоративном, так и в государственном секторе. Целью исследования является проведение анализа успешных современных практик применения наукастинга для статистического прогнозирования социально-экономических показателей. Гипотеза исследования заключается в предположении, что наукастинг как метод макроэкономического анализа может в ближайшей перспективе стать достойной альтернативой традиционным методам анализа и статистического прогнозирования показателей социально-экономического развития, повысив точность их прогнозирования. Методологической основой исследования послужили научные труды и прикладные разработки ведущих отечественных и зарубежных ученых в области экономического прогнозирования с применением статистики поисковых запросов, а также методы сравнительного и статистического анализа, системный подход. Новизна полученных результатов заключается в систематизации и описании успешных практик применения наукастинга и прогнозирования показателей при помощи статистики запросов. В исследовании выделен основной принцип наукастинга, который заключается в получении более точной оценки состояния экономики по мере поступления новых данных. Также описаны ключевые статистические модели, применяемые в качестве инструментария наукастинга в зарубежных странах. В результате проведенного исследования было выявлено, что анализ динамики поисковых запросов, особенно в контексте ее корреляции с классическими метриками и традиционной статистикой, находится на активной фазе развития, особенно в рамках отечественной практики прогнозирования. Полученные результаты могут применяться на практике как в корпоративном, так и в государственном секторе для построения макроэкономических прогнозов.

Ключевые слова: поисковые запросы; экономические исследования; статистика запросов; модели анализа больших данных; наукастинг.

1. Введение

В эпоху цифровой трансформации умение анализировать большие данные становится необходимой компетенцией

для специалистов почти всех областей экономики, и в этой связи одним из основополагающих вопросов является выбор инструментов для проведения

анализа. Такой инструмент, как статистика поисковых запросов, помогает конвертировать данные в информацию, позволяя принимать обоснованные и рациональные решения. Поисковый запрос представляет собой единичные слова, ключевые фразы или предложения, которые отображаются в форме конкретного сайта как конечный результат поиска пользователя. А поисковая система, в свою очередь, это общее название службы, выполняющей поиск.

Анализ статистики поисковых запросов, особенно в контексте их корреляции с классическими метриками опросов и общей статистики, находится еще на очень ранней, но довольно активной ступени развития, особенно в рамках отечественной практики прогнозирования. Можно предположить, что практика применения этих методик анализа будет увеличиваться благодаря нарастающему интересу не только со стороны частного сектора, но и со стороны органов государственной власти.

Главным конкурентным преимуществом анализа статистических запросов является то, что они позволяют с высокой точностью определять экономические изменения на очень коротких временных отрезках. В последнее десятилетие сформировался устойчивый термин этого инструментария – «nowcasting» (далее – наукастинг). *Наукастинг* представляет собой предсказание настоящего, ближайшего будущего и недавнего прошлого состояния экономических индикаторов.

Актуальность проводимого исследования связана с тем, что наукастинг как инструмент позволяет, помимо прочего, прогнозировать поведение пользователей, выявлять ключевые макроэкономические тенденции и, таким образом, может являться основной для принятия управленческих решений как в корпоративном, так и в государственном секторе.

Этим обуславливается необходимость расширения практики применения данного инструмента для статистического прогнозирования социально-экономических показателей.

Целью исследования является проведение анализа успешных современных практик применения наукастинга для статистического прогнозирования социально-экономических показателей. Для достижения поставленной цели были сформулированы следующие задачи: 1) выявление ключевых областей экономики, в которых применение метода наукастинга наиболее целесообразно; 2) определение степени практического применения данного инструмента в отечественной и зарубежной практике экономического прогнозирования.

Гипотеза исследования заключается в предположении, что наукастинг как метод макроэкономического анализа может в ближайшей перспективе стать достойной альтернативой традиционным методам анализа и статистического прогнозирования показателей социально-экономического развития, повысив точность их прогнозирования.

Исследование состоит из нескольких основных частей. Первая часть содержит теоретико-методологические аспекты и инструментарий наукастинга в современных социально-экономических реалиях. Вторая часть исследования раскрывает лучшие мировые и отечественные практики применения наукастинга для прогнозирования социально-экономических показателей.

2. Новая парадигма экономических исследований

2.1. Эволюция применения статистики поисковых запросов для прогнозирования социально-экономических показателей

Говоря об эволюции поисковых систем в мировой и отечественной

практике, отметим, что в конце 90-х гг. XX в. такие поисковые системы, как MSN, Lycos, Excite и Yahoo!, довольно быстро стали набирать популярность среди пользователей по всему миру [1]. Для того чтобы пользователи проводили больше времени на этих сайтах-порталах, было создано множество новых сервисов на базе данных платформ. Однако с появлением в 1998 г. Google, который в отличие от других поисковиков предоставлял широкий инструментарий для анализа взаимосвязей между разными сайтами и применял более совершенные технологические решения, чем вышеперечисленные сайты-порталы, на рынке онлайн-поисковиков появился явный лидер [2].

Говоря о том, какую долю рынка занимает та или иная поисковая система в настоящее время, отметим, что

по состоянию на январь 2021 г. онлайн-поисковик Bing занимает 6,84 % мирового рынка, в то время как лидер рынка Google имеет рыночную долю в 85,86 %. Доля рынка китайской поисковой системы Baidu составила 0,55 %, а российского Yandex – 0,59 %. Несмотря на то, что «Яндекс» и Baidu имеют потенциал для того, чтобы расти несколько более быстрыми годовыми темпами по сравнению с Google, разрыв между Google и другими поисковыми системами с точки зрения занимаемого объема рынка крайне велик (рис. 1).

Лидирующие позиции Google можно объяснить тем, что, помимо того, что Google является поисковой системой, он также предоставляет множество других услуг: Gmail (популярный почтовый сервис), Google news, Google shopping, инструменты повышения

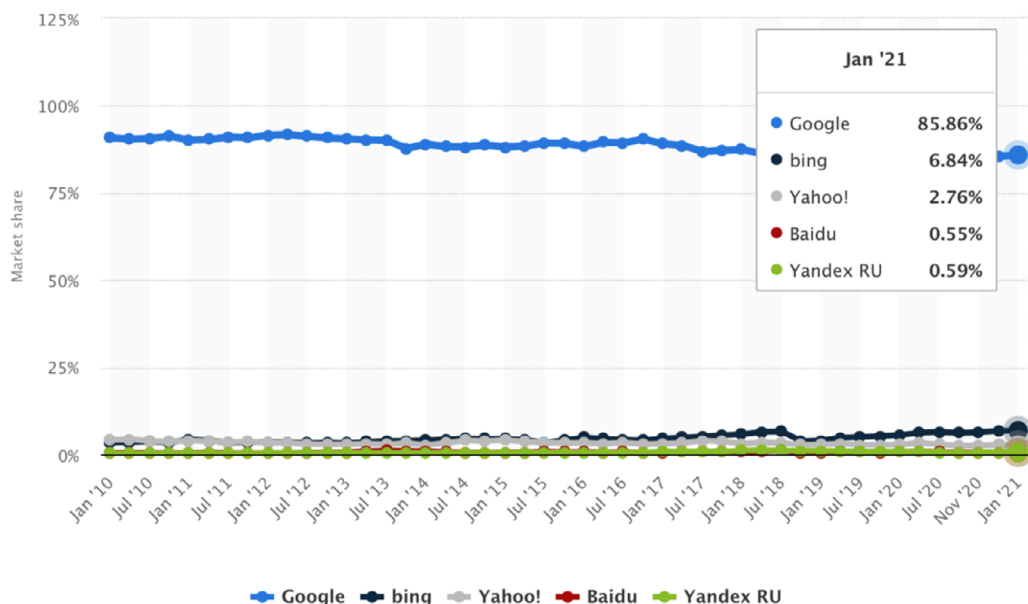


Рис. 1. Доля мирового рынка ведущих поисковых систем (за период с января 2010 г. по январь 2021 г.)

Fig. 1. Global market share of leading search engines (for the period from January 2010 to January 2021)

Источник: составлено авторами на основе данных сайта Statista. URL: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>

Source: compiled by the authors based on data from the Statista website. URL: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>

производительности, платформу для видео и изображений и ряд других инструментов. Всемирно известным аналитическим сервисом является Google Trends, который представляет собой инструментарий, показывающий популярность какого-либо поискового термина в Google. С его помощью можно легко проанализировать, находится тот или иной тренд на подъеме или на спаде. Также данный сервис позволяет найти необходимые демографические данные, смежные тематики, связанные с запросом, и иную информацию, которая поможет лучше понять актуальные тенденции Google.

Google Trends предоставляет доступ к нефилтрованной (в значительной степени) выборке реальных поисковых запросов, сделанных в Google. К особенностям данного сервиса можно отнести то, что он является анонимным (то есть ни один пользователь не идентифицируется), классифицируемым (происходит определение темы для поискового запроса) и агрегируемым (то есть данные группируются вместе) [3]. Существует два типа данных Google Trends, к которым можно получить свободный доступ:

1) данные в режиме реального времени (это выборка, охватывающая данные за последние 7 дней);

2) данные, не относящиеся к реальному времени (представляют собой отдельную выборку данных начиная с 2004 г.).

Отметим, что, хотя в Google Trends используется только выборка поисковых запросов Google, можно утверждать, что этого вполне достаточно для качественного анализа, поскольку в настоящее время Google обрабатывает более 5,6 млрд поисковых запросов в день. Используя выборку конкретных данных, представляется возможным посмотреть на набор данных, репрезентативный для

всех поисковых запросов Google, и найти информацию, которая может быть обработана в течение нескольких минут после события, произошедшего в мире в режиме реального времени [4].

Также стоит сказать о том, что современный инструментарий прогнозирования и анализа на основе поисковых систем не ограничивается анализом определенных языковых форм. В 2021 г. коллектив американских ученых провел исследование, в котором они попытались выявить и проанализировать наличие взаимосвязи между фотографиями Google street view и уровнем преступности в регионе [5].

Исследователи Steven Scot и Hal Varian в своей работе представили систему для прогнозирования экономических временных рядов на основе данных Google Trends и Google Correlate. Система комбинирует структурные модели временных рядов с байесовской регрессией. Результаты анализа показали, что представленная модель позволяет значительно повысить точность официальных статистических методик [6].

Бразильские ученые Marcelo Medeiros и Henrique Pires одними из первых исследовали не только закономерности прогнозирования на основе поисковых запросов, но и закономерности их регионального распределения. Рассматривая поисковые запросы Google в США и Бразилии, они доказали различия в статистических закономерностях разных стран. Исследование показало, что в проведении сравнительного анализа на основе поисковых запросов необходимо обращать внимание не только на общие закономерности, но и на межстрановую специфику [7].

Говоря о социально-экономических показателях (индикаторах), отметим, что они представляют собой важнейший ресурс для органов государственной

власти, руководителей предприятий и представителей широкой общественности. Они помогают принимать решения, основанные на фактических данных, проводить сравнения между планируемыми и фактическими результатами, а также способствуют повышению прозрачности и подотчетности деятельности государственных органов [8].

Наукастинг в современных реалиях может стать достойной альтернативой традиционным статистическим показателям. Анализировать и предсказывать тенденции потребления населения при помощи методик наукастинга и простых поисковых запросов относительно просто, чуть сложнее ситуация становится, когда происходит уход от товарно-сервисной составляющей и исследователи пытаются применить методы наукастинга в рамках анализа безработицы, делового климата и иных показателей, которые зачастую не имеют прямого выражения в сетевом потоке данных.

2.2. Статистические модели, применяемые в качестве инструментария наукастинга

Современный наукастинг не ограничивается анализом только поисковых запросов. За последние несколько лет практический инструментарий наукастинга расширился до анализа всех возможных источников информации, включая данные о покупках населения, макроэкономические большие данные и даже данные с камер видеонаблюдения. Современные модели и методики экономического наукастинга используют разнородный набор предикторов, куда входят как количественные данные (такие как статистика безработицы и объем промышленного производства), так и качественные данные (социальные опросы населения и мнения экспертов).

Устоявшийся методический инструментарий наукастинга включает в себя

ряд статистических моделей, таких как авторегрессионный анализ, анализ ведущих индикаторов, динамическая факторная модель, байесовская векторная авторегрессия и регрессия смешанной выборки данных. Проанализируем особенности каждого из названных методов анализа.

1. *Авторегрессионный анализ*, несмотря на свою простоту, зарекомендовал себя как один из самых действенных инструментов наукастинга и линейного прогностического моделирования [9]. Авторегрессия используется для прогнозирования в тех случаях, когда существует некоторая корреляция между значениями во временном ряду и значениями, которые предшествуют им и следуют за ними. Авторегрессионные модели удивительно гибки при обработке широкого спектра различных моделей временных рядов.

2. *Анализ ведущих индикаторов* – это анализ переменных, движение которых имеет прямое отношение к движению исследуемой переменной. Движение ведущего индикатора может характеризоваться как однонаправленно, то есть двигаться в том же направлении, что и интересующая переменная, так и реверсивно, то есть в противоположном направлении. Главным условием данной модели является наличие причинно-следственной и логической связи, аргументирующей первоначальный расчет. Зачастую найти закономерности между двумя индикаторами не так сложно, как найти между ними логическую взаимосвязь. Ведь если отсутствует логическая аргументация, то взаимосвязь вообще может не существовать, и, следовательно, вполне вероятно, что наблюдаемые отношения являются ложными [10]. Количественный характер отношений должен исходить из симбиоза анализа исторических данных и практического мышления. Некоторые ведущие

индикаторы будут иметь кумулятивный эффект с течением времени (например, количество осадков как индикатор воды, доступной для использования на гидроэлектростанции), поэтому их необходимо суммировать или усреднять.

3. *Байесовская векторная авторегрессия* использует методы Байеса для оценки вектора авторегрессии модели [11]. Байесовская векторная авторегрессия отличается от стандартных моделей авторегрессии тем, что параметры модели рассматриваются как случайные величины с априорными вероятностями, а не как фиксированные значения.

Факторная модель, в свою очередь, присваивает веса рядам, оптимально используя динамические отношения между ними. Суть модели динамических факторов заключается в том, что несколько скрытых динамических факторов управляют совместным перемещением многомерного вектора переменных временного ряда, на который также влияет вектор среднего значения. Скрытые факторы следуют процессу временных рядов, которые обычно считаются посредством векторной авторегрессии. Важной составляющей факторных моделей является тот факт, что при определении точных латентных факторов можно делать эффективные прогнозы практически во всех сферах экономики и финансов.

У динамических факторных моделей есть две главные характеристики, позволившие им занять доминантное положение в практике статистического наукастинга [12]: их способность описать эмпирические макроэкономические данные, основываясь на динамической макроэкономической теории, и их расчетная точность, при которой небольшой ряд факторов объясняют дисперсию макроэкономических показателей.

4. *Регрессия смешанной выборки данных*. Одной из самых часто

применяемых моделей в современной практике наукастинга является модель MIDAS, также известная как регрессия смешанной выборки данных [13]. Макрофинансовые показатели прогнозирования требуют объединения низкочастотных и высокочастотных временных рядов. Регрессии смешанной выборки данных объясняют низкочастотную переменную на основе высокочастотных переменных и их лагов. Например, зависимой переменной может быть квартальный ВВП, а независимыми переменными – ежемесячная активность или ежедневные рыночные данные.

На практике MIDAS используется для прогнозирования волатильности финансовых рынков, роста ВВП, тенденций инфляции и бюджетных тенденций [14]. Однако применение модели MIDAS не ограничивается только лишь экономическим анализом. Первоначальной практикой его применения был анализ краткосрочных погодных изменений в попытке предсказать общие погодные тенденции. А впоследствии эта модель стала применяться практически во всех сферах. Так, одним из примеров ее популяризации в качестве управленческого индикатора выступает применение в работе налоговой службы США, которая в 2003 г. начала использовать массивы административных данных по доходам домохозяйств для продвижения идеи прогрессивной шкалы налогообложения.

3. Процедура исследования

Фундаментом исследования выступают научные труды, размещенные в электронных реестрах публикаций и онлайн-библиотеках (Web of Science, Google Scholar и elibrary.ru). Выборка исследуемых работ проводилась по принципу общего тематического соответствия. Временные рамки работ

ограничиваются исследованиями за последние 10 лет. Систематизация работ проводилась по областям исследований и их отраслевому соответствию.

4. Обзор успешных практик использования наукастинга

4.1. Применение наукастинга в деятельности финансовых органов зарубежных стран

Наукастинг, используемый в качестве метода формирования и прогнозирования экономических показателей, уже сейчас позволяет прийти к выводу, что точность и эффективность прогнозов формируется и зависит от двух основных факторов: совершенства аналитического инструментария и полноты используемых данных (иными словами, от точности применяемых математических и статистических методов прогнозирования) и данных, которые в этих моделях применяются. Оба этих элемента нашли логическое развитие в применении статистики поисковых

интернет-запросов в качестве основного источника получения данных.

За последнее десятилетие наукастинг зарекомендовал себя в качестве надежного инструмента прогнозирования, способного конкурировать с уже существующими моделями статистического прогноза экономических изменений. Академические исследования [15, 16] показывают, что прогнозы наукастинга представляют качественно новый уровень по сравнению со стандартными моделями статистического прогнозирования, которые используют только официальные методы сбора и обработки данных. Ярким примером практического использования наукастинга является его применение в деятельности мировых центральных банков (табл. 1).

Главным объединяющим элементом, характеризующем все приведенные в табл. 1 организации, выступает факт применения в том или ином виде факторных моделей. Несмотря на то, что приведенные модели отличаются друг

Таблица 1. Применение наукастинга в деятельности центральных банков зарубежных стран

Table 1. The use of nowcasting in the activities of central banks in foreign countries

№ п/п	Страна / Country	Центральный банк / Central Bank	Направления применения наукастинга / Areas of application of nowcasting
1	Великобритания	Банк Англии	Комитет по денежно-кредитной политике Банка Англии использует компиляцию прогнозов наукастинга по трем различным моделям, чтобы сформировать свое первоначальное представление о текущем состоянии экономики. В частности: 1) на основе различных отраслей (например, розничных услуг, производства, строительства и производства, строительства и т. д.) для имитации производственного подхода к расчету ВВП; 2) модели выборки смешанных данных; 3) модели динамических факторов. Затем они собираются вместе для формирования прогноза по денежно-кредитной политике

Окончание табл. 1

End of table 1

№ п/п	Страна / Country	Центральный банк / Central Bank	Направления применения наукастинга / Areas of application of nowcasting
2	США	Федеральный резервный банк Атланты	Центр количественных экономических исследований Атланты публикует частые общедоступные прогнозы ВВП США. Он рассчитывает 13 отдельных компонентов расходов ВВП (например, потребительские расходы, инвестиции и т. д.), чтобы имитировать расходный подход к расчету ВВП используется динамическая факторная модель, на основе больших данных
3	США	Федеральный резервный банк Нью-Йорка	Федеральный резервный банк Нью-Йорка публикует свой собственный общедоступный прогноз роста ВВП. Он также использует динамический факторный подход, основанный на принципе поисковых запросов
4	Еврозона	Европейский центральный банк	ЕЦБ использует модели наукастинга и динамических факторов для обоснования своих управленческих решений. Вместе с этим ЕЦБ на постоянной основе публикует рабочие документы по совершенствованию наукастинга в сфере экономического прогнозирования
5	Норвегия	Центральный банк Норвегии	Банк Норвегии использует различные статистические модели наукастинга и краткосрочного прогнозирования ВВП и инфляции для обоснования своих решений по ключевой ставке. Используя несколько взаимозависимых моделей, банк формирует сводный прогноз SAM (Система моделей усреднения), дающую средневзвешенное значение результатов различных моделей

Источник / Source: составлено авторами / compiled by the authors.

от друга, различия эти очень несущественны и в большинстве своем выражаются в типах датасетов на основе которых проводится анализ.

Особо хочется отметить модели и методы наукастинга Федерального резервного банка Нью-Йорка¹, данные

¹ Macroeconomic Nowcasting and Forecasting with Big Data. URL: https://www.researchgate.net/publication/325612519_Macroeconomic_Nowcasting_and_Forecasting_with_Big_Data (дата обращения: 08.03.2021).

которого обновляются ежедневно в 10 часов утра. Ежедневное обновление позволяет пользователям анализировать поток данных в реальном времени и количественно определять, как каждая публикация данных влияет

[net/publication/325612519_Macroeconomic_Nowcasting_and_Forecasting_with_Big_Data](https://www.researchgate.net/publication/325612519_Macroeconomic_Nowcasting_and_Forecasting_with_Big_Data) (дата обращения: 08.03.2021).

на общее положение прогнозов всех других переменных. Вместе с этим ежедневно проводится не только публикация новых аналитических данных, но и происходит пересмотр уже опубликованных данных за предыдущие периоды. На данный момент США является одним из самых активных регионов развития наукастинга в качестве управленческого инструмента, с самой развитой базой обеспечения.

Наукастинг Федеральной резервной системы США в основном фокусируется на нескольких макроэкономических индикаторах, таких как ВВП, номинальное потребление населения, показатели инвестиций, промышленности [17]. Эти данные являются результатом сложных и систематических усилий по измерению всей экономической активности и сопутствующих процессов. Концептуально современная система сбора официальных данных основана на принципах бухгалтерского учета, а не на статистических или экономических моделях и решает скорее проблему общего контроля отдельных базовых частей, не предполагая расширения в плоскость самостоятельного аналитического инструментария. Причина в данном случае кроется в источниках самой информации: из-за своего бухгалтерского происхождения, формирование данных происходит на основе отчетности почти 11 млн юридических лиц и организаций, каждая из которых является генератором новых данных, которые должны быть оформлены и переданы для последующей обработки.

Но, несмотря на это, будет ошибочно не применять эти данные в целях наукастинга. У ФРС США есть значительный опыт в формировании предиктивных моделей наукастинга не только для своей экономики, но и для экономик других государств. В рабочих докладах аналитической группы ФРС

Daniela Bragoli и Michele Modugno предложили факторную модель прогнозирования темпов роста реального ВВП Канады [18]. Модель дает более точные прогнозы, чем те, которые публикуются институциональными участниками, такими как Банк Канады или Организация экономического сотрудничества и развития (далее – ОЭСР). Представленная модель показала, что включение данных США в модель наукастинга экономики Канады резко повышает ее точность. Сравнение прогноза ВВП Канады с использованием переменных США (модель наукастинга), прогноза текущего периода ВВП без переменных США (бенчмарк), фактического значения ВВП и профессиональных прогнозов ОЭСР и Банка Канады представлено на рис. 2 [13].

Спустя два года после исследования D. Bragoli и M. Modugno Центральный банк Канады совместно с Tony Chernis и Rodrigo Sekkel [19] опубликовал свой аналитический отчет «Прогноз экономической активности в Канаде в условиях неопределенности». Отчет посвящен краткосрочному прогнозированию реального ВВП при помощи комбинирования различных методик наукастинга. По сравнению с более ранними исследованиями авторы проводят анализ не только самого валового внутреннего продукта, но и его компонентов. Объединяя такие виды моделей, как авторегрессия, динамичные факторные модели и модели смешанной выборки, авторы обнаружили, что эффективность моделей и их комбинированная точность улучшается по мере применения новых статистических данных [20]. Вес, присваиваемый различным моделям, меняется в течение квартала в соответствии с потоком данных, показывая, что комбинации адаптивно придают больший вес моделям с новой и более актуальной информацией.

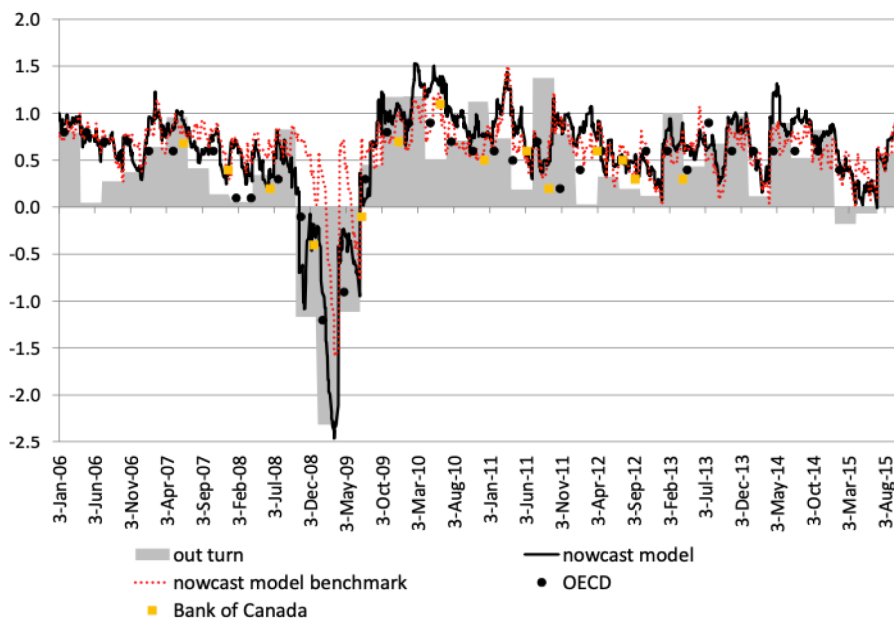


Рис. 2. Сравнение моделей экономического прогнозирования ВВП Канады

Fig. 2. Comparison of economic forecasting models for Canada's GDP

Также авторы обнаружили, что объединение прогнозов из разных классов моделей (более 100 моделей) приводит к более точному прогнозу. Вместе с этим комбинации всегда являются одним из самых эффективных методов прогнозов, но не всегда самым точным. Поскольку эффективность прогнозов отдельных моделей варьируется в зависимости от времени, горизонта прогноза и целевой переменной, комбинация прогнозов позволяет прогнозистам хеджировать свои «ставки». И хотя ни одна из моделей явно не доминирует, наиболее стабильно работают на практике именно факторные модели. Однако это не отменяет вывода о том, что разные модели работают лучше на разных временных горизонтах и в разные периоды времени.

4.2. Применение наукастинга для анализа макроэкономических показателей

Отметим, что большая часть исследований и литературы, в которых

анализируется применение модели MIDAS, носит скорее экспериментальный характер. Одной из самых известных работ является труд китайского коллектива ученых Xing Li, Wei Sheng, Sheng Wang и Jian Ma [21]. Работа посвящена применению ранее рассмотренной модели MIDAS для прогнозирования инфляции в Китае, основываясь на статистике поисковых запросов. Исследователи произвели выборку поисковых запросов в сети Интернет по базе данных слов ассоциированных с повышением цен. Полученные результаты не только превосходят все ранее применяемые методики с точки зрения своевременности данных, но и позволяют сократить среднюю квадратичную ошибку на 32,9% (рис. 3 [21]).

Похожее исследование проводилось тремя годами ранее коллективом из Пекинского университета менеджмента [22]. Фундаментом исследования стал анализ выборки ключевых слов имеющих прямую связь с индексом потребительских цен. При помощи

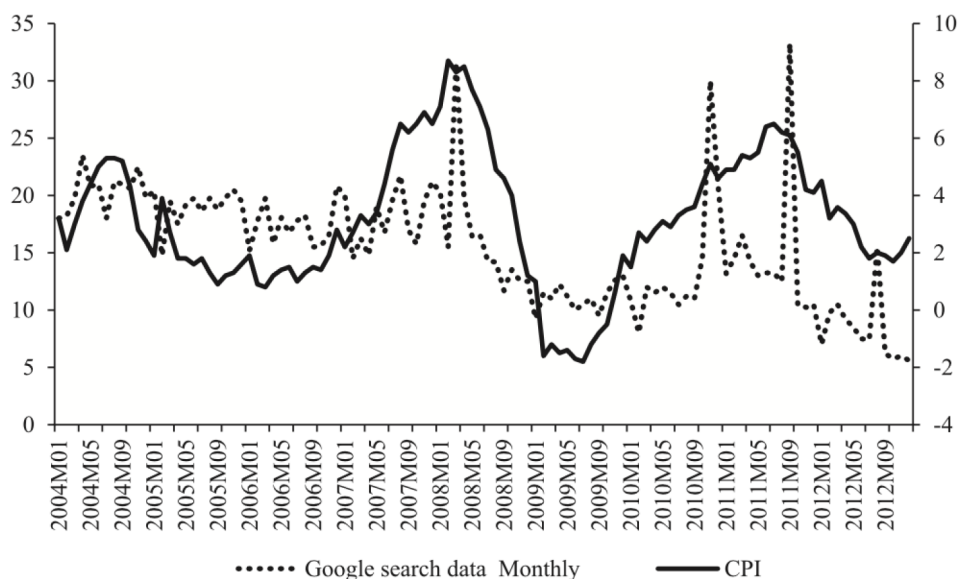


Рис. 3. Ежемесячные данные поисковых запросов Google и индекс потребительских цен

Fig. 3. Monthly Google search query data and Consumer Price Index

корреляции Пирсона исследователи отобрали 44 ключевых слова, индекс корреляции которых превышал 0.4. Их последующая разбивка на 2 временных ряда и коинтеграционный анализ позволил сделать следующие выводы:

1) Наличие интеграционного отношения между данными поисковых запросов и индекса потребительских цен (далее – ИПЦ).

2) Аккуратность прогноза представленной модели равняется 0.978. Ошибка прогноза равняется 0.48.

3) По сравнению с традиционными методами мониторинга ИПЦ, представленная модель своевременно отслеживает изменение ИПЦ и позволяет формировать точный прогноз примерно на месяц раньше, чем национальное бюро статистики.

Помимо рассмотренных работ исследовательских коллективов из Китая, проблематика наукастинга инфляции при помощи новых источников данных рассматривается в аналитических материалах Национального бюро

экономических исследований в США. В своих материалах исследователи Austan Goolsbee и Peter Klenow проводят анализ онлайн-транзакций миллионов пользователей Интернета с целью определить возможность прогнозирования темпов роста инфляции на основе данных интернет-торговли [24]. Результаты исследований показали, что реальный уровень инфляции был значительно ниже, чем предполагалось ранее.

Авторский коллектив из Университета Чикаго под руководством David Coble предложил новый способ прогнозирования выдачи разрешений на строительство в США, используя данные в режиме реального времени из поисковых запросов Google [25]. Временные ряды разрешений на строительство обычно рассматриваются как опережающий индикатор экономической активности в строительном секторе. Тем не менее новые данные о разрешениях на строительство публикуются с задержкой около двух месяцев. Поэтому желательно точное отображение этого

опережающего индикатора. Результаты показали, что модели, использующие поисковые запросы Google, показывают текущее состояние и прогноз выписки разрешений на строительство лучше, чем официальные статистические модели. А поскольку Google предоставляет информацию бесплатно, этот подход представляет собой простой и недорогой способ повышения качества информации. На временных отрезках в два-три месяца точность прогнозов увеличивается почти на 10 % относительно ранее применяемых моделей.

В рамках анализа возможного применения поисковых запросов с целью прогноза макроэкономических показателей особый интерес представляет работа Gary Коор и Luca Onorante [26]. В своем исследовании авторы разрабатывают новые методологические инструменты применения ранее рассмотренных моделей. Авторами были сформулированы 2 основных вывода:

1) Имплементация данных из Google trends позволяет значительно повысить точность моделей наукастинга при работе с дезагрегированными данными.

2) Отдельные переменные, взятые из Google trends, дают значительно более точные результаты, если они позиционируются не как переменные регрессии, а как вероятностные переменные.

Raulin Cadet расширил практику применения поисковых запросов в целях прогноза индекса потребительских цен Гаити [27]. Применяя интегрированные модели авторегрессии, он делает вывод о том, что применение этих моделей может заменить официальные данные в экономически не развитых регионах и странах, где точность официальных методов сбора статистики оставляет желать лучшего. Результаты показывают, что применение поисковых запросов позволило повысить точность

всех представленных моделей. Помимо этого, автор доказывает, что векторные модели авторегрессии значительно превосходят интегрированные модели авторегрессии не только с точки зрения точности полученных данных, но и своевременности.

Исследование, проведенное Daniel Borup и Christian Schutte, показывает, что поисковые запросы Google являются очень стабильным предиктором роста занятости населения в США в период 2004–2019 гг. (как на краткосрочном, так и на долгосрочном горизонте анализа) [28]. Начав с простых поисковых запросов по коррелирующим словам, например таким, как «вакансии», авторы создали панель из 172 переменных, используя собственные алгоритмы Google для поиска семантически связанных поисковых запросов. Полученная модель достигает показателя R^2 вне выборки между 29 и 62 % на горизонте от одного месяца до одного года, что значительно превосходит модели, основанные на одном наборе макроэкономических, финансовых и эмоциональных показателей.

Исследование [29] направлено на прогнозирование уровня безработицы во время пандемии COVID-19 с использованием доли запросов данных Google Trends для ключевого слова «phk» («прекращение работы») и предыдущих рядов из официального опроса трудоспособного населения, проведенного Статистическим управлением Индонезии. Используя векторную авторегрессию, исследования показали, что использование поисковых запросов Google Trends в качестве экзогенной переменной может значительно повысить точность прогнозирования индонезийских макроэкономических показателей.

Sourav Kundu и Rajshekhar Singhanian рассматривают проблематику прогнозирования безработицы смежным образом [30]. Для получения более точного

прогноза авторы предложили использовать новую архитектуру с долгосрочной-краткосрочной памятью (далее – LSTM) рекуррентных нейронных сетей для прогнозирования заявок на пособие по безработицы при помощи запросов Google Trends. Коррелируя набор ключевых слов, они проанализировали 15-летние еженедельные данные с января 2004 г. по январь 2019 г. и создали две модели для анализа: модель векторной авторегрессии (далее – VAR), объединяющую серию официальных заявлений по безработице с тенденциями поиска по ключевым словам (например «предложения о работе»), и модель LSTM, включающую только данные временных рядов тенденций для полного набора идентифицированных ключевых слов. Последующий анализ показал, что модель LSTM значительно превосходит модель VAR в прогнозировании заявок на пособие по безработице на разных временных горизонтах прогноза.

Наукастинг также может быть применен для анализа и прогнозирования цен на нефть. Ting Yao и Yue-Jun Zhang исследуют зависимость поисковых запросов Google и цен на сырую нефть путем включения индекса Google в качестве экзогенной переменной в интегрированной модели авторегрессии скользящего среднего и авторегрессионной условной гетероскедастичной модели [31]. Эмпирические результаты показали, что существует отрицательная зависимость поисковых запросов Google на цены на сырую нефть. Индикативный метод прогнозирования применили также в своей работе Akos Barta и Mark Molnar [32]. Теоретизируя зависимость между человеческим поведением, медиавоздействиями движением цен на нефть, авторы рассмотрели зависимость между поисковыми запросами Google, связанных с ОПЕК и изменением цен на сырую нефть. Результаты

показали, что в 58,7% случаев рост поисковых запросов сопровождается ростом цен на нефть.

4.3. Применение наукастинга для анализа медицинской деятельности

Одним из самых интересных прикладных сфер применения статистики поисковых запросов является медицина. Будучи «первоисточником» данных о человеческом поведении, поисковые запросы могут выступать в качестве доказательной базы зависимости появления тех или иных заболеваний и частотой интернет-запросов по связанным тематикам.

Интересным исследованием в данной сфере является работа «Динамическое прогнозирование эпидемий Зика при помощи Google Trends» [33]. В данной работе международный коллектив авторов разработал динамичную прогнозную модель вируса Зика. Результаты показали, что данные, полученные при помощи анализа поисковых запросов и интегрированных моделей авторегрессии, очень похожи на фактические данные эпидемии Зика. Целочисленная авторегрессия позволила сформировать базовую прогностическую модель для случаев Зика. А за счет включения данных Google Trends, подтверждающих прогностическую полезность наблюдения на основе поисковых запросов, модель может применяться независимо от региональной специфики. Эта доступная и гибкая динамическая модель прогноза может использоваться при мониторинге вирусных заболеваний для обеспечения заблаговременного предупреждения о будущих вспышках.

Похожее исследование провели Amaryllis Mavgarani и Konstantinos Gkillas [34]. В качестве первичного анализа авторы изучают корреляции взаимосвязи между данными Google

Trends и данными COVID-19 как в случаях заболевания, так и смертности. Последующий анализ корреляции строится на основе квантильной регрессии, которая корректируется смещением посредством Бутстрэпинга. Результаты исследования показывают, статистически значимые корреляции между Google Trends и данными COVID-19.

Интересный подход к анализу распространения COVID-19 применяет международный коллектив авторов, сформированный в рамках их независимого исследования [35]. В данной работе авторская методика была разработана для оценки данных Google Trends и традиционных показателей COVID-19 с помощью интерактивной платформы, основанной на анализе основных компонентов и моделировании временных рядов. Данный метод облегчает поиск симптомов, связанных с COVID-19, в 188 странах. Для прогнозирования показателей COVID-19 использовались три различных модели временных рядов (сезонность, интегрированное скользящее среднее авторегрессии и авторегрессия нейронной сети с прямой связью). Модели сравнивались с точки зрения предсказательной способности с использованием среднеквадратичной ошибки первого главного компонента. Прогностическая способность моделей, апробированных как на основании данных Google Trends, так и на основе официальных данных о COVID-19, сравнивалась с моделями, оснащенными только традиционными метриками COVID-19. Результаты показали, что степень корреляции варьируется в зависимости от страны, в целом задержка наблюдалась в пределах 15 дней. Проведенный в данном исследовании анализ показал, что прогнозы, основанные как на данных поисковых запросов, так и на традиционных показателях COVID-19, работали лучше,

чем те, которые не включают поиск в Google, а качество прогнозов варьировалось в зависимости от выбранной страны и временных рамок.

Чуть более узконаправленное исследование по распространению коронавируса было проведено в работе «Улучшенное прогнозирование передачи COVID-19 и смертности с использованием тенденций поиска симптомов в Google в США» [36]. Цель исследования – разработать модели прогнозирования передачи COVID-19 и смертности в США на 7 и 14 дней, используя поисковые запросы в Google на основе симптомов, связанных с COVID-19. В частности, авторы предложили архитектуру многослойной долгосрочной-краткосрочной памяти (далее – SLSTM) для прогнозирования подтвержденных случаев COVID-19 и случаев смерти с использованием данных исторических временных рядов в сочетании с данными вспомогательных временных рядов из набора данных симптомов Google COVID-19.

Принимая во внимание сети SLSTM, обученные с использованием только исторических данных в качестве базовых моделей, базовые модели для прогнозирования случаев COVID на 7 и 14 дней имели значения средней абсолютной процентной ошибки (далее – MAPE) 6,6 и 8,8 % соответственно. С другой стороны, предлагаемые модели улучшили значения MAPE на 3,2 и 5,6 % соответственно. Для прогнозирования смертей от COVID-19 на 7 и 14 дней вперед значения MAPE базовых моделей составили 4,8 и 11,4 %, в то время как улучшенные значения MAPE предложенных моделей составили 4,7 и 7,8 % соответственно. Авторы также обнаружили, что поисковые запросы в Google по словам «пневмония», «одышка» и «лихорадка» являются наиболее точными для прогнозирования передачи COVID-19. Авторы также

обнаружили, что поисковые запросы «гипоксия» и «лихорадка» были наиболее информативными для прогнозирования смертности от COVID-19.

4.4. Российский опыт применения наукастинга для прогнозирования социально-экономических показателей

Российский опыт применения поисковых интернет-запросов с целью наукастинга экономических показателей является довольно ограниченным. Анализ релевантной литературы показал, что настоящий интерес в этих исследованиях начал появляться только в 2018–2020 гг. На сегодняшний день работа Д. Петровой и П. Трунина [23] является одним из ключевых исследований, посвященного данной тематике. В данной работе проводится анализ возможности применения поисковых запросов для прогнозирования инфляции, уровня безработицы, реальных темпов роста ВВП и курса рубля к доллару за период с января 2004 г. по июль 2019 г. В работе используются поисковые запросы, связанные с финансовыми рынками, инфляционными ожиданиями и макроэкономическими условиями. Результаты исследования показывают, что включение в модель интенсивностей поисковых запросов позволяет повысить точность прогнозов инфляции, уровня безработицы и курса рубля к доллару по сравнению с наивным прогнозом.

Отечественный опыт применения наукастинга в целях прогнозирования экономических индикаторов и общих экономических изменений также представлен в серии докладов Центрального банка Российской Федерации².

² Серия докладов об экономических исследованиях Центрального банка Российской Федерации. URL: https://www.cbr.ru/Content/Document/File/16739/wps_2.pdf (дата обращения: 04.04.2021).

Последнее десятилетие характеризовалось периодами как благоприятных, так и неблагоприятных условий для экономического роста России, что было вызвано одновременно и внутренними, и внешними факторами. За ростом цен на нефть, начавшимся в 2002 г., последовал резкий спад во второй половине 2008 г. и большей части 2009 г., который наступил после глобального экономического и финансового кризиса. Дальнейшие тенденции экономического роста не были стойкими и зачастую сопровождались колебаниями ВВП.

В 2020 г. российская экономика пережила очередной спад, природа и перспективы которого могут быть одновременно охарактеризованы ухудшением внешнеэкономических условий, а также повышенной неопределенностью с учетом сложившейся эпидемиологической ситуации (последствия глобальной пандемии, политической напряженности и риск введения новых санкций). Существенные колебания в динамике макроэкономических переменных и структурные изменения в экономике в целом вносят дополнительные сложности в процесс прогнозирования экономической деятельности.

В серии докладов рабочая группа ЦБ РФ использовала для анализа динамическую факторную модель. Главным преимуществом данной модели является то, что она может быть использована в рамках краткосрочной оценки за текущей квартал, а также прогнозирования квартальной динамики ВВП с помощью большого количества доступных наблюдений временных данных более высокой частотности, которые при этом зачастую являются несбалансированными по количеству доступных наблюдений.

Полученная модель решает задачу по оценке российского ВВП как в текущем квартале, так и при его прогнозировании на временных горизонтах в один

и два квартала по сравнению с известными альтернативными эконометрическими спецификациями. Кроме того, ошибки прогноза динамической факторной модели, вычисленные в псевдореальном времени, оказались ниже по сравнению с аналогичными ошибками по другим применяемым в практике прогнозирования ВВП эконометрическим моделям.

На основании проведенного анализа нами была составлена сводная таблица успешных практик применения наукастинга и прогнозирования традиционных социально-экономических показателей при помощи статистики запросов (табл. 2).

5. Заключение

Проведенный нами анализ показал, что в настоящий момент применение статистики поисковых запросов в качестве официального инструмента статистического прогнозирования находится на относительно ранней ступени развития в связи с рядом ограничений для его существования в качестве самостоятельного инструмента (к примеру, отсутствие нормативных регламентов и стандартов осуществления статистического мониторинга, невысокая степень верифицируемости ряда данных). При этом применение данного метода в целях анализа и прогнозирования экономических индикаторов и общих

Таблица 2. Успешные практики применения наукастинга

Table 2. Successful practices in the application of nowcasting

№ п/п	Область исследования / Area of research	Кейс / Case study	Источник / Source
1	Макроэкономический анализ	Прогнозирование ВВП	Brandy Bok, et al. (2018)
2	Финансовая сфера	Прогнозирование инфляции	Bragoli D., Michele M. (2016)
3	Макроэкономический анализ	Прогнозирование ВВП	Tony Chernis, Rodrigo Sekkel (2018)
4	Финансовая сфера и макроэкономический анализ	Прогнозирование ВВП, инфляции	Stock, J. H., and M. W. Watson (1989)
5	Макроэкономический анализ	Прогнозирование ВВП	Banbura M., Modugno M. (2013)
6	Финансовая сфера	Прогнозирование инфляции	Li, X., Shang, W., Wang, S., Ma, J. (2015)
7	Финансовая сфера	Прогнозирование ИПЦ	Chong Zhang, Benfu Lv., Geng Peng, Ying Liu, Qingyu Yuan (2018)
8	Финансовая сфера	Прогнозирование инфляции	Austan D. Goolsbee, Peter J. Klenow (2018)
9	Макроэкономический анализ	Прогнозирование ВВП	Paolo Fornaro, Henri Luomaranta (2016)
10	Макроэкономический анализ	Прогнозирование ВВП	Laurent Ferrara, Anna Simoni (2020)

Окончание табл. 2

End of table 2

№ п/п	Область исследования / Area of research	Кейс / Case study	Источник / Source
11	Туристическая отрасль	Прогнозирование туристических потоков	Antolini, F., Grassini, L. (2019)
12	Строительство	Прогнозирование разрешений на строительство	Coble D, Pincheira P. (2017)
13	Занятость и безработица	Прогнозирование безработицы	Pavlicek J, Kristoufek L. (2015)
14	Макроэкономический анализ	Прогнозирование макроэкономических показателей	Gary Koop, Luca Onorante (2013)
15	Макроэкономический анализ	Прогнозирование макроэкономических показателей	Steven Scot, Hal Varian (2013)
16	Макроэкономический анализ	Прогнозирование макроэкономических показателей	Marcelo Medeiros, Henrique Pires (2021)
17	Медицина	Прогнозирование вирусных заболеваний	Raulin Cadet (2018)
18	Медицина	Прогнозирование вирусных заболеваний	Teng Y, Bi D, et al. (2017)
19	Медицина	Прогнозирование вирусных заболеваний	Mavragani A., Gkillas K. (2020)
20	Медицина	Прогнозирование вирусных заболеваний	Alessandro Rabiolo, Eugenio Alladio, et al. (2021)
21	Медицина	Прогнозирование вирусных заболеваний	Meshrif Alruily, Mohamed Ezz (2021)
22	Занятость и безработица	Прогнозирование занятости	Daniel Borup, Erik Christian (2020)
23	Занятость и безработица	Прогнозирование безработицы	Singhania, Rajshekhar, Kundu, Sourav (2020)
24	Занятость и безработица	Прогнозирование безработицы	Fajar M. Prasetyo O. R., S. Nonalisa (2020)
25	Товарные рынки	Прогнозирование цен на нефть	Ting Yao, Yue-Jun Zhang (2017)
26	Товарные рынки	Прогнозирование цен на нефть	Akos Barta, Mark Molnar (2021)
27	Макроэкономический анализ	Прогнозирование макроэкономических показателей	Петрова Д. А., Трунин П. В. (2020)

экономических изменений в перспективе бесспорно имеет потенциал и ряд преимуществ по сравнению с традиционными показателями (огромная выборка, высокочастотные наблюдения, оперативная статистика).

Важно подчеркнуть, что используемые в настоящее время в Российской Федерации индикаторы в области потребительского спроса, состояния делового климата, инфляции, безработицы, биржевых индексов и ряда других областей не должны являться «зацементированными», раз и навсегда установленными. Они должны меняться по мере трансформации внутренних и внешних условий среды и обуславливаться сдвигами в социально-экономическом развитии страны и трансформации национальных приоритетов.

Использование больших данных, полученных в ходе анализа статистики поисковых запросов, может и должно стать одним из направлений использования дополнительных источников данных для расчета статистических показателей. Большие данные, полученные от операторов мобильной связи, банков, агрегаторов такси и услуг гостиничного бизнеса заключают в себе очень большой потенциал и обладают высокой степенью информативности. Но, как отмечалось ранее, использовать их на практике сложнее, поскольку они в настоящее время не регулируются и не регламентируются со стороны государства: сегодня данные доступны, завтра они хранятся уже в другом виде, послезавтра их вообще нет.

При этом ввиду огромного потенциала больших данных не использовать их просто неразумно: они могли бы способствовать

совершенствованию всей системы статистических наблюдений в России, снижению количества допускаемых статистических ошибок. Так, примером использования статистики запросов для анализа макроэкономических индикаторов может стать использование данных операторов мобильной связи при проведении переписи населения или данных онлайн-касс магазинов для анализа информации о текущих ценах (сейчас их собирают регистраторы ведомства, обходя магазины). Это, мнению авторов, является тем самым недостающим звеном, которое позволит получать первичные корректные данные сразу в электронном виде в систему Росстата. При этом, безусловно, немногие из существующих поисково-аналитическим систем могут выступить источником надежных данных для статистики. Ключевая задача заключается в том, чтобы обеспечить сопоставимость новых показателей с предыдущими данными, в противном случае они будут бесполезны.

В результате проведенного анализа была подтверждена гипотеза, что применение наукастинга позволит повысить точность статистического прогнозирования показателей социально-экономического развития и может в ближайшей перспективе стать достойной альтернативой традиционным методам макроэкономического анализа. Дальнейшие исследования в данной области должны быть направлены на систематизацию существующих направлений применения наукастинга на практике с тем, чтобы разработать комплексную методологию его использования для построения высокоточных макроэкономических прогнозов.

Список использованных источников

1. *Fan J., Han F., Liu H.* Challenges of big data analysis // *National Science Review*. 2014. Vol. 1, Issue 2. Pp. 293–314. DOI: <https://doi.org/10.1093/nsr/nwt032>.
2. *Seymour T., Frantsvog D., Kumar S.* History of search engines // *International Journal of Management & Information Systems*. 2011. Vol. 15, No. 4. Pp. 47–58. DOI: 10.19030/ijmis.v15i4.5799.
3. *Mavragani A., Tsagarakis K. P.* YES or NO: predicting the Referendum results using Google Trends // *Technological Forecasting and Social Change*. 2016. Vol. 109. Pp. 1–5. DOI: 10.1016/j.techfore.2016.04.028.
4. *Hyenjo K.* The Analysis of innovation Management in Google // *Journal of Economics and Management of Sciences*. 2020. Vol. 3, Issue 4. Pp. 10–19. DOI: 10.30560/jems.v3n4p10.
5. *Khorshidi S., Carter J., Mohler G., Tita G.* Explaining Crime Diversity with Google Street View // *Journal of Quantitative Criminology*. 2021. DOI: 10.1007/s10940-021-09500-1.
6. *Scott S., Varian H.* Predicting the Present with Bayesian Structural Time Series // *SSRN Electronic Journal*. 2013. DOI: 10.2139/ssrn.2304426.
7. *Medeiros M., Pires H.* The Properly Use of Google Trends in Forecasting Models. 2021. [Электронный ресурс]. Режим доступа: https://www.researchgate.net/publication/350720260_The_Properly_Use_of_Google_Trends_in_Forecasting_Models.
8. *Julie J., François H.* A package for handling missing values in multivariate data analysis // *Journal of Statistical Software*. 2016. Vol. 70, Issue 1. Pp. 1–31. DOI: 10.18637/jss.v070.i01.
9. *Paolo F.* Predicting Finnish economic activity using firm-level data // *International Journal of Forecasting*. 2016. Vol. 32, Issue 1. Pp. 10–19. DOI: 10.1016/j.ijforecast.2015.04.002.
10. *Knut A. A., Tørres T.* Estimating the output gap in real time: A factor model approach // *The Quarterly Review of Economics and Finance*. 2014. Vol. 54, Issue 2. Pp. 180–193. DOI: 10.1016/j.qref.2013.09.003.
11. *Melo L. F., Loaiza R. A., Villamizar-Villegas M.* Bayesian combination for inflation forecasts: the effects of a prior based on central banks estimates // *Economic Systems*. 2016. Vol. 40, Issue 3. Pp. 387–397. DOI: 10.1016/j.ecosys.2015.11.002.
12. *Chernis T., Sekkel R.* A Dynamic Factor Model for Nowcasting Canadian GDP Growth // *Bank of Canada Staff Working Paper*. 2017. No. 2. Bank of Canada, 2017. 26 p. [Электронный ресурс]. Режим доступа: <https://www.bankofcanada.ca/wp-content/uploads/2017/02/swp2017-2.pdf>.
13. *Colin J. H., Redding S. J., David E. W.* Quantifying the sources of firm heterogeneity // *The Quarterly Journal of Economics*. 2016. Vol. 131, Issue 3. Pp. 1291–1364. DOI: 10.1093/qje/qjw012.
14. *Groshen E. L., Brian C. M., Ana M. A., Ralph B., David M. F., Friedman D. M.* How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches // *Journal of Economic Perspectives*. 2017. Vol. 31, Issue 2. Pp. 187–210. DOI: 10.1257/jep.31.2.187.
15. *Carlo T. C., Marcal E. F.* Forecasting brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon // *Applied Economics*. 2016. Vol. 48, Issue 50. Pp. 4846–4860. DOI: 10.1080/00036846.2016.1167824.
16. *Rüegg J., Gries C., Bond-Lamberty B., Bowen G. J. et al.* Completing the data life cycle: using information management in macrosystems ecology research // *Frontiers in Ecology and the Environment*. 2014. Vol. 12, Issue 1. Pp. 24–30. DOI: 10.1890/120375.
17. *Simionescu M., Gavurova B., Smrcka L.* Combined forecasts to improve survey of profession forecasters predictions for quarterly inflation in the USA // *Economic Research-Ekonomska Istraživanja*. 2017. Vol. 30, Issue 1. Pp. 789–805. DOI: 10.1080/1331677X.2017.1314826.
18. *Bragoli D., Michele M.* A Nowcasting Model for Canada: Do U. S. Variables Matter? // *Finance and Economics Discussion Series*. 2016–036. Washington: Board of Governors of the Federal Reserve System, 2016. 33 p. DOI: 10.17016/FEDS.2016.036.

19. *Chernis T., Sekkel R.* Nowcasting Canadian Economic Activity in an Uncertain Environment // Bank of Canada Staff Discussion Paper. No. 2018–9. Bank of Canada, 2018. 39 p. [Электронный ресурс]. Режим доступа: <https://www.econstor.eu/bitstream/10419/200462/1/1029915407.pdf>.
20. *Binette A., Chang J.* CSI: A Model for Tracking Short-Term Growth in Canadian Real GDP // Bank of Canada Review. 2013. Summer. Pp. 3–12 [Электронный ресурс]. Режим доступа: <https://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-binette.pdf>.
21. *Li X., Shang W., Wang S., Ma J.* A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data // Electronic Commerce Research and Applications. 2015. Vol. 14, Issue 2. Pp. 112–125. DOI: 10.1016/j.elerap.2015.01.001.
22. *Chong Z., Benfu L., Geng P., Ying L., Qingyu Y.* A study on Correlation between Web Search Data and CPI // Recent Progress in Data Engineering and Internet Technology. Vol. 2 / Edited by F. L. Gaol. Springer-Verlag, 2012. Pp. 269–274. DOI: 10.1007/978-3-642-28798-5_36.
23. *Petrova D., Trunin P.* Population Forecasting and Analysis of Demographic Heterogeneity of Russia // SSRN Electronic Journal. 2020. DOI: 10.2139/ssrn.3594521.
24. *Goolsbee A. D., Klenow P. J.* Internet rising, prices falling measuring inflation in a world of E-commerce // NBER Working Paper. No. 24649. National bureau of economic research, 2018. 30 p. DOI: 10.3386/w24649.
25. *Coble D., Pincheira P.* Nowcasting Building Permits with Google Trends // SSRN Electronic Journal. 2017. DOI: 10.2139/ssrn.2910165.
26. *Koop G., Onorante L.* Macroeconomic nowcasting using Google probabilities // Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling. Part A. Vol. 40A / Edited by I. Jeliazkov, J. L. Tobias. Emerald Publishing, 2019. Pp. 17–40 [Электронный ресурс]. Режим доступа: <https://econpapers.repec.org/bookchap/emeaeco2/s0731-90532019000040a003.htm>.
27. *Raulin C.* Forecasting Consumer Price Index in Haiti using Google Trends. 2018 [Электронный ресурс]. Режим доступа: https://www.researchgate.net/publication/323665693_Forecasting_Consumer_Price_Index_in_Haiti_using_Google_Trends.
28. *Borup D., Schütte E.C.M.* In Search of a Job: Forecasting Employment Growth Using Google Trends // Journal of Business & Economic Statistics. 2020. DOI: 10.1080/07350015.2020.1791133.
29. *Fajar M., Prasetyo O. R., Nonalisa S., Wahyudi* Forecasting Unemployment Rate in the Time of COVID-19 Pandemic Using Google Trends Data (Case of Indonesia) // International Journal of Scientific Research in Multidisciplinary Studies. 2020. Vol. 6, Issue 11. Pp. 29–33.
30. *Rajshekhar S., Kundu S.* Forecasting the United States Unemployment Rate by Using Recurrent Neural Networks with Google Trends Data // International Journal of Trade, Economics and Finance. 2020. Vol. 11, No. 6. DOI: 10.18178/ijtef.2020.11.6.679.
31. *Yao T., Zhang Yue-J.* Forecasting Crude Oil Prices with the Google Index // Energy Procedia. 2017. Vol. 105. Pp. 3772–3776. DOI: 10.1016/j.egypro.2017.03.880.
32. *Barta A., Molnar M.* Forecasting oil prices based on online occurrence volume // Journal of Modern Science. 2021. No. 1 [Электронный ресурс]. Режим доступа: https://www.researchgate.net/publication/350358726_Forecasting_oil_price_based_on_online_occurrence_volume.
33. *Teng Y., Bi D., Xie G., Jin Y., Huang Y., Lin B., et al.* Dynamic Forecasting of Zika Epidemics Using Google Trends // PLoS ONE. 2017. Vol. 12, No. 1. DOI: 10.1371/journal.pone.0165085.
34. *Mavragani A., Gkillas K.* COVID-19 predictability in the United States using Google Trends time series // Scientific Reports. 2020. Vol. 10. Article number: 20693. DOI: 10.1038/s41598-020-77275-9.
35. *Rabiolo A., Alladio E., Morales E., McNaught A., Bandello F., Afifi A., Marchese A.* Forecasting the COVID-19 epidemic integrating symptom search behavior: an infodemiology study // MedRxiv. 2021. DOI: 10.1101/2021.03.09.21253186.

36. *Mostafa M., Yanes N., Abbas M., El-Manzalawy Y.* Improved Prediction of COVID-19 Transmission and Mortality Using Google Search Trends for Symptoms in the United States. MedRxiv. 2021. DOI: 10.1101/2021.03.14.21253554.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Фирсов Дмитрий Владимирович

Младший научный сотрудник Института экономической политики и проблем экономической безопасности Финансового университета при Правительстве Российской Федерации, г. Москва, Россия (125993 (ГСП-3), г. Москва, Ленинградский просп., 49); ORCID 0000-0001-5985-7285; e-mail: dvfirsov@fa.ru.

Чернышева Татьяна Константиновна

Стажер-исследователь Института экономической политики и проблем экономической безопасности Финансового университета при Правительстве Российской Федерации, г. Москва, Россия (125993 (ГСП-3), г. Москва, Ленинградский просп., 49); ORCID 0000-0002-4744-7198; e-mail: tkchernysheva@fa.ru.

БЛАГОДАРНОСТИ

Статья подготовлена в рамках государственного задания Правительства РФ Финансовому университету на 2021 г. по теме «Направления модернизации российской университетской системы с учетом запросов реального сектора экономики и мировых технологических трендов».

ДЛЯ ЦИТИРОВАНИЯ

Фирсов Д. В., Чернышева Т. К. Обзор успешных практик применения наукастинга в социально-экономическом прогнозировании // Journal of Applied Economic Research. 2021. Т. 20, № 2. С. 269–293. DOI: 10.15826/vestnik.2021.20.2.012.

ИНФОРМАЦИЯ О СТАТЬЕ

Дата поступления 14 марта 2021 г.; дата поступления после рецензирования 24 апреля 2021 г.; дата принятия к печати 16 мая 2021 г.

Review of Successful Practices of Applying Nowcasting in Socio-Economic Forecasting

D. V. Firsov  , T. C. Chernysheva 

*The Financial University under the Government of the Russian Federation
Moscow, Russia*

 dvfirsov@fa.ru

Abstract. A necessary competence in the present-day reality is the ability to analyze big data in the economy, and therefore one of the key issues is the choice of tools for such analysis. One of the most promising tools is nowcasting, which allows you to accurately determine economic changes in very short time periods. The aim of the study is to analyze successful modern practices of using nowcasting for statistical forecasting of socio-economic indicators. The hypothesis of the research lies in the assumption that nowcasting as a method of macroeconomic analysis can in the near future become a worthy alternative to traditional methods of analysis and statistical forecasting of indicators of socio-economic development, increasing the accuracy of their forecasting. The methodological basis of the study was the scientific works and applied developments of leading domestic and foreign scientists in the field of economic forecasting using statistics of search queries, as well as methods of comparative and statistical analysis, and the systematic approach. The novelty of the results obtained lies in the systematization and description of successful practices of using nowcasting and forecasting indicators using query statistics. The study highlights the basic principle of nowcasting, which is to obtain a more accurate assessment of the state of the economy as new data becomes available. It also describes the key statistical models used as tools for testing in foreign countries. As a result of the study, we highlight the importance of the analysis of statistical search queries, especially in the context of their correlation with classical survey metrics and general statistics. It is in an active phase of development, especially within the framework of the domestic forecasting practice. The results obtained can be applied both in a corporate environment and in the public sector to build macroeconomic forecasts.

Key words: search queries; economic research; query statistics; big data analysis models; nowcasting/

JEL C55, O21

References

1. Fan, J., Han, F., Liu, H. (2014). Challenges of big data analysis. *National Science Review*, Vol. 1, Issue 2, 293–314. DOI: <https://doi.org/10.1093/nsr/nwt032>.
2. Seymour, T., Frantsvog, D., Kumar, S. (2011). History of search engines. *International Journal of Management & Information Systems*, Vol. 15, No. 4, 47–58. DOI: 10.19030/ijmis.v15i4.5799.
3. Mavragani, A., Tsarakakis, K. P. (2016). YES or NO: predicting the Referendum results using Google Trends. *Technological Forecasting and Social Change*, Vol. 109, 1–5. DOI: 10.1016/j.techfore.2016.04.028.
4. Hyenjoo, K. (2020). The Analysis of innovation Management in Google. *Journal of Economics and Management of Sciences*, Vol. 3, Issue 4, 10–19. DOI: 10.30560/jems.v3n4p10.
5. Khorshidi, S., Carter, J., Mohler, G., Tita, G. (2021). Explaining Crime Diversity with Google Street View. *Journal of Quantitative Criminology*. DOI: 10.1007/s10940-021-09500-1.

6. Scott, S., Varian, H. (2013). Predicting the Present with Bayesian Structural Time Series. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2304426.
7. Medeiros, M., Pires, H. (2021). *The Properly Use of Google Trends in Forecasting Models*. Available at: https://www.researchgate.net/publication/350720260_The_Properly_Use_of_Google_Trends_in_Forecasting_Models.
8. Julie, J., François, H. (2016). A package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, Vol. 70, Issue 1, 1–31. DOI: 10.18637/jss.v070.i01.
9. Paolo, F. (2016). Predicting Finnish economic activity using firm-level data. *International Journal of Forecasting*, Vol. 32, Issue 1, 10–19. DOI: 10.1016/j.ijforecast.2015.04.002.
10. Knut, A.A., Tørres, T. (2014). Estimating the output gap in real time: A factor model approach. *The Quarterly Review of Economics and Finance*, Vol. 54, Issue 2, 180–193. DOI: 10.1016/j.qref.2013.09.003.
11. Melo, L. F., Loaiza, R. A., Villamizar-Villegas, M. (2016). Bayesian combination for inflation forecasts: the effects of a prior based on central banks estimates. *Economic Systems*, Vol. 40, Issue 3, 387–397. DOI: 10.1016/j.ecosys.2015.11.002.
12. Chernis, T., Sekkel, R. (2017). A Dynamic Factor Model for Nowcasting Canadian GDP Growth. *Bank of Canada Staff Working Paper*, No. 2. Bank of Canada, 26 p. Available at: <https://www.bankofcanada.ca/wp-content/uploads/2017/02/swp2017-2.pdf>.
13. Colin, J. H., Redding, S. J., David, E. W. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, Vol. 131, Issue 3, 1291–1364. DOI: 10.1093/qje/qjw012.
14. Groshen, E. L., Brian, C. M., Ana, M. A., Ralph, B., David, M. F., Friedman, D. M. (2017). How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches. *Journal of Economic Perspectives*, Vol. 31, Issue 2, 187–210. DOI: 10.1257/jep.31.2.187.
15. Carlo, T. C., Marcal, E. F. (2016). Forecasting brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon. *Applied Economics*, Vol. 48, Issue 50, 4846–4860. DOI: 10.1080/00036846.2016.1167824.
16. Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G.J. et. al. (2014). Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, Vol. 12, Issue 1, 24–30. DOI: 10.1890/120375.
17. Simionescu, M., Gavurova, B., Smrcka, L. (2017). Combined forecasts to improve survey of profession forecasters predictions for quarterly inflation in the USA. *Economic Research-Ekonomska Istraživanja*, Vol. 30, Issue 1, 789–805. DOI: 10.1080/1331677X.2017.1314826.
18. Bragoli, D., Michele, M. (2016). A Nowcasting Model for Canada: Do U. S. Variables Matter? *Finance and Economics Discussion Series*, 2016–036. Washington: Board of Governors of the Federal Reserve System, 33 p. DOI: 10.17016/FEDS.2016.036.
19. Chernis, T., Sekkel, R. (2018). Nowcasting Canadian Economic Activity in an Uncertain Environment. *Bank of Canada Staff Discussion Paper*, No. 2018–9. Bank of Canada, 39 p. Available at: <https://www.econstor.eu/bitstream/10419/200462/1/1029915407.pdf>.
20. Binette, A., Chang, J. (2013). CSI: A Model for Tracking Short-Term Growth in Canadian Real GDP. *Bank of Canada Review*, Summer, 3–12. Available at: <https://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-binette.pdf>.
21. Li, X., Shang, W., Wang, S., Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data. *Electronic Commerce Research and Applications*, Vol. 14, Issue 2, 112–125. DOI: 10.1016/j.elerap.2015.01.001.
22. Chong, Z., Benfu, L., Geng, P., Ying, L., Qingyu, Y. (2012). A study on Correlation between Web Search Data and CPI. *Recent Progress in Data Engineering and Internet Technology*. Vol. 2. Edited by F. L. Gaol. Springer-Verlag, 269–274. DOI: 10.1007/978-3-642-28798-5_36.
23. Petrova, D., Trunin, P. (2020). Population Forecasting and Analysis of Demographic Heterogeneity of Russia. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3594521.

24. Goolsbee, A. D., Klenow, P. J. (2018). Internet rising, prices falling measuring inflation in a world of E-commerce. *NBER Working Paper*, No. 24649. National bureau of economic research, 30 p. DOI: 10.3386/w24649.
25. Coble, D., Pincheira, P. (2017). Nowcasting Building Permits with Google Trends. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2910165.
26. Koop, G., Onorante, L. (2019). Macroeconomic nowcasting using Google probabilities. *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling*. Part A. Vol. 40A. Edited by I. Jeliazkov, J. L. Tobias. Emerald Publishing, 17–40. Available at: <https://econpapers.repec.org/bookchap/emeacozz/s0731-90532019000040a003.htm>.
27. Raulin, C. (2018). *Forecasting Consumer Price Index in Haiti using Google Trends*. Available at: https://www.researchgate.net/publication/323665693_Forecasting_Consumer_Price_Index_in_Haiti_using_Google_Trends.
28. Borup, D., Schütte, E. C. M. (2020). In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business & Economic Statistics*. DOI: 10.1080/07350015.2020.1791133.
29. Fajar, M., Prasetyo, O. R., Nonalisa, S., Wahyudi (2020). Forecasting Unemployment Rate in the Time of COVID-19 Pandemic Using Google Trends Data (Case of Indonesia). *International Journal of Scientific Research in Multidisciplinary Studies*, Vol. 6, Issue 11, 29–33.
30. Rajshekhar, S., Kundu, S. (2020). Forecasting the United States Unemployment Rate by Using Recurrent Neural Networks with Google Trends Data. *International Journal of Trade, Economics and Finance*, Vol. 11, No. 6. DOI: 10.18178/ijtef.2020.11.6.679.
31. Yao, T., Zhang, Yue-J. (2017). Forecasting Crude Oil Prices with the Google Index. *Energy Procedia*, Vol. 105, 3772–3776. DOI: 10.1016/j.egypro.2017.03.880.
32. Barta, A., Molnar, M. (2021). Forecasting oil prices based on online occurrence volume. *Journal of Modern Science*, No. 1. Available at: https://www.researchgate.net/publication/350358726_Forecasting_oil_price_based_on_online_occurrence_volume.
33. Teng, Y., Bi, D., Xie, G., Jin, Y., Huang, Y., Lin, B., et al. (2017). Dynamic Forecasting of Zika Epidemics Using Google Trends. *PLoS ONE*, Vol. 12, No. 1. DOI: 10.1371/journal.pone.0165085.
34. Mavragani, A., Gkillas, K. (2020). COVID-19 predictability in the United States using Google Trends time series. *Scientific Reports*, Vol. 10, Article number: 20693. DOI: 10.1038/s41598-020-77275-9.
35. Rabiolo, A., Alladio, E., Morales, E., McNaught, A., Bandello, F., Afifi, A., Marchese, A. (2021). Forecasting the COVID-19 epidemic integrating symptom search behavior: an infodemiology study. *MedRxiv*. DOI: 10.1101/2021.03.09.21253186.
36. Mostafa, M., Yanes, N., Abbas, M., El-Manzalawy, Y. (2021). Improved Prediction of COVID-19 Transmission and Mortality Using Google Search Trends for Symptoms in the United States. *MedRxiv*. DOI: 10.1101/2021.03.14.21253554.

INFORMATION ABOUT AUTHORS

Firsov Dmitry Vladimirovich

Junior Researcher, Institute of Economic Policy and Economic Security Problems, The Financial University under the Government of the Russian Federation, Moscow, Russia (125993, Moscow, Leningradsky Prospect, 49); ORCID 0000-0001-5985-7285; e-mail: dvfirsov@fa.ru.

Chernysheva Tatiana Constantinovna

Research-Trainee, Institute of Economic Policy and Economic Security Problems, The Financial University under the Government of the Russian Federation, Moscow, Russia (125993, Moscow, Leningradsky Prospect, 49); ORCID 0000-0002-4744-7198; e-mail: mailtchernysheva@fa.ru.

ACKNOWLEDGMENTS

The study was conducted with the support of the state task of the Government of the Russian Federation to the Financial University for 2021 on the topic «Directions of modernization of the Russian university system, taking into account the demands of the real sector of the economy and global technological trends».

FOR CITATION

Firsov D. V., Chernysheva T. C. Review of Successful Practices of Applying Nowcasting in Socio-Economic Forecasting. *Journal of Applied Economic Research*, 2021, Vol. 20, No. 2, 269–293. DOI: 10.15826/vestnik.2021.20.2.012.

ARTICLE INFO

Received March 14, 2021; Revised April 24, 2021; Accepted May 16, 2021.

