Федеральное государственное бюджетное образовательное учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

Петрова Д.А., Трунин П.В.

Прогнозирование основных макроэкономических показателей с использованием поисковых интернет запросов

Москва 2020

Аннотация. В данном исследовании проводится анализ предсказательной способности такого типа интернет данных, как интенсивность поисковых запросов, для прогнозирования инфляции, уровня безработицы, реальных темпов роста ВВП и курса рубля к доллару в период с января 2004 г. по июль 2019 г. В работе используются поисковые запросы, связанные с финансовыми рынками, инфляционными ожиданиями и макроэкономическими условиями. Результаты 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.

Abstract. This study examines the usefulness of Google Trends intensity search queries data as a measure of economic expectations in predicting inflation, unemployment, real gdp growth and exchange rate during the period between January 2004 and July 2019. We use search queries related to financial markets, inflation expectations and macroeconomic conditions. The results show that the addition of Google search queries improveы out-of-sample forecasts of inflation, unemployment and exchange rate over naïve forecast.

Петрова Д.А., научный сотрудник Центра изучения проблем центральных банков ИПЭИ Российской академии народного хозяйства и государственной службы при Президенте РФ

Трунин П.В., директор научно-исследовательского Центра изучения проблем центральных банков ИПЭИ Российской академии народного хозяйства и государственной службы при Президенте РФ

Данная работа подготовлена на основе материалов научно-исследовательской работы, выполненной в соответствии с Государственным заданием РАНХиГС при Президенте Российской Федерации на 2019 год

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Прогнозирование макроэкономических показателей является важной задачей для экономических властей. Одним из возможных способов получения опережающих индикаторов для предсказания инфляции, безработицы и экономической активности и т.д. выступают опросы населения, предприятий и профессиональных аналитиков. Но данный способ квантификации ожиданий экономических агентов обладает рядом недостатков: слишком затратный, низкочастотность и задержка в раскрытии данных, а также зависимость результата от формулировок вопросов в анкете, способах обработки информации и выборки экономических агентов.

В связи с тем, что в современных условиях все больше ежедневная активность человека связана с онлайн сервисами, появляется возможность непосредственного исследования поведения пользователей без проведения социологических опросов. В условиях неопределенности экономические агенты не могут принимать решения об объемах потребления, сбережений и инвестиций на основе имеющегося эмпирического опыта и знаний, накопленных в прошлом, и начинают поиск информации во внешних источниках. Неполнота знаний будет усиливать спрос на информацию и стимулировать поиск. В таком случае спрос на информацию на определенную экономическую тему будет тесно связан с рыночными ожиданиями. Чем выше обеспокоенность о текущем состоянии экономики среди экономических агентов, тем больше спрос на информацию в Интернете, где без значительных затрат и в доступной форме можно собрать сведения о происходящих событиях на текущий момент. Получение доступа к новой информации о текущей экономической ситуации позволяет населению лучше оценивать риски, принимать более взвешенные решения и повышать рыночную эффективность. Таким образом, данные о поведении пользователей в Интернете становятся ключевым источником информации о предпочтениях и ожиданиях экономических агентов.

По этой причине среди исследователей становится популярным использование широкого спектра методов для квантификации общественного мнения на основе больших данных, которые стали доступны в цифровую эпоху за счет усовершенствованных информационных технологий сбора, хранения и обработки структурированных и неструктурированных данных. Как компании, так и государственные органы занимаются анализом больших массивов данных для повышения эффективности принимаемых решений, выявления желаний и интересов экономических агентов. С помощью методов машинного обучения строятся высокочастотные индикаторы ожиданий экономических агентов на основе анализа новостей в СМИ, комментариев в социальных сетях, обсуждения различных тем в

микроблогах или поисковых запросов, и исследуется их предсказательная способность в прогнозировании макроэкономических показателей. Как показывает международный опыт, использование интернет данных дает возможность получать оперативную информацию о текущем состоянии экономики, а также отслеживать настроения и ожидания экономических агентов. На основе поисковых запросов появляется возможность построения регулярно обновляемых прогнозов основных макроэкономических показателей. Для российской экономики также актуальным является вопрос о том, содержится ли в поисковых интернет-запросах информация, позволяющая улучшать прогноз макроэкономических показателей. Ответ на него имеет высокую практическую значимость и может помочь при проведении денежно-кредитной политики.

Целью данной работы является оценка предсказательной способности поисковых запросов в прогнозировании основных макроэкономических показателей: инфляции, валютного курса, потребления и безработицы.

В первом разделе данного исследования представлены основные теоретические концепции причин и последствий асимметрии информации, включая неблагоприятный отбор и риск недобросовестного поведения, а также рассмотрен ряд моделей с несовершенной информацией, на основе которых выявлены ключевые механизмы влияния неопределенности на общественное благосостояние. Во втором разделе приведены результаты обзора эмпирических исследований, посвященных прогнозированию макроэкономических показателей на основе интернет данных поиска в развитых и развивающихся странах, а также выявлены подходящие для эмпирического анализа методы машинного обучения. В третьем разделе приведены результаты исследования предсказательной способности поисковых запросов Google Trends при прогнозировании макроэкономических показателей на российских данных.

1 Основные подходы к анализу неопределенности и информационной асимметрии при принятии решений экономическими агентами

Информация является одним из самых важных ресурсов, которые оказывают влияние на процесс принятия решения экономическим агентом. В условиях неопределенности (почти всегда) существуют два возможных типа рыночных игроков – информированные и неинформированные агенты. Асимметрия информации в таком случае не позволяет принимать наилучшие из возможных решения, а следовательно, становится причиной потерь в благосостоянии экономических агентов. В данном разделе подробно остановимся на возможных примерах таких ситуаций, а также рассмотрим ряд теоретических подходов, посвященных анализу влияния общедоступной информации (например, появление новости об экономической ситуации, заявления представителей денежных и фискальных властей) на общественное благосостояния.

1.1 Обзор теоретических концепций причин и последствий асимметрии информации

Асимметрия информации – это состояние неопределенности, при котором разные стороны, заключающие сделку или взаимодействующие на рынке, имеют различные наборы информации[CITATION Ake70 \l 1049]. Согласно Акерлофу [CITATION Ake70 \l 1049], асимметрия информации возникает, когда у продавца больше информации о качестве или ценности товара, чем у покупателя. Различия в качестве товаров или услуг создают возможность продавать низкокачественный товар как высококачественный. Такое разнообразие в характеристиках товара становится причиной неспособности покупателя оценить стоимость товара с полной определенностью.

1.1.1

Асимметрия информации и ее источники

Акерлоф [CITATION Ake70 \l 1049] утверждает, что владение товаром позволяет продавцу быть более осведомленным относительно качества товара и обеспечивает информационное преимущество, что создает асимметрию информации в экономике. Несовершенство в доступности информации делает рискованным выбор товара для потенциального покупателя. Неполнота информации подразумевает, что каждый игрок имеет личную информацию о своих стратегиях [CITATION Kul00 \l 1049]. Различие в объемах

доступной информации может предоставить конкурентное преимущество одной из сторон в обмене [CITATION Nay90 \l 1049].

Неполноте информации особое внимание уделяется в области экономики и финансов[CITATION Ake70 \l 1049], [CITATION Mil87 \l 1033], [CITATION Ven86 \l 1033], [CITATION Abo00 \l 1033]. Различные характеристики информации, например, количество, качество и тип влияют на процесс распространения информации среди экономических агентов. Наличие разного количества информации, качества и содержания в информационном множестве каждого агента приводит к асимметрии информации и, следовательно, индивидуальному восприятию проблемы принятия решения.

Исследование информации как фундаментального фактора снижения неопределенности в экономике было впервые проведено в рамках теоретических подходов экономики информации. Важными задачами этой теории являются изучение затрат на поиск информации [CITATION Sti61 \l 1049], проблемы дефицита информации и связанные с этим неблагоприятный отбор и риск недобросовестного поведения [CITATION Ake70 \l 1049], [CITATION Sti00 \l 1049]. В работах выделяются следующие ключевые моменты анализа влияния асимметрии информации на экономику: наличие информации при различных типах взаимодействий на товарных или финансовых рынках, конкурентные преимущества обладания информацией и влияние на поведение человека при принятии решений [CITATION Kul00 \l 1049], [CITATION Nay90 \l 1049].

Стиглиц предполагает [CITATION Sti00 \l 1049], что доступная информация несовершенна и получение информации стоит дорого, а степень информационной асимметрии зависит от действий фирм и потребителей. Асимметрия информация представляет собой неравномерное распределение информации на момент совершения сделки. Каждая сторона сделки хотела бы получить максимальную ценность в обмен на свои вложенные средства. При совершенной конкуренции фирмы хотели бы снизить асимметрию информации, чтобы потенциальные покупатели были способны с небольшими издержками выбрать подходящий товар. В то же время покупатели стремятся получить информацию из различных источников для принятия эффективных решений при ограниченных ресурсах.

В условиях монополистической конкуренции будет существовать асимметрия информации за счет уникальности производимого фирмой-инноватором товара [CITATION Abo00 \l 1049]. Инвестиции в НИОКР могут привести к разработке нового продукта или к улучшению качества уже существующего товара. Однако потребители не будут знать об истинной ценности этого товара до приобретения.

Кроме того, у фирмы будет информационное преимущество, например, у инсайдеров, таких как менеджеры и директора, которые лучше знают о возможных последствиях предпринимаемых ими действий и, таким образом, лучше информированы о будущих перспективах фирмы. По словам Chiang и Venkatesh [CITATION Chi88 \l 1033], инсайдеры имеют доступ к информации, которая не является общедоступной, и эта привилегия может дать преимущество инсайдерам во время торговли. В этом случае акционеры и покупатели вынуждены будут полагаться на различные источники, чтобы получить информацию о возможных изменениях и потенциале компании. Эти источники могут включать репутацию фирмы на рынке [CITATION Nay90 \l 1049], [CITATION Sti00 \l 1049], финансовые показатели и реакцию конкурентов на изменение стратегии данной фирмы.

Различные характеристики товара также могут создавать асимметрию информации. Характеристики поиска (цена и качество) играют важную роль в формировании восприятия ценности [CITATION Kul00 \l 1049], поскольку они могут быть исследованы покупателем до приобретения товара. О качестве товара покупатели могут быть информированы с помощью раскрытия различной информации о характеристиках товара. Однако покупатели могут не использовать доступную информацию для оценки качества, а будут основываться на своем опыте или вкусах, которые значительно отличаются от информации, раскрываемой продавцом [CITATION Mor85 \l 1033]. Это несовпадение критериев оценки качества может вызвать асимметрию информация и наличие различных наборов информации каждой из сторон при оценке ценности товара.

Асимметрия информации, связанная с качеством, может также возникать из-за неспособности потребителя изучить материальные и нематериальные характеристики товара. Например, сложно оценить ценность и качество до покупки товаров опыта [CITATION Nel70 \l 1049] или доверия [CITATION Dar73 \l 1033], включая лекарства, витамины, маркетинговые услуги, высшее образование и прочие услуги. Тип и разнообразие характеристик товара затрудняют получение всей информации в условиях неопределенности.

Неполнота информации, возникающая из-за разброса цен [CITATION Sti61 \l 1049], может появляться за счет затрат на поиск, например, вследствие нежелания или неспособности покупателя искать различные альтернативы, доступные на рынке. Это приводит к ограниченности информации у потребителя.

В теоретической литературе выделяют две основные проблемы асимметрии информации – неблагоприятный отбор и риск недобросовестного поведения. Проблема риска недобросовестного поведения [CITATION Ake70 \l 1049] возникает, когда продавец располагает большей информацией, чем покупатель. Несовершенство информации в таком

случае не позволяет покупателю наблюдать за скрытыми действиями продавца и, следовательно, невозможно определить с полной определенностью, собирается ли продавец вести себя правдиво при заключении сделки.

Неблагоприятный отбор происходит из-за неспособности потребителя должным образом изучить характеристики товара [CITATION Sti00 \l 1049]. Это вызывает затруднения у покупателя при оценке истинной ценности товара, поскольку невозможно определить качество товара и тип продавца. Неблагоприятный отбор и риск недобросовестного поведения вынуждают покупателя собирать дополнительную информацию для принятия решения. Иначе будет существовать высокая вероятность купить «лимон» – некачественный товар [CITATION Sti00 \l 1049].

Фирмы и потребители рассматривают различные механизмы для решения проблемы недобросовестного поведения и неблагоприятного отбора. Фирмы используют сигналы для распространения информации и снижения асимметрии информации. Согласно Спенсу [CITATION Spe73 \l 1049], сигналы являются действиями, которые передают информацию и могут изменить ожидания других экономических агентов на рынке. Подача сигналов может быть важным источником информации о фирме для потенциальных покупателей. Финансовые сигналы, например, рыночные индикаторы (коэффициент цена/прибыль, отношения долга к собственному капиталу и т.д.), цена товара [CITATION Sti00 \l 1049] и репутация [CITATION Nay90 \l 1049] играют важную роль в распространении информации об операционной деятельности фирмы и предполагаемом качестве товара или услуги.

Домохозяйства также используют определенные механизмы снижения асимметрии информации. Скрининг, проверка, и поиск являются одними из многих возможных способов, которые используются неосведомленными покупателями для сбора информации о деятельности фирмы [CITATION Sti00 \l 1049]. Скрининг и поиск тесно связаны: поиск – это процесс скрининга фирмы по существенным признакам релевантности – цене продукта компании[CITATION Sti00 \l 1049]. Поиск включает в себя сбор информации о тех характеристиках товара (цены и качество продукта данной фирмы и конкурентов), которые могут быть определены до покупки [CITATION Kul00 \l 1049]. Еще одним способом снижения асимметрии информации является появление фирмы посредника как связующего звена между покупателем и продавцом.

Большие объемы данных, а также отсутствие информации влияют на то, как человек оценивает альтернативы выбора. Понятие неопределенности занимает важное место в процессе принятия решений. Экономический агент для снижения неопределенности вынужден заниматься поиском и обработкой доступной информации при принятии решений

о потреблении, сбережениях и инвестициях. Weaver полагает [CITATION Wea63 \l 1049], что информация связана с доступными альтернативами и возможностью выбора из этих вариантов. Cole [CITATION Col931 \l 1033] изучил теорию информации Шеннона и развил две концепции информации – неопределенности и энтропии[1]. Авторы определяют практическую информацию [CITATION Art73 \l 1049], которая способствует снижению асимметрии информации и обеспечивает распространение информации среди экономических агентов. Появление такой информации снижает неопределенность и позволяет людям сделать выбор в соответствии со своим опытом и знаниями [CITATION Art73 \l 1049]. Доступ к имеющейся информации помогает людям объединять различные источники информации для принятия решения [CITATION Smi93 \l 1049] и снизить асимметрию информации, которая могла бы затруднить покупателю оценку характеристик товара (неблагоприятный отбор) или анализ скрытых действий агента после заключения сделки (риск недобросовестного поведения).

Чтобы понять связь между асимметрией информации и принятием решений, важно изучить поведение потребителя в условиях неопределености и процесс распространения информации. По словам Блэквелла, Миниарда и Энгеля [CITATION Bla01 \l 1033], процесс принятия решения включает в себя пять этапов: выявление потребностей, поиск информации, обработку данных и определение ценности товара, покупку и оценку полезности после покупки. Поиск обычно включает в себя сбор информации о характеристиках товара, например, цене и качестве [CITATION Kul00 \l 1033]. По мере того, как увеличивается объем информации о характеристиках до совершения покупки, потребители формируют свои ожидания относительно качества и ценности продукта [CITATION Zei88 \l 1049]. В процессе обработки информации потребитель может как повысить, так и уменьшить интерес к конкретному товару [CITATION Han05 \l 1049].

1.1.2

Механизм формирования рыночных ожиданий на основе интернет данных

В настоящее время при принятии решений об объемах потребления, сбережений и инвестиций экономические агенты формируют свои рыночные ожидания с использованием различных источников информации, как показано на рисунке . Когда происходит колебание экономической активности в стране, домохозяйства и фирмы могут это наблюдать непосредственно за счет изменения процентных ставок, цен на товары и услуги и т.д.

1 Энтропия Шеннона определяется как непредсказуемость появления какого-либо символа в полученном сообщении, то есть при условии отсутствия информационных потерь равна объему информации на символ передаваемого сигнала.

[CITATION Sun141 \l 1033]. Помимо этого, информацию о текущем состоянии экономики можно найти в средствах массовой информации (в новостях, по радио и журналах).

Примечание – Источник: [CITATION Sun141 \l 1033]

Рисунок 1 – Схема распространения информации среди экономических агентов

После того как экономические агенты получили общедоступную информацию из экономических и медийных каналов, они начинают формировать индивидуальные системы обработки знаний с двумя возможными состояниями – уверенности и неопределенности относительно будущего. Состояние уверенности возникает в процессе обработки собранной информации, когда экономический агент верно интерпретирует свершившееся событие (например, повышение ключевой процентной ставки сигнализирует об увеличении риска повышения инфляции в будущем, следуя предпосылкам новокейнсианской теории). Состояние неопределенности подразумевает, что экономические агенты не могут сделать однозначный вывод на основе имеющейся информации и начинают поиск информации во внешних источниках. В качестве таких внешних источников информации можно выделить поиск в Интернете или обмен информацией в социальных сетях. Неполнота знаний будет

усиливать спрос на информацию и стимулировать поиск. Спрос на информацию на определенную экономическую тему будет тесно связан с рыночными ожиданиями [CITATION Sun141 \l 1033]. Чем выше обеспокоенность о текущем состоянии экономики среди экономических агентов, тем больше спрос на информацию в Интернете, где без значительных затрат и в доступной форме можно собрать сведения о происходящих событиях на текущий момент. Таким образом, интенсивность поиска по ключевым словам может служить непосредственной мерой выявленных ожиданий экономических агентов.

В следующих подразделах более подробно рассмотрим теоретические подходы к анализу влияния процесса распространения информации и экономических ожиданий на благосостояние экономических агентов.

1.2 Влияние неопределенности и несовершенства информации на благосостояние экономических агентов

Экономика информации предполагает, что повышение степени раскрытия данных денежными и фискальными властями вызывает снижение информационной асимметрии и улучшает эффективность проводимой экономической политики в условиях неопределенности. Появление новой информации о текущей экономической ситуации до принятия решений агентами способствует лучшей оценке риска и принятию более информированных решений, повышающих рыночную эффективность.

В рамках данной концепции большинство теоретических работ исследует проблему влияния неопределенности и несовершенства частной и общедоступной информации на благосостояние экономических агентов. Morris и Shin [CITATION Mor02 \l 1033] утверждают, что общедоступная информация содержит сигналы о будущих намерениях экономических властей и делает их политику более предсказуемой.

В данной работе [CITATION Mor02 \l 1049] для анализа благосостояния частного сектора авторы рассматривают игру принципала (центральный банк или правительство) и агента (рыночные игроки). В модели экономические власти раскрывают некоторую общедоступную информацию, например, через публикацию прогнозов, интервью в средствах массовой информации или на пресс-конференциях. Рыночные игроки формируют свои инфляционные ожидания с учетом всей доступной информации. В модели с несовершенной информацией раскрытие общедоступных данных о состоянии экономики имеет важное значение по следующим причинам: доступная информация способствует получению участниками рынка представления о целях денежно-кредитной или бюджетно-налоговой политик и выступает ориентиром для инфляционных ожиданий частного сектора.

В модели существует континуум экономических агентов в интервале [0, 1]. Функция потерь *i*-ого агента состоит из двух частей. Первая часть представляет собой стандартную квадратичную функцию потерь. Она показывает, в какой мере a_i (инфляционные ожидания) *i*-ого агента отличаются от состояния экономики θ. Вторая часть показывает, на сколько отклоняются ожидания экономического агента от среднего значения (консенсус прогноза будущей инфляции) частного сектора. Возникает внешний эффект, при котором часть экономических агентов пытаются оказывать влияние на решения других. Функция потерь L_i возрастает при увеличении разброса инфляционных ожиданий *i*-ого агента. Чем выше фактор «конкурса красоты»[2] *r*, тем более важен эффект координационного мотива, поскольку основой ожиданий экономических агентов становятся не их собственные представления, а среднее мнение других участников. При этом при больших значениях *r* исход будет социально неэффективным из-за игры с нулевой суммой, когда победители могут получить выгоду только за счет потерь проигравших. Таким образом, функция потерь экономического агента имеет следующий вид:

$$u_i(a,\theta) = -(1-r)(a_i-\theta)^2 - r\left(L_i - \acute{L}\right), \tag{1}$$

где *r* – постоянная величина, $0 < r < 1$, учитывающая в функции потерь «эффект от конкурса красоты» в частном секторе и

$$L_i = \int_0^1 (a_j - a_i)^2 dj \quad , \quad \acute{L} = \int_0^1 L_j dj, \tag{2}$$

Следует отметить, что экономические агенты при формировании ожиданий используют публикуемую центральным банком или правительством общедоступную информацию *y* и личную информацию x_i относительно текущей экономической ситуации θ. Но в условиях неполной информации [CITATION Mor02 \l 1033] возникают независимые и нормально распределенные случайные ошибки при обработке доступной информации. По этой причине информационные сигналы экономических властей и участников рынка принимают вид:

2 Понятие «конкурса красоты» ввел Д. Кейнс по аналогии с газетным конкурсом отбора не самой красивой женщины на фото для каждого индивида, а предугадывания среднего вкуса опрашиваемых. Таким образом, под выигрышем в «конкурс красоты» понимается наиболее точный прогноз общественного мнения (среднее значение) других участников. В качестве примера обычно приводятся трейдеры, формирующие свои ожидания на основе средней ставки по активам других участников, а не на основе фундаментальной стоимости.

$$x_i=\theta+\varepsilon_i \quad , \quad \varepsilon_i \ \ N\left(0,\sigma_\varepsilon^2\right), \tag{3}$$

$$y=\theta+\eta_i \quad , \quad \eta_i \ \ N\left(0,\sigma_\eta^2\right). \tag{SEQ Формула 4}$$

Индивидуальный информационный сигнал x_i i-ого агента неизвестен другим агентам, а общедоступная информация распространяется среди всех агентов. Обозначив за α качество публикуемой публичной информации и β как качество индивидуальной информации, получим: $\alpha=\frac{1}{\sigma_\eta^2}$ и $\beta=\frac{1}{\sigma_\varepsilon^2}$.

В равновесии инфляционные ожидания a_i i-ого экономического агента имеют вид:

$$a_i=\theta+\frac{\varepsilon_i\beta(1-r)+\alpha\eta}{\alpha+\beta(1-r)} \quad . \tag{4}$$

При $r=0$ личная и общедоступная информация влияют на ожидания агентов в зависимости от их точности. Это предполагает, что вес публичной информации η равен $\alpha/(\alpha+\beta)$, а вес индивидуальной информации ε_i – $\beta/(\alpha+\beta)$. Чем больше фактор «конкурса красоты» r , тем больше вес информации, раскрываемой экономическими властями. Это обусловлено мотивом к координации представлений экономических агентов относительно состояние экономики и показывает неоднозначное воздействие публичной информации на ожидания частного сектора.

Ценность общедоступной информации зависит лишь от ее точности и должна быть равной $\alpha/(\alpha+\beta)$, в то время как в равновесии получаем $\alpha/(\alpha+\beta(1-r))$, которое всегда больше ценности публичной информации за счет координации экономических агентов. Рыночные игроки придают большее значение информации, поступающей от центрального банка или правительства, поскольку в таком случае она содержит сведения об ожиданиях других агентов.

Далее определим влияние общедоступной и индивидуальной информации на благосостояние экономических агентов. Ожидаемое общественное благосостояние при условии реализации состояния экономики θ будет иметь вид:

$$E\left[W(a,\theta)\middle|\theta\right]=\frac{-\alpha+\beta(1-r)^2}{[\alpha+\beta(1-r)]^2} \quad . \tag{5}$$

Из уравнения (5) можно увидеть, что благосостояние агентов всегда повышается при увеличении точности индивидуальной информации, поскольку $\frac{\partial E(W\vee\theta)}{\partial\beta}>0$.

Однако общественное благосостояние растет в ответ на повышение точности общедоступной информации ($\frac{\partial E(W\vee\theta)}{\partial\alpha}>0$) только при условии, что $r<0.5$. Это связано с тем, что более точная информация денежных или фискальных властей имеет преимущества тогда и только тогда, когда личная информация недостоверна (агенты не способны сами строить прогнозные модели или делать выводы о текущих изменениях в экономике) или ее слишком мало для принятия решений.

Это означает, что при высокой точности сведений, предоставляемых экономическими властями ($\alpha\to\infty$), экономические агенты не принимают во внимание индивидуальную информацию и принимают решению исключительно на основе имеющейся публичной информации. Кроме того, если в заявлениях центрального банка или правительства не содержится новой информации ($\alpha\to 0$), то такая информация игнорируется и не играет координирующую роль. По мнению авторов, в целом при больших объемах раскрываемой общедоступной информации происходит чрезмерная реакция участников финансового рынка либо по причине высоких издержек обработки большого количества данных, либо из-за неопределенности трактовки заявлений представителей власти. Этот результат показывает, по мнению авторов, что центральный банк или правительство должны понимать, какую информацию следует раскрывать, а какую нет, чтобы не допустить повышение уровня неопределенности среди экономических агентов.

Модель Морриса и Шина критикуют за вывод о снижении благосостояния при повышении транспарентности властей. Например, Свенсон [CITATION Sve06 \l 1033] утверждает, что, за исключением некоторых случаев, доступность публичной информации повышает общественно благосостояние. Более того, даже при одинаковой точности частной и общедоступной информации, общественное благосостояние выше, чем в случае отсутствия публичной информации. Он объясняет свои предположения с помощью расширения предпосылок модели Морриса-Шина [CITATION Mor02 \l 1033]. Пусть ожидаемое общественное благосостояние $V(\alpha)$ при заданном состоянии экономики имеет вид

$$E\left[W(a,\theta)\middle|\theta\right]=\frac{-\alpha+\beta(1-r)^2}{\left[\alpha+\beta(1-r)\right]^2}\equiv V(\alpha) \tag{6}$$

Транспарентность властей определяется как точность общественной информации α . Это означает, что более высокий уровень раскрытия информации соответствует более высокой точности сигнала центрального банка (при меньшей дисперсии σ_η^2). Эффект от предельного повышения транспарентности на общественное благосостояние характеризуется производной благосостояния по точности общедоступной информации

$$V'(\alpha)=\frac{\alpha-f(r)\beta}{\left[\alpha+\beta(1-r)\right]^3}, \tag{7}$$

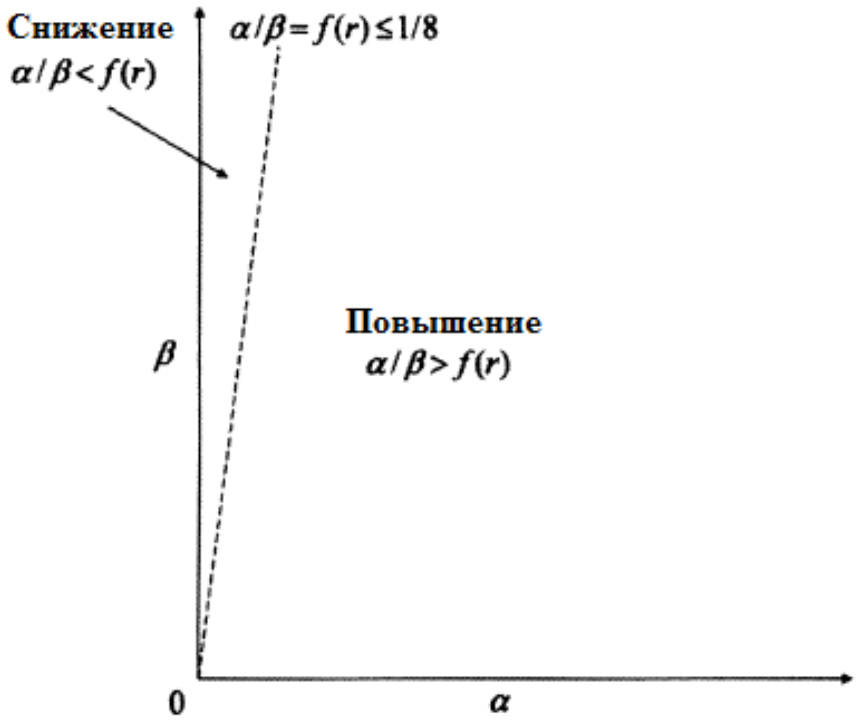
где $f(r)\equiv(2r-1)(1-r)$. Если $V'(\alpha)<0$ и $\frac{\alpha}{\beta}<f(r)$, то благосостояние снижается в ответ на повышение точности общедоступной информации согласно модели Морриса-Шина.

Свенсон предполагает [CITATION Sve06 \l 1033], что это условие нарушается при достаточно разумных предпосылках. Во-первых, как и в модели Морриса-Шина [CITATION Mor02 \l 1033], накладывается ограничение, что $r\in(0.5,1)$ для выполнения условия $\frac{\alpha}{\beta}<f(r)$. Напротив, при $r\in[0,0.5]$ или $r=1$ получаем $f(r)\le 0$ и условие $\frac{\alpha}{\beta}<f(r)$ всегда нарушается. Таким образом, каждый агент должен придавать больший вес r , чем связанной с полезностью компонентой (экономические агенты будут полагаться на среднее мнение других участников рынка). Если этого не происходит, то общественное благосостояние повышается при транспарентности.

Во-вторых, при $f(r)\le f\left(\frac{3}{4}\right)=\frac{1}{8}$, следовательно, $f(r)$ достигает максимума в снижении благосостояния в $\frac{1}{8}$ при $r=\frac{3}{4}$. Даже если $r\in(0.5,1)$, то нарушается условие снижения благосостояния в ответ на повышение уровня раскрытия информации. Таким образом, всякий раз, когда точность информации властей выше $\frac{1}{8}$ от точности частной информации, общественное благосостояние увеличивается. Автор объясняет этот эффект тем, что центральный банк и правительство располагают большими ресурсами для

сбора, обработки и анализа данных об экономической ситуации, чем любой экономический агент.

Рисунок показывает, какими должны быть α и β, при которых общественное благосостояние растет или снижается при повышении уровня транспарентности властей. Как видно из рисунка, благосостояние экономических агентов увеличивается части «Повышение» от пунктирной линии до горизонтальной оси.

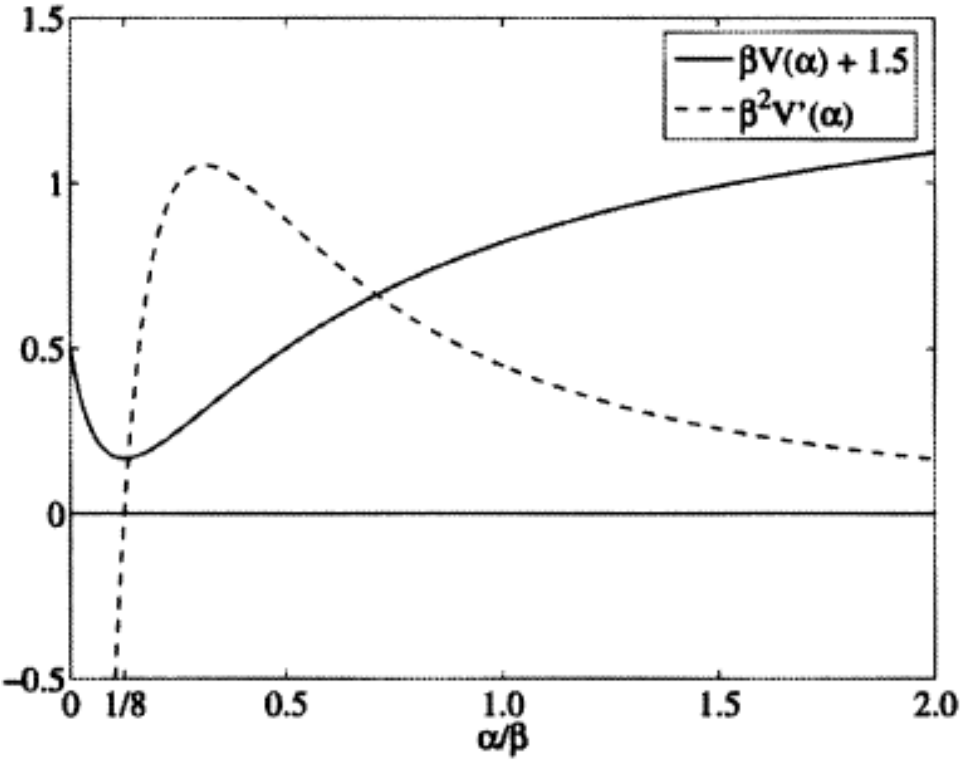


Примечание – Источник: [CITATION Sve06 \l 1033]

Рисунок 2 – Области увеличения или снижения общественного благосостояния при повышении уровня транспарентности

Кроме того, Свенсон [CITATION Sve06 \l 1033] проводит анализ взаимосвязи общественного благосостояния и степени открытости экономических властей при различных соотношениях α и β (см. рисунок). Функция благосостояния для наглядности масштабируется с помощью β и со сдвигом 1.5. Вес фактора «конкурса красоты» задается на уровне $r=\frac{3}{4}$, когда благосостояние агентов снижается при $\frac{\alpha}{\beta}<\frac{1}{8}$, минимальное значение $\frac{\alpha}{\beta}=\frac{1}{8}$ и увеличивается благосостояние при $\frac{\alpha}{\beta}>\frac{1}{8}$ в ответ на повышение степени прозрачности центральным банком или правительством. Пунктирная линия показывает предельную общественную выгоду от транспарентности (производную

общественного благосостояния, умноженную на β^2). Таким образом, предельные общественные издержки возникают при $\frac{\alpha}{\beta}<\frac{1}{8}$, нулевые при $\frac{\alpha}{\beta}=\frac{1}{8}$ и предельные общественные выгоды $\frac{\alpha}{\beta}>\frac{1}{8}$. Аналогичные результаты были получены для $r\in(0.5,1)$.



Примечание – Источник: [CITATION Sve06 \l 1033]

Рисунок 3 – Общественное благосостояние как функция от степени прозрачности центрального банка

Результаты показывают, что исход с одинаковой точностью общедоступной и частной информации при $\frac{\alpha}{\beta}=1$ также входит в область с положительной реакцией благосостояния на повышение уровня транспарентности. Кроме того, автор показал, что, несмотря на достижение локального максимума благосостояния при нераскрытии информации, существует глобальный максимум при максимальном уровне открытости экономических властей с бесконечной точностью общедоступного сигнала. По этой причине, в промежуточных случаях может наблюдаться снижение общественного благосостояния (по сравнению с $\alpha=0$) из-за незначительного объема раскрываемой информации.

Следует отметить, что в реальности невозможна ситуация с полным скрытием данных властями. И при достижении достаточно высокого уровня транспарентности возможно

достичь более высокого общественного благосостояния. В целом Свенсон [CITATION Sve06 \l 1049] делает вывод, что при разумном выборе параметров общественное благосостояние растет при повышении открытости денежных и фискальных властей в модели Морриса-Шина [CITATION Mor02 \l 1049]. Эталонный исход с одинаковой точностью частной и общедоступной информации обеспечивает более высокий уровень общественного благосостояния, чем если бы сведения не раскрывались.

Angeletos и Pavan [CITATION Ang04 \l 1033] и Hellwig [CITATION Hel05 \l 1033] также проводили анализ в рамках модели Морриса-Шина. Рассматривая экономики с возрастающей отдачей от масштаба [CITATION Ang04 \l 1033] и монополистической конкуренцией [CITATION Hel05 \l 1033], авторы продемонстрировали, что более точная общедоступная информация всегда повышает благосостояние экономических агентов. Причина неоднозначности эффекта от транспарентности заключается в различных взглядах на вид индивидуальной функции полезности, например, рассматривая в функции потерь влияние отклонения прогноза индивида от других участников и разброса мнений всех игроков.

В модели Морриса-Шина [CITATION Mor02 \l 1049] потери благосостояния снижаются с увеличением отклонения ожиданий индивидуума от действий других, но этот разброс не играет особой роли для общественного благосостояния. Angeletos и Pavan [CITATION Ang04 \l 1049] утверждают, что увеличение объема публичной информации способствует более эффективной координации, которая имеет ценность для рынка, но не для общества. На самом деле, они рассматривали функцию потерь, учитывая отклонение от действий других экономических агентов и влияние пространственной дисперсии (экономику с комплементарностью на социальном уровне для рассмотрения эффекта разброса мнений частного сектора), поэтому такая координация является социально значимой. Авторы уверены, что финансовые рынки подвержены «конкурсу красоты», когда трейдеры получают выгоды от предсказания консенсусных ожиданий частного сектора, поскольку эффективность рынков предполагает реакцию цен на фундаментальные факторы. В любой экономике общественно желаемым результатом будет избегание любой формы чрезмерной реакции на новости или заявления властей.

Cornand и Heinemann [CITATION Cor08 \l 1033] также делают вывод о повышении благосостояния экономических агентов в случае раскрытия новой информации при высоком уровне транспарентности. Это результат получен при допущении о более эффективном распространении информации в обществе через средства массовой информации, где доступ к данным есть только у части информированных трейдеров. Например, такое может

происходить, если представители центрального банка дают интервью или приглашают лишь небольшую группу журналистов на пресс-конференцию. Заявления в таких экономиках менее широко распространяются, чем формальные решения по экономической политике, или требуют больше времени для распространения внутри частного сектора. Низкая скорость проникновения таких заявлений делает их общим знанием в любой момент (общественность узнает практически одновременно о произошедшем событии), поскольку распространение информации может повысить степень публичности и увеличить долю информированных инвесторов.

Результаты проведенного обзора теоретических концепций причин и последствий асимметрии информации показывают, что проблема нехватки информации является чрезвычайно важной в условиях неопределенности. Несовершенство информации возникает, когда у одной из сторон больше информации на момент совершения сделки или принятия решений.

В теоретической литературе выделяют две основные проблемы асимметрии информации – неблагоприятный отбор и риск недобросовестного поведения. Несовершенство информации при риске недобросовестного поведения не позволяет принципалу наблюдать за скрытыми действиями агента и, следовательно, в этих условиях невозможно оценить стоимость товара до совершения сделки, например, при страховании, выдаче кредита для инвестиционного проекта и тд. Решения данной проблемы можно добиться за счет раскрытия информации фирмой (сигналлинга или данных о внутренних характеристиках товара) или появления на рынке посредника, располагающего информацией об истинной ценности и характеристиках товара. Неблагоприятный отбор происходит из-за неспособности покупателя должным образом изучить характеристики товара, например, при покупке подержанного автомобиля или другого товара длительного пользования. Это вызывает затруднения у потребителя при оценке стоимости товара, поскольку сложно распознать качество товара и тип продавца. Неблагоприятный отбор и риск недобросовестного поведения вынуждает потребителя собирать дополнительную информацию для принятия решения.

Также был рассмотрен ряд теоретических подходов, предполагающих что динамика макроэкономических показателей подвержена влиянию ожиданий экономических агентов. Распространение новой информации способствует корректировке ожиданий и настроений участников рынка относительно экономических условий текущего и будущего периодов.

Кроме того, важным аспектом теории информации является объем доступной информации и влияние его на благосостояние экономических агентов. В условиях

неопределенности повышение точности частной информации (совершенствование моделей прогнозирования, навыки и знания для интерпретации происходящих событий и т.д.) всегда повышает благосостояние экономических агентов. Что касается общедоступной информации, то ее воздействие на благосостояние будет зависеть от точности и объема раскрываемой информации.

Для дальнейшего анализа влияния процесса распространения информации на принятие решений рыночными игроками на практике перейдем к анализу эмпирических методов к прогнозированию макроэкономических показателей на основе прокси переменной к общедоступной информации – поисковых запросов.

Обзор основных эмпирических подходов к прогнозированию макроэкономических показателей

Прогнозирование макроэкономических и финансовых показателей является важной задачей для экономических властей. Одним из возможных способов получения опережающих индикаторов для предсказания инфляции, безработицы, потребительского доверия, экономической активности и т.д. выступают опросы населения, предприятий и профессиональных аналитиков. Но данный способ квантификации ожиданий экономических агентов обладает рядом недостатков: высокий уровень затрат[3], низкочастотность и лаг в раскрытии данных, а также изменение результата при различных формулировках вопросов социологического опроса, способах обработки информации и выборки экономических агентов.

В последнее время становится популярным среди исследователей использование широкого спектра методов для квантификации общественного мнения на основе больших данных. Строятся высокочастотные индикаторы ожиданий экономических агентов на основе новостей в СМИ, комментариев в социальных сетях, обсуждения различных тем в микроблогах или поисковых запросов, и исследуется их предсказательная способность в прогнозировании макроэкономических показателей. В данной работе будут рассмотрены в качестве прокси-переменной для выявленных ожиданий только поисковые запросы Google Trends.

3 К примеру, Фонд общественного мнения проводит опрос населения (по заказу Банка России) на ежемесячной основе для измерения инфляционных ожиданий и опрашивает 2000 человек в 55 регионах России и привлекает к работе штат, проводящий личные интервью по месту проживания. Таким образом, для организации такого типа социологических опросов требуются колоссальные финансовые ресурсы.

1.3 **Описание основных эконометрических подходов и методов машинного обучения**

В настоящее время онлайн сервисы и поисковые системы в Интернете дают возможность получить без значительных дополнительных издержек информацию о состоянии экономики до публикации официальных данных. Раскрытие статистики правительством и центральным банком во всех странах происходит с запаздыванием от нескольких недель до нескольких месяцев, и экономические агенты получают доступ к данным по экономическим показателям в основном на ежемесячной или ежеквартальной основах. В свою очередь интенсивность поисковых запросов Google Trends по экономической и финансовой тематикам предоставляется с еженедельной и ежемесячной частотой данных (на небольших горизонтах есть сопоставимые дневные данные[4]) и может характеризовать обеспокоенность населения относительно макроэкономической ситуации в экономике. Как правило, официальные статистические данные позволяют наилучшим образом строить прогноз по основным макроэкономическим показателям, но поскольку они публикуются с временным лагом, статистика не доступна на момент принятия решений населением. По этой причине актуальность приобретает использование еженедельных и ежемесячных данных по запросам в Интернете с целью получения прокси переменной для выявленных ожиданий относительно экономических показателей.

Сервис Google Trends публикует временные ряды интенсивности поисковых запросов по различным ключевым словам в поисковой системе Google для заданного географического местоположения. Индекс интернет запроса сервиса Google Trends по определенному ключевому слову дан не в абсолютном выражении, а в виде интенсивности в диапазоне от 0 до 100 баллов, измеренной как:

$$GSVI_{K,t} = \frac{S_{i,t}}{\sum_{K_{1,t}}^{K_{m,t}} S_{i,t}}, \quad (8)$$

где $S_{i,t}$ – количество интернет-запросов в Google Trends по определенному ключевому слову/фразе на момент времени t, $\sum_{K_{1,t}}^{K_{m,t}} S_{i,t}$ – совокупное количество поисковых запросов в

4 Можно выгружать со скользящим окном, начиная с 2004 г. Однако, данные несопоставимы и имеют свои максимальные значения. При агрегации в один ряд таких данных можно получить сильно зашумленный временной ряд, так как мы не можем корректировать на объем поисковых запросов, а усреднение вызовет изменение процесса порождения данных.

момент времени t и для заданного местоположения, $GSVI_{K,t}$ (Google Search Volume Index) – интерес к определенной теме относительно максимального значения ряда в 100 баллов для заданного региона. 100 баллов присваивается максимальному значению интенсивности $GSVI_{K,t}$, 50 баллов – интенсивности поискового запроса вдвое ниже, чем в первом случае, 0 – в случае отсутствия интереса к определенной тематике (не превышает 1% от максимума).

Несмотря на все преимущества показателей интенсивности интернет-запросов возникает ряд сложностей с их непосредственным использованием в эмпирическом анализе. Во-первых, повсеместное использования Интернета для сбора информации произошло не так давно, поэтому временные ряды поисковых запросов доступны только с января 2004 г. и являются более короткими рядами по сравнению с другими макроэкономическими показателями. Поиск в сети Интернет довольно часто коррелирует с такими демографическими характеристиками, как возраст и доход, поэтому выборка пользователей нерепрезентативна.

1.3.1 Поисковые запросы как мера инфляционных ожиданий при прогнозировании инфляции

Прогнозирование инфляции является важной задачей при принятии решений экономическими агентами относительно потребления, инвестиций и сбережений. Статистические данные по инфляции на ежемесячной основе публикуются с временным лагом и на момент принятия решений нет доступа к официальным данным. Население может лишь наблюдать повышение цен на продукты питания или колебания цен. На основании имеющихся данных экономические агенты могут делать вывод о возможных в будущем краткосрочных колебаний макроэкономических показателей. Для подтверждения или опровержения гипотез об изменении цен экономические агенты будут предъявлять спрос на информацию в Интернете до того, как данные по фактической инфляции станут им доступны. В этой связи с распространением Интернета многие решения стали приниматься на основании полученной во время поиска информации. В этом случае интернет данные могут служить прокси переменной для инфляционных ожиданий.

В таблице представлены основные результаты эмпирических исследований по данной теме. Большинство исследований показывает, что использование поисковых запросов повышает предсказательную способность моделей прогнозирования инфляции. Важно отметить, что в эмпирических исследованиях рассматриваются различные стратегии отбора ключевых слов для поисковых запросов.

Таблица 1 –Результаты эмпирических работ по анализу предсказательной способности поисковых запросов в прогнозировании инфляции

Авторы	Выборка	Метод	Выводы
Guzman (2011)	США, февраль 2004 - октябрь 2008	МНК	Запрос «инфляция» позволяет улучшить точность прогноза инфляции
Koop, Onorante (2013)	США, январь 2004 – июль 2012 г.	DMA, DMS	Включение вероятностей Google Trends в DMS обеспечивает улучшение прогноза инфляции по сравнению с альтернативными моделями с фундаментальными факторами
Seabold et al (2015)	Коста-Рика, Сальвадор и Гондурас 2004-2014 гг	МНК, лассо, эластичная сеть	Индекс GT повышает точность прогноза инфляции: фасоль, образование, продукты питания в Коста-Рике и в Сальвадоре мука, алкоголь, развлечения.
Li et al (2015)	Китай, январь 2004 -декабрь 2012 г.	MIDAS, ADL	MIDAS имеет более высокую предсказательную силу по сравнению с ADL на большинстве прогнозных горизонтов, за исключением прогноза на 1 месяц
Niesert и др. (2019)	5 развитых стран, февраль 2004 – декабрь 2016 г.	STS, BSTS	Модель с поисковыми запросами прогнозирования ИПЦ имеет не высокую предсказательную силу

Примечание – Источник: составлено авторами

Например, в работе [CITATION Guz113 \l 1033] для прогноза инфляции в следующие 12 месяцев в США используется непосредственно интенсивность интернет-запроса «инфляция» и сравнивается с другими мерами инфляционных ожиданий.

В существующих исследованиях предсказательной способности поисковых запросов при прогнозировании макроэкономических показателей уделяется недостаточное внимание специфике интернет данных и проблеме проклятья размерности. В стандартном случае авторы оценивают линейные модели с постоянными во времени коэффициентами, которые не учитывают структурные изменения в параметрах модели, а также степень воздействия в различные периоды времени поисковых запросов по заданным экономическим ключевым фразам.

В исследовании [CITATION Koo13 \l 1033] рассматривается большое количество поисковых запросов по экономическим тематикам: «цены на сырье», «предложение денег», «индекс финансовых условий», «промышленное производство», «инфляция», «временной спред», «безработица», «цена на нефть», «заработная плата».

В различных спецификациях в качестве объясняющих переменных динамическое усреднение моделей (DMA) и динамический выбор модели (DMS) включаются как интенсивность поисковых запросов по заданным тематикам, так и вероятность $p_{t,j}$ того, что существует значимая взаимосвязь между макроэкономическими показателями и интенсивностью поисковых запросов. Авторы предполагают, что включение в анализ вероятности $p_{t,j}$ позволяет учесть в модели, например, не сам факт увеличения или снижения цены на нефть после увеличения поисковых запросов, но сигнализировать о целесообразности включения цены на нефть как предиктора в модель с переключениями при прогнозировании макроэкономических показателей. Данный подход позволяет учесть специфику поисковых запросов и структурные изменения в параметрах модели, а также степень воздействия в различные периоды времени поисковых запросов по заданным экономическим ключевым фразам.

В другой работе [CITATION LiX15 \l 1033] для отбора ключевых слов применяется анализ экономических и финансовых новостей. Авторы предполагают, что пользователи выбирают ключевые слова перед тем, как они начинают поиск новостей об изменении цен в экономике. Этот подход позволяет смоделировать поведение экономических агентов при условии, что ключевые слова удовлетворяют двум предположениям. Во-первых, ключевые слова содержатся в новостных статьях, которые они ищут. Во-вторых, ключевые фразы не должны включаться в текст новостей, которые не представляют интереса для пользователей. Когда эти предположения выполнены, ключевые слова могут быть эффективно извлечены из новостей, а значит, будут использоваться при поиске в Интернете. Это следует из принципа рационального поиска информации: пользователи склонны объективно оценить актуальность и достоверность сведений без предвзятости при сборе информации [CITATION Bir07 \l 1033], [CITATION Par10 \l 1033].

Выгрузка поисковых запросов осуществляется по отобранным ключевым словам. Авторы рассматривают как поисковые запросы по отдельным ключевым словам, так и композитные индексы, которые построены на алгоритме негативной эмоциональной окраски [CITATION Ant04 \l 1033], [CITATION Dan10 \l 1033], учитывающем вероятность повышения инфляционных ожиданий при восприятии негативных ключевых слов[5]. Это допущение основано на том, что общественность сильнее реагирует на плохие новости, чем на положительные новости. Негативные поисковые запросы могут представлять собой пассивное отношение к текущей экономической ситуации со стороны экономических

5 Экономические агенты сильнее реагируют на плохие новости, поэтому авторы строят индекс инфляционных ожиданий на основе предположения о таких негативных ключевых словах/словосочетаниях, как рост цен, цены выросли, резкий скачок цен и т.д.

агентов. Например, пользователи ищут «цены растут в июле», а значит они уверены, что цены в июле растут быстро.

После проведенного анализа текста и создания базы ключевых слов на основе экспертного мнения исключались ключевые слова, не относящиеся к тематике «инфляции», например, «международный рынок», «труд», «импорт» и т.д. В результате было отобрано 20 наиболее релевантных терминов, включая синонимы китайского языка, с учетом эмоциональной окраски направленности изменения цен: потребительская цена, ИПЦ, увеличение цен, рост цен, цена увеличилась, резкий скачок цен, цены упали, цены снизились, снижение цен, поднимать цены и т.д.

В работе [CITATION Sea15 \l 1033] ключевые слова были выбраны из предположения, что они содержат важную информацию о потребительских настроениях и предпочтениях. По мнению авторов [CITATION Sea15 \l 1033], получение информации о поведении потребителей в режиме реального времени дает возможность лучше предсказывать изменения цен при прочих равных условиях. Были отобраны следующие поисковые запросы, характеризующие потребительские настроения: рис, сахар, мясо, дорогой, свинина, топливо, благотворительность, дизель, фасоль, газ, бензин, инфляция, доход, кукуруза, платеж, хлеб, цена, цены, пропан, оклад, заработная плата, пшеница.

Niesert и др. [CITATION Nie19 \l 1033] для прогнозирования ИПЦ в США, Великобритании, Канаде, Германии и Японии применяет две стратегии отбора поисковых запросов – наиболее связанных с макроэкономическими показателями и предлагаемых Google Correlate[6]. Для каждого макроэкономического ряда подбираются 60 релевантных категорий. Из Google Correlate выгружается статистика по интернет-запросам, наиболее положительно или отрицательно коррелирующих с макроэкономической переменной.

В целом можно сделать вывод, что интернет данные позволяют повысить точность прогнозов инфляции. Это обусловлено тем, что повышение обеспокоенности среди населения относительно текущей инфляции стимулирует экономических агентов к поиску и сбору информации в Интернете в условиях неопределенности.

1.3.2

Прогнозирование валютного курса на основе поисковых запросов

Начиная с 1970-х гг., модели прогнозирования обменного курса строятся на предположении о том, что номинальный валютный курс подвержен влиянию ожиданий экономических агентов [CITATION Eng05 \l 1049], [CITATION Fre85 \l 1049] и [CITATION

6 Сервис Google Correlate предоставляет перечень поисковых запросов, наиболее коррелирующих с исходным временным рядом или заданным ключевым словом.

Dyb17 \l 1049]. Как отмечалось ранее, в цифровую эпоху в качестве меры ожиданий на валютном и фондовом рынке часто используются поисковые запросы. Многие авторы сравнивают предсказательную способность моделей, включающих интернет данные, с альтернативными структурными моделями или случайным блужданием.

Более подробно остановимся на работах, показанных в таблице . Одной из первых работ по исследованию влияния интенсивности поисковых запросов на волатильность валютного курса стала [CITATION Smi12 \l 1033]. Поскольку Интернет является одним из ключевых источников информации, то повышение интереса пользователей к определенной теме будет отражать процесс распространения информации и ожидания экономических агентов. По этой причине наблюдаемые колебания валютного курса могут сильно зависеть от частоты появления информации на финансовых рынках.

Автор проводит анализ для курсов валют Австралии, Канады, Европы, Японии, Новой Зеландии, Швейцарии и Великобритании к доллару США в течение 2004-2010 гг. на еженедельных данных. Отбор ключевых слов осуществляется исходя из экономических предположений о возможном влиянии объема поиска на глобальные потоки информации. Автор собирает данные по интенсивности поисковых запросов по ключевым словам: «экономический кризис+финансовый кризис», «инфляция» и «рецессия». Выбор таких ключевых слов обусловлен возможностью проверить, происходит ли в кризисный период переток средств из более рисковых активов в менее рискованные (в кризис инвесторы начинают покупать больше активов с минимальным риском, таких как казначейские облигации США или другие долговые инструменты, номинированные в долларах США), а, следовательно, повышение волатильности национальной валюты к доллару США.

Таблица 2 –Результаты эмпирических работ по анализу предсказательной способности поисковых запросов в прогнозировании валютного курса

Авторы	Выборка	Метод	Выводы
Smith (2012)	7 развитых стран, январь 2004 - декабрь 2010	GARCH	Поисковый запрос «экономический +финансовый кризисы» оказывает повышательное давление на волатильность валютного курса
Bulut (2017)	11 развитых стран, января 2004 г. по июнь 2014 г	МНК	Модели с поисковыми запросами превосходят по точности случайное блуждание только для Австралии, Канады, Дании и Сингапура
Dybka и др. (2017)	Польша, январь 2004 - май 2016	SVAR с PCA	Модель с поисковыми запросами имеет более высокую предсказательную силу на горизонтах прогноза 6 и 12 месяцев

Примечание – Источник: составлено авторами

В другой работе [CITATION Dyb17 \l 1049] осуществляется прогнозирование курса польского злотого к евро с учетом настроений экономических агентов на кредитном рынке, ожиданий относительно цен и настроений на финансовом рынке в период. Показатели настроений и ожиданий инвесторов извлекаются с помощью метода главных компонент из поисковых запросов. Информационное множество для кредитного рынка состоит из интернет запросов, связанных с крупными польскими банками, финансовыми институтами и товарами, приобретаемыми за счет кредита (недвижимость или автомобили). Инфляционные ожидания идентифицируются из поисковых запросов, включающих названия крупных торговых сетей, популярных товаров и энергетических компаний (для учета цен на тарифы ЖКХ). Настроения на финансовом рынке оцениваются исходя из предположения, что поисковые запросы, содержащие информацию об интересе инвесторов к польским фондовым индексам и курсу злотого, оказывают влияние на выбор инвестиционной стратегии розничным инвестором и отражает выявленные ожидания относительно будущих колебаний ценных бумаг или валютных курсов.

При прогнозировании валютного курса зачастую также используется отбор ключевых слов из предположений экономической теории. Например, в работе [CITATION Bul151 \l 1049] рассматриваются следующие теории формирования фундаментального валютного курса: монетарная, паритет покупательной способности и паритет процентных ставок.

Помимо оценки традиционных моделей, основанных на теории, автор в качестве альтернативных моделей использует спецификации только с поисковыми запросами, следующие из теории паритета покупательной способности и монетарной теории. С этой целью автор отбирает ключевые слова по ценовой тематике («инфляция», «цены», «ИПЦ», «дешевый») для паритета покупательной способности, а для анализа влияния спроса на деньги на валютный курс собираются данные интернет запросов по тематике дохода («купить», «потратить», «сбережения», «благотворительность», «работа», «вакансии», «реализация залога», «безработица») и ликвидности («наличность», «кредит», «банкомат») в период с января 2004 г. по июнь 2014 г. (ежемесячно) для Австралии, Великобритании, Гонконг, Дании, Еврозона, Израиля, Канады, Сингапура, Швеции, Швейцарии и Японии.

В целом результаты исследований показывают, что структурные модели и модели с интернет запросами не всегда позволяют получить более точный прогноз номинального валютного курса в развитых странах, поскольку для развитых стран характерно выполнение гипотезы эффективного рынка. Следует отметить, что это также может возникать из-за неправильной спецификации моделей и не учета важных факторов валютного курса.

1.3.3

Прогнозирование безработицы на основе поисковых запросов Google Trends

Поиск более точных прогнозов динамики рынка труда всегда был важной задачей исследователей и экономических властей. Понимание того, какие изменения будут происходить на рынке труда, приобретает особую значимость в периоды спадов, когда наблюдается замедление экономической активности. В последние десятилетия все более популярным среди экономических агентов различных стран становится поиск работы в Интернете. Это обусловлено развитием онлайн сервисов-агрегаторов, на которых работодатели размещают вакантные должности и возможные требования для кандидатов. Такой способ поиска работы позволяет снизить издержки сбора информации о существующих вакансиях и сделать процесс подбора персонала более открытым. Следовательно, доступные данные по интернет запросам обеспечивают исследователей высокочастотными индикаторами активности пользователей при поиске работы. Эмпирические результаты исследований, посвященных данной проблеме представлены в таблице .

Таблица 3 – Результаты эмпирических работ по анализу предсказательной способности поисковых запросов в прогнозировании безработицы

Автор	Страна	Метод	Выводы
M’Claren (2011)	Великобритания, июнь 2004 - январь 2011	AR, ARX	«JSA» позволяет повысить точность прогноза безработных и содержит дополнительную информацию относительно опросных мер
Bughin (2011)	Бельгия [CITATION Bug11 \l 1033]	ECM	интернет запросы объясняют 15 % дисперсии безработицы
Choi, Varian (2012)	США, 2004-2011	AR, ARX	Результаты показывают улучшение точности прогноза первоначальных требований в поворотных точках
Vicente et al (2015)	Испания, 2004-2012	ARIMAX	Модели с поисковыми запросами «предложение о трудоустройстве» или «предложение работы» имеют более высокую предсказательную способность
Smith (2016)	Великобритания январь 2004 - ноябрь 2014	MIDAS	Преимущество в прогнозировании GT только на горизонте 3-4 недель. На горизонтах 1-2 и 5-8 недель точность сопоставима с индикаторами на основе опросов
Brake (2017)	Нидерланды [CITATION Bra17 \l 1049]	AR, ARX	Повышение интереса пользователей в Интернете к поиску информации о характеристиках статуса безработного

			ведет к увеличению уровня безработицы.
D’Amuri, Marcucci (2017)	США, февраль 2004-февраль 2014	AR, ARX	Включение в модель интернет-запроса «вакансии» улучшает точность прогноза на следующие 1-12 месяцев

Примечание – Источник: составлено авторами

В работе [CITATION McL11 \l 1049] предполагается, что важными индикаторами настроений на рынке труда в Великобритании вероятно выступают интенсивности поисковых запросов по ключевым словам: «вакансии», «пособие по безработице», «льготы для безработных», «безработный» и «уровень безработицы». После проведенного первичного анализа авторы получили, что поиск по интернет-запросу «вакансии» скорее всего осуществляется как безработными, так и трудоустроенными экономическими агентами, поскольку всегда к этой теме высокий уровень интереса и динамика не соотносится с безработицей. Поисковый запрос «безработный» также практически не отражает изменения в фактической безработице, за исключением периода рецессии. И только интенсивность интернет-запроса «JSA», характеризующий интерес экономических агентов, которые в скором времени могут стать безработными, к пособию по безработице. По мнению авторов «JSA» является хорошим индикатором ожиданий на рынке труда, поскольку сильно коррелирует с количеством безработных.

Choi и Varian [CITATION Cho123 \l 1033] отмечают важность прогнозирования числа первоначальных требований выплаты пособия по безработице, которое является опережающим индикатором снижения уровня занятости. Так, периоды рецессии США характеризовались тем, что первоначальные требования были на пике за 12-18 месяцев до начала спада в экономике и роста безработицы.

Авторы используют для прогноза числа первоначальных требований выплаты пособия по безработице поисковые запросы Google Trends для США в течение 2004-2011 гг. Отбор ключевых слов строится на предположении о возможной активности безработного в Интернете. Когда кто-то становится безработным, стоит ожидать, что экономический агент будет искать: «документы для безработного», «биржа труда», «пособие по безработице», «требования к безработному», «вакансии», «резюме» и т.д. Авторы отмечают, что сервис Google Trends обеспечивает всеми необходимыми данными по интернет-запросам в двух категориях «Вакансии» и «Благосостояние и безработица».

В другой работе [CITATION Bug11 \l 1049] для прогнозирования заявок на пособие по безработице в Бельгии по тем же причинам рассматривалась категория «Вакансии», но в качестве предиктора использовался наиболее популярный поисковый запрос из этой категории – «безработица».

Brake [CITATION Bra17 \l 1033] исследует возможность улучшения качества прогнозов безработицы на основе интернет данных Google Trends в период с января 2004 г. по апрель 2017 г. для Нидерландов. Автор выделил восемь поисковых запросов Google Trends, связанных с тематикой безработицы и характеризующихся максимальной интенсивностью поиска. Агрегированный индикатор строился следующим образом:

$$Google\ Indicator = \left[\frac{I_{w,t} * SV_{w,t}}{\sum SV_t} \right], \quad (9)$$

где $I_{w,t}$ – интенсивность интернет запроса по определенному ключевому слову w в момент времени t, $SV_{w,t}$ – объем поисковых запросов по определенному ключевому слову w в момент времени t и $\sum SV_t$ – совокупный объем интернет-запросов для восьми выбранных ключевых слов («институт социального страхования», «пособие по безработице», «закон о безработице», «пособие малоимущим», «безработный, werkloosheid», «безработный, werkloos», «подать заявку для безработного», «размер пособия по безработице»).

В работе [CITATION Vic15 \l 1033] проводится анализ поисковых запросов Google Trends, которые непосредственного связаны с трудоустройством – «предложение о трудоустройстве» и «предложение работы». Авторы предполагают, что такого типа поисковые запросы отражают активность пользователей при поиске работы в Интернете. Важность таких индикаторов обуславливается усилением цифровизации (более половины населения страны имели доступ в Интернет к 2008 году) и популярностью среди работодателей публикации вакансий на своих сайтах или на онлайн сервисах-агрегаторах.

В работе [CITATION Smi16 \l 1049] использовались следующие поисковые запросы: «пособие по безработице», «увольнение работников», «выходное пособие», «прописанный в законе размер выходного пособия», «закон об увольнении работников», «период уведомления о сокращении», «увольнение по собственному желанию», «отдел выплат пособий по сокращению», «выплата при сокращении работника», «увольнение работников в Великобритании», «уведомление об увольнении», «выплата выходного пособия», «письмо о причинах увольнения», «расчет выходного пособия», «права уволенных работников», «расчет выплаты по сокращению», «соответствующий закону порядок увольнения», «сколько уволенных», «налог на сокращение работников», «увольнение работников Великобритания», «налог с выходного пособия», «консультация уволенных».

С помощью отобранных поисковых запросов строились три типа показателей настроений населения относительно безработицы. Первый показатель GoogJSA – интернет запрос «пособие по безработице», показывающий интерес населения к выплачиваемым безработным пособиям. Однако, по мнению автора, такой подход определения настроений населения не лишен недостатков. Во-первых, не для всех категорий безработных предусмотрено такое пособие, а значит, индикатор не в полной мере отражает изменения на рынке труда. Помимо этого, использование единственного ключевого слова для прогнозирования уровня фактической безработицы ограничено анализом специфической информации и колебаниями активности пользователей, не связанной с безработицей, например, при устаревании понятия, так как правительство Великобритании в рамках национальной реформы социального обеспечения объединило в системе универсального кредита пособия по безработице с другими льготами, включающими детский налоговый кредит, пособие по нетрудоспособности, жилищное пособие и поддержку малообеспеченных.

GoogR – поисковый запрос по ключевому слову «увольнение работников». GRI – взвешенный индекс поисковых запросов (перечень был представлен выше), относящихся к тематике «увольнения работников». Веса каждого из поисковых запросов на еженедельной основе задаются как:

$$W_{i,t} = \frac{GST_{i,t}}{\sum_{i=0}^{p} GST_{i,t}}, \tag{10}$$

где $W_{i,t}$ – вес -го интернет запроса в момент времени t, $GST_{i,t}$ – i -ый интернет запрос в момент времени t и $\sum_{i=0}^{p} GST_{i,t}$ – суммарный объем данных поиска по p ключевым словам.

Авторами исследования [CITATION DAm17 \l 1033] используется ключевое слово «вакансии» для прогнозирования уровня безработицы в США из предположения, что такой поисковый запрос является наиболее популярным среди ищущих работу, а также подходит для широкого круга соискателей, а значит не чувствителен к шокам спроса или предложения, свойственных различным группам работников.

Подводя итог, важно отметить, что существует обширная литература, посвященная прогнозированию уровня безработицы с помощью интернет данных Google Trends в

развитых и развивающихся странах. Международный опыт показывает, что активность экономических агентов в Интернете, связанная со сбором информации о пособии по безработице, об условиях увольнения работников и поиском работы, дает возможность выявить информацию о будущем изменении уровня безработицы. Таким образом, содержащаяся в индексах поисковых запросов информация обеспечивает более высокую предсказательную способность даже в странах с небольшим уровнем проникновения Интернета. Это в свою очередь дает возможность экономическим властям и инвесторам делать более точный прогноз такого важного показателя делового цикла, как безработица. Возможные преимущества в прогностической способности достигается за счет своевременной публикации временных рядов поисковых запросов Google Trends по сравнению с данными опроса, а также доступностью высокочастотных данных в открытом доступе. Кроме того, данные Google Trends являются недорогостоящим альтернативным источником экономической информации. Кроме того, интернет данные особенно полезны для краткосрочного прогнозирования в периоды рецессии, когда происходят значительные структурные сдвиги в экономике.

1.3.4 Прогнозирование экономических показателей с помощью поисковых запросов в России

При анализе российских данных авторы также полагают, что на основе интернет запросов можно получить опережающие индикаторы, способные заблаговременно выявить нестабильность в экономике.

В работе [CITATION Сто11 \l 1049] проводится анализ влияния поисковых запросов Google Trends, которые, по мнению автора, являются мерой финансовых настроений населения, на депозиты физических лиц с января 2004 г. по июнь 2011 г. для России. Столбов [CITATION Сто11 \l 1049] полагает, что интернет запросы могут быть применены для анализа текущей финансовой конъюнктуры и краткосрочного прогнозирования (nowcasting), поскольку население, прежде чем принять решение о вкладах, осуществляет поиск по наиболее выгодным предложениям. По этой причине существует лаг между поиском информации в Интернете и открытием депозита в банке.

Автор показывает в исследовании, что произошло изменение предпочтений населения в период кризиса 2008-2009 гг. (структурный сдвиг был выявлен в октябре 2008 г., поэтому выборка разбивается на два подпериода: с января 2004 г. по октябрь 2008 г. и с ноября 2008 г. по июнь 2011 г.). Показатель настроений населения строится на основе интернет запроса

Google Trends по ключевому слову «вклады (депозиты)». Из-за существования временного лага между сбором информации и влиянием на экономические показатели в качестве объясняющей переменной берется интенсивность запроса «вклады» за первую неделю текущего месяца.

В период с января 2004 г. по октябрь 2008 г. настроения населения (данные по поисковому запросу «вклады») не оказывали значимого влияния на депозиты физических лиц. Автор связывает этот результат с незначительной аудиторией российских пользователей в поисковой системе Google (интенсивность запроса была нулевой в этот период). Фактически в течение 2004-2008 гг. не был налажен процесс сбора финансовой информации в Интернете.

С ноября 2008 г. по июнь 2011 г. интенсивность поискового запроса «вклады» оказывала понижательное давление на прирост вкладов в российских банках. В целом при повышении интереса населения к «вкладам» происходило снижение депозитов на 2.8 млрд рублей. Автор допускает, что повышение обеспокоенности населения относительно вкладов говорит о возможных проблемах в банковском секторе или возникших факторах неопределенности.

Кроме того, в работе автор попытался получить барометр финансовой конъюнктуры, оцененный с помощью поисковых запросов Google Tends по финансовой тематике. Под интегральным показателем финансовой конъюнктуры в данном случае понимается опережающий индикатор, характеризующий уверенность населения в устойчивости финансовой системы. С этой целью отбираются наиболее важные поисковые запросы в период кризиса 2008-2009 гг. из тематической категории «Финансы и страхование». Поскольку финансовый кризис затронул практически все сектора российской экономики, то настроения населения могут быть полезны при прогнозировании макроэкономических показателей.

Были отобраны следующие поисковые запросы в системе Google Trends: «ЦБ», «банк», «РТС», «курс доллара», «ММВБ», «акции», «девальвация», «дефолт», «финансовый кризис», «евро», «ПИФ», «взять кредит», «залог», «ипотека», «банкротство». В данном случае барометр финансовой конъюнктуры строится как линейная комбинация всех отобранных поисковых запросов с определенными весами:

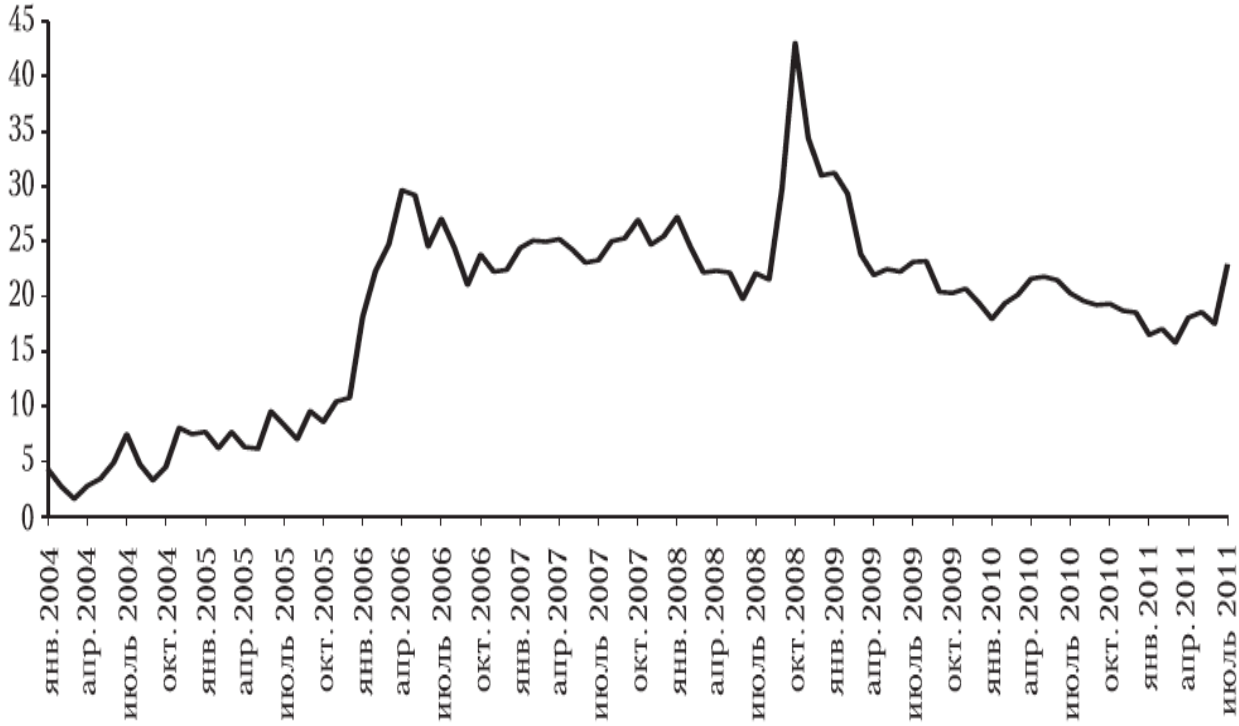
$$Y_t = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_{15} x_{15}, \qquad (11)$$

где Y_t – показатель устойчивости финансовой системы, x_i – i-ый поисковый запрос за месяц, α_i – вес i-ого поискового запроса в барометре финансовой конъюнктуры, которые определяются следующим образом:

$$\alpha_i = \frac{\sum_{i=1}^{14} r_{ij}}{\sum_{i=1}^{14}\sum_{j=1}^{14} r_{ij}}, \quad (12)$$

где $\sum_{i=1}^{14} r_{ij}$ – сумма коэффициентов парной корреляции i-ого поискового запроса с остальными поисковыми запросами, $\sum_{i=1}^{14}\sum_{j=1}^{14} r_{ij}$ – сумма коэффициентов в матрице парной корреляции, характеризующая взаимосвязь между всеми показателями. Автор получает в итоге удельный вес каждого поискового запроса в совокупной величине показателя устойчивости финансовой системы: «ЦБ» (0.016), «банк» (0.049), «РТС» (0.110), «курс доллара» (0.045), «ММВБ» (0.120), «акции» (0.035), «девальвация» (0.067), «дефолт» (0.101), «финансовый кризис» (0.078), «евро» (0.079), «ПИФ» (0.044), «взять кредит» (0.004), «залог» (0.089), «ипотека» (0.069), «банкротство» (0.094). Автор полагает, что при одинаковом изменении всех интернет запросов, наибольшее влияние на показатель уверенности экономических агентов в устойчивости финансовой системы будет оказывать интерес инвесторов к фондовому рынку (фондовые индексы РТС и ММВБ с весами 0.11 и 0.12, соответственно) и ожидания населения относительно будущей неопределенности в экономике (банкротство и дефолт с весами 0.101 и 0.094, соответственно).

Опережающий индикатор уверенности экономических агентов в устойчивости финансовой системы представлен на рисунке . Автор выделяет несколько ключевых изменений в российской финансовой системе. Например, повышение интереса населения к паевым инвестиционным фондам, ипотеке и акциям в феврале 2006 г., период стабильности с мая 2006 г. по август 2008 г. и значительное повышение индикатора в период глобального финансового кризиса (сентябрь 2008 – апрель 2009 г.). Под уверенностью в данном случае понимается снижение интереса к активам или активности по сбору информации в Интернете. В период глобального кризиса уверенность падала, а обеспокоенность населения повышалась и увеличивался интерес к информации о текущем состоянии экономики.



Примечание – Источник: [CITATION Сто11 \l 1049]

Рисунок 4 – Динамика индикатора финансовой конъюнктуры

Период 2010-2011 гг. характеризуется снижением интереса к финансовой тематике. Автор интерпретирует это явление, во-первых, как отсутствие среди экономических агентов ожиданий кризиса, а, во-вторых, как недоверие инвесторов к вложениям в российские финансовые активы.

Кроме того, автор [CITATION Сто11 \l 1049] для подтверждения полученных результатов на первом этапе сравнивает показатель финансовой конъюнктуры с индексом финансовых настроений, рассчитываемый Сбербанком на основе социологических опросов. Корреляция между этими показателями оказалась не слишком высокой (коэффициент корреляции 0.48), но при этом динамика отражает основные изменения в финансовой конъюнктуре. Повышение оптимизма населения (увеличение индекса финансовых настроений Сбербанка) часто совпадает со снижением показателя финансовой конъюнктуры, полученного из поисковых запросов.

На втором этапе для выявления предсказательной способности альтернативных мер доверия населения к финансовой системе при прогнозировании спроса на ипотеку, депозиты и кредиты со стороны физических лиц, изменения активов банковской системы и индекса РТС проводится оценка эконометрических моделей в период с мая 2009 г. по июнь 2011 г. Результаты показывают, что значимое влияние индекс финансовой конъюнктуры, предложенный в работе, оказывает только на фондовый индекс РТС и спрос на кредит.

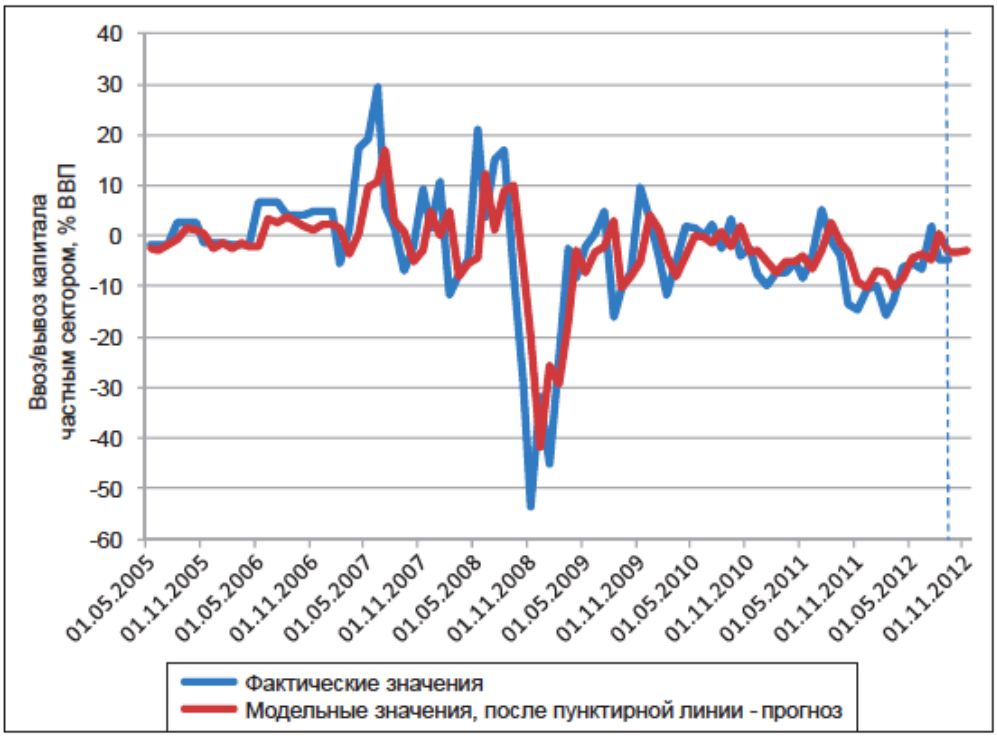
На российских данных также исследовалась возможность прогнозирования оттока капитала с помощью интенсивности поисковых запросов по более общим темам, интерес к которым вызывает значительные изменения в экономике. В работе [CITATION Бор13 \l 1049] было выявлено, что важным фактором при прогнозировании макроэкономических показателей является интенсивность интернет запроса по ключевому слову «кризис». При прогнозировании чистого ввоза/вывоза капитала частным сектором в течение 2005-2012 гг. на квартальных данных используется модель следующего вида:

$$Cap\,F_t = \alpha + \beta_1 Cap\,F_{t-1} - \beta_2 Crisis\,R_{t-1} + \varepsilon_t, \quad (13)$$

где $Cap\,F_t$ – объем чистого ввоза/вывоза капитала (в % ВВП), $Crisis\,R_{t-1}$ – запаздывание интенсивности поискового запроса «кризис» системы Google Trends и ε_t – случайная ошибка.

Результаты свидетельствуют о наличии значимой зависимости между оттоком капитала и обеспокоенностью населения экономическим кризисом в России. Автор отмечает, что включение в модель интенсивности интернет запроса «кризис» дает возможность прогнозировать отток капитала с двухмесячным опережением. Это объясняется тем, что данные по поисковым запросам публикуются с минимальной задержкой в 3 дня, а статистика по чистому ввозу/вывозу капитала частным сектором раскрывается раз в квартал.

Автор проводит графический анализ сравнения прогнозных и фактических значений чистого ввоза/вывоза капитала (см. рисунок). Прогнозные значения хорошо описывают динамику фактических данных о движении капитала. Стоит отметить высокую точность прогноза модели оттока капитала в кризисные периоды 2008-2009 гг.



Примечание – Источник: [CITATION Бор13 \l 1049]

Рисунок 5 – Динамика фактических и прогнозных значений чистого ввоза/вывоза капитала в России

В работе [CITATION Бор13 \l 1049] прогнозируется уровень безработицы для России в течение 2005-2012 гг. на ежемесячных данных. При построении модели для прогнозирования безработицы предполагается, что экономическая активность России сильно зависит от цен на нефть, так как высока доля экспорта сырьевых товаров в ВВП. Совокупный спрос также оказывает влияние на уровень безработицы. Когда происходит спад в экономике, фирмы начинают снижать спрос на труд из-за падения выпуска товаров и услуг, что приводит в конечном счете к росту безработицы. В период экономического кризиса наблюдается снижение совокупного спроса, и рост интереса населения к кризису (повышение интенсивности запросов по ключевому слову «кризис») может служить индикатором ожиданий экономических агентов. Таким образом, оценивается модель вида:

$$Unemp_t = \alpha - \beta_1 Brent_{t-3} - \beta_2 CrisisR_{t-4} + \varepsilon_t, \quad (14)$$

где $Unemp_t$ – уровень безработицы, $Brent_{t-3}$ – цена на нефть марки Brent, $CrisisR_{t-4}$ – запаздывание интенсивности поискового запроса «кризис» системы Google Trends и ε_t – случайная ошибка.

Фактические и прогнозные значения уровня безработицы представлены на рисунке . Эмпирический анализ показал, что обеспокоенность кризисными явлениями и лаг цены на нефть вызывают в последующие три месяца повышение безработицы. Более высокой точностью прогноза модель обладает в период с октября 2008 г. по апрель 2009 г.

Примечание – Источник [CITATION Бор13 \l 1049]

Рисунок 6 – Динамика фактических и прогнозных значения уровня безработицы в России

В целом автор заключает, что статистика по поисковым запросам позволяет повысить точность моделей краткосрочного прогнозирования.

В работе [CITATION Фан18 \l 1049] используются данные поиска Google Trends для краткосрочного прогнозирования показателя социального самочувствия российского населения, публикуемого ВЦИОМ, за период 2009-2016 гг. на квартальных данных с помощью байесовской модели усреднения.

Авторы [CITATION Фан18 \l 1049] выделяют ряд преимуществ использования показателя интенсивности поисковых запросов для анализа благосостояния российского населения. Во-первых, интенсивность интернет запросов в отличие от социологических опросов дает возможность изучать непосредственно поведение россиян и выявить предпочтения и ожидания населения. Во-вторых, данные интернет запросов публикуются с

минимальной задержкой, что представляет интерес для денежных и фискальных властей, принимающих меры социально-экономической политики. В-третьих, данные Google Trends доступны на региональном уровне и отражают широкий спектр запросов, что нехарактерно для показателей социологических опросов, для которых слишком дорогостоящим является процесс достижения репрезентативности на местном уровне.

Отбор групп ключевых слов осуществляется с учетов различных аспектов жизни: материальных условий (экономика страны, финансовая безопасность, рынок труда и поиск работы), физическое и ментальное здоровье (состояние здоровья, поддержка тонуса и здоровые привычки[7], образование, идеалы и ценности, отдых и досуг), социальные условия и окружающая среда (семья, семейные проблемы, личная безопасность, политическая обстановка и действия государства, экология). Используемые данные Google Trends нормализуются (z-оценка), и исключается сезонность. В результате в анализе авторы выделили 382 ключевых слова. Авторы проводят факторный анализ для построения индексов благосостояния по категориям из всего набора данных интернет запросов.

В работе [CITATION Фан18 \l 1049] для исследования предсказательной способности поисковых запросов в прогнозировании индексов социального самочувствия ВЦИОМ используется модель байесовского усреднения вида:

$$y = \alpha_\gamma + \beta_\gamma X_\gamma + \epsilon \quad , \tag{15}$$

где y – индекс социального самочувствия (индекс самооценок материального положения, индекс удовлетворенности жизнью, индекс социального оптимизма, индекс оценок экономической ситуации, индекс оценок общего вектора развития страны), X_γ – подмножество категорий поисковых запросов и $\epsilon \ N(0, \sigma^2)$ – случайная ошибка.

Модель байесовского усреднения [CITATION Фан18 \l 1049] решает проблему «проклятья размерности» с помощью апостериорной модели вероятностей, основанной на байесовском выводе, где априорное распределение данных уточняется при поступлении новых данных следующим образом:

$$PMP_y = p(M_y | y, X) = \frac{p(y | M_y, X)\, p(M_y)}{p(y | X)} \quad , \tag{16}$$

7 Категории отбора поисковых запросов определены авторами данного исследования на основе опроса ВЦИОМ.

где $p(y|X)$ – зависит от модели и является константой, $p(M_y)$ – задаваемое экспертным путем априорное распределение (в данном случае равномерное распределение), $p(y|M_y, X)$ – условная вероятность данных при заданных параметрах модели, M_y – модель с подмножеством регрессоров X_γ (всего возможно построение 2^K моделей, поскольку, если X содержит K регрессоров, то требуется оценить 2^K комбинаций объясняющих переменных).

В итоге апостериорные вероятности усредняются для получения апостериорных вероятностей включения переменных в модель. Анализ таких вероятностей показывает, насколько важны регрессоры для прогнозирования различных индексов социального самочувствия. Критерием отбора моделей служит предположение о превышении апостериорных вероятностей значения 0.7. Авторы отмечают, что такой подход не требует значительных вычислительных мощностей, так как не подсчитывает предельные вероятности и апостериорное распределение для всех моделей, а идентифицирует лишь подмножество моделей с большим весом апостериорной вероятности на основе алгоритма Метрополиса-Гастингса. [CITATION Met \l 1033], [CITATION Has70 \l 1033]. Авторы применили модификацию этого алгоритма – «birth-death». Такой алгоритм позволяет случайно выбирать одну из K переменных на каждом шаге и на последующем шаге исключать уже имеющиеся в модели переменные.

На основании полученных результатов авторы делают следующие выводы. Материальное положение российского населения зависит от состояния рынка труда и способности найти работу, отдыха, личной безопасности, поддержания тонуса и здоровых привычек, образования, идеалов и ценностей. Удовлетворенность жизнью населения зависит от поисковых запросов, связанных с тематикой семьи, поиска работы и рынка труда, уровня образования, личной безопасности, отдыха, политической обстановки и действий государства и здоровых привычек. Социальный оптимизм формируется на основе ожиданий и предпочтений относительно рынка труда, семьи, здоровых привычек, семейных проблем и финансовой безопасности.

Проверка достоверности полученных выводов осуществляется с помощью регрессионных моделей. В качестве объясняющих переменных использовались категории поисковых запросов Google Trends, отобранных в модели байесовского усреднения. Повышение интереса к рынку труда приводило к ухудшению материального положения, социального оптимизма и снижению удовлетворенности жизнью. При этом запросы, связанные с семьей, оказывают положительное влияние на индекс самооценок материального положения, индекс социального оптимизма и индекс удовлетворенности жизнью.

Высокий уровень личной безопасности приводит к улучшению текущей экономической обстановки. Увеличение обеспокоенности населения относительно поиска работы говорит об ухудшении экономической ситуации. Снижение интереса к тематикам «семья» и «здоровье» сигнализирует о большей уверенности в будущем страны и повышении индекса оценок общего вектора развития страны.

Подводя итог, можно утверждать, что эмпирические исследования на российских данных указывают на то, что интенсивность поисковых запросов Google Trends повышает точность краткосрочных прогнозов по сравнению с эталонными моделями. Данные по поисковым запросам Google Trends содержат ценную информацию, характеризующую предпочтения и интерес населения к определенной теме. Однако необходимо отметить, что существует недостаточно исследований, изучающих предсказательную способность поисковых запросов в прогнозировании макроэкономических и финансовых показателей на российских данных.

1.4 **Преимущества и недостатки методов машинного обучения при обработке больших объемов данных**

Как было показано в предыдущем подразделе, большинство эконометрических методов прогнозирования подходят лишь для случаев малой размерности, когда число объясняющих переменных в разы меньше числа наблюдений. В цифровую эпоху за счет новых информационных технологий стало возможным собирать и обрабатывать большие объемы интернет данных. Методы машинного обучения в целом позволяют решить основные проблемы традиционных статистических методов [CITATION Has09 \l 1033]: переобучения на обучающей выборке, мультиколлинеарности и выбора между смещением и дисперсией.

Проблема переобучения появляется при идеальной аппроксимации данных применяемой моделью. В итоге при тестировании подобранной модели на новых данных, имеющих совсем другую структуру, получается большая ошибка прогноза. При большом количестве предикторов также может произойти переобучение из-за оценки множества параметров. Тесно связанной с идеальной аппроксимацией данных проблемой является выбор между смещением и дисперсией. Например, метод наименьших квадратов обеспечивает минимальное смещение оцененных параметров, но при этом имеет более высокую дисперсию оценок, что также может стать причиной незначительной предсказательной способности модели.

Наличие линейной зависимости объясняющих переменных (частичной мультиколлинеарности) может стать причиной неустойчивости оцененных параметров, повышения дисперсии оценок и возможного несоответствующего экономической теории знака перед объясняющей переменной. Более того, размерность предикторов может быть завышена искусственно[8], а для эмпирического анализа требуется лишь небольшая часть признаков.

На основе результатов международных и российских исследований можно сделать вывод о том, что методы отбора (лассо, гребневая регрессия, эластичная сеть [CITATION Zou05 \l 1033], модели динамического отбора и прочие байесовские модели [CITATION Koo13 \l 1033]-[CITATION Sco13 \l 1033]) и выделения признаков (метод главных компонент и его модификации и другие факторные модели) успешно справляются с вышеописанными проблемами.

Однако модели с регуляризацией и стандартные факторные модели не лишены недостатков. Во-первых, метод лассо при высоких значениях коэффициента перед штрафом может обнулять большое количество объясняющих переменных, что снижает предсказательную способность модели. Для решения этой проблемы используется метод наименьших углов [CITATION Efr04 \l 1033] или эластичная сеть [CITATION Zou05 \l 1033] (с взвешенными штрафами лассо и гребневой регрессии).

Во-вторых, стандартные факторные модели, такие как метод главных компонент, не способны выделить интерпретируемые экономически признаки. Это следует из основной идеи метода главных компонент, когда мы получаем линейную комбинацию всех возможных объясняющих переменных. В эмпирической литературе для получения более интерпретируемых факторов [CITATION Zou06 \l 1033] применяются кластеризация с последующим выделением для каждого кластера главной компоненты, регрессия на главные компоненты с регуляризацией, метод главных компонент с нелинейным ядром и автокодировщики.

В-третьих, модели отбора и выделения признаков не учитывают изменение оцениваемых коэффициентов во времени, поэтому некоторые авторы используют модели динамического выбора моделей [CITATION Koo13 \l 1049] или динамическую факторную модель [CITATION Sto12 \l 1033].

8 При анализе больших данных, например, если важно учесть более 100 экономических факторов для прогнозирования инфляции, как в работе Стока и Уотсона, может возникнуть ситуация мультиколлинеарности. Решением такой проблемы является использование факторных методов, когда мы можем без значительных потерь снизить размерность и учесть информацию из данных для построения прогноза временного ряда.

Подводя итог, отметим, что остается актуальным проведение исследования с более устойчивыми методами, использование которых не так распространено на текущий момент при прогнозирования макроэкономических показателей с помощью поисковых запросов.

Результаты эмпирических исследований показывают, что интернет запросы являются важным источником информации при прогнозировании различных макроэкономических показателей. Это связано с тем, что интернет данные могут быть полезными в исследовании поведения экономических агентов по широкому кругу социально-экономических вопросов за счет своевременного получения информации об интересах и предпочтениях населения.

Повышение точности прогнозов инфляции и безработицы, сделанных на основе поисковых запросов, достигается вследствие пересмотра ожиданий населения, когда растет обеспокоенность относительно текущей экономической ситуации, что и стимулирует экономических агентов к поиску и сбору информации в Интернете. Таким образом, содержащаяся в индексах поисковых запросов информация обеспечивает более высокую предсказательную способность даже в странах с небольшим уровнем проникновения Интернета.

В исследованиях, посвященных финансовым рынкам, предполагается, что интернет запросы могут быть использованы в качестве меры настроений и ожиданий экономических агентов, поскольку отражают уровень оптимизма на финансовых рынках. Более того, информация из открытых источников может содержать некоторую общедоступную информацию, которая не отразилась в ценах рыночных активов. Поиск информации в Интернете может сделать инвесторов более информированными о трендах на фондовом и валютном рынках.

Особую роль поисковые запросы играют при прогнозировании экономических показателей в период финансового кризиса. Обеспокоенность экономических агентов проблемами финансового кризиса или рецессии вызывали рост безработицы, оттока капитала и высокой волатильности валютных курсов на краткосрочном горизонте прогноза.

На основе обзора международных и российских исследований были определены наиболее подходящие для эмпирического анализа методы отбора и выделения признаков при большом объеме данных: лассо, гребневая регрессия, эластичная сеть, случайный лес, градиентный бустинг и метод главных компонент.

На наш взгляд, несмотря на значительное число исследований в этой области, задача прогнозирования макроэкономических показателей с помощью интернет данных остается актуальной для России. Применение современных и более устойчивых методов машинного обучения для анализа больших данных даст возможность исследовать структуру поведения

пользователей в Интернете и учесть при прогнозировании экономических показателей ожидания и настроения населения. В этой связи в следующих разделах будут представлены результаты исследования предсказательной способности поисковых запросов в прогнозировании основных макроэкономических показателей в России.

2 Прогнозирование макроэкономических показателей на основе интернет-запросов на российских данных

В данном разделе будут рассмотрены основные результаты эмпирического исследования предсказательной способности интенсивности интернет-запросов в прогнозировании таких макроэкономических показателей, как инфляция, уровень безработицы, курс рубля к доллару и реальные темпы роста ВВП с помощью методов машинного обучения и эконометрических моделей.

2.1 Построение прогнозов инфляции и безработицы с применением данных интернет-запросов на основе методов машинного обучения

В рамках данного исследования для учета всей информации из поисковых запросов и решения проблемы «проклятья размерности», а также отбора ключевых запросов из заданных тематических групп, как и в работе Байбузы при прогнозировании инфляции на основе больших данных по широкому кругу макроэкономических показателей[CITATION Бай18 \l 1049], для прогнозирования инфляции, уровня безработицы и курса рубля к доллару в период с января 2004 г. по июль 2019 г. были использованы следующие методы машинного обучения для анализа больших данных: лассо, метод наименьших углов, гребневая регрессия, эластичная сеть, случайный лес, градиентный бустинг и линейная модель с главными компонентами.

Метод лассо [CITATION Tib96 \l 1049] предполагает включение в модель в модель L1-штрафа, накладывающего ограничения на абсолютные значения коэффициентов и позволяющего получить разреженную матрицу объясняющих переменных (исключить из модели факторы с небольшой предсказательной способностью). Оптимизационная функция со штрафом для модели лассо имеет вид:

$$\min_{\beta} \frac{1}{2n} \left\| y - X\beta \right\|_2 + \alpha \left\| \beta \right\|_1 \quad , \tag{17}$$

где y – исходный ряд или сезонно-скорректированная инфляция к предыдущему месяцу, X – поисковые запросы Google Trends, α – коэффициент регуляризации, подбираемый на кросс-валидации на скользящем окне.

LARS [CITATION Efr04 \l 1049] (метод наименьших углов): метод выбора такого набора факторов, который имел бы наиболее значимую статистическую связь с зависимой переменной.

$$S(\beta)=|y-\mu|^2 \tag{18}$$

LARS на каждом шаге строит оценку коэффициентов $\hat{\beta}$. Далее алгоритм вычисляет приближение вектора значений зависимой переменной $\mu=X\beta$. Для приближения используется вектор корреляций столбцов X с регрессионными остатками:

$$c(\mu)=X^T(y-\mu) \tag{19}$$

На k-ом шаге приближенное значение y вычисляется:

$$\hat{\mu}_k=\hat{\mu}_{k-1}+\gamma_k u_k \tag{20}$$

где u_k - нормированный вектор, задающий биссектрису между добавляемым вектором-столбцом матрицы X и вектором остатков.

Таким образом, веса подбираются так, чтобы добиться максимальной корреляции объясняющей переменной с регрессионными остатками.

Гребневая регрессия определяется следующим образом:

$$\min_{\beta}\frac{1}{2n}\|y-X\beta\|_2+\alpha\|\beta\|_2 \tag{21}$$

Введение в линейную модель L2-штрафа накладывает ограничения на оцениваемые параметры так, что они могут принимать большие значения при пропорциональном снижении среднеквадратичной ошибки. Как и лассо происходит сжатие коэффициентов к нулю, но при этом веса не зануляются, а лишь приближаются к нулю. Кроме того, гребневая регрессия позволяет учитывать взаимную информацию из коррелирующих факторов, в то время как лассо отбирает один из них.

Эластичная сеть [CITATION Zou05 \l 1049] представляет собой линейную комбинацию L_1-регуляризатора (лассо) и L_2-регуляризатора (гребневой регрессии) и имеет целевую функцию:

$$\min_{\beta} \frac{1}{2n} \|y - X\beta\|_2 + \alpha\rho \|\beta\|_1 + \frac{1}{2}\alpha(1-\rho)\|\beta\|_2 \quad , \tag{22}$$

где α и ρ выбираются с помощью кросс-валидации на скользящем окне.

Случайный лес [CITATION Bre01 \l 1033] представляет собой ансамблевый алгоритм с решающими деревьями и призван снизить дисперсию прогноза, а также решить проблему переобучения базовой модели. Для этого на основе фактической выборки X генерируется N искусственных подвыборок длины исходной выборки. В искусственную подвыборку входят не все признаки, а только случайный набор. Далее по каждой получившейся искусственной выборке строится регрессионное решающее дерево $b_i(x)$.

Итоговым ответом алгоритма является среднее значение результатов построения отдельных решающих деревьев:

$$a(x) = \frac{1}{N}\sum_{i=1}^{N} b_i(x) \tag{23}$$

Градиентный бустинг [CITATION Fri01 \l 1033] представляет собой следующую модель. На 1 шаге оценивается дерево решений на всей выборке

$$b_1(x) = arg\min_{b} \sum_{i=1}^{l} \left(b(x_i) - y_i\right)^2 \tag{24}$$

где $b_1(x)$ – обученная на 1 шаге модель.

Следующая модель на 2 шаге дообучается на остатках, полученных на предыдущем этапе. Итоговый ансамбль моделей принимает вид

$$a(x) = \sum_{i=1}^{N} \gamma_i b_i(x) \tag{25}$$

где $a(x)$ – итоговый прогноз ансамбля из 100 регрессионных деревьев.

Нельзя сделать однозначного вывода какая из моделей с регуляризацией или ансамблевые алгоритмы будут давать прогноз лучше для различных типов данных. Однако часто предполагается, что ансамблевые методы хорошо работают при нелинейных связях между факторами и объясняющей переменной.

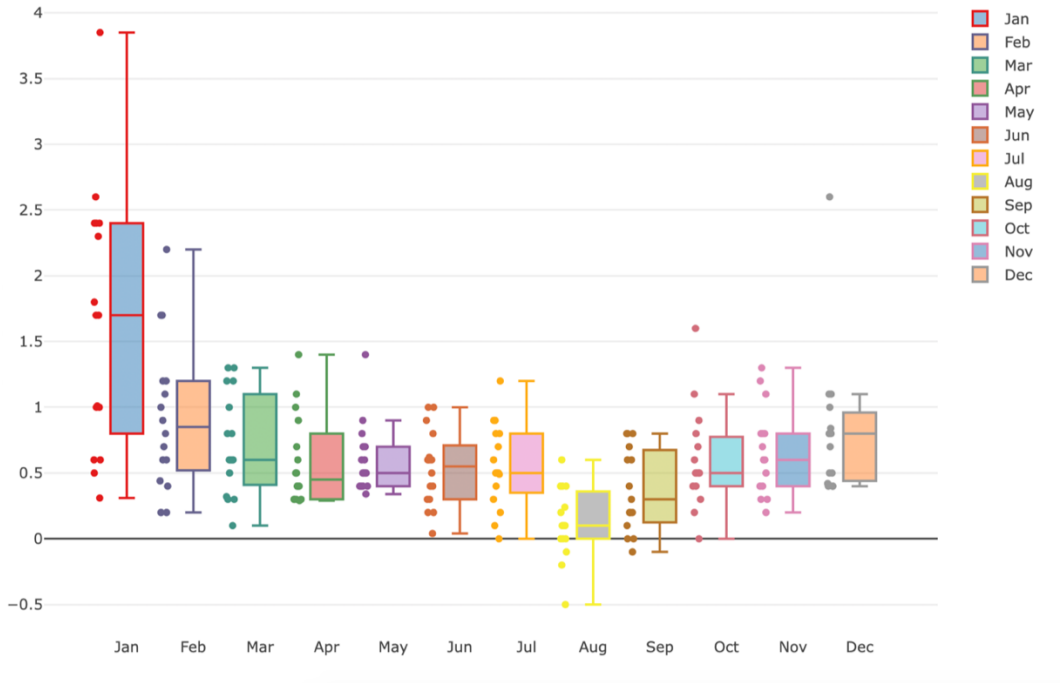
В рамках эмпирического анализа используется 75 поисковых запроса для инфляции, 48 интернет-запроса для безработицы и 60 поисковых запроса для курса рубля (см. таблицу), связанных с финансовыми рынками, интересом населения к текущей экономической ситуации и инфляционными ожиданиями.

Таблица 4 – Интернет запросы для инфляции и безработицы

Макроэкономический показатель	Перечень поисковых запросов
Инфляция	курс доллара, курс рубля, евро, ММВБ, акция, девальвация, санкции, ОФЗ, цена на нефть, brent, котировка, московская биржа, курс валют, нефть, рост доллара, девальвация рубля, падение доллара, доллар, банк, ЦБ, взять кредит, залог, дефолт, зарплата, импорт, экспорт, рецессия, сбережения, облигации, трудовое законодательство, банкомат, налог, экономика России, процентная ставка, ндфл, ипотека, банкротство, Сбербанк, ВТБ, деньги, оклад, безработица, инфляция, продажа, купить, розничная торговля, интернет-магазин, смартфон, цена на газ, сеть магазинов, компьютер, услуги, оборудование, обслуживание, страхование, транспорт, осаго, кино, квартира, авиабилет, билет на поезд, билет, аренда, товар, косметика, отель, ресторан, кафе, овощи, цены на квартиры, ЖКХ, дизель, бензин, рост цен, повышение цен
Безработица	финансовый кризис, экономический кризис, банк, взять кредит, залог, экономический рост, счет, сбережения, работа, вакансии, hh.ru, депозит, увольнение, налог, ВВП, экономика, экономика России, социальный пакет, ндфл, компенсация, потребление, доход, премия, вознаграждение, ипотека, банкротство, центр занятости, пенсия, расходы, вклад, биржа труда, пособие по безработице, заработная плата, безработица, оплата труда, уровень инфляции, индекс потребительских цен, индекс цен, инфляция, кредит онлайн, алкоголь, рост инфляции, сигареты, цены, квартира, автомобиль, аренда, цены на квартиры
Курс рубля	курс доллара, курс рубля, евро, ЦБ, РТС, ММВБ, акция, взять кредит, банкротство, залог, дефолт, финансовый кризис, санкции, США, Китай, Сбербанк, ОФЗ, цены на квартиры, экспорт, фондовый индекс, форекс, цена на нефть, фьючерсы, котировка, московская биржа, квартира, кредит онлайн, курс валют, ведомости, индекс цен, тасс, экономический рост, падение доллара, инфляция, экономический кризис, доходность, счет, падение, валютный рынок, ЖКХ, продукты питания электроэнергия, инвестор, денежно-кредитная политика, ФРС, инвестиции, купить, потратить, сбережения, банкомат, налог, облигации, рост цен, повышение цен, цена, цены, индекс потребительских цен, ВВП, экономика, экономика России

Примечание – Источник: составлено авторами

Временной ряд инфляции имеет выраженную сезонность (рис.). По этой причине в данном исследовании строятся модели двух типов: с включением сезонных дамми-переменных для исходного ряда или сезонно-скорректированной инфляцией к предыдущему месяцу.



Примечание – Источник: расчеты авторов

Рисунок 7 – Временной ряд инфляции к предыдущему месяцу

Далее рассмотрим полученные результаты применения различных методов машинного обучения и выберем наилучшую модель по минимальному корню из среднеквадратичной ошибки прогноза для обоих показателей инфляции, уровня безработицы и курса рубля к доллару.

Результаты оценки точности прогноза моделей для исходного временного ряда и сезонно-скорректированной инфляции представлены в таблицах -, соответственно. В качестве эталонной модели рассматривается наивный прогноз. Вневыборочный прогноз в псевдореальном времени строился с июля 2016 по июль 2019 г. В последнем столбце таблиц и представлены результаты проведенного модифицированного теста Diebold-Mariano на одинаковую предсказательную способность [CITATION Har97 \l 1033] для сравнения качества моделей наивного прогноза и наилучших по RMSE из моделей машинного обучения. Как видно из таблицы , линейная модель с 5 главными компонентами имеет более высокую предсказательную способность (отношение RMSE меньше 1) по сравнению с

эталонной моделью на горизонтах прогноза 1-12 месяцев. Если в качестве критерия качества в рассмотрение брать только отношение RMSE, то можно увидеть, что для прогноза на следующий месяц наилучшей моделью является лассо и эластичная сеть. Однако, модифицированный тест Diebold-Mariano показал, что нет значимого отличия в предсказательной способности эталонной модели и лассо/эластичной сети. По этой причине в качестве наилучшей модели берется линейная модель с 5 главными компонентами, для которой гипотеза отвергается на 5 % уровне значимости.

Более точный прогноз инфляции 5 месяца получается на основе модели лассо, которая оказывается по качеству прогноза примерно такой же, как и эластичная сеть. Градиентный бустинг показывает хороший результат при прогнозе на 6 и 12 месяца.

Таблица 5 – Отношение RMSFE моделей с поисковыми запросами к наивному прогнозу для инфляции

Шаг	AR	Случайный лес	Градиентный бустинг	Лассо	lars	enet	ridge	PCA	D-M статистика
h=1	1.00	1.25	1.15	0.68	0.89	0.70	0.88	**0.79**	2.16**
h=2	1.00	0.99	0.75	1.18	1.20	1.18	0.83	**0.68**	4.38***
h=3	1.00	1.23	1.05	0.66	0.98	0.76	0.87	**0.63**	5.32***
h=4	1.00	1.18	0.94	0.74	0.85	0.76	1.11	**0.71**	6.30***
h=5	1.00	0.78	1.03	**0.53**	0.80	0.53	1.03	0.73	3.83***
h=6	1.00	0.96	**0.64**[9]	1.62	0.81	1.74	1.54	0.76	3.31***
h=7	1.00	1.06	0.80	1.65	0.96	1.74	1.55	**0.76**	5.36***
h=8	1.00	1.33	1.27	1.23	0.92	1.36	1.95	**0.73**	4.84***
h=9	1.00	1.00	1.13	0.73	0.94	0.79	1.70	**0.72**	5.29***
h=10	1.00	0.95	0.92	1.08	1.04	1.08	1.50	**0.66**	6.51***
h=11	1.00	0.66	0.53	1.03	0.99	1.03	1.34	**0.66**	4.18***
h=12	1.00	0.53	**0.49**[10]	0.94	0.91	0.94	1.03	0.71	4.30***

Примечание – D-M статистика – t-статистика модифицированного теста Диболда-Мариано для проверки гипотезы на одинаковую предсказательную способность. Lars – метод наименьших углов, enet – эластичная сеть, ridge – гребневая регрессия, PCA – линейная модель с главными компонентами. Уровень значимости «***» – на 1 %, «**» – на 5% и «*» – на 10%. Источник: расчеты авторов

Для сезонно-скорректированной инфляции наилучшей по качеству прогноза (см. таблица) на всех горизонтах оказалась линейная модель с главными компонентами. Результаты показывают, что при прогнозе на 1-12 месяцев гребневая регрессия имела наихудшую предсказательную способность из-за проблем с переобучением. Аналогичная

9 Градиентный бустинг и линейная модель с главными компонентами имеют одинаковую предсказательную способность.

10 Случайный лес и градиентный бустинг имеют одинаковую предсказательную способность.

проблема наблюдалась и для других моделей с регуляризацией: лассо, метода наименьших углов и эластичной сети. Случайный лес и градиентный бустинг также были склоны к переобучению на горизонтах прогноза на 1-6 месяца.

Таблица 6 – Отношение RMSFE моделей с поисковыми запросами к наивному прогнозу для сезонно-скорректированной инфляции

Шаг	AR	Случайный лес	Градиентный бустинг	Лассо	lars	enet	ridge	PCA	D-M статистика
h=1	1.00	1.35	1.63	1.44	1.39	1.51	2.02	**0.78**	2.34**
h=2	1.00	1.12	1.43	1.51	1.16	1.51	1.41	**0.56**	6.84***
h=3	1.00	2.77	2.47	1.21	1.23	1.20	1.27	**0.52**	8.57***
h=4	1.00	1.53	1.37	1.21	1.17	1.20	1.57	**0.58**	11.19***
h=5	1.00	1.13	1.29	1.11	1.13	1.11	1.05	**0.61**	11.66***
h=6	1.00	1.22	1.48	1.09	1.09	1.09	1.43	**0.66**	11.21***
h=7	1.00	0.66	0.78	1.11	1.11	1.11	1.73	**0.62**	9.71***
h=8	1.00	0.96	0.82	1.10	1.10	1.10	1.65	**0.57**	10.02***
h=9	1.00	0.95	1.16	0.99	1.12	1.00	1.54	**0.59**	12.25***
h=10	1.00	0.94	0.86	1.13	1.13	1.13	1.61	**0.58**	10.91***
h=11	1.00	**0.43**[11]	0.44	1.02	1.02	1.02	1.39	0.51	10.80***
h=12	1.00	0.51	0.52	0.89	0.89	0.89	1.22	**0.50**	12.08***

Примечание – D-M статистика – t-статистика модифицированного теста Диболда-Мариано для проверки гипотезы на одинаковую предсказательную способность. Lars – метод наименьших углов, enet – эластичная сеть, ridge – гребневая регрессия, PCA – линейная модель с главными компонентами. Уровень значимости «***» – на 1 %, «**» – на 5% и «*» – на 10%. Источник: расчеты авторов

Таким образом, линейная модель с главными компонентами является наилучшей для прогнозирования исходного временного ряда и сезонно-скорректированной инфляции. Тест Диболда-Мариано подтверждает полученные результаты, показывая, что предсказательная сила линейной модели с 5 главными компонентами всегда выше эталонной модели и не хуже, чем у отобранных моделей по минимальному отношению RMSFE.

Точность моделей прогнозирования уровня безработицы представлена в таблице . На горизонтах прогноза на 1-7 месяцев более точный прогноз дает линейная модель с 5 главными компонентами. Прогноз на следующие 8 месяцев оказывается лучше у метода наименьших углов, что в целом подтверждается при проведении модифицированного теста

11 Дополнительно были проведены тесты Harvey и др. для прогноза на 11 месяц гипотеза об одинаковой предсказательной способности не была отвергнута для случайного леса, градиентного бустинга и линейной модели с главными компонентами.

Диболда-Мариано на одинаковую предсказательную способность. Метод наименьших углов превзошел по качеству прогноза линейную модель с главными компонентами. Наилучший прогноз по критерию отношения RMSFE на 9-10 месяцев стала эластичная сеть. Однако дополнительно проведенный тест Диболда-Мариано на сравнение точности прогноза на 10 месяцев лассо, эластичной сети и линейной модели с главными компонентами показал, что нет значимого различия в точности прогноза этих моделей.

Таблица 7 – Отношение RMSFE моделей с поисковыми запросами к наивному прогнозу для уровня безработицы

Шаг	AR	Случайный лес	Градиентный бустинг	Лассо	lars	enet	ridge	PCA	D-M статистика
h=1	1.00	2.13	2.14	1.05	1.63	1.20	2.24	**0.80**	3.97***
h=2	1.00	1.90	1.98	0.66	1.81	0.63	0.87	**0.60**	4.36***
h=3	1.00	1.50	1.30	0.70	0.74	0.64	0.71	**0.50**	4.25***
h=4	1.00	0.91	0.86	0.87	0.47	0.80	0.67	**0.43**	6.85***
h=5	1.00	0.73	0.74	1.16	0.56	1.12	0.89	**0.38**	7.66***
h=6	1.00	0.87	0.81	1.57	0.40	1.58	1.09	**0.36**	7.92***
h=7	1.00	0.88	0.84	1.04	1.24	0.99	1.02	**0.41**	8.52***
h=8	1.00	1.05	0.92	0.92	**0.40**	0.87	0.95	0.55	8.62***
h=9	1.00	0.98	1.01	0.68	2.85	**0.65**	0.79	0.75	5.54***
h=10	1.00	1.10	1.05	0.58	1.36	**0.53**[12]	0.84	0.79	8.94***
h=11	1.00	1.16	1.12	**0.48**	1.51	0.52	0.86	0.68	11.31***
h=12	1.00	1.10	1.04	0.59	0.78	0.65	0.71	**0.58**	12.44***

Примечание – D-M статистика – t-статистика модифицированного теста Диболда-Мариано для проверки гипотезы на одинаковую предсказательную способность. Lars – метод наименьших углов, enet – эластичная сеть, ridge – гребневая регрессия, PCA – линейная модель с главными компонентами. Уровень значимости «***» – на 1 %, «**» – на 5% и «*» – на 10%. Источник: расчеты авторов

Результаты сравнительного анализа по отношению RMSFE наивного прогноза и методов машинного обучения при прогнозировании курса рубля к доллару показаны в таблице .

12 Одинаковая предсказательная способность у моделей лассо, эластичная сеть и линейная модель с главными компонентами

Таблица 8 – Отношение RMSFE моделей с поисковыми запросами к наивному прогнозу для курса рубля

Шаг	AR	Случайный лес	Градиентный бустинг	Лассо	enet	ridge	PCA	D-M статистика
h=1	**1.00**	2.08	2.02	1.58	1.33	1.15	1.19	-1.27
h=2	1.00	2.02	2.19	1.41	1.07	1.27	**0.97**	1.14
h=3	1.00	1.57	1.95	1.29	1.00	2.88	**0.92**	2.56**
h=4	1.00	1.38	1.93	0.99	0.87	1.48	**0.83**	2.75***
h=5	1.00	1.35	1.20	0.90	0.87	1.68	**0.82**	3.01***
h=6	1.00	1.77	1.73	0.88	0.85	1.58	**0.83**	2.76***
h=7	1.00	1.07	1.68	0.92	0.88	1.39	**0.82**	2.13**
h=8	1.00	0.91	1.14	1.50	1.03	1.14	**0.85**	1.79*
h=9	1.00	0.77	1.05	1.28	0.97	1.02	**0.84**	1.86*
h=10	1.00	0.71	**0.72**	1.16	0.95	0.95	0.84	1.75*
h=11	1.00	0.81	**0.65**	1.23	1.00	1.04	0.86	2.38**
h=12	1.00	0.87	**0.87**	1.36	1.07	1.00	0.95	1.11

Примечание – D-M статистика – t-статистика модифицированного теста Диболда-Мариано для проверки гипотезы на одинаковую предсказательную способность модели наивного прогноза и наилучшей модели с минимальной RMSFE. enet – эластичная сеть, ridge – гребневая регрессия, PCA – линейная модель с главными компонентами. Уровень значимости «***» – на 1 %, «**» – на 5% и «*» – на 10%. Источник: расчеты авторов

Результаты показывают, что значимым преимуществом в предсказательной силе линейной модели с 3 главными компонентами можно наблюдать только на горизонтах прогноза на следующие 3-9 месяцев. Повышение точности прогноза на следующие 10-11 месяцев наблюдалось в модели градиентного бустинга, что согласуется с результатами модифицированного теста Диболда-Мариано на одинаковую предсказательную способность. На всех горизонтах наихудший прогноз был получен для метода наименьших углов и гребневой регрессии.

Кроме того, проводился сравнительный анализ предсказательной способности моделей с поисковыми запросами и прогнозами модели ARIMA Института экономической политики им. Е.Т. Гайдара[13]. Результаты тестов на одинаковую предсказательную способность приведены для исходного ряда инфляции и курса рубля к доллару в таблице .

13 Научный вестник ИЭП им. Гайдара.ру

Таблица 9 – Сравнение модели ARIMA и линейной модели с главными компонентами с интернет-запросами

	Сравнение моделей	D-M статистика	p_value
Инфляция	h=1 ARIMA против PCA	1.545	0.131
	h=2 ARIMA против PCA	2.910	0.006
	h=3 ARIMA против PCA	2.550	0.015
	h=4 ARIMA против PCA	1.465	0.152
	h=5 ARIMA против lasso	2.142	0.040
	h=6 ARIMA против градиентного бустинга	1.479	0.149
	h=7 ARIMA против PCA	1.396	0.173
	h=8 ARIMA против PCA	1.667	0.106
Курс рубля к доллару	h=1 ARIMA против PCA	-0.615	0.543
	h=2 ARIMA против PCA	2.346	0.025
	h=3 ARIMA против PCA	2.529	0.016
	h=4 ARIMA против PCA	2.251	0.031
	h=5 ARIMA против PCA	2.930	0.006
	h=8 ARIMA против PCA	1.381	0.178

Примечание – Источник: по расчетам авторов

Для инфляции сравнительный анализ с помощью теста Harvey [CITATION Har97 \l 1049] на одинаковую предсказательную способность показывает, что более высокая точность методов машинного обучения достигается при прогнозировании на следующие 1, 2 и 4 месяца. Для остальных горизонтов прогноза не было выявлено значимого различия.

Для курса рубля к доллару более высокая точность прогноза у линейной модели с главными компонентами наблюдается для прогнозов на следующие 2-5 месяцев. В остальных случаях гипотеза об одинаковой предсказательной способности не отвергается.

В данном разделе были рассмотрены различные методы отбора переменных и метод главных компонент, позволяющий выделять из большого количества переменных наиболее важную информацию. Как показали результаты, выделение признаков из большого числа интернет-запросов работает наилучшим образом для построения прогноза. В этой связи в следующем разделе рассмотрим результаты оценки структурных моделей и проведем сравнение моделей с фундаментальными факторами и ожиданиями экономических агентов, полученных методом главных компонент из поисковых запросов.

2.2 Сравнение качества традиционных моделей и использующих интернет-запросы для российских данных

Как и в работе [CITATION Dyb17 \l 1049], прогнозирование номинального курса рубля к доллару осуществляется с помощью структурной векторной авторегрессионной модели, построенной из предположений экономической теории по степени эндогенности переменных в период с января 2010 г. по июль 2019 г. Однако рассмотрение исключительно фундаментальных переменных в качестве детерминант валютного курса не позволяет учесть поведенческие характеристики экономических агентов. Как обсуждалось ранее, альтернативной спецификацией может стать модель с поисковыми, включающая как фундаментальные факторы, так и настроения экономических агентов на кредитном рынке, инфляционные ожидания и настроения на финансовом рынке[14]. Показатели настроений и ожиданий инвесторов извлекаются с помощью метода главных компонент из поисковых запросов, представленных в таблице .

Таблица 10 – Перечень поисковых запросов для трех типов ожиданий в структурной векторной авторегрессионной модели

Ожидания на кредитном рынке	Ожидания на финансовом рынке	Инфляционные ожидания
банк, банкротство, Сбербанк, деньги, кредит онлайн, вклад	курс доллара, курс рубля, ЦБ, ММВБ, дефолт, девальвация, санкции, цена на нефть, brent, московская биржа, курс валют, рецессия, инвестор, нефть, рост доллара, девальвация рубля, процентная ставка, падение доллара, доллар, рбк, тасс	инфляция, авито, продажа, купить, смартфон, алиэкспресс, аренда, ресторан, газ, электричество

Примечание – Источник: расчеты авторов

Информационное множество для кредитного рынка состоит из интернет запросов, связанных с крупнейшим российским коммерческим банком – Сбербанком, коммерческими банками и банковскими услугами (кредит и вклад). Инфляционные ожидания идентифицируются из поисковых запросов по тематике «инфляция». Настроения на финансовом рынке оцениваются исходя из предположения, что поисковые запросы, содержащие информацию об интересе инвесторов к российскому фондовому и валютному рынку. Ожидания на финансовом рынке в таком случае должны быть тесно связаны с

14 Интернет-запросы отобраны по критериям значимой корреляции с курсом рубля и согласно ранее проведенному анализу с помощью методов машинного обучения для инфляции.

выбором инвестиционной стратегии розничным инвестором и отражать выявленные ожидания относительно будущих колебаний цен финансовых активов или валютных курсов.

Для прогнозирования курса рубля к доллару используется структурная векторная авторегрессионная модель вида:

$$Y_t = c + \sum_{i=1}^{p} A_i Y_{i,t-i} + A_0^{-1} B_0 u_t \qquad (26)$$

где Y_t – вектор эндогенных переменных, u_t – случайные ошибки.

В модели на основе экономической теории вектор эндогенных переменных включает: $y_{\prod_t}$ – прирост индекса промышленного производства, $M2_t$ – прирост денежной массы М2, $inflation_t$ – инфляцию, $MIACR1D_t$ – однодневную ставку денежного рынка МИАКР, $usdrub_t$ – прирост логарифма номинального курса рубля к доллару. В качестве экзогенной переменной используется цена на нефть марки Brent. Это обусловлено тем, что Россия – страна-экспортер нефти, и наша экономика в значительной степени зависит от изменения нефтяных цен на международных рынках. B_0 – единичная матрица, A_0^{-1} – нижняя треугольная матрица, имеющая вид:

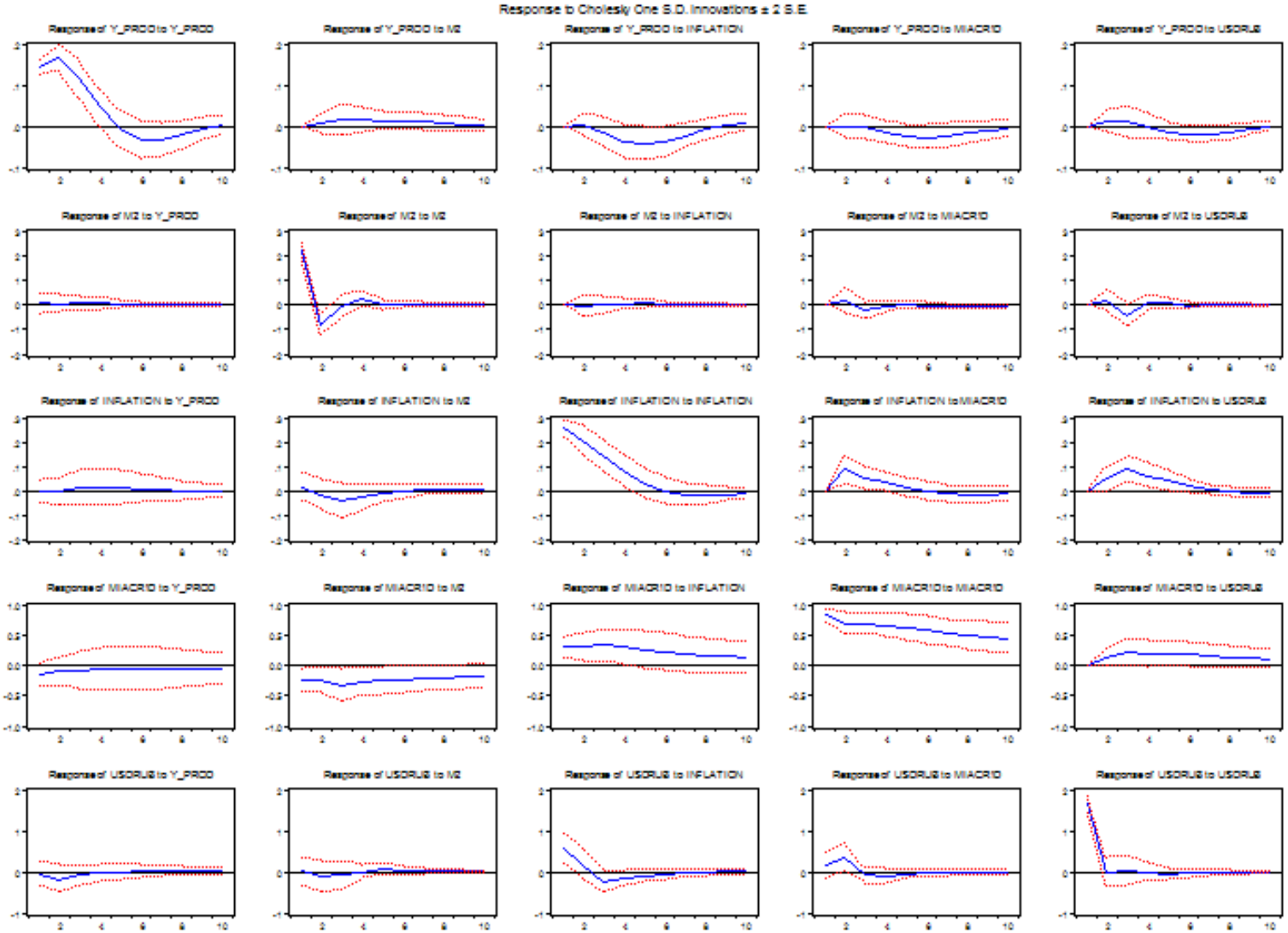
$$\begin{matrix} y_{\prod_t} \\ y_{\prod_t} \end{matrix}$$

$$\begin{bmatrix} \varepsilon_t^{y_{\prod_t}} \\ \varepsilon_t^{M2} \\ \varepsilon_t^{inflation} \\ \varepsilon_t^{MIACR1D} \\ \varepsilon_t^{usdrub} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 1 & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & 1 \end{bmatrix} \begin{bmatrix} u_t^{y_{\prod_t}} \\ u_t^{M2} \\ u_t^{inflation} \\ u_t^{MIACR1D} \\ u_t^{usdrub} \end{bmatrix} \qquad (27)$$

Идентификационная стратегия модели строится на следующих предположениях. Проводится нормализация к 1 реакции эндогенных переменных на собственные шоки. Реальный сектор (выпуск) не реагирует мгновенно на шоки монетарного сектора. Второе уравнение (27) показывает, что существует одновременная реакция предложения денег только на шоки экономической активности. Цены на потребительские товары реагируют одновременно на шоки реального сектора и предложения денег.

Четвертое уравнение (27) отражает функцию реакции центрального банка (стандартное правило Тейлора) на изменения в реальном секторе и монетарном секторе. Последнее уравнение (27) показывает, что номинальный курс реагирует мгновенно на все вышеуказанные шоки.

Функции импульсных откликов эндогенных переменных показаны на рисунке . Важными детерминантами курса рубля в модели с фундаментальными факторами выступают инфляция и однодневная ставка МИАКР. Повышение инфляции вызывает обесценение курса рубля к доллару за счет эффекта переноса в цены.



Примечание – Источник: расчеты авторов

Рисунок 8 – Функции импульсных откликов эндогенных переменных в модели с фундаментальными факторами

Повышение ставки денежного рынка также становится причиной обесценения курса рубля. Возможно, это происходит на фоне повышения уровня неопределенности на денежном рынке на развивающемся рынке как за счет проводимой денежно-кредитной политики, так и в случае проблемы рационирования кредитов.

Для учета влияния поведения экономических агентов на валютный курс, как в работе [CITATION KoH15 \l 1049], модель расширяется за счет добавления ненаблюдаемых настроений и ожиданий экономических агентов, построенные на основе запросов в Интернете. В данном случае ограничения для идентификации принимают следующий вид:

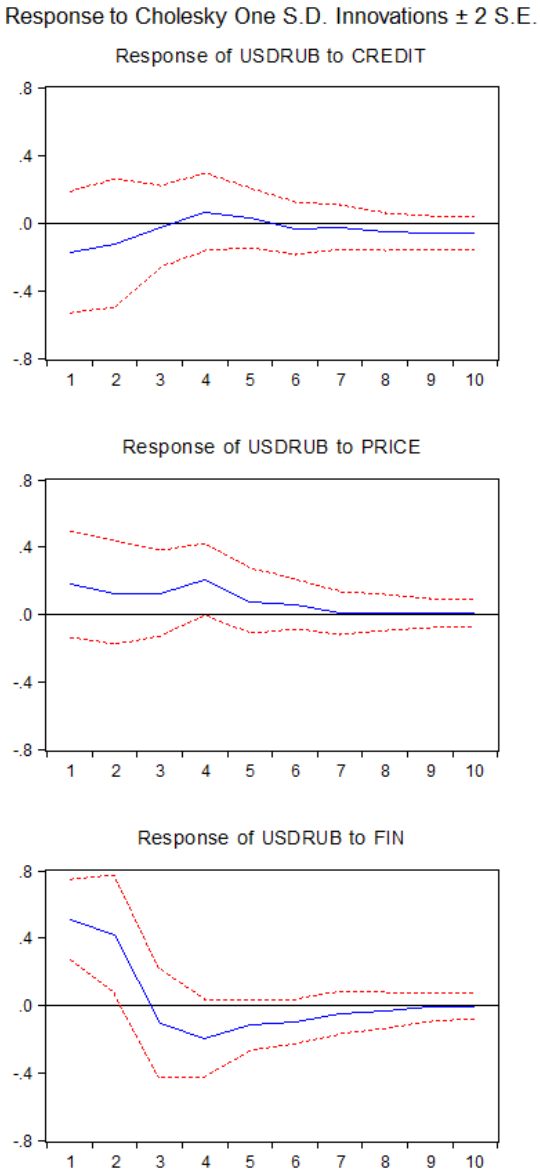
$$\begin{bmatrix} \varepsilon_t^{y_{\prod}} \\ \varepsilon_t^{M2} \\ \varepsilon_t^{inflation} \\ \varepsilon_t^{MIACR1D} \\ \varepsilon_t^{credit} \\ \varepsilon_t^{price} \\ \varepsilon_t^{fin} \\ \varepsilon_t^{usdrub} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & a_{25} & 0 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 & 0 & 0 & a_{37} & 0 \\ a_{41} & a_{42} & a_{43} & 1 & 0 & a_{46} & 0 & 0 \\ a_{51} & 0 & 0 & a_{54} & 1 & a_{56} & a_{57} & 0 \\ a_{61} & 0 & a_{63} & 0 & 0 & 1 & 0 & 0 \\ a_{71} & 0 & a_{73} & a_{74} & a_{75} & a_{76} & 1 & 0 \\ a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & a_{87} & 1 \end{bmatrix} \begin{bmatrix} u_t^{y_{\prod}} \\ u_t^{M2} \\ u_t^{inflation} \\ u_t^{MIACR1D} \\ u_t^{credit} \\ u_t^{price} \\ u_t^{fin} \\ u_t^{usdrub} \end{bmatrix} \quad (28)$$

где перечень эндогенных переменных: $y_{\prod_t}$ – прирост индекса промышленного производства, $M2_t$ – прирост денежной массы М2, $inflation_t$ – инфляция, $MIACR1D_t$ – однодневная ставка денежного рынка МИАКР, $credit_t$, $price_t$ и fin_t – ненаблюдаемые факторы, настроения на кредитном рынке, инфляционные ожидания и настроения на финансовых рынках (поисковые запросы Google Trends) и $usdrub_t$ – прирост номинального курса рубля к доллару. Экзогенная переменная – прирост логарифма цены на нефть марки Brent.

Идентификационная стратегия (28), как говорилось выше, основана на теоретических предположениях. Наиболее экзогенной переменной является индекс промышленного производства. По мнению авторов [CITATION Dyb17 \l 1049], не происходит мгновенной реакции реального сектора на изменения предложения денег и рыночных настроений. Предложение денег реагирует одновременно на изменения промышленного производства и смену настроений на кредитном рынке, а цены подвержены влиянию шоков спроса, предложения денег и настроений на финансовом рынке. Правило денежных властей теперь принимает во внимание не только шоки реального сектора, предложения денег, инфляции, но

и также изменения инфляционных ожиданий. Настроения на кредитном рынке и финансовом рынке изменяются в ответ на шок выпуска, изменение настроений и инфляционных ожиданий экономических агентов. Настроения на кредитном рынке также зависят от изменения процентной ставки из-за изменения спроса на деньги. А настроения на финансовом рынке реагируют на шоки инфляции и монетарной политики. В результате на номинальный валютный курс оказывают воздействие все шоки рассматриваемых эндогенных переменных.

Как показывают результаты (см. рисунок), шоки настроений на кредитном рынке и инфляционных ожиданий не оказывают значимое влияние на курс рубля (хотя и имеет положительный знак). Обеспокоенность населения относительно финансовых рынков приводит к обесценению курса рубля. Это можно объяснить за счет следующих положений.

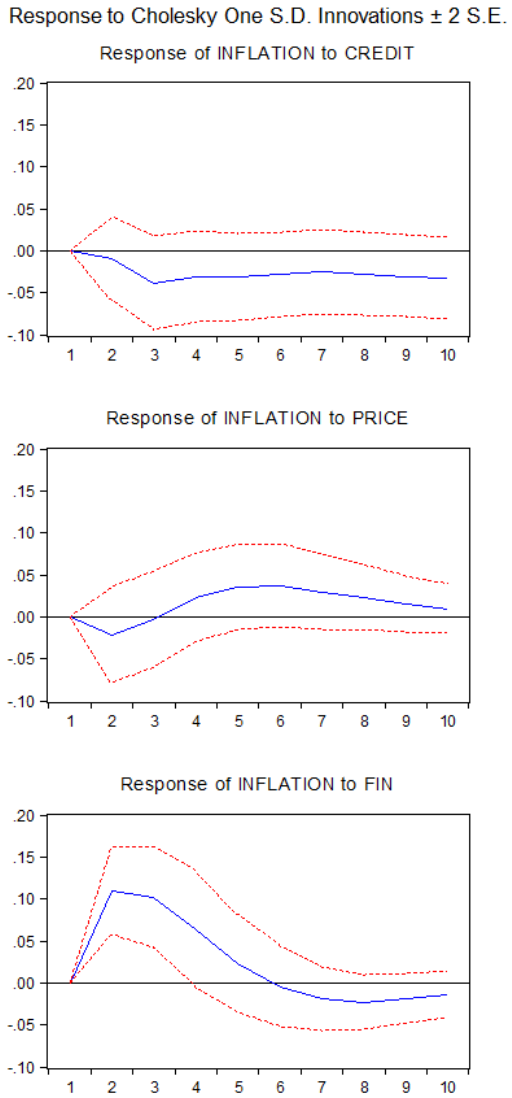


Примечание – Источник: расчеты авторов

Рисунок 9 – Функции импульсных откликов курса рубля на шоки настроений экономических агентов

Поиск и сбор информации пользователями в Интернете происходит наиболее часто в условиях неопределенности, когда на финансовых рынках произошло негативное событие. Как мы знаем, экономические агенты в большей степени чрезмерно реагируют на появление плохих новостей. Это согласуется с теоретическим предположением о том, что колебания валютного курса пропорциональны периодичности поступления информации на финансовые рынки[15]. Таким образом, при пессимистичных ожиданиях на фондовом или валютном рынке будет происходить обесценение курса рубля.

Функции импульсных откликов инфляции на шоки в настроениях экономических агентов представлены на рисунке .



Примечание – Источник: расчеты авторов

15 Процесс распространения информации неоднороден. Появление новой информации на какое-то время разделяет участников финансовых рынков на информированную и неинформированную группы.

Рисунок 10 – Функции импульсных откликов инфляции на шоки настроений населения

Как и для курса рубля, не было выявлено значимого влияния ожиданий на кредитном рынке и инфляционных ожиданий на инфляцию. Однако ожидания на финансовом рынке вызывают рост инфляции течение следующих 4 месяцев. Экономические агенты при повышении неопределенности на финансовых рынках будут пересматривать свои инфляционные ожидания, что вызовет рост инфляции в будущем.

Что касается однодневной ставки МИАКР (см. рисунок), то на нее оказывают влияние ожидания на финансовом рынке и инфляционные ожидания. Повышение интереса к финансовой тематике оказывает повышательное давление на процентную ставку МИАКР в течение 1-2 месяцев. Инфляционные ожидания оказывают положительное влияние на ставку МИАКР только через 5 месяцев.

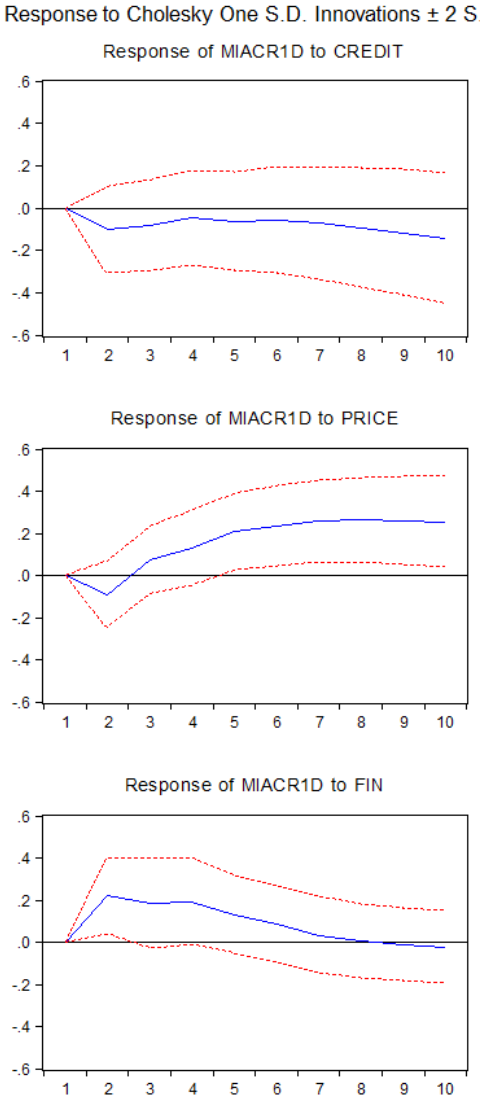


Рисунок 11 – Функции импульсных откликов однодневной ставки МИАКР на шоки настроений экономических агентов

Примечание – Источник: расчеты авторов

В целом в условиях неопределенности происходит ужесточение денежно-кредитной политики при инфляционных рисках и нестабильности на финансовых рынках.

Можно сделать вывод, что ожидания экономических агентов являются важными детерминантами курса рубля. Кроме того, корень из среднеквадратичной ошибки меньше в модели с поисковыми запросами и фундаментальными факторами составляет 1.36, а в модели с фундаментальными факторами – 1.38.

Подводя итог, следует отметить, что ожидания и настроения экономических агентов, построенных на основе интернет-запросов, являются важными детерминантами курса рубля и инфляции. Включение в модель подобного рода показателей дает возможность учесть поведенческие характеристики экономических агентов при прогнозировании.

2.3 Прогнозирование реальных темпов ВВП на основе модели со смешанной периодичностью данных

Повышение предсказательной способности эконометрических моделей является ключевым фактором проведения успешной экономической политики. Существование различных внешних и внутренних лагов, связанных с формулированием и реализацией мер экономической политики, не позволяют властям ориентироваться на текущие показатели экономики. Кроме того, некоторые ключевые макроэкономические переменные доступны только с низкой частотой и с определенной задержкой, что вынуждает при принятии решений ориентироваться на прогнозные значения. А это означает, что особый интерес представляют эконометрические модели, способные делать прогноз низкочастотных макроэкономических показателей при доступной лишь частично информации о текущем состоянии экономики.

Кроме того, временные ряды с более высокой частотностью (например, безработица, цена на нефть, индекс промышленного производства) доступны в периоды между публикацией таких данных, как ВВП или компоненты ВВП, и могут быть хорошими предикторами при прогнозировании.

До недавнего времени для решения данной проблемы проводилось усреднение высокочастотных временных рядов до периодичности низкочастотных данных, а после чего использовались для прогнозирования. В работе (Ghysels et al., 2004) была предложена модель MIDAS (mixed data sampling), позволяющая включать в модель переменные со смешанной периодичностью.

Важно отметить, что агрегирование высокочастотных временных рядов может отрицательно сказаться на качестве оцениваемой модели и при прогнозировании. Во-первых, агрегирование данных становится причиной потери важной информации о динамике высокочастотного временного ряда. Во-вторых, происходит изменение процесса порождения

данных, поэтому динамика агрегированных данных в эконометрической модели будет значительно отличаться от моделей с высокой или смешанной периодичностью.

Из этого следует, что модель со смешанной периодичностью имеет много возможных применений в макроэкономических исследованиях. В данной части исследования будет показано, как MIDAS может быть использована для прогнозирования ВВП в России.

Базовая версия модели MIDAS может быть записана в следующем виде:

$$Y_t = \alpha + \sum_{1}^{p} \beta_i L^i Y_t + \gamma \sum_{k=1}^{m} \Phi(k;\theta) L_{HF}^{k} X_t + \varepsilon_t, \quad (29)$$

где L обозначает лаговый оператор, Y_t – это низкочастотная переменная (реальные темпы роста ВВП в годовом выражении), X_t – независимые переменные (M2 – прирост денежной массы в широком определении, brent – прирост логарифма цены на нефть, unemployment – уровень безработицы, rub – прирост логарифма курса рубля, fin, credit и price – ожидания на финансовых рынков, настроения на кредитном рынке и инфляционные ожидания, построенные, как было показано ранее, с помощью поисковых запросов) с 3 лагами.

В данном случае при прогнозировании реальных темпов роста ВВП (квартальные данные) использовалась нормализованная экспоненциальная функция Алмона. Экспоненциальное взвешивание Алмона принимает следующий вид:

$$\Phi(k;\theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=1}^{m} \exp(\theta_1 j + \theta_2 j^2)} \quad (30)$$

В качестве эталонной модели оценивалась рассматривается наивный прогноз для прогноза на 1-8 кварталов вперед. Результаты относительных RMSFE для MIDAS модели представлены в таблице .

Таблица 11 – Относительная ошибка вневыборочного прогноза MIDAS к AR модели

	naive	MIDAS_M2	MIDAS_ brent	MIDAS_ unemloyment	MIDAS_ rub	MIDAS_ fin	MIDAS_ price	MIDAS_ credit
h=1	**1.00**	1.11	1.18	1.17	1.17	1.00	1.07	1.70
h=2	1.00	**0.69**	0.75	1.07	1.06	1.03	0.93	1.49
h=3	1.00	**0.81**	0.91	1.32	0.96	0.94	1.03	1.63

h=4	1.00	0.93	0.99	1.33	**0.83**	0.93	0.99	1.42
h=5	1.00	0.99	0.99	1.36	**0.86**	1.02	0.99	1.41
h=6	1.00	**0.94**	0.96	1.41	1.02	1.09	0.98	1.38
h=7	1.00	**0.96**	1.00	1.46	1.04	1.12	1.00	1.21
h=8	1.00	0.97	1.05	1.46	1.07	**0.91**	1.02	1.17

Примечание – Источник: расчеты авторов

Результаты свидетельствуют о превышении предсказательной способности модели со смешанной частотой данных с приростом денежной массы М2 над авторегрессионной моделью, начиная с 2 квартала. Отношение корня из среднеквадратичной ошибки прогноза модели MIDAS к AR меньше 1. Наилучший прогноз модели с М2 для реальных темпов роста ВВП был получен для 2-3 кварталов и 6-7 кварталов. На горизонте прогноза 4-5 кварталов наименьшие отношение RMSFE было получено в модели с курсом рубля к доллару. Следует отметить, что ожидания и настроения экономических агентов, оцененные с помощью интернет-запросов, имеет меньшую предсказательную способность на всех горизонтах прогноза, за исключением прогноза реальных темпов роста ВВП на 8 кварталов.

В целом эмпирический анализ на российских данных показал, что поисковые запросы могут быть использованы как опережающие индикаторы для прогнозирования инфляции, безработицы и курса рубля. Кроме того, применение методов машинного обучения при прогнозировании макроэкономических показателей позволяет получить более точный прогноз и набор предикторов. Модель со смешанной периодичностью данных с макроэкономическими переменными дала более точный прогноз реальных темпов роста, по сравнению с поисковыми запросами.

ЗАКЛЮЧЕНИЕ

Обзор теоретических подходов показал, что асимметрия информации является важной проблемой при принятии решений экономическими агентами в условиях неопределенности. В рамках теоретических концепций выделяют две возможные причины существования асимметрии информации – неблагоприятный отбор и риск недобросовестного поведения, которые вынуждают рыночного игрока собирать дополнительную информацию для принятия наилучшего решения. Однако, как показывает ряд исследований, эффект на общественное благосостояние будет зависеть от объема и точности доступной информации. Информационное множество экономических агентов состоит из двух возможных источников информации – частного сектора и общедоступного источника (заявления центрального банка или фискальных властей). В условиях неопределенности повышение точности частной информации всегда приводит к росту благосостояния экономических агентов. В свою очередь влияние общедоступной информации на общественное благосостояние будет зависеть от точности и объема раскрываемой информации. При высокой степени прозрачности действий центрального банка или правительства, а также согласованности во времени и предсказуемости их действий, возможно будет добиться более высокого уровня общественного благосостояния.

Результаты обзора эмпирических исследований свидетельствует о том, что поисковые запросы, как прокси переменные выявленных ожиданий, являются ключевым источником информации при прогнозировании различных макроэкономических показателей. Международный опыт показывает, что включение в эконометрические модели интернет запросов по экономическим тематикам обеспечивают улучшение качества краткосрочных прогнозов инфляции, валютного курса, безработицы и потребительских настроений. Исследователи предполагают, что данный результат обусловлен влиянием процесса распространения информации в Интернете на принимаемые решения через пересмотр ожиданий экономических агентов. Когда происходит какое-то негативное событие, вызывающее нестабильность в экономике, будет повышаться интерес населения к текущей экономической ситуации и спрос на информацию в Интернете. В таком случае содержащаяся информация в интенсивности интернет запросов может обеспечить улучшение прогноза макроэкономических показателей даже в странах с небольшим уровнем проникновения Интернета.

Следует отметить, что важную роль поисковые запросы играют при прогнозировании экономических показателей в период финансового кризиса или рецессии. Повышение

обеспокоенности пользователей относительно кризисных явлений становилось причиной повышения безработицы, оттока капитала и волатильности валютных курсов в будущем.

На основе обзора международных и российских исследований были определены наиболее подходящие для эмпирического исследования методы отбора и выделения признаков из больших объемов данных: лассо, гребневая регрессия, эластичная сеть, случайный лес, градиентный бустинг и метод главных компонент.

Результаты эмпирического анализа на российских данных позволили сделать несколько ключевых выводов. Во-первых, поисковые интернет-запросы Google Trends являются важным источником информации о предпочтениях и ожиданиях экономических агентов в России, поскольку доля пользователей поисковика Google на протяжении десятилетия приближалась к 50%. Во-вторых, добавление интенсивности поисковых запросов в модель улучшает качество прогноза инфляции и безработицы, полученных на основе методов машинного обучения. В-третьих, наиболее подходящим инструментом для снижения размерности поисковых запросов может выступать метод главных компонент. Методы отбора с регуляризацией или ансамблевые методы имеют более низкую предсказательную способность, чем линейная модель с главными компонентами для инфляции, безработицы и курса рубля.

В-четвертых, оценка структурной векторной авторегрессионной модели также показала, что добавление настроений на кредитном рынке, инфляционных ожиданий и ожиданий на финансовых рынках, полученных из интернет-запросов, позволяют улучшить качество подгонки и учесть поведенческие характеристики экономических агентов в модели.

В-пятых, при прогнозировании реальных темпов роста ВВП не было выявлено преимуществ в предсказательной силе модели со смешанной периодичностью, включающую какую-либо переменную, характеризующую ожидания экономических агентов. Наилучшими оказались модели со смешанной периодичностью с денежной массой М2 или курсом рубля на горизонтах прогноза от 1 до 8 кварталов.

Таким образом, в целом результаты исследования показывают, что интенсивность поисковых запросов, связанных с экономической тематикой, позволяют повысить точность прогнозов макроэкономических показателей по сравнению со стандартными методами. Перспективным остается исследование предсказательной способности более высокочастотных интернет-запросов в прогнозировании макроэкономических показателей, например, в рамках подхода со смешанной периодичностью данных.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

x

1. Akerlof, ,G. A., "The Market for “Lemons”: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, Vol. 84, No. 3, 1970. pp. 488-500.
2. Kulkarni, , "The Influence of Information Technology on Information Asymmetry in Product Markets," *Journal of Business and Economic Studies*, Vol. 6, No. 1, 2000. pp. 55-71.
3. Nayyar, P. R. ;, "Information Asymmetries: A Source of Competitive Advantage for Diversified Service Firms," *Strategic Management Journal*, Vol. 11, No. 7, 1990. pp. 513-519.
4. Milgrom, P.; Roberts, J.;, "Informational Asymmetries, Strategic Behavior, and Industrial Organization," *American Economic Review*, Vol. 77, No. 2, (987. pp. 184-193.
5. Venkatesh, P. C.; Chiang, R.;, "Information Asymmetry and the Dealer’s Bid-Ask Spread: A Case Study of Earnings and Dividend Announcements," *Journal of Finance*, Vol. 41, No. 5, 1986. pp. 1089-1102.
6. Aboody, D.; Lev, B.;, "Information Asymmetry, R&D, and Insider Gains," *Journal of Finance*, Vol. 55, No. 6, 2000. pp. 2747-2766.
7. Stigler, G. J.;, "The Economics of Information," *Journal of Political Economy*, Vol. 69, No. 3, 1961. pp. 213-225.
8. Stiglitz, J. E., "The Contributions of the Economics of Information to Twentieth Century Economics," *Quarterly Journal of Economics*, Vol. 115, No. 4, 2000. pp. 1441-1478.
9. Chiang, R.; Venkatesh, P. C.;, "Insider Holdings and Perceptions of Information Asymmetry: A Note," *Journal of Finance*, Vol. 43, No. 4, 1988. pp. 1041-1048.
10. Morgan, L. A. . The Importance of Quality // In: Perceived Quality / Ed. by Jacoby J., Olson J. Lexington, MA. 1985.
11. Nelson , "Information and Consumer Behaviour," *Journal of Political Economy*, Vol. 78, No. 2, 1970. pp. 311-329.
12. Darby, M.; Karni, E. ;, "Free competition and the optimal amount of fraud," *Journal of Law and Economics*, Vol. 16, No. 1, 1973. pp. 67-88.
13. Spence, A. M. ;, "Job Market Signalling," *Quarterly Journal of Economics*, Vol. 87, No. 3, 1973. pp. 355-374.
14. Weaver, W. ;, "Recent Contributions to the Mathematical Theory of Communication," *The Mathematical Theory of Communication*, 1963.
15. Cole, C. ;, "Shannon Revisited: Information in Terms of Uncertainty," *Journal of the American Society for Information Science*, Vol. 44, No. 4, 1993. pp. 204-211.
16. Artandi S., "Information Concepts and their Utility," *Journal of the American Society for Information Science*, Vol. 24, No. 4, 1973. pp. 242-245.
17. Smith, R. E. ;, "Integrating Information from Advertising and Trial: Processes and Effects on Consumer Response to Product Information," *Journal of Marketing Research*, Vol. 30, No. 2, 1993. pp. 204-219.
18. Blackwell R.D., Miniard P.W., and Engel J. Consumer Behaviour (9th ed.). Harcourt, Dryden. 2001.
19. Zeithaml V.A., "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, Vol. 52, No. 3, 1988. pp. 2-22.
20. Hansen, T. ;, "Perspectives on consumer decision making: An integrated approach," *Journal of Consumer Behaviour*, Vol. 4, No. 6, 2005. pp. 420-437.
21. Sun, Y.; Lv, B.; Xue, T.. 11th International Conference on e-Business (ICE-B) // A Research on Inflation Expectations Measurements and Applications. A View Based on Network Behavior. Vienna, Austria. 2014. pp. 76-83.
22. Morris S., Shin H., "Social value of public information," *American Economic Review*, Vol. 92, 2002. pp. 1521–1534.
23. Svensson, L. E.;, "Social value of public information: Morris and Shin (2002) is actually pro transparency, not con," *American Economic Review*, Vol. 96, No. 1, 2006. pp. 448-451.

2 Angeletos, G.; Pavan, A.;, "Transparency of information and coordination in economies with 4. investment complementarities," *American Economic Review*, Vol. 94, No. 2, 2004. pp. 91-98.
2 Hellwig, C.;, "Heterogeneous information and the benefits of transparency," 2005.
5.
2 Cornand, C.; Heinemann, F.;, "Optimal Degree of Public Information Dissemination," *The* 6. *Economic Journal*, Vol. 118, No. 528, 2008. pp. 718-742.
2 Guzman, G.;, "Internet search behavior as an economic forecasting tool: the case of inflation 7. expectations," *Journal of Economic and Social Measurement*, Vol. 36, No. 3, 2011. pp. 119–167.
2 Koop, G.; Onorante, L.;, "Macroeconomic nowcasting using Google probabilities," University of 8. Strathclyde, 2013.
2 Li, X.; Shang, W.; Wang, S.; Ma, J., "A MIDAS modelling framework for Chinese inflation index 9. forecast incorporating Google search data," *Electronic Commerce Research and Applications*, Vol. 14, 2015. pp. 112–125.
3 Birchler, U.; Bütler, M.;, "Information Economics," Routledge, New York., 2007.
0.
3 Park, J.; Konana, P.; Gu, B.; Kumar, A.; Raghunath, R.;, "Confirmation bias, overconfidence, and 1. investment performance: evidence from stock message boards," McCombs Research Paper Series, IROM-07-10, 2010.
3 Antweiler, W.; Frank, M. Z.;, "Is all that talk just noise? The information content of internet stock 2. message boards," *The Journal of Finance*, Vol. 59, No. 3, 2004. pp. 1259–1294.
3 Dang Y., , Zhang Y., and Chen H., "A lexicon-enhanced method for sentiment classification: an 3. experiment on online product reviews," *Intelligent Systems*, Vol. 25, No. 4, 2010. pp. 46–53.
3 Seabold, S.; Coppola, A.;, "Nowcasting Prices Using Google Trends. An Application to Central 4. America," World Bank Group, Policy Research Working Paper 7398, 2015.
3 Niesert, R. F., Oorschot, J. A.; Veldhuisen, C. P.; Brons, K.; Lange, R. J. , "Can Google search data 5. help predict macroeconomic series?," *International Journal of Forecasting.*, 2019.
3 Engel, C.; West, K. ;, "Exchange Rates and Fundamentals," *Journal of Political Economy*, Vol. 113, 6. No. 3, 2005. pp. 485-517.
3 Frenkel J.A., Mussa M.L. Asset Markets, Exchange Rates, and the Balance of Payments. Vol 2. // 7. In: Handbook of International Economics / Ed. by Jones R.W., Kenen P.B. Amsterdam: North-Holland. 1985.
3 Dybka, P.; Chojnowski, M.;, "Is Exchange Rate Moody? Forecasting Exchange Rate with Google 8. Trends Data," *Econometric Research in Finance*, Vol. 2, 2017. pp. 1-21.
3 Smith, J. P., "Google Internet search activity and volatility prediction in the market for foreign 9. currency," *Finance Research Letters*, Vol. 9, 2012. pp. 103–110.
4 Bulut, L., "Google Trends and the forecasting performance of exchange rate models," *Journal of* 0. *Forecasting*, Vol. 37, No. 3, 2018. pp. 303-315.
4 Bughin, J.;, ""Nowcasting" the belgian economy," *SSRN Electronic Journal*, 2011.
1.
4 Brake, G.;, "Unemployment? Google it! Analyzing the usability of Google queries in order to 2. predict unemployment," 2017.
4 McLaren, N.; Shanbhogue, R.;, "Using internet search data as economic indicators," *Bank of* 3. *England Quaterly Bulletin*, Vol. 51, No. 2, 2011. pp. 134–140.
4 Choi, H.; Varian, H.;, "Predicting the present with Google Trends," *Economic Record*, Vol. 88, 4. 2012. pp. 2-9.
4 Vicente, M.; López-Menéndez, A.; Pérez, R.;, "Forecasting unemployment with internet search 5. data: Doet it help to improve predictions when job destruction is skyrocketing?," *Technological Forecasting Social Change*, Vol. 92, 2015. pp. 132–139.
4 Smith, P.;, "Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data," 6. *Journal of Forecasting*, Vol. 35, 2016. pp. 263–284.
4 D'Amuri, F.;Marcucci, J., "The predictive power of Google searches in forecasting US 7. unemployment," *International Journal of Forecasting*, Vol. 33, No. 4, 2017. pp. 801-816.
4 Столбов, М.;, "Статистика поиска в Google как индикатор финансовой конъюнктуры," 8. *Вопросы экономики*, Т. 11, 2011. С. 79-93.

4 Борочкин, А. А.;, "Использование статистики поисковых запросов в сети Интернет для 9. краткосрочного прогнозирования макроэкономических переменных," *Деньги и кредит*, Т. 8, 2013. С. 27-32.

5 Фантаццини, Д.; Шаклеина, М.; Юрас, Н.;, "Big Data в определении социального 0. самочувствия населения России," *Прикладная эконометрика*, Т. 50, 2018. С. 43-66.

5 Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.;, "Equation of state calculations by fast computing 1. machines," *The Journal of Chemical Physics*, Vol. 21, No. 6. pp. 1087–1092.

5 Hastings, W.;, "Monte Carlo sampling methods using Markov chains and their applications," 2. *Biometrika*, Vol. 57, No. 1, 1970. pp. 97–109.

5 Hastie, T.; Friedman, J. H.; Tibshirani, R.. The elements of statistical learning (2nd ed.). Springer, 3. 2009.

5 Zou, H.; Hastie, T.;, "Regularization and variable selection via the elastic net," *Journal of the Royal* 4. *Statistical Society*, Vol. 67, No. 2, 2005. pp. 301-320.

5 Scott, S. L.; Varian, H. R.;, "Bayesian variable selection for nowcasting economic time series," 5. National Bureau of Economic Research, w19567, 2013.

5 Efron, B.; Hastie, T.; Johnstone, L.; Tibshirani, R.;, "Least angle regression," *Annals of Statistics*, 6. Vol. 32, 2004. pp. 407–499.

5 Zou, H.; Hastie, T.; Tibshirani, R.;, "Sparse principal component analysis," *Journal of* 7. *Computational and Graphical Statistics*, Vol. 15, No. 2, 2006. pp. 262–286.

5 Stock, J. H.; Watson, M. W.;, "Generalized shrinkage methods for forecasting using many 8. predictors," *Journal of Business and Economic Statistics*, Vol. 30, No. 4, 2012. pp. 481–493.

5 Байбуза, И. , "Прогнозирование инфляции с помощью методов машинного обучения," *Деньги* 9. *и кредит* , Т. 77, № 4, 2018. С. 42-59.

6 Tibshirani , R.;, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical* 0. *Society*, Vol. Series B (Methodological), 1996. pp. 267-288.

6 Breiman, L., "Random Forests," *Machine Learning*, Vol. 45 , No. 1, 2001. pp. 5-32.
1.

6 Friedman, J., "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of* 2. *Statistics*, Vol. 29, No. 5 , 2001.

6 Harvey, D.; Leyborne, S.; Newbold, P.;, "Testing the Equality of Prediction Mean Squared Errors," 3. *International Journal of Forecasting*, Vol. 13, 1997. pp. 281-291.