

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

Препринт

ТЕОРЕТИЧЕСКИЕ И ЭМПИРИЧЕСКИЕ АСПЕКТЫ ВЫЯВЛЕНИЯ ОЖИДАНИЙ
ЭКОНОМИЧЕСКИХ АГЕНТОВ НА ОСНОВЕ ТЕКСТОВОГО АНАЛИЗА

ГРАЧЁВА В.А., Институт прикладных экономических исследований, младший научный
сотрудник, ORCID: 0000-0002-9692-8588, vgracheva-18@edu.ranepa.ru

ПЕТРОВА Д.А., Институт прикладных экономических исследований, научный сотрудник,
ORCID: 0000-0003-1030-4932, petrova-da@ranepa.ru

Москва 2021

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

Preprint

THEORETICAL AND EMPIRICAL ASPECTS OF DETERMINING THE EXPECTATIONS
OF ECONOMIC AGENTS BASED ON TEXT ANALYSIS

GRACHEVA V.A., Institute of Applied Economic Research, junior researcher, ORCID: 0000-
0002-9692-8588, vgracheva-18@edu.ranepa.ru

PETROVA D.A., Institute of Applied Economic Research, researcher, ORCID: 0000-0003-
1030-4932, petrova-da@ranepa.ru

Moscow 2021

Аннотация

Интернет является одним из общедоступных источников информации, где можно найти информацию с минимальными издержками. Социальные сети становятся все более популярными у интернет-пользователей для поиска и анализа информации о текущей экономической ситуации. Это связано с тем, что пользователи получают возможность обмениваться мнениями или обсуждать различные проблемы в новостных сообществах социальных сетей. Полученная информация может использоваться при принятии решений экономическими агентами. Таким образом, исследование поведения пользователей в социальных сетях делает возможным выявить ожидания и предпочтения населения.

Целью данного исследования является оценка ожиданий и настроений экономических агентов на основе текстового анализа материалов социальных сетей.

Для достижения поставленной цели в исследовании решается ряд **задач**:

- Анализ механизмов влияния процесса распространения информации и сетевых эффектов на поведение экономических агентов;
- Систематизация результатов теоретических и эмпирических работ по анализу ожиданий экономических агентов;
- Обзор методов машинного обучения, используемых при обработке больших текстовых данных;
- Разработка алгоритма определения источников информации для последующего веб-скрапинга и правила отбора текстовой информации для создания корпуса комментариев;
- Сбор базы данных и подготовка постов и комментариев к текстовому анализу;
- Применение методов тематического моделирования для определения тем и ключевых слов в данных социальных сетей;
- Построение высокочастотных индикаторов настроений экономических агентов.

Предметом исследования является количественная оценка настроений интернет-пользователей на российских данных.

Новизна научного исследования заключается в оценке индикаторов инфляционных ожиданий, настроений на валютном рынке и индексов экономических условий с помощью структурированных и неструктурированных интернет-данных.

Метод и методология исследования основываются на тематическом моделировании; методах машинного обучения и эконометрических методах анализа временных рядов.

Исследование выполнено на российских данных за 2014-2021 гг. В работе показано, что данные социальных сетей, поисковые запросы и новостные статьи интернет-изданий могут быть прокси-переменными для ожиданий экономических агентов. Авторы анализируют три типа индикаторов настроений экономических агентов на основе интернет-данных:

инфляционные ожидания; настроения на валютном рынке и индекс экономических условий. **Результаты** эконометрического анализа свидетельствуют о том, что качество моделей с индикаторами настроений для макроэкономических показателей выше, чем без них. Кроме того, индикаторы на основе постов социальной сети «ВКонтакте», новостных статей РБК и поисковых запросов Google Trends являются более информативными по сравнению с комментариями к постам. **Основным выводом** проведенного исследования является то, что интернет-данные позволяют улучшить качество моделирования макроэкономических показателей. **В дальнейшем** планируется расширение перечня индикаторов настроений экономических агентов и оценка продвинутых моделей временных рядов.

Ключевые слова: текстовый анализ; машинное обучение; инфляционные ожидания; настроения экономических агентов; интернет-данные; тематическое моделирование; индекс экономических условий; социальные сети; поисковые запросы.

Коды JEL Classification: E37, E30, C50, D83

Abstract

The Internet is a public source of information, where information can be found at minimum search cost. Social media are becoming increasingly popular among web users trying to find and analyze information about the current economic situation. Web users get the opportunity to exchange views or discuss various issues in the news communities of social networks. This information can be used by economic agents to make decisions. Thus, the study of user behavior in social networks makes it possible to identify the expectations and preferences of economic agents.

The **goal** of this study is to assess the expectations and sentiments of economic agents based on textual analysis of social media data.

The study addresses the following objectives:

- Analysis of the mechanisms of influence of the information dissemination and networking effects on the behavior of economic agents;
- Systematization of the results of theoretical and empirical analysis of the economic agents' expectations;
- An overview of machine learning methods used in text processing;
- Development of an algorithm for identifying sources of information for web scraping and rules for selecting text information to create a body of posts and comments;

- Collecting a database and preparing posts and comments for text analysis;
- Application of topic modeling to the identification of topics and keywords in social media data;
- Assessment of high-frequency indicators of the public sentiment.

The **subject** of the research is a quantitative assessment of the sentiment of web users based on Russian data.

The **novelty** of the study is the assessment of inflation expectations, sentiments in the foreign exchange market and indices of economic conditions using structured and unstructured internet data. **Methods:** topic modeling; machine learning methods and econometric methods of time series analysis.

The study is based on data for Russia in 2014-2021. The study shows that social media posts, search queries and online news articles can be good proxy variables for the economic agents' expectations. We construct three types of public confidence indicators based on internet data: inflation expectations; sentiment in the foreign exchange market and index of economic conditions. The results of econometric analysis indicate that the quality of macroeconomic performance models with sentiment indicators is higher than without these indicators. Additionally, indicators based on VK posts, RBC news articles and Google Trends search queries are more informative compared to comments. The main conclusion of the study is that internet data can improve the quality of macroeconomic performance models. In a further study, we plan to expand the list of indicators of the sentiment of economic agents and to evaluate advanced time series models.

Keywords: textual analysis; machine learning; inflation expectations; sentiment of economic agents; internet data; topic modeling; index of economic conditions; social networks; search queries.

JEL Classification: E37, E30, C50, D83

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
1 Обзор теоретических концепций по анализу влияния информации на поведение экономических агентов	10
1.1 Анализ механизмов влияния распространения информации в социальных сетях на ожидания населения	10
1.2 Обзор теоретических работ по влиянию информации на настроения экономических агентов	18
2 Обзор эмпирических исследований по анализу социальных сетей.....	24
2.1 Описание методов получения количественных показателей из данных социальных сетей.....	24
2.2 Эконометрические подходы к моделированию экономических показателей на основе социальных сетей	32
3 Анализ социальных сетей на российских данных.....	47
3.1 Описание статистических данных и первичная подготовка данных социальных сетей	47
3.2 Построение ожиданий населения на основе данных социальных сетей.....	47
4 Эконометрический анализ на основе высокочастотных индикаторов ожиданий населения.....	56
4.1 Анализ влияния настроений экономических агентов на инфляцию и экономическую активность.....	57
4.2 Анализ влияния настроений на валютном рынке на курс рубля.....	60
ЗАКЛЮЧЕНИЕ.....	64
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	68

ВВЕДЕНИЕ

В течение последнего десятилетия популярность социальных сетей у пользователей возросла. При этом посты и комментарии в социальных сетях отражают настроения экономических агентов относительно текущей ситуации в экономике, обеспокоенность в условиях финансового кризиса или отношение к ключевым решениями в сфере экономической политики. Следовательно, анализ большого количества постов и комментариев пользователей социальных сетей позволяет исследовать ожидания экономических агентов. По этой причине построенные на основе анализа социальных сетей индикаторы могут быть использованы при прогнозировании экономических и финансовых показателей.

Можно выделить ряд преимуществ данных, полученных на основе анализа социальных сетей, для выявления настроений и ожиданий экономических агентов по сравнению с опросными и рыночными индикаторами. Во-первых, посты и комментарии пользователей социальных сетей доступны в реальном времени, что дает возможность выявлять настроения экономических агентов до момента публикации официальных статистических данных. Во-вторых, социальные сети могут быть использованы для анализа поведения пользователей, в том числе изменения их предпочтений, доступных оперативнее и на более высокочастотной основе, чем опросные индикаторы.

Целью данного исследования является оценка ожиданий и настроений экономических агентов на основе текстового анализа материалов социальных сетей.

Для достижения поставленной цели в исследовании решается ряд задач:

- Анализ механизмов влияния процесса распространения информации и сетевых эффектов на поведение экономических агентов;
- Систематизация результатов теоретических и эмпирических работ по анализу ожиданий экономических агентов;
- Обзор методов машинного обучения, используемых при обработке больших текстовых данных;
- Разработка алгоритма определения источников информации для последующего веб-скрапинга и правила отбора текстовой информации для создания корпуса комментариев;

- Сбор базы данных и подготовка постов и комментариев к текстовому анализу;
- Применение методов тематического моделирования для определения тем и ключевых слов в данных социальных сетей;
- Построение высокочастотных индикаторов настроений экономических агентов.

Объектом исследования являются данные социальных сетей и индикаторы настроений экономических агентов, построенные на их основе.

Метод и методология исследования основываются на тематическом моделировании, методах машинного обучения и эконометрических методах анализа временных рядов.

Гипотезы исследования состоят в следующем. Во-первых, предполагается, что распространяемая в социальных сетях информация может влиять на ожидания и настроения населения в условиях неопределенности. Во-вторых, мудрость толпы или стадное поведение в социальных сетях может оказывать влияние на процесс принятия решений экономическими агентами. В-третьих, индикаторы настроений экономических агентов, полученные на основе анализа социальных сетей, могут стать альтернативой для традиционных статистических показателей. В-четвертых, данные социальных сетей могут конкурировать при анализе и прогнозировании с альтернативными интернет-данными – поисковыми запросами и новостными статьями интернет-изданий.

Новизна научного исследования заключается в оценке индикаторов инфляционных ожиданий, настроений на валютном рынке и индексов экономических условий с помощью структурированных и неструктурированных интернет-данных.

Практическая значимость исследования состоит в том, что полученные меры настроений экономических агентов могут быть альтернативой опросным и рыночным индикаторам, а также позволить повысить качество моделей макроэкономических показателей.

Результаты исследования базируются на опыте зарубежных и российских работ по аналогичной тематике. Во-первых, обзор теоретических исследований позволил выделить механизмы влияния ожиданий населения на процесс принятия решений, а также рассмотреть причины более сильной реакции экономических агентов на плохие новости. Во-вторых, был проведен обзор методов получения

количественных показателей на основе данных социальных сетей, используемых в российских и зарубежных исследованиях, а также рассмотрены эконометрические подходы к прогнозированию экономических и финансовых показателей на основе индикаторов настроений пользователей в Интернете.

1 Обзор теоретических концепций по анализу влияния информации на поведение экономических агентов

Интерес экономистов к анализу коллективных действий людей имеет достаточно продолжительную историю. Более ранние исследования были посвящены изучению формирования мнений людей при голосовании, а также воздействия на этот процесс разнообразия мнений в группе индивидов и подверженности влиянию информационных лидеров [1].

Как показывают исследования, использование агрегированных мнений или прогнозов отдельных людей может способствовать повышению точности анализа и прогнозов на финансовых рынках [2]. В работе [3] утверждается, что принятое решение на основе агрегированного мнения (мудрости толпы) экономических агентов в среднем будет лучше для всей группы, чем отдельного индивида, в частности эксперта или информационного лидера [4].

Особую значимость мудрость толпы имеет в контексте анализа поведения пользователей социальных сетей в процессе принятия решения. В социальной сети экономические агенты получают доступ к информации о текущей ситуации с минимальными затратами на ее сбор и обработку, а скорость распространения информации находится на достаточно высоком уровне. Из этого следует, что важно понять, при каких условиях мудрость толпы может быть использована для повышения общественного благосостояния. В данном разделе будут рассмотрены различные теоретические концепции, используемые для анализа влияния мудрости толпы на настроения экономических агентов. Кроме того, обсуждается асимметрия влияния плохих и хороших новостей на благосостояние экономических агентов.

1.1 Анализ механизмов влияния распространения информации в социальных сетях на ожидания населения

Экономические агенты формируют свои ожидания на основе всей доступной информации на момент принятия решения. В современных условиях получение информации в Интернете позволило снизить издержки по ее поиску и сбору. Одним из ключевых информационных ресурсов стали социальные сети, где пользователи

могут обмениваться информацией и обсуждать интересующие их вопросы, включая экономические и социальные проблемы.

Теоретические концепции влияния распространения информации в социальных сетях на поведение экономических агентов основываются на анализе результатов исследований по голосованию и другим коллективным действиям [5]. Предполагается, что максимизации благосостояния общества можно добиться при реализации решений на основе «мудрости толпы». Под мудростью толпы понимается достигнутый консенсус внутри группы.

Действительно, как показывают ряд исследований, коллективные решения могут давать лучший исход в контексте общественного благосостояния, чем решения отдельного экономического агента [2]. Однако для принятия обоснованных экономических решений группой должны выполняться следующие условия. Во-первых, существует разнообразие мнений внутри группы, когда индивиды принимают свои решения на основе имеющегося опыта, навыков и образования. Во-вторых, мнения экономических агентов должны быть независимыми. Под независимостью мнений понимается принятие отдельным экономическим агентом решений на основе собственной частной информации, что снижает возможность одинаковых ошибок при формировании ожиданий. В-третьих, допускается децентрализация мнений, когда возможность влиять на мнение группы равномерно распределяется между всеми участниками группы. В-четвертых, наблюдается согласованность решений, при которой индивиды будут корректировать ожидания в соответствии с наилучшим решением для группы. Наконец, экономические агенты должны доверять друг другу и верить в систему принятия решений группой.

В то же время принятие решений на основе мудрости толпы может вызвать снижение общественного благосостояния по следующим причинам. Экономические агенты могут отнестись несерьезно к рассматриваемому вопросу, намеренно сделать плохой прогноз или обсуждать интересующий их вопрос в отдельных группах, например, поддерживая мнение информационного лидера, что искажает итоговый результат. Формирование мнения экономического агента по результатам общения и обсуждения с другими участниками группы может привести к манипулированию коллективным мнением и дезинформации. Кроме того, может возникать безумие толпы или стадное поведение при выборе большинства наиболее популярной стратегии, что делает невозможным разнообразие мнений.

Проверка гипотезы о наилучшем решении при коллективных действиях обычно проводится с помощью трех типов теоретических моделей – эпидемиологических, агентоориентированных и теоретико-игровых.

В рамках эпидемиологических подходов рассматриваются модель SIR¹ и ее модификации. В работе [6] проводится анализ процесса распространения информации в социальной сети в рамках SIR модели для однородной социальной сети. В социальной сети есть три группы экономических агентов: игнорирующие слухи; распространители информации и группа, осведомленная о существовании слуха, но не распространяющая его.

В анализе для сравнения скорости распространения новостей рассматриваются два возможных источника слухов – устаревшие и новые способы общения. Авторы к устаревшим способам общения относят устные сообщения, звонки по телефону и СМС сообщения. В качестве новых способов общения рассматривают переписку по электронной почте, блоги, подкасты, микроблоги и социальные сети. Новые способы обмена информацией позволяют любому человеку создавать, изменять и обмениваться контентом, используя бесплатные или недорогие онлайн сервисы. Таким образом, индивид может мгновенно обмениваться информацией с большим количеством людей через онлайн ресурсы. Следовательно, при новых способах общения выше скорость распространения информации. При этом чем больше информации, тем в большей степени экономические агенты склонны забывать или терять интерес к слухам, а значит, δ больше в новых медиа.

Авторы получили следующие результаты. При прочих равных, чем больше значение вероятности потери интереса к информации δ , тем меньше влияние слухов на экономических агентов. Это происходит потому, что сам слух не содержит важной информации. В результате уменьшается количество распространителей информации и влияние на всех агентов. Авторы показали, что слухи оказывают большее влияние при традиционном способе общения. Высокая вероятность потери интереса к слухам δ дает ослабляет влияние слухов, например, при обмене информации в социальной сети. Это согласуется с реально наблюдаемыми явлениями, когда на онлайн сервисах средства массовой информации могут публиковать достоверную и своевременную

¹ Название модели произошло от деления группы на S (susceptible) – людей, подверженных влиянию эпидемии, I (infected) – заболевших и R (recovered) – выздоровевших.

информацию, чтобы развеять слухи, тогда вероятность δ возрастет, что приводит к меньшему влиянию слухов.

При остальных фиксированных параметрах, чем больше значение η , тем меньше влияние слуха на экономических агентов. Высокая вероятность η означает, что подаватели слухов, например, эксперты или власти, не доверяют полученной информации. Распространение недостоверной информации быстрее прекращается в социальных сетях.

Следует отметить, что у такой постановки теоретической модели есть недостатки. Предпосылка о том, что экономические агенты имеют равные шансы получить слух, а скорость передачи информации фиксирована, является малореалистичной, так как взаимодействие между различными людьми происходит по-разному. Скорость распространения информации зависит от устойчивости взаимодействия между индивидами социальной сети, репутации индивида, среды распространения и т.д. К примеру, чем устойчивее взаимосвязь между игнорирующими слухи и распространителями информации, тем выше будет скорость передачи информации в социальной сети.

Кроме того, если контакты распространителей информации ограничены, то индивид не может одновременно взаимодействовать со всеми соседями, а лишь с определенной долей участников социальной сети. Вместе с тем, разные индивиды демонстрируют различную степень активности в социальной сети. Если игнорирующие слухи активно взаимодействуют с распространителями информации, то увеличивается вероятность распространить слух дальше. Из этого следует, что чем больше распространителей информации вокруг игнорирующего слух, тем выше вероятность изменить статус на распространителя.

В модификации SIR модели, как сделано в работе [7], три ранее определенные группы делятся на 6 типов агентов в зависимости от активности взаимодействия с другими агентами, а также вводится предпосылка о нелинейности скорости распространения информации.

Если игнорирующие слухи активно ищут сведения о происходящих событиях, то они сами находят распространителей информации. Напротив, игнорирующие слухи, придерживающиеся позиции не проявлять инициативу, узнают о происходящих событиях только тогда, когда с ними свяжутся активные распространители информации.

Активные распространители информации всегда делятся новостями с игнорирующими слухами. Неактивные распространяют информацию только тогда, когда игнорирующие слухи активно занимаются поиском информации.

Полученные авторами выводы согласуются с вышеупомянутыми результатами простой модели. Однако в модификации SIR при более правдоподобных предпосылках было выявлено, что при увеличении доли активных пользователей или скорости распространения информации будет расти степень взаимодействия пользователей, и больше появляется распространителей информации. В итоге вырастет количество пользователей, получивших информацию внутри социальной сети.

Другим подходом к анализу поведения пользователей социальных сетей является построение теоретико-игровой модели, где предполагается обмен информацией между агентами в результате байесовского обучения [8]. Стандартными предположениями в таких моделях являются возможность наблюдать все прошлые действия других участников социальной сети и неограниченность ожиданий. Если условия выполняются, то будет происходить агрегирование информации и выбор правильного решения за счет мудрости толпы [9]. В противоположном случае при ограниченных² ожиданиях невозможно достичь консенсуса при агрегировании информации, и агенты будут придерживаться стратегии стадного поведения или информационных каскадов, повторяя прошлые действия и/или полностью игнорируя собственные частные сигналы о текущем состоянии мира [10], [11]

Однако предпосылка о доступности всех прошлых действий пользователей социальных сетей является нереалистичной, поскольку индивиду сложно обрабатывать большие объемы информации. В работе Асемоглу и др. [8] представлена теоретическая модель, не предполагающая доступности полной информации и позволяющая выявить условия для возникновения мудрости толпы и повышения общественного благосостояния.

В целом авторы получили следующие результаты. В социальных сетях с влиятельной группой, являющейся источником информации для всех остальных агентов (нерасширяющиеся наблюдения), не возникает асимптотического обучения. Этот результат согласуется с предположением об отсутствии агрегирования

² Под ограниченностью здесь понимается существование барьеров при передаче информации в социальной сети.

информации, когда набор информации, на основе которой бесконечное подмножество индивидов принимают свои решения, всегда остается ограниченным.

Когда ожидания экономических агентов неограниченны, а топология сети предполагает разнообразие мнений, то будет происходить асимптотическое обучение. Этот результат также подразумевает, что некоторые агенты, такие как «информационные лидеры», являются соседями будущих агентов, но не могут оказывать сильного воздействия на их ожидания, асимптотическое обучение может замедляться, но будет продолжаться до тех пор, пока ожидания не ограничены.

Следует отметить, что для большинства распространенных детерминированных и стохастических сетей ограниченные ожидания агентов несовместимы с асимптотическим обучением. Это согласуется с существующими результатами по асимптотическому обучению, например, в работах [11], [10] и [9].

Асимптотическое обучение возможно с ограниченными ожиданиями агентов для некоторых стохастических топологий сетей. В этих случаях поступает достаточное количество новой информации для достижения коллективного консенсуса, потому что некоторые агенты принимают решения на основе ограниченных наблюдений за действиями других агентов.

В работе [12] также рассматривается теоретическая модель с байесовским обучением для изучения мудрости толпы. Авторы показали, что мудрость толпы позволяет принять наилучшее для группы решение, если есть разнообразие мнений и отсутствуют какие-либо структурные барьеры для передачи информации внутри социальной сети.

Golub и Jackson [13] получили аналогичные результаты при наивной схеме обновления ожиданий экономическими агентами во времени. Авторы утверждают, что для получения наилучшего решения при коллективных действиях в социальной сети должны соблюдаться следующие условия. Ни одна маленькая группа индивидов в социальной сети не пользуется большим доверием и не имеет возможности воздействовать на мнения остальных участников социальной сети. Кроме того, большое сообщество пользователей социальной сети будет сколь угодно близко к правильному решению, если будет высока степень разнообразия мнений. Таким образом, возникает чрезмерное внимание к мнению небольшой группы лидеров, способных повлиять на общественное мнение, что сделает невозможным схождение в равновесии к наилучшему для группы решению. Авторы полагают, что ограниченно

рациональная структура социальной сети может удовлетворять условиям для возникновения мудрости толпы.

Третьим подходом для анализа влияния процесса распространения информации на процесс принятия решения в социальной сети и достижения мудрости толпы является агентоориентированная модель. Типичная теоретическая модель предложена в работе [14], где рассматривается возможность принятия правильного решения в рамках режима с полной информацией или при наблюдаемом консенсусе других участников социальной сети (режим агрегирования информации).

В результате на основе оценки построенных моделей были сделаны следующие выводы. Если первоначальное мнение ($x_i(0)$) сильно отличается от истинного значения, то есть коллективная ошибка высока в момент времени 0, то повышение общественного влияния α значительно уменьшает коллективную ошибку ϵ_{LT} , а значит, приводит к консенсусу всех агентов, тогда как индивидуальные убеждения β мало влияют на результат. Этот эффект намного сильнее в модели с полной информацией, в которой агенты имеют возможность наблюдать действия остальных участников социальной сети. Снижение коллективной ошибки ϵ_{LT} меньше при получении экономическими агентами только информации о консенсусе остальных пользователей социальной сети. Это означает, что мудрость толпы позволяет достичь правильного решения только при агрегировании информации. При доступности полной информации о прошлых действиях экономических агентов можно прийти к общему мнению в группе, но оно будет отличаться от оптимального решения.

Если начальные условия заданы так, что усредненное мнение экономических агентов в момент времени 0 близко к правильному решению, то есть коллективная ошибка невысока, то общественное влияние может ухудшить итоговый результат. Такая ситуация возникает из-за значительного отклонения усредненного мнения пользователей социальной сети в момент времени 0 от истинного значения. Это приведет к согласованным коллективным действиям и консенсусу, но в итоге к неверному решению и снижению общественного благосостояния, что согласуется с проблемой стадного поведения. Данный эффект будет выше в модели с полной информацией. В модели с агрегированием информации этого может не произойти, если определить диапазон параметров с низким уровнем общественного влияния (α), где мудрость толпы не сильно искажается. Иначе при высокой степени общественного влияния невозможно достичь мудрости толпы.

В модели с полной информацией разнообразие мнений резко снижается при высокой степени восприимчивости экономических агентов к общественному воздействию. Это делает невозможным достижения мудрости толпы в случаях, когда отсутствие разнообразия мнений участников социальной сети сочетается с высокой коллективной ошибкой. В итоге агенты коллективно сходятся к принятию неправильного решения. В модели с агрегированием информации можно добиться разнообразия мнений при заданном диапазоне параметров и достичь консенсуса близкого к правильному решению.

Что касается попыток непосредственного измерения мудрости толпы, то в большинстве случаев увеличение подверженности мнений экономических агентов к воздействию общественности приводит к снижению мудрости толпы. Особенно сильным эффект оказывается в модели с полной информацией. Важно отметить, что можно добиться снижения искажения мнений экономических агентов при высоком уровне убежденности в правильности собственных решений β . Авторы отмечают, что в реальности для общего случая нельзя сделать вывод о воздействии общественного мнения на консенсус участников социальной сети при заранее неизвестных параметрах. Результат можно заключаться как в повышении, так и снижении коллективной ошибки в принятии решений.

Подводя итог, отметим, что теоретические исследования предполагают возможность существования наилучшего решения для группы индивидов, если отдельные индивиды имеют различные и независимые мнения. При этом для выбора наилучшего решения не нужна вся доступная информация, как в модели с полной информацией, поскольку экономическим агентам сложно обрабатывать большие объемы информации, а достаточно наблюдать консенсус мнений группы. Кроме того, существование информационных лидеров, например, политиков, экспертов и т.д., может искажать мнения остальных участников группы, что в итоге делает невозможным принятие правильного решения.

В дополнение к анализу коллективных действий в социальных сетях, важным вопросом остается то, как процесс принятия решений экономическими агентами зависит от получаемой информации. В этой связи в следующем подразделе обсудим влияние позитивных и негативных новостей о состоянии экономики на благосостояние экономических агентов.

1.2 Обзор теоретических работ по влиянию информации на настроения экономических агентов

В теоретической литературе анализ поведения экономических агентов играет важную роль, поскольку принимаемые потребителями и фирмами решения непосредственно влияют на экономические процессы и деловой цикл. В современных условиях экономические агенты могут собирать информацию в Интернете с минимальными издержками, в том числе на сайтах интернет-изданий или в социальных сетях. В таком случае процесс пересмотра экономических ожиданий также упрощается. При этом снижаются транзакционные издержки по поиску информации.

Если экономический агент правильно предсказал направление делового цикла на основе имеющейся информации о повышении экономической активности в будущем, полученной из новостей или социальных сетей, то он может повысить свое благосостояние. Такая же логика справедлива для фирм. В работе [15] предполагается, что изменения промышленного производства связаны с ожиданиями компаний. Следовательно, источником экономических колебаний являются изменения в ожиданиях бизнеса.

Однако, как отмечается в [16], существует асимметричная реакция на плохие и хорошие новости. Плохие новости о кризисных явлениях или экономических проблемах оказывают более сильное воздействие на поведение экономических агентов, поскольку их последствия могут быть сопряжены со значительным ухудшением благосостояния [17]. Потребители резко снижают расходы на потребление после получения новости о снижении экономической активности, чтобы не допустить значительного снижения уровня потребления в случае кризиса [18].

Рассмотрим теоретическую модель для более подробного анализа влияния эмоциональной окраски информации из социальных сетей, в частности новостей в интернет-изданиях, на поведение людей. Одна из наиболее подходящих для этого моделей предложена в работе [19]. В статье осуществляется построение теоретической модели для анализа влияния информации из газет на благосостояние экономических агентов. Постановка и выводы данной модели также могут быть применены к исследованию процесса распространения информации интернет-изданиями в социальных сетях.

Пусть СМИ публикует два типа статей – хорошие (z_g) и плохие (z_b) новости. Потребители заинтересованы в получении такой информации по целому ряду причин. Авторы утверждают, что домохозяйства воспринимают хорошие новости, содержащие информацию о том, как принять решение о потреблении, сбережениях и инвестициях, чтобы повысить свое благосостояние. Когда они получают плохие новости, то могут использовать эту информацию для предотвращения неправильного выбора и, как следствие, потерь в благосостоянии. Для сравнения делается предположение о том, что влияние на доход симметрично и потребители имеют начальный уровень дохода Y .

Ожидаемая полезность потребителя после прочтения плохой новости, позволившего снизить потери благосостояния, таким образом, равна $u_k(Y) - u_k(Y - z_b)$. Ожидаемая полезность от прочтения хорошей новости равна $u_k(Y + z_g) - u_k(Y)$.

Количество положительных и негативных событий, освещаемых i -ым новостным источником равно, соответственно, q_b^i и q_g^i для $i \in \{A, B\}$. $p(q)$ – это вероятность того, что потребитель узнает об определенном событии из новостного источника при увеличении количества плохих или хороших новостей. Потребитель предпочитает получать такие новости с уменьшающейся скоростью, то есть $p'(\cdot) > 0$ и $p''(\cdot) < 0$.

Когда потребитель читает только одну газету, то вероятности того, что потребитель прочитает о негативном или положительном событиях, соответственно, будут равны $p(q_b^i)$ и $p(q_g^i)$. Ожидаемая полезность от чтения газеты i (1) для потребителя k за вычетом цены и затрат на обработку информации будет равна $w_k(q_g^i, q_b^i)$:

$$w_k(q_g^i, q_b^i) = p(q_g^i)[u_k(Y + z_g) - u_k(Y)] + p(q_b^i)[u_k(Y) - u_k(Y - z_b)], \quad (1)$$

где $w_k(q_g^i, q_b^i)$ – ожидаемая полезность от чтения газеты i ,

$p(q_g^i)$ – вероятность того, что потребитель прочитает хорошую новость в i -ой газете,

q_g^i – количество хороших новостей в i -ой газете,

$p(q_b^i)$ – вероятность того, что потребитель прочитает плохие новости в i -ой газете,

q_b^i – количество плохих новостей в i -ой газете,

$u_k(Y)$ – ожидаемая полезность потребителя при отсутствии негативных и положительных событий,

$u_k(Y + z_g)$ – ожидаемая полезность потребителя в случае реализации хорошей новости,

$u_k(Y - z_b)$ – ожидаемая полезность в случае реализации плохой новости.

Оценка ценности газеты и содержания новостей для потребителя зависит от других ее характеристик, например, редакционной позиции или сферы интересов. Ценность этих аспектов для потребителя k равна, соответственно, a_k для газеты A и b_k для газеты B . Таким образом, газеты A и B могут принести полезность потребителю k , равную $w_k(q_g^A, q_b^A) + a_k$ и $w_k(q_g^B, q_b^B) \geq b_k$, соответственно. Потребитель k покупает газету A при условии (2), что

$$\Delta w_k = w_k(q_g^A, q_b^A) - w_k(q_g^B, q_b^B) \geq b_k - a_k, \quad (2)$$

где Δw_k – прирост полезности для потребителя k ,

$w_k(q_g^A, q_b^A) + a_k$ – полезность от газеты A ,

Действительно, если ценность газеты A выше, чем газеты B , то лучше принимать решения на основе более релевантного источника информации – A . Иначе потребителю k более выгодно купить газету B . Разница в ценности газет $b_k - a_k$ имеет распределение $\Gamma(\cdot)$ и плотность распределения $\gamma(\cdot)$. В таком случае вероятность, что потребитель k будет читать газету A равна $\Gamma(\Delta w_k)$.

С точки зрения стороны предложения информации, для i -ой газеты ожидаемая функция издержек на поиск и сбор информации (3) выглядит следующим образом:

$$C^i(q_g^i, q_b^i) = c_q(q_g^i + q_b^i) + n^i c_s, \quad (3)$$

где $C^i(q_g^i, q_b^i)$ – ожидаемая функция издержек на поиск и сбор информации для i -ой газеты,

c_q – стоимость создания одной новостной статьи на основе собранной информации,

q_g^i – количество хороших новостей в i -ой газете,

q_b^i – количество плохих новостей в i -ой газете,

c_s – стоимость доставки газеты потребителю,

n^i – ожидаемое количество потребителей, купивших i -ую газету.

Ожидаемое количество покупателей газеты A пропорционально вероятности покупки данной газеты и равно $n^A = n\Gamma(\Delta w_k)$, по аналогии для газеты B равно $n^B = n[1 - \Gamma(\Delta w_k)]$. n – общее количество покупателей газет A и B .

Газета i максимизирует ожидаемую прибыль (4) следующим образом:

$$E(\Pi^i) = n^i(p - c_s) - c_q(q_g^i + q_b^i), \quad (4)$$

где $E(\Pi^i)$ – ожидаемая прибыль газеты i ,

n^i – ожидаемое количество потребителей, купивших i -ую газету,

p – цена газеты,

c_q – стоимость создания одной новостной статьи на основе собранной информации,

q_g^i – количество хорошие новости в i -ой газете,

q_b^i – количество плохих новостей в i -ой газете,

c_s – стоимость доставки газеты потребителю.

В равновесии условия первого порядка для максимизации прибыли (5)-(6) принимают вид:

$$\frac{dE(\Pi^i)}{dq_g^i} = n(p - c_s)\gamma(\Delta w_k) \frac{dw_k(q_g^i, q_b^i)}{dq_g^i} - c_q, \quad (5)$$

$$\frac{dE(\Pi^i)}{dq_b^i} = n(p - c_s)\gamma(\Delta w_k) \frac{dw_k(q_g^i, q_b^i)}{dq_b^i} - c_q, \quad (6)$$

где n^i – ожидаемое количество потребителей, купивших i -ую газету,

p – цена газеты,

c_q – стоимость создания одной новостной статьи на основе собранной информации,

q_g^i – количество хороших новостей в i -ой газете,

q_b^i – количество плохих новостей в i -ой газете,

c_s – стоимость доставки газеты потребителю (в социальной сети она будет равна нулю),

Δw_k – прирост полезности от покупки газеты A .

Эти условия предполагают, что в оптимуме (7) для газеты i должно выполняться

$$\frac{dw_k(q_g^i, q_b^i)}{dq_g^i} = \frac{dw_k(q_g^i, q_b^i)}{dq_b^i} \text{ в равновесии при } (q_g^{i*}, q_b^{i*}), \quad (7)$$

Условие первого порядка для потребителя (8):

$$\frac{dp}{dq_g^i} [u_k(Y + z_g) - u_k(Y)] = \frac{dp}{dq_b^i} [u_k(Y) - u_k(Y - z_b)], \quad (8)$$

$u_k(Y)$ – ожидаемая полезность потребителя при отсутствии негативных и положительных событий,

$u_k(Y + z_g)$ – ожидаемая полезность потребителя в случае реализации хорошей новости,

$u_k(Y - z_b)$ – ожидаемая полезность в случае реализации плохой новости.

Авторы полагают, что новостные газеты публикуют больше негативных новостей, следовательно, в равновесии $q_b^{i*} > q_g^{i*}$. Это следует из условия первого порядка для потребителя, где при строго вогнутой функции полезности u_k получаем $u_k(Y + z_g) - u_k(Y) < u_k(Y) - u_k(Y - z_b)$, а $\frac{dp}{dq_g^i} > \frac{dp}{dq_b^i}$ при $p''(\cdot) < 0$.

Этот результат может быть интерпретирован следующим образом. Когда плохие новости оказывают более сильное влияние на полезность, чем хорошие новости, спрос потребителей, при прочих равных условиях, выше на плохие новости. Для максимизации прибыли новостные источники реагируют на эту разницу в спросе, публикуя больше информации о негативных событиях. Преобладание плохих новостей может иметь важные последствия, поскольку будет происходить искажение

общественного мнения. Этот эффект может быть усилен, если негативная информация оказывает большее влияние на настроения людей, чем хорошие новости.

Помимо этого, результаты показали, что существует асимметричная реакция экономических агентов на плохие и хорошие новости. Это обусловлено тем, что люди в большей степени обеспокоены возможными потерями благосостояния при реализации кризисных явлений. При этом авторы отмечают, что экономические агенты могут не обращать внимания на хорошие новости, когда уровень экономического оптимизма и так высок.

* * *

Таким образом, обзор теоретических работ позволяет сделать следующие выводы. Настроения потребителей и фирм могут быть одним из ключевых факторов экономических колебаний. Особую роль в формировании ожиданий и настроений экономических агентов играет общедоступная информация, в частности, социальные сети. В социальных сетях пользователи могут обсуждать различные экономические события и обмениваться информацией, что увеличивает скорость распространения сведений и снижает транзакционные издержки по поиску и сбору информации.

Мудрость толпы может приводить к наилучшему решению для группы индивидов при предпосылках о различных и независимых мнениях. При этом делегирование процесса принятия решения влиятельной группе способно привести к снижению общественного благосостояния практически для большинства топологий социальной сети. Кроме того, процесс принятия решения экономическими агентами в значительной степени зависит от типа новости. Потребители сильнее реагируют на плохие новости из-за боязни больших потерь благосостояния. По этой причине интерес в социальных сетях к конкретной теме может быть обусловлен обеспокоенностью населения относительно перспектив при реализации неблагоприятного исхода в периоды неопределенности.

В следующем разделе более подробно остановимся на способах оценки настроений населения с помощью данных социальных сетей, а также рассмотрим эконометрические подходы к анализу влияния такого типа данных на экономические и финансовые показатели.

2 Обзор эмпирических исследований по анализу социальных сетей

За последние два десятилетия развитие социальных сетей способствовало усилению их роли в повседневной жизни человечества. С каждым годом растет количество их пользователей, следовательно, у все большей доли населения появляется возможность обсуждения и изучения онлайн различных политических и экономических событий. А это в свою очередь дает возможность исследователям наблюдать за поведением пользователей в социальных сетях и строить альтернативные опросным индикаторы настроений экономических агентов. Кроме того, исследователи на основе таких данных могут повышать точность прогнозов макроэкономических и финансовых показателей, поскольку выявленные таким образом настроения экономических агентов могут быть получены с минимальными издержками по сбору и обработке данных, а также практически мгновенно в отличие от альтернативных показателей.

В следующих подразделах подробнее остановимся на описании методов количественной оценки настроений экономических на основе данных социальных сетей и рассмотрим эконометрические подходы к анализу их предсказательной способности при прогнозировании макроэкономических и финансовых показателей.

2.1 Описание методов получения количественных показателей из данных социальных сетей

В настоящее время социальные сети привлекают все больше пользователей, так как там люди могут свободно выражать свое мнение. Число веб-сайтов, предоставляющих доступ к данным, увеличивается, что позволяет их использовать для анализа общественного мнения, а также строить индикаторы настроений на высокочастотной основе в реальном времени. Обобщая, в эмпирических исследованиях можно выделить несколько подходов для оценки настроений: частотный подход, когда подсчитывается количество постов; анализ эмоциональной окраски с помощью словаря или моделей машинного обучения и тематическое моделирование для выявления популярности конкретных тем в отдельные моменты времени. В данном подразделе приведем примеры реализации таких подходов к текстовому анализу социальных сетей.

В работе [20] исследуется связь между мнениями, опубликованными в социальной сети для трейдеров StockTwits³ и внутридневными ценами акций, которые были взяты из источника финансовых и экономических данных Quandl. Индикаторы настроений инвесторов рассчитывались в рамках частотного анализа с tf-idf.

Модели тестировались на 17 акциях, средняя точность прогнозов моделей составила 52.45%. Следовательно, возможно существование корреляции между настроениями трейдеров и ценой акции.

Люди привыкли обсуждать новую информацию в социальных сетях, делясь эмоциями и мнениями. Пандемия COVID-19 стала важной темой, охватившей весь мир. Существует гипотеза, что, получая и анализируя данные из социальных сетей, можно прогнозировать развитие пандемии.

Исследование [21] показывает, что результаты анализа твитов коррелируют с реальной ситуацией с COVID-19, а значит, подход может быть использован в дальнейшем. Пандемия связана с высокими негативными настроениями, в частности, страхом. Однако, были зафиксированы всплески положительных настроений, связанные с введением новых мер или улучшением ситуации в больницах. Поэтому социальные сети могут быть использованы для измерения настроений населения и уровня доверия к новым мерам.

В работе [22] также изучаются настроения людей во время пандемии COVID-19 в социальной сети Twitter. Данные были получены из массива, опубликованного Zhou и др. [23] по данным Twitter и основанного на ключевых словах, относящихся к коронавирусу, за период с 1 по 14 апреля 2020 года на английском языке.

Для тематического моделирования выбрана модель латентного размещения Дирихле (LDA) с 70 темами, предложенная в [24].

Оценка тональности каждого твита с помощью словаря VADER, который адаптирован для социальных сетей, предложенный в [25]. Преимущество метода заключается в возможности его использования на необработанных данных и без обучения моделей. Этот метод классифицирует эмоциональную окраску каждого твита как положительную, негативную или нейтральную. Кроме эмоциональной

³ StockTwits – социальная сеть для инвесторов, позволяющая выгрузить достаточный объем необходимых данных – мнения индивидуальных инвесторов о фондовом рынке или конкретных акциях.

окраски, также анализируются темы твитов за каждый день с помощью динамической тематической модели (DTM).

В результате авторы получают распределение настроений для каждой темы, относящейся к COVID-19 за период с 1 по 14 апреля 2020 года. С помощью динамической тематической модели было получено 70 тем. Также к каждой теме указаны наиболее важные релевантные слова для ее характеристики. Для каждого из 14 дней выборки был получен список тем по убыванию количества твитов. В топ-3 тем вошли: темы «оставайтесь дома»; «отчет»; «болезнь». Эти темы показывают, что в первой половине апреля 2020 люди обсуждали количество новых случаев заболеваний COVID-19, количество смертей из-за вируса, а также то, что следует оставаться дома для обеспечения безопасности себя и близких.

Анализ настроений показал, что тема «оставайтесь дома» в среднем имеет положительную эмоциональную окраску. Причиной этому может быть вера людей в то, что такая мера принесет успехи в борьбе с вирусом. Тема «болезнь» показывает противоположную тенденцию, количество негативных твитов по этой теме преобладает над положительными. Преобладание пессимистичных настроений может говорить о страхе людей во время пандемии, недовольстве последствиями игнорирования правительством вируса и отсутствием эффективных мер борьбы. По теме «отчет» в некоторые дни наблюдается преобладание позитивных настроений, а в другие – негативных. Это может быть связано с количеством новых случаев и смертей от вируса, а также с тем, в какой мере реальная ситуация соответствует ожиданиям людей.

Во время пандемии возрос уровень тревожности и увеличилось число случаев депрессии. Это может быть связано с неполной и неточностью информации в новостях и других информационных источниках. В работе Samuel и др. [26] исследуется динамика изменения настроений в США по мере приближения заболеваний к локальному пиковому уровню с помощью данных Twitter. В дополнение, авторы сравнивают эффективность двух основных методов классификации твитов – наивный байесовский классификатор и логистическая регрессия. Данные получены по ключевому слову «corona» («корона») с февраля по март 2020 года. Авторы выявили увеличение страха в настроениях людей, который можно объяснить обеспокоенностью населения из-за COVID-19.

В работе [27] проводится оценка настроений населения Китая относительно коронавируса (COVID-19) в период с 9 января по 10 февраля 2020 г. в социальной сети Weibo. Отбор постов пользователей осуществляется по ключевым словам «пневмония» и «коронавирус».

Авторы строят индикатор интереса населения к COVID-19 в Китае как количество постов по данному вопросу за день. Авторы показали, что полученный индикатор настроений имеет схожую динамику с количеством подтвержденных случаев заболевания коронавирусом. Корреляционный анализ показал, что количество подтвержденных случаев заражения COVID-19 и индикатор настроений имеют коэффициент корреляции 0.84.

Авторы также исследуют тематику настроений по данным социальной сети Weibo с помощью широко используемых методов текстового анализа – модели латентного размещения Дирихле для кластеризации и случайного леса для задачи классификации корпуса постов.

В рамках латентного размещения Дирихле была получена оптимальная тематическая модель с 20 темами. Но для получения более интерпретируемых результатов похожие темы были объединены, а нерелевантные тематические группы были исключены из анализа. Таким образом были определены 7 тем о COVID-19: «предоставление сведений о событиях»; «популяризация профилактики и лечения»; «реакция правительства»; «личная реакция»; «мнения и настроения»; «обращение за помощью» и «пожертвования».

Случайный лес применяется в задаче классификации для разбиения тем «личная реакция», «мнения и настроения» и «обращение за помощью» на подтемы для выявления социального самочувствия населения в период начала распространения коронавируса. Были получены следующие подтемы: «страх и беспокойство»; «вопросы правительства и СМИ»; «осуждение вредных привычек»; «объективное комментирование ситуации»; «принятие научно обоснованных мер защиты»; «обращение за помощью пациентами»; «готовы вернуться на работу»; «оставаться дома и принимать необходимые меры предосторожности»; «популяризация противоэпидемических знаний в семье»; «обращение за медицинской помощью» и «прочее».

По результатам исследования был сделан ряд выводов. На ранней стадии пандемии из-за COVID-19 в Китае самой обсуждаемой темой стала «мнения и

настроения» с долей 34.42%. Это можно объяснить тем, что социальные сети являются важным источником информации, через который население обсуждало риски и делилось мнениями во время вспышки COVID-19. Доля других популярных тем «предоставление сведений о событиях», «популяризации профилактики и лечения» и «реакция правительства» составила более 60%, что говорит о том, что своевременное информирование правительством населения позволяло повысить эффективность борьбы с вспышкой коронавируса, устранить панику и стабилизировать настроения экономических агентов.

В статье Malla и Alphonse [28] использовалась MVEDL модель для определения твитов, относящихся к тематике COVID-19. Majority Voting-based Ensemble Deep Learning (MVEDL) – модель ансамблевого глубокого обучения на основе голосования большинством. В ансамбле были использованы предварительно обученные модели RoBERTa, BERTweet и CT-BERT. Далее на полученной выборке был проведен анализ тональности с помощью алгоритма TextBlob.

Авторы выделяют информативные твиты, содержащие информацию о количестве новых предполагаемых или подтвержденных случаев, количестве выздоровлений или смертей, а также о симптомах болезни пациентов. Твиты, содержащие ложную информацию о вирусе, называют фейковыми, а твиты, не относящиеся к COVID-19 – неинформативными. Задача авторов – отобрать информативные твиты из общего числа твитов. Корпус твитов был взят из работы Nguyen и др. [29]. Он построен с декабря 2019 по май 2020 на основе списка из 10 ключевых слов, связанных с коронавирусом.

Пандемия унесла миллионы жизней и привела к кризису в области здравоохранения. Для предотвращения этих последствий можно использовать информацию из социальных сетей. Например, для определения фейковых новостей, которые быстро распространяются среди пользователей, чтобы остановить их распространение. Дезинформация приводит к панике.

Когда эпидемия коронавируса начала распространяться все быстрее, Всемирная организация здравоохранения объявила, что изоляция людей и уменьшение числа социальных контактов – эффективный метод замедления распространения вируса. Аналогично предыдущей статье, в работе Dubey [30] данные Twitter использовались для исследования эмоций людей во время локдауна, связанного с вспышкой COVID-19.

Исследование проводилось в 12 странах, которые в марте 2020 года уже столкнулись с распространением вируса: США; Италия; Испания; Германия; Китай; Франция; Великобритания; Швейцария; Бельгия; Нидерланды и Австралия. Было сделано предположение, что Индия с населением 1,3 миллиарда человек может стать следующим эпицентром развития эпидемии после Китая, поэтому страна также была включена в исследование.

Ученые сделали предположение, что деление эмоций людей на два типа – положительные и отрицательные, не может хорошо описать настроения людей, поэтому расширили количество классов: страх; радость; ожидание; гнев; отвращение; печаль; удивление; доверие. NRC Word-Emotion Association Lexicon – словарь для определения слов, ассоциирующихся с эмоциями – содержит 10170 лексических единиц для выявления в текстах эмоций. Итак, в работе [30] использовался этот словарь для анализа тональности (положительная/отрицательная) и настроений людей (8 эмоций, описанных выше) на основе твитов, опубликованных с 11 по 31 марта 2020 года по ключевым словам «коронавирус», «COVID-19» и «оставайтесь дома».

По результатам анализа тональности авторы получили, что из этих 12 стран в Китае больше всего преобладала негативная тональность твитов – 55%, а в Бельгии (40 % негативных твитов), Индии и Австралии – позитивная. В остальных странах был примерный баланс. Это могло быть связано с тем, что Китай уже приближался к пику кривой распространения в то время, как в других странах распространение вируса только началось.

Далее проведен анализ настроений по 8 типам эмоциональной окраски. США, Франция и Китай имели наибольшее количество твитов с эмоциональной окраской «гнев», Швейцария – с тональностью печали и страха. Страна с наибольшим количеством твитов с доверием и удивлением – Бельгия, тогда как эмоций ожидания – Германия. Радость преобладала в постах людей, связанных с шутками и мемами по теме COVID-19.

Одна из задач анализа тональности текстов – облегчить процесс принятия решений с помощью учета настроений экономических агентов, реакций на какую-то меру и т.д.. Методы обработки естественного языка эффективно работают для структурированных текстов. А вопрос анализа твитов, которые содержат символы эмоджи и много URL-ссылок, остается открытым. Работа Pota и др. [31] посвящена задаче изучения тональности постов в социальной сети Twitter. Также, символы

эмоджи и хештеги могут указывать на настроение пользователя, поэтому их нужно учитывать при обработке. В статье изучалась обработка с помощью модели BERT. Также, твиты рассматривались на двух языках – английском и итальянском, чтобы оценить зависимость от языка. Корпус английского текста с классифицированными твитами на положительные, отрицательные или нейтральные был взят из работы Rosenthal и др. [32].

Первым делом, для анализа твитов использовались показатели количества лайков и репостов. На пики по максимальным значениям этих показателей приходились такие события, как введение локдауна по всей стране, максимальное количество смертельных случаев за период и экстренная помощь других стран, например, гуманитарная помощь от Китая для Индии. Далее были построены три облака слов: позитивные, негативные и нейтральные. Были слова, возникающие во всех темах: коронавирус; случаи заболевания; статистика. Позитивными словами, относящимися к коронавирусу, получились – здоровье, вакцина, больницы, а негативными – пандемия, смерть, страх. Модель BERT классифицировала твиты на 6 групп слов с разной эмоциональной окраской: грусть; злость; любовь; удивление; страх; радость. Анализ настроений проводится для выявления успешности мер, принятых правительством страны.

Другим кризисным событием, часто обсуждаемым как в СМИ, так и в социальных сетях, стал выход Великобритании из Европейского союза (Brexit). Например, в работе [33] были проанализированы 16 миллионов твитов британцев с 11 сентября по 31 октября 2019 на английском языке. Для построения индекса неопределенности в экономике проводилась оценка тональности корпуса твитов с помощью VADER⁴ и усреднялся полученный индикатор на ежедневной основе.

Далее авторы для анализа экономической ситуации в Великобритании в течение сентября-октября 2019 года сравнивают динамику полученного индекса неопределенности с доходностью фондового индекса FTSE 100 и фунта стерлингов 100. Авторы получили, что доходность фондового индекса FTSE 100 слабо связана с индикатором настроений (коэффициент корреляции (-0.169)).

Для фунта стерлинга была выявлена положительная и значимая связь с настроениями относительно Брекзита с коэффициентом корреляции 0.589.

⁴ Пакет в Python для определения тональности текста на основе заранее заданных словарей и правил.

Повышение обеспокоенности британцев относительно последствий Брекзита было сопряжено с обесценением фунта стерлингов.

На российских данных в работе Голощапова и Андреев [34] строится индикатор инфляционных ожиданий населения, и проводится текстовый анализ комментариев для оценки степени доверия населения к достижению цели по инфляции на основе постов в крупнейших новостных СМИ в социальных сетях «ВКонтакте» и Facebook в период 2014-2016 гг. Авторы производили отбор комментариев к новостным сообщениям об инфляции⁵ в сообществах «Ведомости», «Коммерсант», «РБК» и «Эксперт Онлайн». Всего было собрано 18537 новостей по инфляции. Интенсивность инфляционных ожиданий рассчитывается как количество комментариев о росте цен за месяц. Поскольку период исследования составил всего три года, авторы не проводят регрессионный анализ, а сравнивают полученную меру инфляционных ожиданий с опросным показателем – медианными оценками инфляционных ожиданий населения в следующие 12 месяцев.

Авторы получили, что меры инфляционных ожиданий высоко коррелированы и показатель на основе данных социальных сетей является опережающим на 1 месяц для опросного индикатора. Также были выявлены ключевые факторы⁶ инфляционных ожиданий на основе данных социальных сетей «ВКонтакте» и Facebook. В комментариях в качестве факторов формирования инфляционных ожиданий в 2014-2016 гг. были выделены: валютный курс; цены на сырьевые товары; денежно-кредитная политика; экономическая неопределенность; экономический кризис; внутренняя политика и мировая конъюнктура. В целом индикаторы инфляционных ожиданий показывают повышательную тенденцию в периоды валютного кризиса 2014-2015 гг.

В результате зарубежные и российские исследования показали, что данные социальных сетей являются ценным источником данных для анализа поведения интернет-пользователей. Настроения экономических агентов, полученные на основе текстового анализа постов в социальных сетях, становятся альтернативой для опросных и рыночных индикаторов. Следует отметить, что в рассмотренных исследованиях практически не проводилось сравнение прогностической силы

⁵ Перечень ключевых слов задан на основе типичных выражений, связанных с ростом цен, в том числе «инфляция», всевозможные сочетания «цена»+глагол.

⁶ Оценивается как доля комментариев по заданной теме к общему числу комментариев в корпусе.

настроений пользователей социальных сетей, настроений населения и участников финансовых рынков при прогнозировании различных экономических показателей.

В следующем подразделе будет приведен обзор эмпирических исследований, в которых проводится анализ влияния индикаторов настроений пользователей в социальных сетях на экономические и финансовые показатели, а также сравнивается качество моделей с традиционными индикаторами настроений или другими интернет-данными.

2.2 Эконометрические подходы к моделированию экономических показателей на основе социальных сетей

Как обсуждалось ранее, настроения экономических агентов могут быть получены за счет определения тональности постов или тематического моделирования. Однако ценность таких индикаторов может быть подтверждена только в рамках эконометрического анализа их влияния на экономические и финансовые показатели. По этой причине в данном разделе остановимся на конкретных эмпирических подходах, используемых для этих целей, и определим преимущества и недостатки интернет-данных, полученных из социальных сетей, по сравнению с опросными и рыночными индикаторами.

В последние несколько лет в социальных сетях все в большей степени обсуждаются действия и решения центральных банков в сфере денежно-кредитной политики. На текущий момент большинство центральных банков как развитых, так и развивающихся стран, имеют официальные аккаунты в Twitter. Остановимся более подробно на эмпирических исследованиях, в которых были выявлены информационные сигналы денежных властей и приведены оценки их влияния на поведение экономических агентов.

В работе [35] изучается реакция фондового рынка США на посты в Twitter о принятых FOMC решениях по денежно-кредитной политике за период 2007-2014 гг. Отбор постов проводится по ключевым словам «FOMC», «Федеральная резервная система», «Бернанке» и Йеллен.

Тональность постов определяется в интервале от -1 (сильно негативные) до 1 (очень позитивные) на основе словаря эмоциональной окраски, предложенного в [36]. Авторы для измерения охвата аудитории конкретного твита проводят взвешивание

полученного индикатора тональности поста по количеству подписчиков пользователя, разместившего пост.

Авторы получили следующие результаты. В дни заседаний Комитета по операциям на открытом рынке в течение 2007-2014 гг. доходность в среднем оказалась выше на 0.3 п.п. При позитивных настроениях в твитах доходность фондового рынка США также росла на 0.05 п.п. Важно отметить, что эффект от твитов по тематике ДКП усиливается в дни заседаний Комитета по операциям на открытом рынке. Авторы отмечают, что твиты в день заседания FOMC могут содержать дополнительную информацию, не учтенную в рыночных индикаторах. Рост количества твитов на одно стандартное отклонение будет сопряжено с ростом доходности на 0.62 п.п. Но при этом индикатор настроений пользователей оказывается незначимым.

Как было показано в статье, мудрость толпы может формироваться на основе постов в социальных сетях и оказывать влияние на цены финансовых активов. В таком случае текстовый анализ постов по экономической тематике может улучшить прогностическую способность существующих моделей, в том числе факторных моделей Фамы-Френча.

Прозрачность денежно-кредитной политики для экономических агентов – важный фактор эффективности политики и достижения доверия людей. Прозрачность помогает людям лучше понимать текущее состояние экономики и более плавно адаптировать свои экономические решения к применяемой в данный момент политике. Поэтому центральные банки стран всё больше делятся с населением информацией о проводимой политике, в том числе в социальных сетях. Исследование Gorodnichenko и др. [37] направлено на изучение влияния постов Федеральной резервной системы США (ФРС) в Twitter и Facebook на финансовую систему, в том числе анализ влияния на финансовые рынки и инфляционные ожидания, а также вовлеченности аудитории с помощью подсчета количества лайков, репостов и упоминаний.

Авторы провели анализ тональности с помощью словарного подхода, основанного на словаре VADER, который применим к настроениям в контексте социальных сетей. Далее авторы провели анализ сходства между публикациями ФРС в социальных сетях и их официальными заявлениями, а также пресс-релизами Федерального комитета по операциям на открытом рынке ФРС США (FOMC).

Применялось косинусное сходство, которое измеряет косинус угла между двумя векторами слов, содержащихся в каждом тексте. Затем публикации были классифицированы по темам: упоминание экономики; состояния финансовых рынков; фондового рынка; денежно-кредитной политики или заседания FOMC. Если в публикации присутствует хотя бы одно ключевое слово из темы – публикация относится к данной теме. Было отмечено, что ключевые слова в каждой теме являются взаимоисключающими, то есть ключевое слово в одной группе не входит в другую. Однако, одна публикация могла относиться сразу к нескольким темам.

В результате исследования авторы делают несколько выводов. Пользователи Twitter предпочитают упоминать аккаунты ФРС в своих записях, а не репост записи. Также, пользователи Twitter, как правило, упоминают ФРС, когда говорят об экономике. Что касается привлечения внимания пользователей, то посты с прикрепленными фотографиями или видео получают больше лайков и чаще публикуются. По количеству лайков и репостов наиболее популярны посты по теме денежно-кредитной политики.

Также было замечено, что сильная общественная реакция в социальных сетях практически не влияет на фондовый рынок. В частности, доходность S&P 500 растет с увеличением количества твитов по тематике денежно-кредитной политики, опубликованных в часы торговли, хотя подобный эффект не наблюдается в социальной сети Фейсбук. Однако, участники фондового рынка пересматривают свои среднесрочные инфляционные ожидания на основе информации, полученной из постов в социальных сетях, опубликованных с аккаунтов ФРС.

Следует отметить, что международный опыт говорит о том, что особую роль интернет-данные, в частности поисковые запросы, новостные статьи и посты в социальных сетях, могут играть в кризисный период. Экономические агенты проявляют значительный интерес к информации о происходящих событиях в условиях повышенной турбулентности на финансовых рынках, поскольку желают принять обоснованные решения и избежать потерь благосостояния. Так, в работе [38] показаны преимущества использования данных социальной сети Twitter при прогнозировании курса евро к доллару в период европейского долгового кризиса с 1 января 2012 г. по 27 сентября 2013 г. в рамках модели ARIMAX.

Авторы выгружают твиты, связанные с европейским кризисом, по выражениям «евро» + «кризис» + «ключевое слово». Список из 594 ключевых слов/фраз

составляется исходя из необходимости учета всевозможных экономических, социальных и политических проблем в Еврозоне и во всем мире. Например, задействованы ключевые слова, связанные с обеспокоенностью населения, потерей работы, иммиграционной политикой, денежно-кредитной политикой ЕЦБ и ФРС США, протестами, проблемами с долгами и нестабильностью на финансовых рынках, фискальной политикой, рисками дефолта, макроэкономическими показателями, фамилии политических лидеров или отдельных стран и т.д. По каждому выражению подсчитывается частота постов на ежедневной основе.

Далее автор сравнивает прогностическую способность каждого из полученных 594 индикаторов в модели ARIMAX с моделью случайного блуждания. По информационным критериям Акаике и Шварца наилучшими оказались модели с 17 ключевыми словами: банк; банковское дело; банки; долг; ЕЦБ; экономика; евро; Германия; Греция; греческий; Олланд; итальянец; Италия; Moody; риск; SP; Испания. Кроме того, автор для исследования возможного случайного влияния информации, содержащейся в корпусе твитов, на курс евро строят переменную $Rand$, как случайную комбинацию полученных 17 предикторов. На тестовом периоде модели сравниваются по среднеквадратичной ошибке прогноза

Как показывают результаты, наиболее точный вневыборочный прогноз курса евро к доллару был получен на основе частоты твитов по темам «риск», «Италия» и «SP». Это говорит о важности обсуждаемых в социальных сетях экономических рисков и решений по присвоению суверенных рейтингов агентством S&P в течение долгового кризиса. Кроме того, в топ предикторов входят частота твитов, в которых речь идет о долгах сильно закредитованных стран Европейского Союза, – Греции, Италии и Испании.

В целом результаты говорят о хорошей прогностической способности данных социальной сети при прогнозировании курса евро к доллару в течение европейского кризиса. В работе был изучен вклад различных экономических, социальных и политических проблем в колебания курса евро. Автором было показано, что интерес к экономической ситуации и обсуждение кризисных явлений может быть хорошим индикатором настроений населения в периоды нестабильности.

В работе Ozturk и Ciftci [39] также изучается взаимосвязь между настроениями экономических агентов в Twitter и обменным курсом доллара США к турецкой лире (USD / TRY) за период с 1 января по 31 декабря 2013 года.

Из Twitter выгружены сообщения, содержащие ключевые слова USD/TRY, #USD/TRY, Dollar, #Dollar, за 2013 год (с 1 января по 31 декабря). Твиты отсортированы по датам и трем категориям: положительные (обозначают посты, в которых говорится о том, что валюту нужно покупать); отрицательные (продавать) и нейтральные. Затем по каждой из категорий подсчитывается количество твитов для определения настроений инвесторов относительно торговой стратегии.

В качестве метода для оценки направленности изменения валютного курса выбрана логистическая регрессия. Авторы получили, что модель правильно предсказывает направление роста курса турецкой лиры с вероятностью 58%, а снижения курса – 82%. Данная работа показала наличие взаимосвязи между настроениями из социальных сетей и колебаниями валютного курса, а также то, что данные Twitter могут быть использованы для улучшения прогнозов валютного курса.

Написать пост в социальных сетях – один из наиболее простых способов выразить свое мнение на сегодняшний день, поэтому социальные сети являются хорошим источником информации для оценивания настроений инвесторов.

В работе [40] исследуется влияние настроений из социальной сети Twitter на дневную доходность акций, отбор которых делается следующим образом. На предварительном этапе исследования был получен доступ к высокочастотным данным по акциям от норвежского банка Netfonds. Далее авторы выбирают наиболее торгуемые и обсуждаемые в социальных сетях акции, и выгружают данные по этим акциям за август 2018 года. Выбранные акции: Exxon Mobil (XOM); General Electric (GE); Chesapeake Energy (CHK); Ford Motor Company (F); Disney (DIS) и Walmart (WMT). Авторы выгрузили данные по выбранным акциям для различных временных интервалов, чтобы вычислить 1-, 5- и 30-минутные доходности.

Построение индикаторов настроений инвесторов осуществляется на основе определения эмоциональной окраски твитов. Авторы применяют подход Канадского национального исследовательского совета (NRC) к определению тональности текста, что позволяет выделить следующие термины-эмоции: злость; предчувствие; отвращение; страх; радость; грусть; удивление; доверие; позитивное настроение; негативное настроение; индекс SMOG, измеряющий количество бессмыслицы в тексте, т.е. отражающий его читаемость.

Для выявления настроений пользователей в Twitter, связанных с конкретной акцией, авторы выгружают твиты с тикером компании. Тикеры начинаются либо с

тега "#" (хештег), либо с тега "\$" (кэштег). Кэштеги пишут инвесторы, которые определяют себя как «разбирающиеся в финансах лучше, чем среднестатистический человек». Кэштеги предназначены для профессиональных инвесторов и финансовых специалистов.

Для эмпирического анализа используется простая модель ценообразования активов в виде модификации модели CAPM (Capital Asset Pricing Model) с настройками инвесторов, полученными из твитов.

В результате для каждой из шести рассмотренных акций выявлена значимая реакция на настроения пользователей социальной сети Twitter относительно конкретной акции, а также некоторая реакция на настроения на фондовом рынке.

Авторы сделали вывод, что участникам фондового рынка требуется время, чтобы обработать информацию и сформировать инвестиционные стратегии. По этой причине они использовали лаги настроений инвесторов.

Следует отметить, что важной особенностью работы являются различные результаты (в частности, несоответствие знаков) для различных интервалов доходностей. Негативные настроения могут отражать негативные новости, снижая цены акций. Также, негативные настроения могут свидетельствовать о временном падении цены акции и стимулировать спекулятивную торговлю, если профессиональные аналитики не видят серьезных поводов для снижения цены акции.

Кроме цен акций и валютного курса индикаторы настроений, полученные из социальной сети Twitter, дают возможность моделирования колебаний цены на нефть и золота. В исследовании [41], используя объемы поисковых запросов Google и настроения Twitter, авторы моделируют цены на нефть и золото, а также показатели валютного и фондового рынков.

В качестве данных используются твиты со 2 июня 2010 года по 13 сентября 2011 года. Для построения индикаторов настроений экономических агентов определяется тональность каждого твита на основе наивного байесовского классификатора.

Авторы получили, что модели с настройками инвесторов имеют более высокую предсказательную способность для всех рассматриваемых активов. Это свидетельствует о том, что настроения инвесторов из социальных сетей и объемы интернет-запросов имеют хорошую прогностическую силу.

Уменьшение асимметрии информации между инвесторами и фирмами может повлиять на цену акций, особенно во времена неопределенности. В работе Albarrak и др. [42] изучается влияние решения фирм публиковать финансовую информацию в социальной сети Twitter на издержки собственного капитала.

Асимметрия информации вызывает проблему неблагоприятного отбора и снижает ликвидность акций. Поэтому, фирмы выпускают акции с дисконтом. Чтобы уменьшить стимулы инвесторов к получению закрытой информации, фирмы раскрывают свою финансовую информацию. И когда инвесторы меньше обеспокоены асимметрией информации, ликвидность акций растет.

Можно сделать вывод, что фирмы, которые распространяют больше финансовой информации, имеют более низкие издержки на капитал. По результатам работы [42] можно сделать следующий вывод. Фирмы могут улучшить коммуникацию с инвесторами с помощью публикаций новостей в Twitter. Из-за того, что инвесторы будут тратить меньше времени и энергии на получение новостей о фирме, уменьшится асимметрия информации и увеличится узнаваемость фирмы среди инвесторов.

На основе данных социальных сетей можно исследовать не только реакцию участников финансовых рынков на отдельные кризисные эпизоды, но и оценивать влияние неопределенности экономической политики на экономику. Например, в исследовании [43] предложен подход к оценке индекса неопределенности экономической политики⁷ на основе твитов 15 крупнейших СМИ⁸ в Чили в период с января 2012 г. по март 2020 г. Отбор твитов осуществляется по ключевым словам, связанным с экономикой Чили, фискальной, денежно-кредитной и торговой политикой в Чили, неопределенностью или общей экономической ситуацией в Чили. Категория «общая экономическая ситуация» характеризует неопределенность в периоды гражданских протестов 2019-2020 гг. и начало пандемии.

Поскольку авторы используют три типа источника информации – аккаунты 5 новостных каналов, 7 новостных газет и 3 радиостанций, то по каждому из них производится подсчет частоты твитов, в которых встречается хотя бы одно из

⁷ Применяется аналогичная методология построения индекса неопределенности экономической политики, как у Baker и др.[68], где проводился текстовый анализ новостей.

⁸ Рассматриваются аккаунты наиболее популярных СМИ в Twitter – 5 новостных каналов (CHVNoticias, Puntoticias, CNNChile, 24HorasTVN и T13), 7 новостных газет ((Emol, EYN_ELMERCURIO, elmostrador, DFinanciero, pulso_tw, Estrategia.cl и latercera) и трех радиостанций (adnradiochile, biobio и cooperativa).

ключевых слов. Затем три полученных ряда стандартизируются. Агрегированный индекс неопределенности экономической политики (DEPUC) рассчитывается как простое среднее трех индексов по различным типам источника информации и сглаживается 7-дневным скользящим средним для избегания сильных дневных колебаний.

Далее авторы осуществляют интерпретацию полученных результатов и описывают преимущества их методологии. Для этого они также делают оценки индекса неопределенности экономической политики (DEPU), как у Baker и др. [44].

Модификация методологии построения индекса основывается на предположении о том, что важную роль для Чили играет общеэкономическая ситуация в периоды протестов и пандемии коронавируса.

Во-первых, оба индекса неопределенности экономической политики имеют схожую динамику. Корреляция для периода 2012-2020 гг. между DEPU и DEPUC составляет 0,64. Коэффициент увеличивается до 0,77, если исключить период пандемии коронавируса. Это происходит из-за повышения индекса DEPUC более чем на 2 стандартных отклонения, потому что повышается обеспокоенность населения относительно заражения коронавирусом и последствия пандемии активно обсуждаются в социальных сетях.

Во-вторых, оба индекса неопределенности экономической политики имеют пики, согласующиеся с различными эпизодами глобальной и локальной экономической неопределенности: европейским долговым кризисом; обнародованием законопроекта о налоговой реформе в Чили в апреле 2012 г.; президентскими выборами в конце июня 2013 года; вторым туром в середине декабря 2013 года, одобрением законопроекта о налоговой реформе в середине мая 2014 г.; взрывом в сентябре 2014 г. в торговом центре Сантьяго недалеко от станции метро; периодами гражданских протестов 2019 года и пандемией COVID-19 в начале 2020 года. В целом индикатор неопределенности экономической политики с учетом общей экономической ситуации (DEPUC) показывает наилучший результат в периоды протестов и пандемии коронавируса, поэтому он и используется в дальнейшем в эконометрическом анализе.

В рамках эконометрического анализа проводится оценка влияния индекса DEPUC на курс песо к доллару с помощью двухрежимной модели с марковскими переключениями

Индекс неопределенности экономической политики оказался значимым во всех спецификациях. При повышении неопределенности экономической политики происходит обесценение чилийского песо к доллару. Эффект сильнее при переходе в режим высокой изменчивости курса песо.

Социальные сети также могут быть полезными для оценки экономической активности, инфляционных ожиданий и анализа поведения экономических агентов при изменении их благосостояния при потере работы. В исследовании [45] предлагается подход к оценке индекса потери работы на данных социальной сети Twitter для США в период с июля 2011 г. по ноябрь 2013 г. на еженедельной основе. Отбор постов в Twitter основывался на предположении о том, как люди могут комментировать увольнение с работы, и был задан следующий перечень из 10 ключевых слов: «сократили»; «сокращенный»; «аутсорсинг»; «розовый лист»; «потеря работы»; «уволенный с работы»; «быть уволенным»; «потерявший работу» и «безработица». Для каждого ключевого слова рассчитывается количество твитов, в которых они упоминаются. В итоге индекс потери работы строится как первая главная компонента по полученным 10 индикаторам.

В работе полученный индекс потери работы сравнивается с первоначальными требованиями на пособие по безработице, чтобы попытаться определить, могут ли данные социальных сетей стать опережающим индикатором экономической активности. Авторы отмечают, что динамика обоих индикаторов оказалась схожей за счет общего тренда в течение июля 2011г. – ноября 2013 г. и выраженных пиков, например, осенью 2012 г. после урагана Сэнди.

Также авторы приводят результаты регрессионного анализа для моделирования первоначальных требований на основе данных социальных сетей, консенсус прогноза профессиональных аналитиков⁹ и лага первоначальных требований. Эконометрический анализ проводится для предварительных и пересмотренных данных по первоначальным заявкам на пособие по безработице.

При использовании предварительных и пересмотренных данных первоначальных заявок авторы получили, что индекс потери работы не является наилучшим предиктором по сравнению с лагом первоначальных заявок на пособие по безработице или их консенсус-прогноза. Это может свидетельствовать о

⁹ Медианное значение прогнозов 50 профессиональных аналитиков, опрашиваемых Блумбергом.

невозможности улучшить прогноз и так довольно высокочастотного индикатора первоначальных заявок на пособие по безработице, публикуемого на еженедельной основе. Но при этом индекс потери работы значим в модели, включающей все предикторы, следовательно, может содержать дополнительную информацию о рынке труда, не учтенную в консенсус-прогнозах и исторических данных.

Еще одним подходом к измерению экономической активности на основе данных социальных сетей является построение индикаторов уверенности потребителей или потребительских настроений. Daas и Puts в работе [46] предлагают методологию построения настроений населения в Нидерландах с июня 2010 г. по ноябрь 2013 г. по обширной базе постов на различных платформах в Интернете. Базу данных по общедоступным постам собирает компания Coosto. Сбор корпуса постов на голландском языке осуществляется практически на всех онлайн платформах, где население может выразить свое мнение, в том числе Twitter, Facebook и LinkedIn, а также на различных веб-сайтах, форумах и в блогах. Индикатор настроений рассчитывается как разница между долей классифицированных негативных и позитивных постов.

Авторы сравнили полученный индикатор настроений на основе социальных сетей с индексом потребительской уверенности населения. Показатели оказались высоко коррелированными с коэффициентом корреляции 0.9. Наиболее важными источниками информации оказались Facebook и Twitter. Для анализа причинности между индикатором социальных сетей и индексом потребительской уверенности был применен тест Грейнджера. Результаты показали, что изменения индекса потребительской уверенности предшествуют изменениям настроений населения, полученных на основе данных социальных сетей. Из этого авторы делают вывод, что после изменения потребительских настроений происходит повышение интереса к происходящим событиям в социальных сетях.

В другой работе Daas с соавторами [47] используется аналогичный подход для построения индикатора потребительских настроений в Нидерландах с июня 2010 г. по март 2015 г. на основе данных Facebook и Twitter. В данном случае авторы уже на более широком временном интервале 2010-2015 гг. проводят регрессионный анализ

влияния социальных сетей на индекс потребительской уверенности с помощью модели структурных временных рядов¹⁰.

Результаты работы показали, что двумерная модель с индексом социальных сетей позволяет лучше прогнозировать индекс потребительской уверенности по сравнению с одномерной.

Также существуют исследования, где оценка настроений потребителей в социальных сетях делается с помощью словаря позитивных и негативных слов. Orsal и Sturn в исследовании [48] строят индикатор настроений населения на основе постов в Twitter для Германии с мая 2019 г. по август 2020 г.

Отбор твитов проводится по ключевым словам – «экономика», «покупка» и «купить», поскольку они часто употребляются при опросах респондентов для построения индикатора потребительской уверенности. Для определения тональности использовался словарь SentiWS для немецкого языка. В итоге авторы вычисляют 3 индикатора настроений потребителей: агрегированный показатель по трем ключевым словам ($Score^{всe}$); «экономика» ($Score^{экономика}$) и «покупка и купить» ($Score^{покупка}$).

В результате авторы делают вывод о том, что уловить изменения настроений экономических агентов можно в немецких социальных сетях без значительных затрат на опросы населения и на более высокочастотной основе.

Также существуют исследования, в которых предлагаются подходы к оценке инфляционных ожиданий на основе данных социальных сетей. В работе [49] инфляционные ожидания строятся на основе постов в Twitter для Италии с июня 2013 г. по 31 декабря 2019 г. Отбор твитов на итальянском языке проводится по ключевым словам, связанным с динамикой цен, арендной платой, счетами, бензином и ценами на нефть.

Авторы отмечают, что сложно определить, отражают ли выбранные твиты инфляционные ожидания, так как в них могут обсуждаться ожидаемая инфляция или наблюдаемое изменение цен. Воспринимаемая инфляция, конечно, не характеризует напрямую ожидаемую инфляцию, но эта информация учитывается в процессе формирования ожиданий экономическими агентами.

¹⁰ Один из вариантов модели пространства состояний, разработанной Дарбиным и Купманом в 2012 г. Уравнение измерений строится для объясняемой переменной с включением сезонности, тренда и регрессионной компоненты. Переменные состояний задаются как случайное блуждание. Для оценки параметров модели применяется метод максимального правдоподобия и фильтр Калмана.

Кроме того, авторы обнаружили, что твиты из исходного корпуса можно отнести к различным темам, не всегда относящимся к инфляции или дефляции. Например, корпус твитов также включает рекламные объявления, предложения по продаже товаров в интернете или твиты, в которых ключевые слова на итальянском языке могут использоваться в другом контексте. По этой причине применяется трехэтапная процедура для фильтрации и построения индикаторов инфляционных ожиданий. На первом этапе для исключения всей нерелевантной информации применяется тематическое моделирование. Для этих целей оценивается модель латентного размещения Дирихле с 50 заданными темами.

Модель латентного размещения Дирихле позволила выявить две подходящие темы с твитами по инфляции и дефляции. Тема 19 характеризует твиты, в которых речь идет о инфляции, заработных платах, дефляции, евро, цене, ценах и занятости. Тема 36 содержит твиты о ценах, распродаже, инфляции, ценах, нефти и фондовом рынке. В итоге фильтрация с помощью тематического моделирования уменьшает корпус на 86% до 1534743 твитов.

На втором этапе авторы строят на основе финального корпуса твитов два индекса изменения динамики цен – «рост инфляции» и «снижение инфляции». Индексы рассчитываются как частота появления твитов, в которых содержатся биграммы и триграммы¹¹ по тематике «инфляция» и «дефляция». Также накладывается дополнительное условие на выявление наиболее релевантных твитов – биграмма или триграмма должны встречаться в среднем не менее чем в 100 твитах ежегодно в течение 2013-2019 гг. Всего авторы использовали 10691 биграмм и 2853 триграмм. Из них для определения направленности изменения цен рассматривались 121 биграмма и 112 триграмм для индекса «повышение цен», а также 111 биграмм и 126 триграмм для индекса «снижение цен».

Авторы отмечают, что подсчет частоты дает возможность получить представление о том, что чем больше пользователи в социальных сетях публикуют твиты на какую-то тему, тем больше вероятность повлиять на ожидания других экономических агентов. Следовательно, чем значительнее повышение интереса к тематике «изменения цен в будущем», тем сильнее эффект распространения новостей

¹¹ N-грамма – это словосочетания из подряд встречающихся слов в тексте. Так биграмма – это словосочетание из двух слов, а триграмма – словосочетание из трех подряд встречающихся слов, характеризующих в данном случае «инфляцию» или «дефляцию».

в социальных сетях. Кроме того, важную роль в процесс формирования инфляционных ожиданий экономических агентов играет обсуждение возможного направления изменения цен в опубликованных твитах. Когда пользователи сообщают о повышении суммы чека, это может сигнализировать о повышении инфляции. В то же время обсуждение итальянцами падения цен на нефть может говорить об ожиданиях более низкой инфляции.

На третьем этапе на основе имеющихся индикаторов ожиданий по повышению и снижению инфляции строятся 4 агрегированных показателя инфляционных ожиданий населения. Первый индикатор вычисляется как разница между индикаторами «рост инфляции» и «снижение инфляции». Полученный индекс стандартизируется и сглаживается с использованием скользящей средней (МА) за 10, 30 и 60 дней.

Второй индикатор инфляционных ожиданий рассчитывается, как и предыдущий, но корректируется для извлечения информации, не содержащейся в сообщениях ЕЦБ или в предварительных релизах ИПЦ в Италии и Германии. Для этого оценивается регрессионная модель инфляционных ожиданий, где в качестве объясняющих переменных выступают дамми-переменные, принимающие значение 1 в даты релизов предварительных данных по ИПЦ в Италии и Германии, в дни пресс-конференций ЕЦБ и выступлений представителей ЕЦБ, а также фиктивная переменная, принимающая значение 1, 29 августа 2014 г., когда годовая инфляция в Италии впервые после 1959 г. стала отрицательной. Итоговый индикатор представляет собой остатки ранее описанной регрессии. Данные стандартизируются и сглаживаются с помощью скользящей средней за 10, 30 и 60 дней.

Третий индикатор строится как экспоненциальное сглаживание первого индекса инфляционных ожиданий населения.

Четвертый индикатор инфляционных ожиданий рассчитывается как разность натуральных логарифмов индекса «роста цен» и «снижения цен». Индикатор стандартизируется и сглаживается с помощью скользящей средней за 10, 30 и 60 дней.

В дальнейшем авторы проводят эконометрический анализ влияния полученных мер инфляционных ожиданий, построенных на основе твитов, на опросный индикатор инфляционных ожиданий населения (ежемесячная частота) и рыночный индикатор – процентную ставку по годовым инфляционным свопам (ежедневная частота).

Все индикаторы инфляционных ожиданий, рассчитанные по данным Twitter, значимо и положительно связаны с ожиданиями населения, основанными на опросах с ежемесячной периодичностью. Чем выше интерес пользователей в социальной сети к тематике инфляции, тем выше инфляционные ожидания населения.

Рассматриваемые индикаторы Twitter объясняют от 23 до 43% дисперсии инфляционных ожиданий населения, основанных на опросах. Индикаторы 4-го типа имеют самые высокие показатели R^2 при самой большой ширине окна, что свидетельствует о том, что ожидания пользователей социальных сетей могут стать альтернативой для низкочастотных изменений инфляционных ожиданий.

Индикаторы на основе твитов также оказывают положительное влияние и объясняют от 19 до 54% изменчивости ставок по инфляционным свопам на горизонте 1 года. Авторы делают из этого вывод, что индикаторы на основе Twitter позволяют отслеживать изменение инфляционных ожиданий с ежедневной периодичностью, а не только на ежемесячной основе. Как и в случае инфляционных ожиданий на основе опросов, индикатор 4-го типа имеет более высокую объясняющую способность при моделировании процентных ставок по инфляционным свопам.

Авторы показали, что с помощью методов машинного обучения и данных социальных сетей можно построить инфляционные ожидания потребителей. Такой подход позволяет своевременно на высокочастотной основе и без лага публикации наблюдать за инфляционными ожиданиями. Кроме того, эти результаты свидетельствуют о том, что информация, распространяемая экономическими агентами в социальных сетях, может стать ценным ресурсом при формировании инфляционных ожиданий в краткосрочной перспективе.

В российской литературе можно выделить исследование [50], в котором проводится построение индексов экономической активности на основе поисковых запросов Google Trends, новостных статей интернет-изданий и данных социальной сети «ВКонтакте». Отбор данных проводится по регулярным выражениям: курс доллара; евро; девальвация; банк; цб; ртс; ммвб; акции; дефолт; финансовый кризис; банкротство; ипотека; взять кредит; ПИФ; залог. Новостные статьи были выгружены с сайтов «РБК», «РИА Новости», «ТАСС», «Интерфакс» и «Лента.ру». Посты и комментарии были собраны из новостных сообществ социальной сети «ВКонтакте»: «Ведомости»; «Интерфакс»; «Коммерсант»; «Лента»; «Медуза»; «РБК»; «Российская газета»; «РИА Новости»; RT и «ТАСС».

Автор использовал различные индексы экономической активности, построенные на основе интернет-данных, при наукастинге и прогнозировании на следующий месяц различных экономических и финансовых показателей с помощью ARIMAX модели. Автор показал, что в среднем индикаторы настроений на основе новостных статей и интернет-запросов имеют более высокую предсказательную силу при наукастинге и прогнозировании экономических и финансовых показателей, по сравнению с опросными индикаторами экономической активности.

В целом эконометрический анализ влияния настроений и ожиданий на экономические и финансовые показатели позволяет сделать следующие выводы. Социальные сети, как общедоступный онлайн ресурс, могут быть хорошим каналом передачи сигналов денежных властей о будущей направленности денежно-кредитной политики. Следовательно, они могут снижать асимметрию информации для участников финансовых рынков, а также повысить предсказуемость действий монетарных властей и снизить неопределенность в экономике.

Согласно международному опыту, особую важность посты в социальных сетях имеют в периоды кризисов. В такой ситуации инвесторы и население предъявляют повышенный спрос на информацию о происходящих событиях в условиях неопределенности на финансовых рынках, поскольку боятся значительных потерь благосостояния. Социальные сети дают возможность отслеживать поведение пользователей в реальном времени, а также определять причины обеспокоенности и учесть это при прогнозировании экономических и финансовых показателей.

Анализ постов в социальных сетях также позволяет уловить изменения настроений и ожиданий экономических агентов, а следовательно, получить альтернативу традиционным показателям. Как показывают рассмотренные исследования, настроения пользователей социальных сетей могут полностью заменить опросные индикаторы, полученные с помощью опросов населения.

В следующем разделе будут представлены результаты анализа данных социальной сети «ВКонтакте», а также будет проведено сравнение различных мер настроений и ожиданий экономических агентов.

3 Анализ социальных сетей на российских данных

Обзор эмпирических исследований показал, что данные социальных сетей могут стать ценным источником информации для общественности, что непосредственно позволяет изучить поведение экономических агентов. По этой причине в данном разделе будет проведен анализ данных социальной сети «ВКонтакте» и приведены результаты оценки различных индикаторов настроений пользователей Интернета.

3.1 Описание статистических данных и первичная подготовка данных социальных сетей

В рамках данного исследования будет рассматриваться только социальная сеть «ВКонтакте», среди преимуществ которой можно выделить довольно простую структуру API и бесплатный доступ¹², а также высокую степень популярности у российской аудитории в течение последних десяти лет.

Для построения индикаторов настроений на валютном рынке, индекса экономических условий и инфляционных ожиданий экономических агентов в текущем месяце были выбраны новостные сообщества, которые потенциально могут стать общедоступным источником информации для пользователей. Сбор данных с 2010 г. по 9 сентября 2021 г. осуществлялся из новостных сообществ «ВКонтакте»: «РИА Новости»; Forbes; «РБК»; «ТАСС»; «Ведомости»; «Лента.ру» и «Коммерсант». В следующем подразделе будут рассмотрены подходы к расчету индикаторов настроений экономических агентов на основе постов и комментариев новостных сообществ «ВКонтакте», а также сделаем сравнение с альтернативными способами – поисковыми запросами и новостными статьями интернет-изданий.

3.2 Построение ожиданий населения на основе данных социальных сетей

Интернет-данные могут стать хорошими прокси-переменными для анализа поведения экономических агентов в условиях неопределенности. Особый интерес представляет оценка количественных мер настроений на основе общедоступных

¹² В бесплатном доступе к архиву баз данных Twitter было отказано.

источников информации, которые доступны в реальном времени и с минимальными затратами, в отличие от опросных¹³ и рыночных¹⁴ индикаторов. В данном подразделе оцениваются три типа индикаторов на основе интернет-данных: инфляционные ожидания; настроения на валютном рынке и индекса экономических условий.

Инфляционные ожидания в текущем месяце на основе постов новостных сообществ (*inf_posts*) и новостных статей РБК¹⁵ (*inf_rbc*) строятся как доля публикаций по тематике инфляция в общем количестве. Отбор постов/новостных статей осуществляется на основе регулярных выражений по тематике инфляции: инфляция; ипц; покупательная способность; завышение цен; цены на продукты питания и всевозможные словосочетания о динамике цен (рост, снижение, удорожание или прилагательное/глагол).

Выбор поисковых запросов Google Trends проводится по ключевым словам¹⁶: инфляция; рост цен; цены выросли; цены растут; повышение цен; цена на газ; цена на бензин; цены на продукты; индекс потребительских цен; индекс цен; ипц. Мера инфляционных ожиданий на основе интернет-запросов (*inf_gt*) осуществляется с помощью метода главных компонент¹⁷.

Построенные индикаторы инфляционных ожиданий и сезонно скорректированная инфляция к предыдущему месяцу (*inflation_mom_sa*) представлены на (*рисунке 1*).

¹³ Опросы различных экономических агентов являются дорогостоящими и требуют значительного времени для агрегирования и анализа данных.

¹⁴ В развивающихся странах может быть недостаточно данных для расчета рыночных индикаторов, например, довольно сложно с оценками безубыточного уровня инфляции по данным государственных облигаций, индексируемых на инфляцию, который также можно было бы рассчитывать в реальном времени и на высокой частоте с минимальной задержкой.

¹⁵ РБК является одним из популярнейших интернет-изданий по экономической тематике, имеющий простую и доступную структуру для парсинга полнотекстовых новостных статей.

¹⁶ Первоначально список интернет-запросов по ценовой динамике был значительно шире. Из 45 ключевых слов были исключены поисковые запросы с нулевыми значениями больше 10% и отсутствием данных при небольшом количестве запросов.

¹⁷ Первая главная компонента хорошо аппроксимирует набор данных. Интернет-запросы сезонно скорректированы.

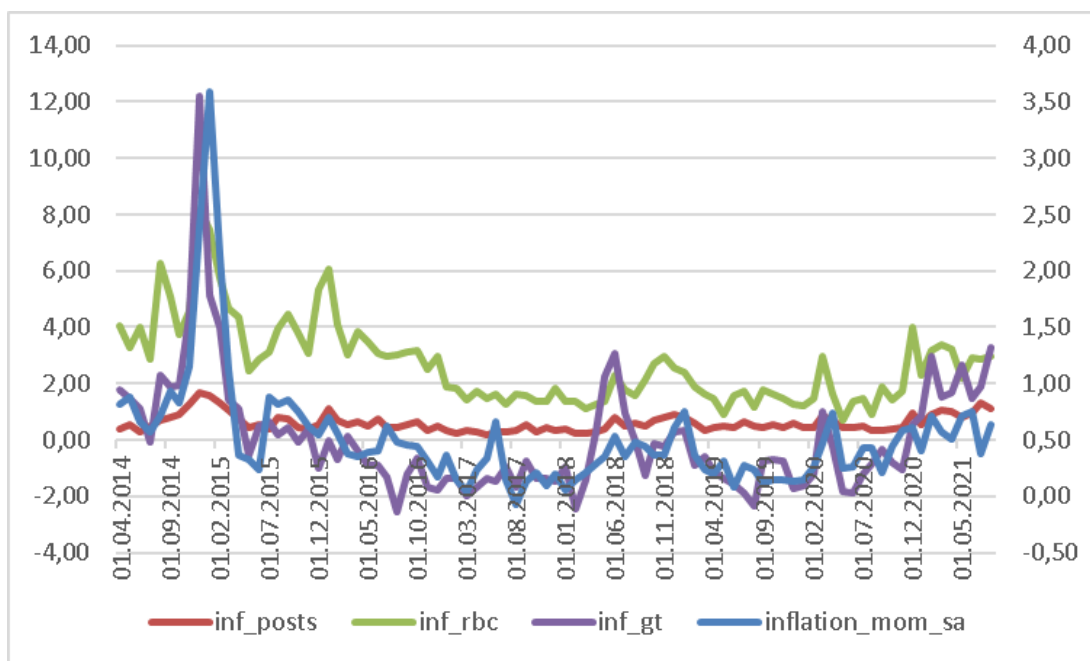


Рисунок 1. Динамика инфляционных ожиданий и сезонно скорректированной инфляции

Примечание. Левая ось – меры инфляционных ожиданий, правая ось – фактическая инфляция. Источник: по расчетам авторов

В (таблице 1) представлены коэффициенты корреляции мер инфляционных ожиданий на основе интернет-данных без лага и с лагом с сезонно скорректированной инфляцией. По сделанным оценкам, полученные индикаторы положительно коррелируют с инфляцией на уровне 0.72-0.76 без лага и 0.7-0.85 с одним лагом.

Таблица 1

Корреляция мер инфляционных ожиданий экономических агентов с сезонно скорректированной инфляцией

Инфляция	inf_vk	inf_gt	inf_rbc
Без лага	0,718	0,749	0,764
Лаг1	0,692	0,847	0,722
Лаг2	0,517	0,620	0,509
Инфляционные ожидания населения инФОМ	inf_vk	inf_gt	inf_rbc
Без лага	0,456	0,388	0,777
Лаг1	0,485	0,465	0,807
Лаг2	0,476	0,524	0,794

Источник: составлено авторами

Следует отметить, что сильно коррелирует с инфляционными ожиданиями населения только индикатор, построенный на основе новостных статей РБК.

Построение инфляционных ожиданий населения проводится при подсчете количества¹⁸ комментариев¹⁹ тремя способами: по всем комментариям к постам по тематике инфляции (inf_count_st); по отобранным комментариям только о динамике цен (inf_count_sort_st) и комментариям о динамике цен по всему корпусу постов (inf_count_all_st). Для дальнейшего анализа данные стандартизируются. Сравнение индикаторов инфляционных ожиданий на основе комментариев социальной сети «ВКонтакте» и медианных значений инфляционных ожиданий населения по данным инфОМ представлены на (рисунке 2).

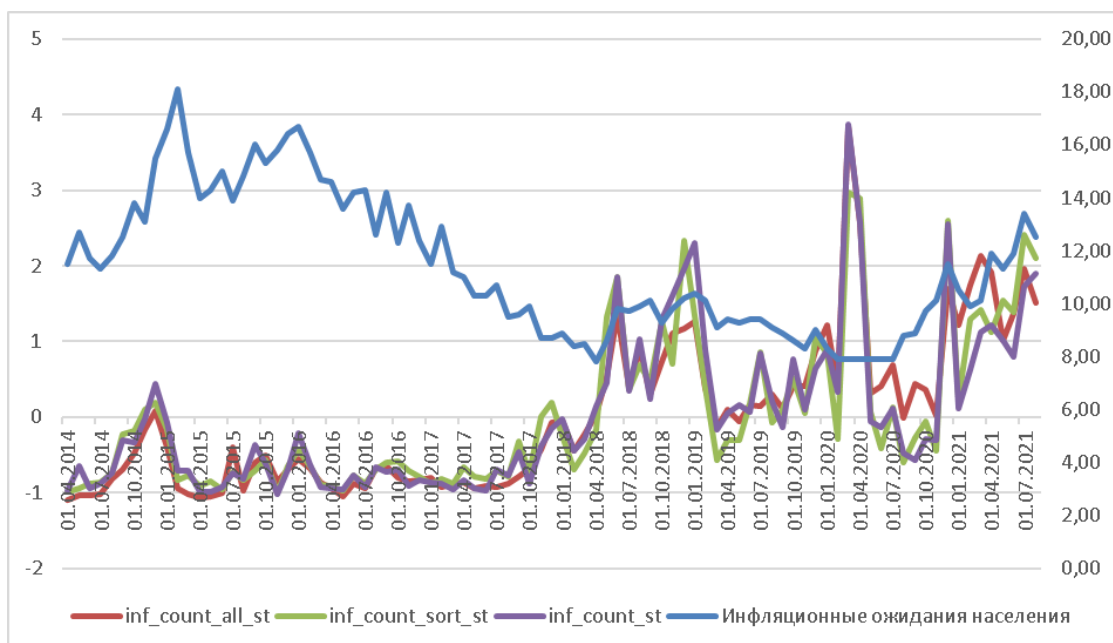


Рисунок 2. Динамика инфляционных ожиданий на основе комментариев и опросов инфОМ

Источник: по расчетам авторов

В (таблице 2) представлены коэффициенты корреляции частотных индексов на основе комментариев по инфляции с сезонно скорректированной инфляцией и инфляционными ожиданиями населения. По проведенным расчетам видно, что индикаторы на основе комментариев не коррелируют с фактической инфляцией и слабо отрицательно коррелируют с инфляционными ожиданиями населения.

¹⁸ Подсчет количества комментариев о динамике цен может отражать беспокойство населения относительно роста цен.

¹⁹ Также проводился анализ тональности комментариев для построения индикаторов настроений с помощью предобученных нейросетевых моделей DeepPavlov: Rusentiment многоязычный BERT, Rusentiment разговорный BERT, Twitter Mokron и Rusentiment CNN Fasttext. Построенные индикаторы слабо коррелируют с соответствующими макроэкономическими показателями, по этой причине не рассматриваются в эконометрическом анализе.

Таблица 2

Корреляция мер инфляционных ожиданий на основе комментариев с инфляционными ожиданиями населения

Инфляция	inf_count_all	inf_count_sort	inf_count
Без лага	-0,042	0,018	0,034
Лаг1	0,000	0,032	0,045
Лаг2	-0,033	-0,006	-0,013
Инфляционные ожидания населения инФОМ	inf_count_all	inf_count_sort	inf_count
Без лага	-0,476	-0,375	-0,394
Лаг1	-0,452	-0,368	-0,396
Лаг2	-0,440	-0,379	-0,405

Источник: составлено авторами

В целом из графического анализа можно сделать вывод, что большая часть индикаторов инфляционных ожиданий отражает обеспокоенность экономических агентов в периоды неопределенности, включая валютный кризис 2014-2015 г. в России, нефтяных шоков и период пандемии. Особенно чувствительными к периоду марта 2020 г. – августа 2021 г. оказались индикаторы, построенные на основе комментариев, что может говорить о высоком уровне обеспокоенности роста цен на продукты питания в условиях пандемии коронавируса²⁰. В результате меры инфляционных ожиданий на основе постов, новостных статей и поисковых запросов, поскольку, по мнению авторов, являются более адекватными с экономической точки зрения, чем на основе комментариев.

Далее рассмотрим подходы к построению настроений на валютном рынке и индекса экономических условий. Для этих целей будет использоваться предобученная модель²¹ LDA²² с 100 темами, построенная на новостных статьях с сайта РБК с января 2010 г. по март 2020 г. Такой подход позволяет без задания перечня ключевых слов определить тематику новостной статьи или поста «ВКонтакте» по заранее определенным темам. Фактически в качестве меры настроений пользователей рассматривается вероятность принадлежности новостной статьи или поста к выбранной теме за определенный период времени. Интерпретация таких показателей

²⁰ Также этот эффект может быть частично связан с увеличением аудитории.

²¹ Проводилось тематическое моделирование корпуса постов, но построенная модель латентного размещения Дирихле оказалась хуже предобученной на новостных статьях РБК.

²² Подготовлена в ходе выполнения научно-исследовательской работы по теме: «Прогнозирование экономических показателей на основе анализа новостей в средствах массовой информации», выполняемой в рамках государственного задания РАНХиГС на 2020 год,

настроений подразумевает, что повышение интереса СМИ к экономической тематике скорее всего связано с плохими новостями, что непосредственно может повлиять на процесс принятия решений экономическими агентами. Поскольку новости на сайтах интернет-изданий и в социальных сетях являются общедоступным источником информации, то и поиск информации в Интернете также увеличится. В результате строятся индикаторы по постам новостных сообществ в «Вконтакте», РБК и поисковых запросов Google Trends с 1 января 2013 г. по 31 августа 2021 г. Облака слов для тем «валютный рынок» и «экономические условия» проиллюстрированы на (рисунках 3 и 4).

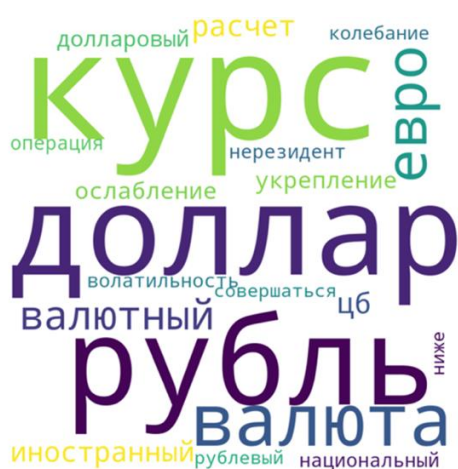


Рисунок 3. Облако слов по теме «Валютный рынок»
Источник: по расчетам авторов

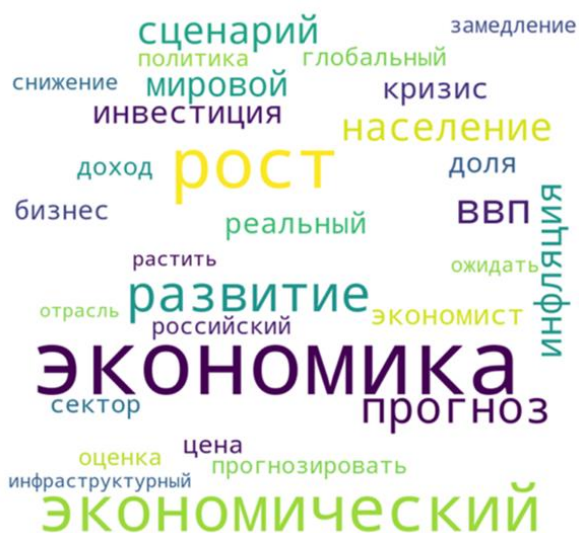


Рисунок 4. Облако слов по теме «Экономические условия»
Источник: по расчетам авторов

Ключевые слова для построения настроений на основе поисковых запросов основываются на полученных распределениях слов по темам. С учетом сортировки поисковых запросов для индекса экономических условий были выбраны следующие: экономика; экономический рост; экономическое развитие; прогноз; ввп; бизнес; доход; отрасль; инфляция; кризис; сектор и цена.

Для построения индикатора настроений на валютном рынке были выгружены поисковые запросы: интервенция; курс доллара; доллар; рубль; валюта; евро; валютный; цб; курс рубля и центробанк. На (рисунке 5) представлены следующие индикаторы: *ex_rate_rbc* – новостной индекс настроений на валютном рынке РБК; *ex_rate_vk* – индекс настроений на валютном рынке, построенный по постам «ВКонтакте»; *ex_rate_gt* – индекс настроений на валютном рынке по интернет-запросам и *usd* – прирост логарифма курса рубля к доллару. Как можно увидеть, различные индикаторы настроений имеют схожую динамику с колебаниями курса рубля к доллару, а пики характеризуют различные кризисные явления, в том числе ужесточение денежно-кредитной политики в декабре 2014 г., когда ключевая ставка поднялась до уровня 17%, нефтяные шоки в 2015-2016 гг., а также начало пандемии коронавируса и стремительного падения цен из-за отсутствия договоренности о снижении добычи странами ОПЕК+ в марте 2020 г.

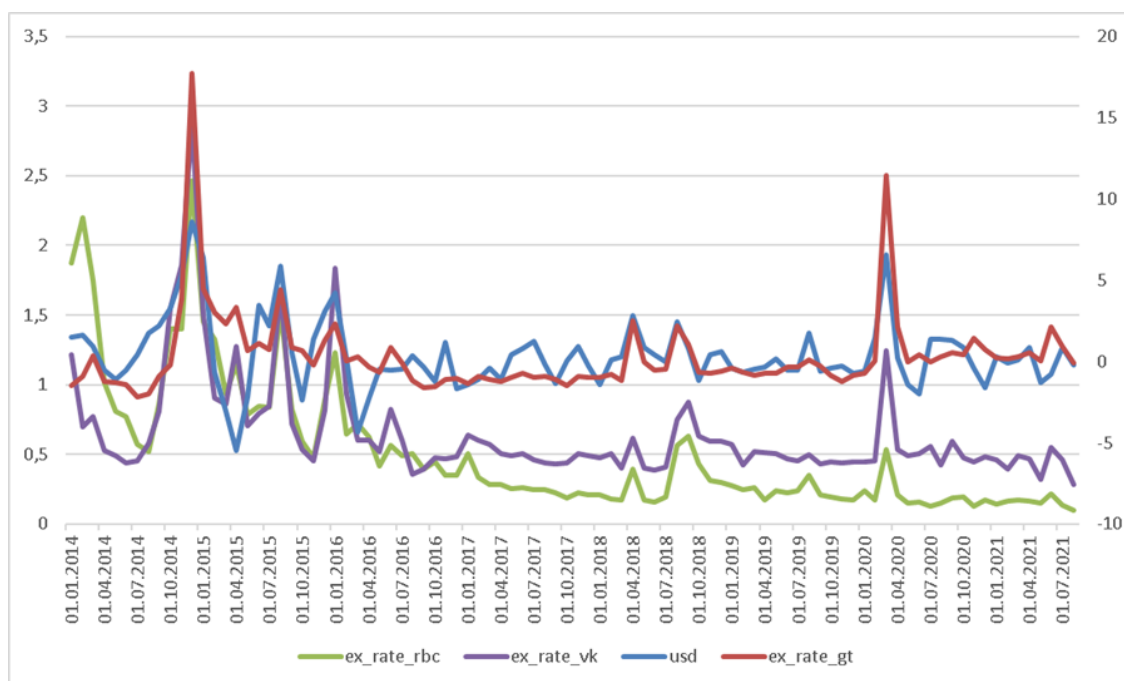


Рисунок 5. Настроения на валютном рынке и прирост курса рубля к доллару

Источник: по расчетам авторов

На (рисунке 6) показаны следующие индикаторы экономических условий: econ_rbc – новостной индекс экономических условий РБК; econ_vk – индекс экономических условий, построенный по постам «ВКонтакте»; econ_gt – индекс экономических условий по поисковым запросам и pmi – индекс деловой активности.

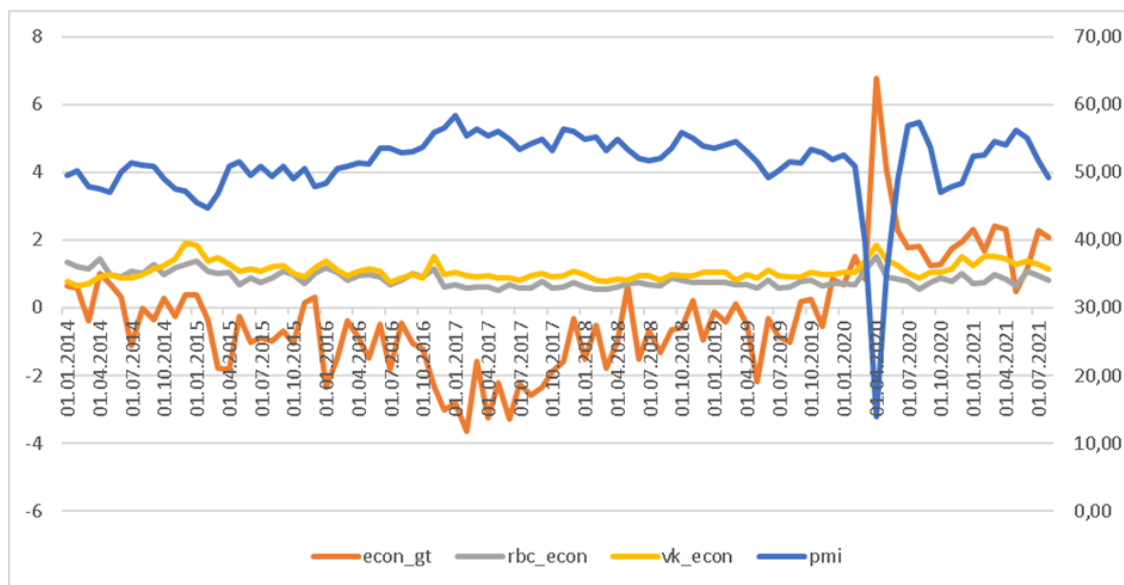


Рисунок 6. Индексы экономических условий и композитный индекс деловой активности PMI

Источник: по расчетам авторов

Индексы экономических условий также позволяют выделить различные кризисные явления в российской экономике. При этом снижение экономической активности соответствует росту индекса экономических условий, что можно трактовать как повышение пессимизма экономических агентов в условиях неопределенности.

На (рисунке 7) изображены индикаторы настроений на валютном рынке и экономических условий на основе комментариев «ВКонтакте». По аналогии проводится подсчет количества комментариев тремя способами: по всем комментариям к постам по тематике «валютный рынок» или «экономические условия» (ex_rate_count_st и econ_count_st, соответственно); по отобранным комментариям только о событиях на валютном рынке или экономическим условиям (ex_rate_count_sort_v2_st и econ_count_sort_st, соответственно) и комментариев по валютному рынку или экономическим условиям по всему корпусу постов (ex_rate_count_all_st и econ_count_all_st, соответственно). Затем данные стандартизируются.

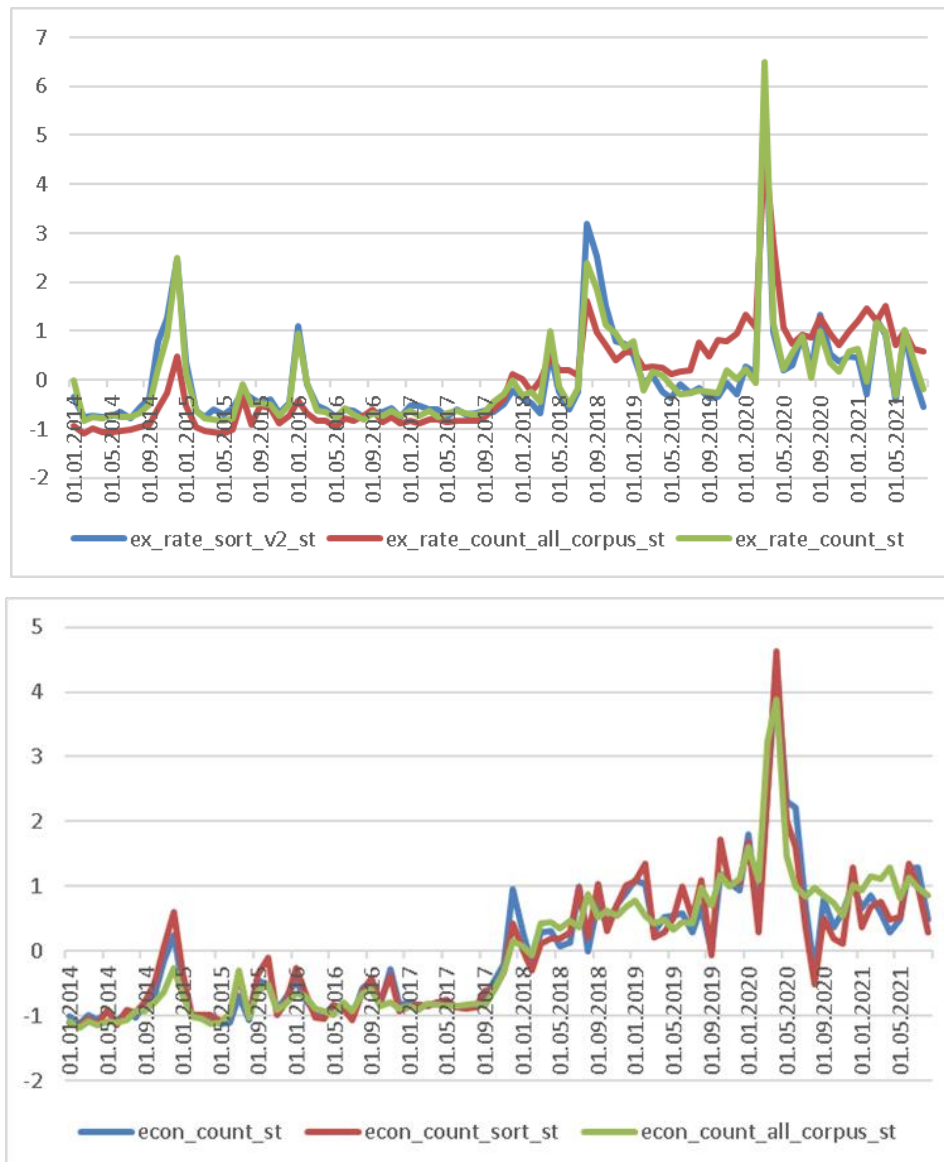


Рисунок 7. Индикаторы настроений на валютном рынке и экономических условий, построенным на основе комментариев

Источник: по расчетам авторов

Индикаторы настроений на валютном рынке и экономических условий, построенные на основе комментариев, обладают схожей динамикой. В данном случае можно предположить, что обеспокоенность населения относительно различных кризисных явлений сопряжено с увеличением активности в социальных сетях. Пользователи социальной сети для пересмотра своих решений используют доступную информацию из новостных сообществ, а также обсуждают в группе происходящие события для лучшего понимания сложившейся экономической ситуации. Также, как и для инфляционных ожиданий, максимальный уровень обеспокоенности наблюдается с началом пандемии коронавируса и нефтяным шоком в марте 2020 г.

В (таблице 3) показаны коэффициенты корреляции настроений на валютном рынке и индексов экономических условий с соответствующим макроэкономическим показателем.

Таблица 3

Корреляция настроений экономических агентов с соответствующими макроэкономическим показателем

Прирост номинального курса рубля к доллару			
	ex_rate_gt	ex_rate_rbc	ex_rate_vk
Без лага	0,588	0,450	0,642
Лаг1	0,216	0,242	0,342
ex_rate_sort ex_rate_count_all_corpus ex_rate_count			
Без лага	0,487	0,194	0,476
Лаг1	0,168	0,007	0,137
Индекс деловой активности PMI			
	econ_gt	rbc_econ	vk_econ
Без лага	-0,567	-0,586	-0,494
Лаг1	-0,380	-0,464	-0,332
econ_count econ_count_sort econ_count_all_corpus			
Без лага	-0,413	-0,442	-0,340
Лаг1	-0,294	-0,318	-0,339

Источник: по расчетам авторов

На основе проведенного графического анализа и корреляции можно сделать вывод о том, что индикаторы настроений на валютном рынке и экономических условий на основе постов, новостных статей и поисковых запросов оказываются более подходящими для последующего эмпирического анализа.

В следующем разделе будет проведен эконометрический анализ влияния настроений и ожиданий на макроэкономические показатели. Кроме того, проведем сравнение качества моделей с различными индикаторами на основе интернет-данных для выявления релевантных высокочастотных мер настроений и ожиданий.

4 Эконометрический анализ на основе высокочастотных индикаторов ожиданий населения

В данном разделе будут обсуждаться результаты анализа влияния прокси-переменных выявленных ожиданий и настроений на макроэкономические показатели.

Это позволит ответить на вопрос могут ли настройки пользователей в Интернете быть полезными для экономического моделирования.

4.1 Анализ влияния настроений экономических агентов на инфляцию и экономическую активность

Эконометрический анализ для инфляции²³ проводится в период с февраля 2015 г. по август 2021 г. на ежемесячной основе. Следуя подходу, предложенному в работе [51], где впервые было предложено использовать поисковые запросы в качестве меры инфляционных ожиданий, будем оценивать регрессионные модели с помощью МНК со стандартными ошибками в форме Ньюи-Веста (9) вида:

$$\pi_t = c + \alpha\pi_{t-1} + \beta infexp_t + \varepsilon_t \quad (9)$$

где c – константа,

π_t – сезонно скорректированная инфляция к предыдущему месяцу (*mom_sa*),

$infexp_t$ – показатель инфляционных ожиданий:

inf_vk_t – доля постов по тематике инфляции в общем количестве постов новостных сообществ в социальной сети «ВКонтакте»,

inf_rbc_t – доля новостных статей об инфляции в общем количестве интернет-издания РБК,

inf_gt_t – индекс инфляционных ожиданий на основе поисковых запросов Google Trends,

$inf_count_all_t$ – количество комментариев об инфляции по всему корпусу постов «ВКонтакте» (данные стандартизированы),

inf_count_t – количество комментариев к постам по инфляции (данные стандартизированы),

$inf_count_sort_t$ – количество комментариев по инфляции к постам по инфляционной тематике (данные стандартизированы),

ε_t – случайные ошибки.

²³ Отметим, что тесты на причинность по Грейнджеру от 1-12 лагов показали, что фактическая инфляция не является причиной по Грейнджеру для всех полученных мер инфляционных ожиданий. Этот результат отличается, например, для прогнозов профессиональных аналитиков, публикуемые Блумбергом, для которых гипотеза об отсутствии причинности отвергается.

По результатам эконометрического анализа для инфляции показаны в (таблице 4).

Таблица 4

Результаты эконометрического анализа влияния инфляционных ожиданий на фактическую инфляцию

Переменные	(1)	(2)	(3)	(4)	(5)	(6)	(7)
c	0.173***	0.056	0.06	0.235***	0.164***	0.167***	0.169***
π_{t-1}	0.538***	0.456***	0.435***	0.447***	0.55***	0.547***	0.547***
inf_vk_t	-	0.272***	-	-	-	-	-
inf_rbc_t	-	-	0.067***	-	-	-	-
inf_gt_t	-	-	-	0.058***	-	-	-
$inf_count_all_t$	-	-	-	-	0.045***	-	-
inf_count_t	-	-	-	-	-	0.047***	-
$inf_count_sort_t$	-	-	-	-	-	-	0.046***
R2 скорр.	0.679	0.712	0.716	0.730	0.699	0.701	0.701
AIC	-0.658	-0.754	-0.768	-0.818	-0.708	-0.715	-0.715
SIC	-0.597	-0.664	-0.678	-0.728	-0.618	-0.625	-0.625

Источник: по расчетам авторов

Как показывают результаты, все индикаторы инфляционных ожиданий в текущем месяце оказывают значимое положительное влияние на инфляцию. Качество подгонки в среднем выше, чем у эталонной модели – AR(1) (см. таблицу 4). Так скорректированный R^2 в моделях с индикаторами инфляционных ожиданий находится в диапазоне 0.7-0.73. По информационным критериям Акаике и Шварца наилучшей оказалась модель с поисковыми запросами. Полученные результаты можно интерпретировать следующим образом. Повышение интереса пользователей к тематике инфляции в Интернете сопряжено с повышением обеспокоенности населения относительно роста цен в экономике. Следовательно, будет происходить пересмотр инфляционных ожиданий экономическими агентами после сбора и анализа полученной информации, что окажет повышательное давление на фактическую инфляцию.

В итоге показатели инфляционных ожиданий на основе постов, новостных статей и поисковых запросов являются более пригодными для моделирования инфляции, по сравнению с индикаторами на основе комментариев. Это может быть связано с тем, что посты и новости, которые люди могут найти в Интернете через поисковую систему содержат ценную информацию о происходящих экономических событиях.

На ежемесячных данных также был проведен регрессионный анализ для композитного индекса PMI²⁴ для оценок на текущий месяц, где также предполагается использование расширенной ARX модели с различными индексами экономических условий вида:

$$pmi_t = c + \alpha pmi_{t-1} + \beta econ_t + \varepsilon_t \quad (10)$$

где c – константа,

pmi_t – композитный индекс деловой активности PMI,

$econ_t$ – индексы экономических условий:

$econ_vk_t$ – индекс экономических условий, построенный по постам новостных сообществ в социальной сети «ВКонтакте»,

$econ_rbc_t$ – индекс экономических условий на основе новостных статей интернет-издания РБК,

$econ_gt_t$ – индекс экономических условий, оцененный по поисковым запросам Google Trends,

$econ_count_all_t$ – количество комментариев об экономических условиях по всему корпусу постов «ВКонтакте» (данные стандартизированы),

$econ_count_t$ – количество комментариев к постам об экономических условиях (данные стандартизированы),

$econ_count_sort_t$ – количество отсортированных комментариев по регулярным выражениям к постам об экономических условиях (данные стандартизированы),

ε_t – случайные ошибки.

Результаты для индекса деловой активности представлены в (таблице 5). Полученные результаты говорят о том, что для индекса деловой активности все индикаторы экономических условий, построенные по интернет-данным, оказались значимыми. Повышение индекса экономических условий в данном случае можно

²⁴ Композитный индекс PMI является опережающим индикатором для экономической активности, в частности для ВВП. В связи с тем, что нет возможности провести анализ для ВВП, поскольку индексы, построенные по постам и комментариям, не имеют пропусков и малых значений только с 2013 г., в качестве индекса экономической активности рассматривается композитный индекс PMI.

интерпретировать как увеличение пессимизма интернет-пользователей относительно текущей экономической ситуации.

Таблица 5

Результаты эконометрического анализа влияния индекса экономических условий на композитный PMI

Переменные	(1)	(2)	(3)	(4)	(5)	(6)	(7)
c	18.041***	35.649***	37.256***	25.545***	21.975***	23.954***	24.247***
pmi_{t-1}	0.652***	0.497***	0.479***	0.50***	0.581***	0.542***	0.536***
$econ_vkt_t$	-	-8.946***	-	-	-	-	-
$econ_rbc_t$	-	-	-12.977***	-	-	-	-
$econ_gt_t$	-	-	-	-1.038***	-	-	-
$econ_count_all_t$	-	-	-	-	-1.822***	-	-
$econ_count_t$	-	-	-	-	-	-1.769***	-
$econ_count_sort_t$	-	-	-	-	-	-	-1.976***
R2 скорр.	0.423	0.508	0.569	0.501	0.516	0.506	0.53
AIC	5.777	5.629	5.497	5.644	5.613	5.634	5.584
SIC	5.837	5.719	5.587	5.734	5.703	5.724	5.674

Источник: по расчетам авторов

Это обусловлено тем, что в условиях неопределенности повышается спрос на информацию о текущем состоянии в экономике среди населения. Таким образом, рост пессимизма сопряжен со снижением экономической активности.

Следует отметить, что расширенные модели имеют более высокое качество подгонки (скорректированный R^2 от 0.5 до 0.57). По информационным критериям наилучшей оказалась модель с новостным индикатором РБК.

В данном случае также было показано, что индексы на основе комментариев оказались не хуже в качестве объясняющих переменных, чем другие индикаторы настроений.

4.2 Анализ влияния настроений на валютном рынке на курс рубля

При проведении эконометрического анализа влияния настроений на валютном рынке на курс рубля к доллару применяется подход экспоненциальной GARCH-модели. EGARCH позволяет оценить эффект от настроений на валютном рынке, как на условное среднее, так и условную дисперсию, а также учесть эффект рычага²⁵ и не накладывает каких-либо ограничений на коэффициенты в уравнении для

²⁵ Возможна асимметричная реакция волатильности валютного курса на плохие и хорошие новости

волатильности по сравнению с простой GARCH. Модели оцениваются на данных с 2 февраля 2015 г. – 31 августа 2021 г. на ежедневной основе.

Оцениваемая EGARCH(1,1) модель (11)-(12) имеет следующий вид

$$dlogusdrub_t = c + \delta_1 ex_rate_{t-1} + \delta_2 dlogbrent_t + \varepsilon_t \quad (11)$$

$$\log(\sigma_t^2) = \omega + \alpha \left[\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right] + \beta \log(\sigma_{t-1}^2) + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \lambda_1 ex_rate_{t-1} + \lambda_2 dlogbrent_t \quad (12)$$

где $dlogusdrub_t$ – прирост логарифма номинального курса рубля к доллару (Источник: CBONDS),

ex_rate_{t-1} – один из индикаторов настроений:

$ex_rate_vk_{t-1}$ – вероятность принадлежности постов новостных сообществ «Вконтакте» к теме «валютный рынок», усредненная за день,

$ex_rate_rbc_{t-1}$ – вероятность принадлежности новостных статей РБК к теме «валютный рынок», усредненная за день,

$ex_rate_gt_{t-1}$ – индекс настроений на валютном рынке, построенной на основе ежедневных данных²⁶,

$dlogbrent_t$ – прирост логарифма цены на нефть марки Brent (Источник: CBONDS),

σ_t^2 – условная дисперсия,

ε_t – случайные ошибки, имеющие t-распределение Стьюдента.

Результаты оцененных моделей представлены в (таблице б). В качестве эталонной модели рассматриваются спецификации (4), в которой включается только цена на нефть марки Brent.

По сделанным оценкам можно сделать следующие выводы. Во-первых, все индикаторы настроений на валютном рынке не оказывают значимое влияние на курс рубля. Поскольку в качестве объясняющей переменной рассматривается первый лаг индикаторов настроений для избегания одномоментной реакции, что в целом довольно ожидаемый результат. Информация на валютном рынке достаточно быстро

²⁶ Индикаторы настроений на валютном рынке на ежедневной и ежемесячной основах строились на различных данных с соответствующей периодичностью данных по интернет-запросам. Для ежемесячной основы сервис Google Trends предоставляет статистику по каждому поисковому запросу единым временным рядом, а для ежедневных данных – скользящим окном, поэтому делалось склеивание на различных промежутках времени.

учитывается инвесторами. Во-вторых, повышение девальвационных настроений в течение 2015-2021 гг. предшествовало увеличению неопределенности на валютном рынке, за исключением настроений на основе поисковых запросов.

Таблица 6

Результаты эконометрического анализа влияния настроений на валютном рынке на прирост курса рубля к доллару

Переменные	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Уравнение условного среднего							
$ex_rate_vk_{t-1}$	0.052	-	-	-	-	-	-
$ex_rate_rbc_{t-1}$	-	0.0004	-	-	-	-	-
$ex_rate_gt_{t-1}$	-	-	0.004	-	-	-	-
$ex_rate_count_all_{t-1}$	-	-	-	-	-0.00	-	-
$ex_rate_count_{t-1}$	-	-	-	-	-	-0.00	-
$ex_rate_count_sort_{t-1}$	-	-	-	-	-	-	-0.00
$dlogbrent_t$	-0.101***	-0.099***	-0.098***	-0.099***	-0.098***	-0.098***	-0.098***
Уравнение условной дисперсии							
γ	0.04***	0.051***	0.039***	0.036***	0.041***	0.039***	0.039***
$ex_rate_vk_{t-1}$	5.115***	-	-	-	-	-	-
$ex_rate_rbc_{t-1}$	-	3.523*	-	-	-	-	-
$ex_rate_gt_{t-1}$	-	-	-0.028	-	-	-	-
$ex_rate_count_all_{t-1}$	-	-	-	-	-0.003	-	-
$ex_rate_count_{t-1}$	-	-	-	-	-	-0.005	-
$ex_rate_count_sort_{t-1}$	-	-	-	-	-	-	-0.005
$dlogbrent_t$	-1.509**	-1.591**	-1.953***	-1.902***	-1.930***	-1.976***	-1.961***
LL	5781.993	5780.515	5777.580	5776.378	5777.50	5777.40	5777.50
AIC	-6.954	-6.952	-6.949	-6.950	-6.949	-6.949	-6.949
SIC	-6.922	-6.920	-6.916	-6.924	-6.916	-6.916	-6.916

Источник: по расчетам авторов

Как и в случае поисковых запросов, показатели на основе комментариев также не оказывают значимого влияния ни на условное среднее, ни на условную дисперсию. В-третьих, коэффициент γ , показывающий существование асимметричной реакции волатильности курса рубля к доллару на плохие и хорошие новости, оказался значимым и имеет положительный знак во всех спецификациях, что свидетельствует о том, что обесценение курса рубля к доллару сопряжено с увеличением волатильности на валютном рынке. В-четвертых, увеличение цены на нефть соответствует укреплению рубля и снижению неопределенности на валютном рынке.

В результате эконометрический анализ показал, что интернет-данные могут быть прокси-переменными для анализа влияния поведения экономических агентов на экономические процессы. Среди мер инфляционных ожиданий наилучший результат

показал индикатор, построенный на основе интенсивности интернет-запросов по тематике инфляции. Вероятно, это обусловлено тем, что экономические агенты через поисковую систему могут найти интересующую информацию в широком перечне информационных источников, включая социальные сети и сайты новостных интернет-изданий. Как было показано ранее, повышение уровня обеспокоенности населения в Интернете относительно роста цен в экономике сопряжено с увеличением инфляции. Этот результат устойчив для различных мер инфляционных ожиданий, рассчитанных по постам в «ВКонтакте», новостным статьям РБК, поисковым запросам и комментариям к постам по тематике инфляции.

В случае индекса деловой активности наилучшей является модель с новостным индикатором РБК. Кроме того, индексы на основе комментариев оказались не хуже в качестве объясняющих переменных, чем другие индикаторы настроений.

При исследовании влияния настроений на валютном рынке на ежедневной основе не было выявлено значимого влияния на прирост курса рубля к доллару. Однако, повышение интереса к тематике валютного рынка, обусловленное кризисными событиями, соответствует росту неопределенности на валютном рынке.

ЗАКЛЮЧЕНИЕ

Новизна текущего исследования заключается в количественных оценках индикаторов настроений экономических агентов с помощью структурированных и неструктурированных интернет-данных. Для обоснования проводимого исследования на данных социальных сетей был сделан обзор теоретических концепций и эмпирических работ. Проведенный обзор теоретических работ по анализу механизмов влияния распространения информации в социальных сетях и эмоциональной окраски новостей на ожидания населения позволяет сделать следующие выводы. В социальных сетях скорость распространения информации выше по сравнению с традиционными каналами передачи информации. Пользователи социальных сетей могут обмениваться информацией и обсуждать интересующие их вопросы, включая политические, экономические и социальные проблемы, а также получать релевантный контент от интернет-изданий. В таком случае экономические агенты формируют свои ожидания с минимальными издержками по поиску и сбору информации.

Помимо этого, может наблюдаться консенсус мнений экономических агентов при принятии решений. Такое консенсусное решение может стать наилучшим в терминах общественного благосостояния. Однако мудрость толпы приведет к правильному решению только при существовании разнообразия мнений, доверии и независимости принимаемых решений. Важно, что существование группы лидеров делает невозможным достижения мудрости толпы из-за искажения общественного мнения.

Следует отметить, что ожидания экономических агентов в значительной степени зависят от типа новости. Люди сильнее реагируют на плохие новости из-за риска больших потерь благосостояния при реализации неблагоприятного исхода, например, снижения экономической активности в течение финансового кризиса.

Результаты обзора эмпирических исследований показали, что посты и комментарии в социальных сетях являются ценным источником данных для оценки настроений и ожиданий пользователей Интернета, и позволяют улучшать прогностическую силу моделей различных экономических и финансовых показателей. Социальные сети является хорошим новостным каналом для передачи информации компаниями о своей экономической активности и сигналов центральных

банков о направленности денежно-кредитной политики, а следовательно, снижают асимметрию информации для экономических агентов.

Как показано в российских и зарубежных исследованиях, посты в социальных сетях играют ключевую роль при формировании ожиданий населением в периоды высокой неопределенности в экономике. В этой связи социальные сети позволяют анализировать поведение населения в реальном времени, а также определять причины обеспокоенности и улавливать изменения настроений и ожиданий экономических агентов.

По результатам рассмотренных эмпирических исследований можно сделать вывод о том, что индикаторы настроений пользователей, построенные на основе данных социальных сетей, становятся полноценным инструментом анализа поведения населения и способны заменить опросные индикаторы. Важно отметить, что информации в социальных сетях недостаточно для принятия решений экономическими агентами, а следовательно, полностью отказаться от рыночных индикаторов или прогнозов профессиональных аналитиков невозможно. При этом прогнозные модели, включающие настроения участников финансовых рынков и данные социальных сетей, имеют более высокую предсказательную силу. Это обусловлено тем, что использование данных социальных сетей позволяет учесть дополнительную информацию по сравнению с традиционными подходами.

В рамках эмпирического анализа различные индикаторы настроений были построены на данных социальной сети «ВКонтакте», новостных статьях РБК и поисковых запросах Google Trends. В целом индикаторы инфляционных ожиданий, настроений на валютном рынке и экономических условий, построенные на основе новостных сообщений в «ВКонтакте», РБК и поисковых запросов, являются более информативными, по сравнению с комментариями к постам. В данном случае можно предположить, что обеспокоенность населения относительно различных кризисных явлений сопряжена с увеличением активности в Интернете. Пользователи социальной сети для пересмотра своих решений используют всю доступную информацию в Интернете, а также обсуждают в социальной сети происходящие события для лучшего понимания сложившейся экономической ситуации. Однако, как показывает анализ комментариев, в отдельные периоды наблюдается паника среди населения, в частности максимальный уровень обеспокоенности относительно роста цен, экономических условий и ситуацией на валютном рынке наблюдается с началом

пандемии коронавируса в марте 2020 г. В случае постов «Вконтакте», новостных статей и поисковых запросов также характеризуют различные кризисные явления, включая валютный кризис 2014-2015 гг., нефтяные шоки, введение санкций и начало пандемии коронавируса, что делает их более интерпретируемыми для эконометрического анализа. Это может быть связано с тем, что при принятии решений экономические агенты больше доверяют публикациям средств массовой информации.

Эконометрический анализ влияния настроений экономических агентов на макроэкономические показатели позволил сделать следующие выводы. Настроения интернет-пользователей могут быть прокси-переменными для анализа влияния поведения экономических агентов на реальные процессы в экономике. Во-вторых, мера инфляционных ожиданий в текущем месяце, построенная с помощью интернет-запросов по тематике инфляции, показала наилучший результат при моделировании инфляции. Это обусловлено тем, что экономические агенты через поисковую систему могут получить доступ к широкому перечню информационных ресурсов с минимальными затратами для принятия решений в условиях неопределенности, включая социальные сети и сайты новостных интернет-изданий. По этой причине повышение уровня обеспокоенности населения в Интернете относительно роста цен в экономике может привести к увеличению инфляционного давления. Этот результат устойчив для различных мер инфляционных ожиданий, рассчитанных по данным социальной сети «Вконтакте», новостным статьям РБК, поисковым запросам.

В случае индекса деловой активности наилучшей является модель с новостным индикатором РБК. Кроме того, индексы на основе комментариев оказались не хуже в качестве объясняющих переменных, чем другие индикаторы настроений. Проинтерпретировать полученные результаты для композитного индекса PMI можно следующим образом. В условиях неопределенности повышается спрос на информацию о текущем состоянии в экономике среди населения. Таким образом, рост пессимизма сопряжен со снижением экономической активности.

При исследовании влияния настроений на валютном рынке на ежедневной основе не было выявлено значимого влияния на прирост курса рубля к доллару. Однако, повышение интереса к тематике валютного рынка, обусловленное кризисными событиями, соответствует росту неопределенности на валютном рынке.

Важно отметить, что в среднем интернет-данные, содержащие непосредственно информацию состоянии экономики, в том числе посты в социальной

сети «ВКонтакте», новостные статьи или поисковые запросы, являются более информативными, чем комментарии к постам. Это может быть связано с тем, что при принятии решений экономические агенты больше доверяют публикациям средств массовой информации.

В целом проведенное исследование показало, что интернет-данные полезны для анализа поведения экономических агентов и позволяют улучшить качество моделей макроэкономических показателей. Меры инфляционных ожиданий пользователей в Интернете могут быть альтернативой опросным индикаторам, что снижат издержки обработки данных и временной лаг для экономических властей.

На практике полученные индикаторы настроений экономических агентов могут быть использованы при прогнозировании макроэкономических показателей и оценке деловой конъюнктуры. В перспективе планируются расширение перечня индикаторов настроений экономических агентов и оценка более комплексных моделей временных рядов.

Благодарности

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

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Katz E., Lazarsfeld P. F. Personal influence: the part played by people in the flow of mass communications//New York: Free Press – 1955
2. Surowiecki J. The wisdom of crowds//New York: Doubleday - 2004.
3. Galton F. Vox populi// Nature - 1907. - Vol. 75. - pp. 450-451.
4. Clemen R. T. Combining forecasts: a review and annotated bibliography// International Journal of Forecasting - Vol. 5. - 1989. - pp. 559-583.
5. Acemoglu D., Bimpikis K., Ozdaglar A. Dynamics of information exchange in endogenous social networks// Theoretical Economics - Vol. 9, No. 1 - 2014. - pp. 41-97.
6. Zhao L., Cui H., Qiu X., Wang X., Wang J. SIR rumor spreading model in the new media age//Physica A: Statistical Mechanics and its Applications - Vol. 392, No. 4 - 2013. - pp. 995-1003.
7. Cheng Y., Liu C., Ding F. Dynamic analysis of rumor spreading model for considering active network nodes and nonlinear spreading rate//Physica A: Statistical Mechanics and its Applications - Vol. 506. - 2018. - pp. 24-35.
8. Acemoglu D., Dahleh M., Lobel I., Ozdaglar A. E. Bayesian Learning in Social Networks//NBER. - Working Paper. - w14040. - 2008.
9. Smith L., Sorensen P. Pathological Outcomes of Observational Learning// Econometrica - Vol. 68, No. 2 - 2000. - pp. 371-398.
10. Banerjee A. A simple model of herd behavior//The Quarterly Journal of Economics - 1992. -Vol. 107 - pp. 797-817.
11. Bikhchandani S., Hirshleifer D., Welch I. A theory of fads, fashion, custom and cultural change as information cascades//The Journal of Political Economy – 1992. - Vol. 100. - pp. 992-1026.
12. Çelen B., Özgür O. Diversity of Opinion in Social Networks// SSRN – 2721614. - 2016.
13. Golub B., Jackson M. O. Naive learning in social networks: convergence, influence, and the wisdom of crowds// Stanford – Preprint. - 2007.
14. Mavrodiev P., Schweitzer F. Enhanced or distorted wisdom of crowds? An agent-based model of opinion formation under social influence//Swarm Intelligence - 2021. - pp. 1-16.
15. Pigou A.C. Industrial fluctuations// London: MacMillan – 1927.

16. Soroka S. N. Good news and bad news: asymmetric responses to economic information//Journal of Politics - 2006. - Vol. 68, - pp. 372-385.
17. Kahneman D., Tversky A. Prospect theory: an analysis of decision under risk// Econometrica -1979 - Vol. 47, No. 2 - pp. 263–292.
18. Doms M., Morin N. Consumer sentiment, the economy and the news media//Federal Reserve Bank of San Francisco - Working Paper. - 2004-9 - 2004.
19. McCluskey J. J., Swinnen J., Vandemoortele T. You get what you want: A note on the economics of bad news//Information Economics and Policy - 2015. - Vol. 30. - pp. 1-5.
20. Coyne S., Madiraju P., Coelho J. Forecasting stock prices using social media analysis//2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress.- IEEE. - 2017. - pp. 1031-1038.
21. Bashar M. A., Nayak R., Balasubramaniam T. Topic, sentiment and impact analysis: COVID19 information seeking on social media//arXiv - preprint arXiv. - 2008.12435 – 2020.
22. Yin H., Yang S., Li J. Detecting topic and sentiment dynamics due to Covid-19 pandemic using social media// International Conference on Advanced Data Mining and Applications - 2020. - pp. 610-623.
23. Zhou J., Yang S., Xiao C., Chen F. Examination of community sentiment dynamics due to COVID-19 pandemic: a case study from a state in Australia//arXiv - preprint arXiv. - 2006.12185 – 2020.
24. Blei D. M., Ng A. Y., Jordan M. I. Latent dirichlet allocation// The Journal of machine Learning research – 2003. - Vol. 3. - pp. 993-1022.
25. Hutto C., Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text// Proceedings of the International AAAI Conference on Web and Social Media - 2014. - Vol. 8.
26. Samuel J., Ali G. G., Rahman M., Esawi E., Samuel Y. Covid-19 public sentiment insights and machine learning for tweets classification//Information – 2020. - Vol. 11, No. 6 - p. 314.

27. Han X., Wang J., Zhang M., Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China// International Journal of Environmental Research and Public Health - 2020. - Vol. 17, No. 8. - P. 2788.
28. Malla, S. J., Alphonse, P. J. A. COVID-19 outbreak: an ensemble pre-trained deep learning model for detecting informative tweets// Applied Soft Computing. – 2021. - Vol. 107. - P. 107495.
29. Nguyen D. Q., Vu T., Rahimi A., Dao M. H., Nguyen L. T., Doan L. WNUT-2020 task 2: identification of informative COVID-19 english tweets// arXiv - preprint arXiv. - 2010.08232 - 2020.
30. Dubey A. D. Twitter sentiment analysis during COVID-19 outbreak// SSRN – 3572023. - 2020.
31. Pota M., Ventura M., Fujita H., Esposito M. Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets// Expert Systems with Applications. – 2021.- Vol. 181. - P. 115119.
32. Rosenthal S., Farra N., Nakov P. SemEval-2017 task 4: sentiment analysis in Twitter// Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). - 2017. - pp. 502-518.
33. Ilyas S. H. W., Soomro Z. T., Anwar A., Shahzad H., Yaqub U. Analyzing Brexit's impact using sentiment analysis and topic modeling on Twitter discussion//The 21st Annual International Conference on Digital Government Research - 2020. - pp. 1-6.
34. Голощапова И. О., Андреев М. Л. Оценка инфляционных ожиданий российского населения методами машинного обучения//Вопросы экономики – 2017. - Т. 6 - С. 71-93.
35. Azar P. D., Lo A. W. The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds//The Journal of Portfolio Management - 2016. - Vol. 42, No. 5 - pp. 123-134.
36. De Smedt T., Daelemans W. Pattern for python//The Journal of Machine Learning Research - 2012. - Vol. 13, No. 1. - pp. 2063-2067.
37. Gorodnichenko Y., Pham T., Talavera O. 'Liked', 'Shared', 'Commented': central bank communication on Facebook and Twitter. - 2021. - No. 21-05.
38. Janetzko D. Using twitter to model the eur/usd exchange rate//arXiv - preprint arXiv - 1402.1624 - 2014.

39. Ozturk S. S., Ciftci K. A sentiment analysis of twitter content as a predictor of exchange rate movements//Review of Economic Analysis - 2014. - Vol. 6, No. 2 - pp. 132-140.
40. Broadstock D. C., Zhang D. Social-media and intraday stock returns: The pricing power of sentiment//Finance Research Letters - 2019. - Vol. 30. - pp. 116-123.
41. Rao T., Srivastava S. Modeling movements in oil, gold, forex and market indices using search volume index and twitter sentiments//Proceedings of the 5th annual ACM Web science conference - 2013. - pp. 336-345.
42. Albarrak M. S., Elnahass M., Papagiannidis S., Salama A. The effect of twitter dissemination on cost of equity: A big data approach// International Journal of Information Management/ - 2020. - Vol. 50. - pp. 1-16.
43. Becerra J. S., Sagner A. Twitter-based economic policy uncertainty index for Chile// Central Bank of Chile. - 2020.
44. Baker S. R., Bloom N., Davis S. J. Measuring economic policy uncertainty//Quarterly Journal of Economics - 2016. - Vol. 131, No. 4 - pp. 1593–1636.
45. Antenucci D., Cafarella M., Levenstein M., Ré C., Shapiro M. D. Using social media to measure labor market flows// National Bureau of Economic Research - w20010 - 2014.
46. Daas P. J. H., Puts M. J. H. Social media sentiment and consumer confidence// ECB - - Statistics Paper 5. - 2014.
47. Van den Brakel J., Söhler E., Daas P., Buelens B. Social media as a data source for official statistics the Dutch consumer confidence index//Survey Methodology - 2017. - Vol. 43, No. 2 - pp. 183-210.
48. Örsal K., Dilan. D., Sturm S. Computing consumer sentiment in Germany via social media data//University of Hamburg, Hamburg Discussion Papers in International Economics 7. – 2021.
49. Angelico C., Marcucci J., Miccoli M., Quarta F. Can we measure inflation expectations using Twitter?//Bank of Italy - Temi di Discussione (Working Paper) - 1318 - 2021.
50. Ульяновкин Ф. Прогнозирование российских макроэкономических показателей на основе информации в новостях и поисковых запросах //Деньги и кредит. – 2020. – Т. 79. – №. 4. – С. 75-97.

51. Guzman G. Internet search behavior as an economic forecasting tool: The case of inflation expectations //Journal of economic and social measurement. – 2011. – Т. 36. – №. 3. – С. 119-167.