

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В. ЛОМОНОСОВА

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

*На правах рукописи*

**Голощапова Ирина Олеговна**

**РАЗРАБОТКА МЕТОДИКИ ПОСТРОЕНИЯ ВЫСОКОЧАСТОТНЫХ  
ИНДИКАТОРОВ ЭКОНОМИЧЕСКИХ ОЖИДАНИЙ НАСЕЛЕНИЯ НА  
ОСНОВЕ БОЛЬШИХ ДАННЫХ (НА ПРИМЕРЕ ИНФЛЯЦИОННЫХ  
ОЖИДАНИЙ)**

Специальность: 08.00.10 – Финансы, денежное обращение и кредит

**ДИССЕРТАЦИЯ**

на соискание ученой степени

кандидата экономических наук

Научный руководитель:

кандидат экономических наук, доцент

Павлова Е.В.

Москва – 2018

## Оглавление

Введение.....	4
Глава 1. Современные подходы к оценке экономических ожиданий населения .....	16
1.1. Ожидания экономических агентов как значимый фактор экономической динамики.....	16
1.1.1. <i>Определение экономических ожиданий населения</i> .....	16
1.1.2. <i>Влияние экономических ожиданий на макроэкономическую конъюнктуру</i> .....	18
1.1.3. <i>Адаптивные экономические ожидания</i> .....	19
1.1.4. <i>Рациональные экономические ожидания</i> .....	24
1.2. Способы измерения экономических ожиданий .....	28
1.2.1. <i>Измерение экономических ожиданий населения с помощью опросов</i> .....	28
1.2.2. <i>Измерение экономических ожиданий населения на основе анализа биржевых данных</i> .....	31
1.2.3. <i>Измерение экономических ожиданий населения с использованием методов анализа больших данных</i> .....	32
1.3. Инфляционные ожидания населения .....	40
1.3.1. <i>Воздействие инфляционных ожиданий на динамику экономики</i> ..	40
1.3.2. <i>Подходы к измерению инфляционных ожиданий</i> .....	45
Глава 2. Методика построения высокочастотных индикаторов экономических ожиданий населения на основе больших данных.....	55
2.1. Обоснование и общее описание методики построения индикаторов.	55
2.2. Определение источников информации и оценка репрезентативности выборки для анализа.....	62
2.2.1. <i>Составление выборки экономических СМИ, репрезентативной для построения индикаторов</i> .....	62
2.2.2. <i>Составление выборки источников информации о мнении читателей относительно материалов СМИ</i> .....	67

2.3. Отбор новостных статей экономических СМИ и комментариев интернет-пользователей для измерения различных видов экономических ожиданий.....	69
2.4. Расчет индикатора интенсивности ожиданий населения для заданной области экономики.....	77
2.4.1. Расчет индикатора интенсивности экономических ожиданий и оценка его релевантности заданному виду ожиданий.....	77
2.4.2. Оценка устойчивости результатов расчета индикатора интенсивности экономических ожиданий.....	82
2.5. Выявление факторов изменения интенсивности экономических ожиданий населения во времени.....	87
2.5.1. Оценка зависимости индикатора интенсивности экономических ожиданий от релевантных экономических параметров.....	88
2.5.2. Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе экспертного анализа.....	90
2.5.3. Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе методов машинного обучения и текстовой обработки.....	92
Глава 3. Верификация методики построения высокочастотных индикаторов экономических ожиданий российского населения на примере инфляционных ожиданий.....	107
3.1. Определение источников информации и данных для анализа.....	107
3.2. Расчет индикатора интенсивности инфляционных ожиданий российского населения.....	109
3.3. Выявление факторов изменения интенсивности инфляционных ожиданий населения во времени.....	129
Заключение.....	168
Библиография.....	177
Приложение 1.....	196
Приложение 2.....	197
Приложение 3.....	199
Приложение 4.....	203

## Введение

### Актуальность темы исследования

Ожидания экономических агентов являются основой для экономических теорий, объясняющих процессы потребления, инвестиций, спроса на деньги, инфляции и др. В научной литературе экономические ожидания определены как прогнозы экономических агентов – домохозяйств, фирм, государства, экспертного сообщества – относительно будущих событий в экономике, которые значимо влияют на принятие решений [Mikolajek-Gosejna, 2014, р. 35]. В диссертации рассматриваются экономические ожидания домохозяйств или населения.

В современной практике экономических исследований выделяют два ключевых подхода к оценке экономических ожиданий: (1) подход, базирующийся на опросах экономических агентов (survey-based approach) и (2) подход, основанный на анализе биржевых параметров (market-based approach). Однако указанные подходы имеют ряд существенных недостатков.

С 2015 года официальным режимом монетарной политики в России является инфляционное таргетирование, переход к которому фактически был начат регулятором ещё с сентября 2013 года. На протяжении этого периода времени изучение инфляционных ожиданий населения является для Банка России важным объектом исследования. Численный ориентир регулятора по инфляции при текущем режиме денежно-кредитной политики призван оказывать значительное влияние на инфляционные ожидания населения. Последние, в свою очередь, способны воздействовать на уровень цен в долгосрочном периоде.

Актуальность разработки методики построения индикаторов экономических ожиданий населения на базе больших массивов данных обусловлена их существенным влиянием на экономические процессы, а также возможностью преодоления ключевых недостатков современных методов измерения ожиданий с помощью методов анализа больших данных.

При этом использование индикатора интенсивности инфляционных ожиданий российского населения в качестве примера применения разработанной методики в целях ее верификации также является актуальным.

### **Степень научной разработанности проблемы**

Одними из первых работ в области влияния ожиданий экономических агентов на экономическую конъюнктуру стали исследования Ф. Найта и Дж. М. Кейнса. А. Харт впервые явно обозначил потребность в измерении экономических ожиданий на историческом периоде.

Дж. Мут акцентировал внимание на необходимости анализа механизмов формирования ожиданий как важных факторов будущей макроэкономической динамики. В научной литературе выделяют два ключевых вида ожиданий: адаптивные и рациональные. Адаптивные ожидания рассматриваются в работах Л. Койка, П. Кагана, М. Фридмана, М. Нерлов и др., находят эмпирическое подтверждение у Г. Чоу, С. Фиглевски, П. Уочтела, К. Кэрролла. Гипотезу рациональных ожиданий рассматривали Т. Саргент, Н. Уоллес, Р. Лукас и др., эмпирические подтверждения гипотезы представили Дж. Мьюэльбауэр, Т. Гудвин, С. Шеффрин, М. Форселлз, Г. Кенни и др.

Л. Мламбо, Р. Хафер, С. Хейн, Л. Томас, М. Бриан, Г. Венкату и др. подчеркивают необходимость проведения опросов для оценки экономических, и в частности инфляционных, ожиданий населения. На недостатки подхода указывают О. Армантьер, В. Бруин де Бруин, Р. Кертин и др.

Метод измерения экономических ожиданий населения, базирующийся на анализе биржевых параметров, рассматривают Дж. Фауст, Э. Свансон, Дж. Райт, Дж. Кохрейн, М. Пиаццесси, Д. Барр, Дж. Кампбелл, М. Жемков, О. Кузнецова и др. Критические замечания представлены С. Миллингтон, М. Пасаогуллари.

Л. Драгер, М. Ламла, Д. Пфаджфар, М. Бауэр, К. Кэрролл, К. Яковлева используют новости экономических СМИ в целях построения индикаторов

инфляционных ожиданий и ожиданий относительно будущей экономической динамики, а также показывают необходимость учета новостей при анализе ожиданий. Вопросы применения методов текстового анализа и машинного обучения в целях создания экономических индикаторов рассматривают С. Бейкер, Н. Блум, С. Дэвис, Р. Ниман, С. Кападиа, Д. Такет, Д. Ардиа, К. Блюто, К. Боудт, А. Торсруд, Н. Бейлк и др. При этом Х. Чои, Х. Вэриан, Т. Шмидт, С. Возен, Г. Гузман утверждают, что активность пользователей в сети Интернет может быть использована для построения индикаторов экономических ожиданий населения, включая инфляционные ожидания.

Настоящее исследование продолжает проработку вопроса применения современных методов машинного обучения и текстовой обработки в целях построения индикаторов экономических ожиданий населения.

Вклад данной работы состоит в предложении методики построения высокочастотных индикаторов экономических ожиданий российского населения на основе больших массивов текстовой информации. Разработанная методика направлена на устранение ключевых недостатков существующих подходов к оценке экономических ожиданий. Использование методики возможно для аналогичных целей на данных развитых и развивающихся рынков других стран.

### **Объект и предмет исследования**

*Объектом исследования* являются экономические ожидания населения. *Предметом исследования* выступают высокочастотные индикаторы экономических ожиданий российского населения на основе больших наборов текстовых данных.

### **Цель и задачи исследования**

*Целью* диссертации является формирование системы правил построения высокочастотных индикаторов для оценки экономических ожиданий российского населения на основе больших массивов текстовой информации.

Для достижения поставленной цели в работе поставлены и решены следующие задачи:

1. Систематизировать результаты теоретических и эмпирических исследований влияния экономических, и в частности инфляционных, ожиданий населения на экономическую динамику, а также методов их измерения и возможностей по усовершенствованию с помощью алгоритмов анализа больших массивов текстовых данных.

2. Разработать алгоритм определения источников информации как первый этап методики, правил отбора новостных статей СМИ как второй этап методики и комментариев интернет-пользователей для расчета индикаторов экономических ожиданий как третий этап методики.

3. Составить алгоритм расчета высокочастотных индикаторов экономических ожиданий российского населения, включая оценку качества результата и его устойчивости к правилам вычисления, как четвертый этап методики.

4. Сформировать алгоритм выявления факторов изменения интенсивности экономических ожиданий населения во времени как пятый этап методики.

5. Провести верификацию разработанной методики на примере построения высокочастотного индикатора инфляционных ожиданий российского населения.

**Научная новизна** диссертационной работы заключается в следующем:

1. Автором разработана оригинальная методика построения высокочастотных индикаторов экономических ожиданий населения на основе больших массивов текстовых данных. Методика опирается на расчет индикаторов экономических ожиданий и комплексную процедуру выявления факторов их изменения, включает оценку релевантности результата и проверку устойчивости результата к правилам расчета.

2. Предложен авторский алгоритм выявления факторов изменения экономических ожиданий населения на основе текстовых данных. В

дополнение к классической статистической проверке теоретических гипотез и экспертному анализу содержания текстовых сообщений, упоминающих целевые экономические ожидания, предложенный алгоритм опирается на современные методы тематического моделирования. Основанная на них часть алгоритма объединяет имеющиеся на данный момент разработки в области текстового анализа и машинного обучения, а именно: стадии предобработки текста, подбора параметров для моделирования и выявления тем в больших массивах текстовой информации.

3. В рамках апробации авторской методики построен высокочастотный индикатор инфляционных ожиданий российского населения. Полученный индикатор позволяет исследовать интенсивность ожиданий на ежедневной основе, тогда как максимальная частотность индикаторов инфляционных ожиданий российского населения, измеренных с помощью опросов, является ежемесячной.

По результатам оценок построенный индикатор опережает динамику ключевого на текущий момент индикатора инфляционных ожиданий российского населения, рассчитываемого на основе опросов, а также является значимым фактором для будущей динамики потребительской инфляции.

Рассматриваемый для анализа набор данных включил около 260 тыс. новостных статей СМИ, 1.5 миллиона комментариев от 140 тыс. интернет-пользователей. Выборка является сбалансированной по медиа-источникам.

4. С учетом высокочастотных данных выявлены новые в контексте исследования инфляционных ожиданий российского населения факторы, с которыми население связывало изменение ожиданий в период с января 2014 г. по февраль 2018 г.

Экспертный анализ, проведённый на пятом этапе методики, позволил выделить такие новые факторы как: неопределенность экономической ситуации, внутренняя политика и мировая конъюнктура за исключением воздействия динамики курса доллара и цен на нефть. Новые факторы, выделенные с помощью тематического моделирования, включили: тарифы,



налоги и платежи; ситуацию в секторе недвижимости; условия розничной торговли.

5. Впервые на высокочастотных данных верифицирована значимость ключевых факторов инфляционных ожиданий российского населения. По результатам оценки подтверждено значимое влияние на инфляционные ожидания динамики цен на нефть, условий денежно-кредитной политики, изменения доходов населения. Кроме того, показано, что динамика валютного курса оказывает влияние на инфляционные ожидания российского населения в большей степени в краткосрочном периоде.

### **Теоретическая и практическая значимость диссертационного исследования**

Теоретическая значимость исследования состоит в следующем:

1) разработанная методика построения высокочастотных индикаторов экономических ожиданий российского населения способствует развитию теоретического направления по изучению экономических ожиданий населения. Построение индикатора инфляционных ожиданий населения, релевантного динамике потребительской инфляции и опережающего ближайший индикатор-аналог, показывает значимость текстовой информации, содержащейся в новостных статьях СМИ и комментариях интернет-пользователей, для изучения экономических ожиданий населения. При этом методика является универсальной и предполагает возможность, с одной стороны, построения индикаторов ожиданий населения для широкого круга областей экономики, с другой – построения индикаторов экономических ожиданий на данных развитых и развивающихся рынков других стран. Полученные индикаторы могут быть использованы как параметр в различных теоретических моделях экономики.

2) предложенный трехшаговый алгоритм выявления факторов изменения экономических ожиданий населения представляет существенный вклад в развитие теории исследования текстовой информации в

экономической науке. Алгоритм учитывает различные аспекты исследования текстовых данных для множества областей экономики.

3) предложенный алгоритм оценки устойчивости результатов расчета индикаторов интенсивности экономических ожиданий к составленным правилам отбора новостных статей СМИ и комментариев интернет-пользователей представляет весомый вклад в развитие процедуры оценки качества результатов поиска информации, релевантной содержательным целям исследователя. Алгоритм позволяет сделать вывод о качестве поиска информации в текстовых данных, а также наличии либо отсутствии негативной специфики рассматриваемого набора данных.

Практическая значимость исследования заключается в следующем:

1) предложенная методика представляет практическую ценность при построении высокочастотных индикаторов экономических ожиданий населения. Она может быть использована центральными банками и министерствами экономики различных стран, реализация политики которых предполагает мониторинг и детальный анализ финансового поведения домохозяйств в зависимости от различных экономических событий и трендов;

2) индикатор интенсивности инфляционных ожиданий представляет существенный вклад в практику оценки инфляционных ожиданий населения в России. Индикатор может быть использован: (1) Банком России при реализации политики инфляционного таргетирования для уточнения и дополнения существующих измерений, в том числе в области высокочастотных оценок; (2) крупным бизнесом, взаимодействующим с конечным потребителем, с целью прогноза его возможных будущих действий или принятия решений, зависящих от инфляционных ожиданий населения, в частности решений о ценообразовании;

3) предложенная комплексная процедура тематического моделирования на основе алгоритма Латентного размещения Дирихле представляет собой значимый вклад в практику выделения ключевых тем в

текстовой информации и может быть использована в различных областях экономической науки.

### **Методы исследования**

В диссертации использовались методы научного познания теоретического и эмпирического уровней.

На теоретическом уровне в качестве ключевых методов исследования применены: анализ, синтез, экономический анализ, включая сравнение, группировку, метод научной абстракции.

На эмпирическом уровне основными методами исследования являлись:

1. Метод информационного поиска на основе комбинирования регулярных выражений с целью выявления заданной исследователем информации в массивах текстовых данных;

2. Метод машинного обучения – тематическое моделирование на базе алгоритма Латентного размещения Дирихле – для выявления кластеров слов и словосочетаний с максимальной содержательной нагрузкой в массивах текстовой информации;

3. Различные статистические и эконометрические методы, включая: вычисление статистических критериев для валидации результатов тематического моделирования; определение коэффициентов корреляции и критериев взаимной информации для оценки степени взаимозависимости временных рядов; проведение теста Гренджера для первичной оценки причинно-следственных эффектов между временными рядами.

### **Информационная база исследования**

В качестве информационной базы для диссертации использовались:

1) исследования отечественных и зарубежных авторов, материалы периодической печати, медиа-источники, представленные в сети Интернет.

2) новостные статьи четырех крупных экономических СМИ России, находящиеся в открытом доступе за период с января 2014 г. по февраль 2018 г.: Ведомости, РБК, Коммерсант, Эксперт;

3) комментарии интернет-пользователей ко всем новостным статьям выборки п. 2, опубликованные и находящиеся в открытом доступе в

следующих источниках информации за период с января 2014 г. по февраль 2018 г.: официальные страницы СМИ в сети Интернет, официальные группы СМИ в социальных сетях ВКонтакте и Facebook.

Все указанные данные собраны с помощью специальной программы-краулера, написанной на языке программирования Python, при осуществлении проекта Big Data Indicators [Голощапова, Андреев, 2017; Андреев, 2017].

### **Положения, выносимые на защиту**

1. Для существующих подходов к оценке экономических, и в частности инфляционных, ожиданий населения, основанных на опросах экономических агентов и на анализе биржевых параметров, характерны значительные недостатки. Последние связаны, в первую очередь, с низкой частотностью и неустойчивостью итоговых индикаторов к правилам расчета, слабым охватом целевой аудитории, высокими издержками расчетов, а также низкой релевантностью конкретным видам рисков. Это свидетельствует о необходимости разработки новых подходов к оценке экономических ожиданий населения.

2. Авторская методика построения высокочастотных индикаторов экономических ожиданий населения на основе больших массивов текстовых данных способствует устранению недостатков существующих методов за счет следующих преимуществ: (1) широкий охват целевой аудитории; (2) высокая частотность анализа; (3) обновление индикаторов в реальном времени, что невозможно для индикаторов на основе опросов; (4) устойчивость результатов построения индикаторов к правилам вычисления и возможность перерасчета при изменении правил; (5) низкие издержки проведения расчетов.

3. Разработанный алгоритм выявления факторов изменения экономических ожиданий населения с помощью анализа текстовых данных способствует углублению знаний о факторах (экономических явлениях и трендах), с которыми население связывает изменение собственных ожиданий. В дополнение к классическим статистическим методам и

экспертным оценкам предложенный алгоритм опирается на современные методы тематического моделирования. Это способствует снижению доли субъективной экспертной оценки при выявлении факторов, а также позволяет на высокочастотном уровне проверить появление новых категорий или тем, существенных для изменения ожиданий.

4. Индикатор инфляционных ожиданий российского населения, построенный для апробации авторской методики, впервые в российской практике позволяет исследовать интенсивность ожиданий на ежедневной основе. По результатам статистических тестов построенный индикатор оказывает значимое влияние на будущую динамику потребительской инфляции. При этом полученный индикатор в среднем на 2 месяца опережает динамику индикатора-аналога, формируемого на базе опросов и являющегося основным для анализа инфляционных ожиданий российского населения на текущий момент.

5. Разработанная методика позволила выявить новые для исследования инфляционных ожиданий российского населения факторы, с которыми население связывало изменение ожиданий с января 2014 г. по февраль 2018 г. Это –неопределенность экономической ситуации, внутренняя политика и мировая конъюнктура за исключением воздействия динамики курса доллара и цен на нефть, а также тарифы, налоги и платежи, ситуация в секторе недвижимости и условия розничной торговли.

6. Апробация авторской методики для инфляционных ожиданий позволила верифицировать и подтвердить на высокочастотном уровне значимость выделяемых в существующих исследованиях факторов: динамики цен на нефть, условий денежно-кредитной политики и изменения доходов населения. Показано влияние на ожидания динамики валютного курса в краткосрочном периоде.

#### **Степень достоверности результатов**

Достоверность полученных результатов обеспечивается соответствием научных положений, рекомендаций и выводов диссертации следующим критериям:

1. Достоверность полученных результатов основана на общепринятых принципах фундаментальных и прикладных наук и корректном использовании строго доказанных выводов, а также на комплексном использовании научных методов при проведении исследования.

2. Достоверность результатов работы обеспечивается полученным высокочастотным индикатором интенсивности инфляционных ожиданий российского населения в период с января 2014 г. по февраль 2018 г., оценками его релевантности, устойчивости результата к правилам расчета, а также выявленными на основе комплексной процедуры факторами изменения индикатора.

3. Достоверность полученных результатов базируется на публикации ключевых результатов исследования в рецензируемых научных журналах.

#### **Соответствие диссертации паспорту научной специальности**

Диссертационное исследование соответствует следующему пункту паспорта специальности 08.00.10 – Финансы, денежное обращение и кредит: 4.8. Развитие финансовых отношений и принятие финансовых инвестиционных решений в домашнем хозяйстве.

#### **Апробация результатов**

Результаты настоящего исследования были представлены на следующих конференциях и научных семинарах:

1. Солнцев О.Г., Сухарева И.О. О проекте «Основных направлений единой государственной денежно-кредитной политики на 2014 г. и период 2015 и 2016 гг.» // Заседание Совета по финансовому регулированию и денежно-кредитной политике Ассоциации региональных банков России. Государственная Дума. – Октябрь 2013 г.

2. Сухарева И.О. Monetary Policy versus Economic Growth // HSE PhD Seminar in Finance. XV Апрельская международная научная конференция «Модернизация экономики и общества» в НИУ ВШЭ. – Апрель 2014 г.

3. Ser-Huang Poon, Goloshchapova Irina, Matthew Pritchard and Phil Reed. Corporate Social Responsibility Reports: Topic Analysis Big Data Approach //

Workshop on “Recent Developments in Econometrics and Financial Data Science”. ICMA Centre, Reading University. – November 2017.

4. Goloshchapova I., Andreev M. Big Data approach to measure inflation expectations: the case of the Russian economy // IFABS 2017, Oxford. – July 2017.

5. Голощапова И. Оценка экономических ожиданий населения на основе big data (на примере инфляционных ожиданий населения России) // Научный семинар ЦМАКП. – Ноябрь 2017 г.

6. Голощапова И. Макроэкономические индикаторы и Big Data. Интернет-экономика. Радио “МедиаМетрикс”. – Март 2018 г.

По результатам диссертационного исследования реализован исследовательский проект «Big Data Indicators» по оценке экономических ожиданий населения на основе больших данных в соавторстве с М. Андреевым [Голощапова, 2017]. Результаты применения разработанной методики для построения высокочастотных индикаторов инфляционных ожиданий и кризисных настроений российского населения размещены на веб-сайте [bigdata-indicators.com](http://bigdata-indicators.com) с ноября 2017 г. Сайт предполагает регулярное, ежемесячное, обновление значений индикаторов и публикацию отчетов с их детальной интерпретацией.

Основные положения диссертации представлены в 7 публикациях автора, (общим объемом – 9,4 п.л., из них личный вклад автора – 4,4), в том числе в 4 работах в научных журналах, входящих в перечень ВАК (общим объемом – 8,1 п.л., из них личный вклад автора – 3,6 п.л.).

### **Структура диссертации**

Цель и задачи определили структуру настоящего исследования. Диссертационная работа содержит введение, три главы, заключение и библиографию общим объемом 205 стр., включая 10 таблиц, 14 рисунков и 4 Приложения.

# Глава 1. Современные подходы к оценке экономических ожиданий населения

## 1.1. Ожидания экономических агентов как значимый фактор экономической динамики

### 1.1.1. Определение экономических ожиданий населения

Первое упоминание экономических ожиданий было зафиксировано в Древней Греции. В трактате «Политика» Аристотель написал о Фалесе из Милета (636-546 до н.э.), который получил значительную выгоду от точного прогноза будущего урожая оливок.

Экономические ожидания в международных научных исследованиях определены как прогнозы экономических агентов относительно будущих событий в экономике, которые значимо влияют на принятие решений [Mikolajek-Gosejna, 2014, p. 35].

С точки зрения ожиданий отдельных категорий экономических агентов, как правило, выделяют следующие ключевые группы: домохозяйства, фирмы, государство и экспертное сообщество. В данной работе наибольший интерес представляет исследование экономических ожиданий домохозяйств или населения. Основная роль домохозяйств в экономике: потребление товаров и услуг, взаимодействие на рынке труда, частные инвестиционные и сберегательные решения и голосование на выборах [Hartley, 1997].

Таким образом, наиболее существенными в экономической теории областями экономических ожиданий являются ожидания населения относительно:

- общей экономической конъюнктуры и стадии бизнес-цикла: наличие или отсутствие кризиса, общая неопределенность экономической ситуации и др. Эта категория ожиданий оказывает непосредственное влияние на принятие решений в рамках всех



ключевых ролей домохозяйств в экономике. Например, в случае сильных ожиданий экономического кризиса, потребление населения, как правило, снижается, инвестиционные и сберегательные решения корректируются, решения о смене работы откладываются и т.д.

- инфляции: общая траектория темпов роста цен и рост цен на отдельные товары и услуги или их группы. При инфляционном таргетировании в качестве режима денежно-кредитной политики инфляционные ожидания экономических агентов играют ключевую роль для принятия решений о динамике ключевой процентной ставки. Кроме того, ожидания относительно инфляции влияют на решения домохозяйств о потреблении, сбережениях и займах, а также на переговоры о заработной плате с работодателем [Tyszka, 1997];
- курса национальной валюты: общая траектория курса относительно ключевых валют. Ожидания слабого или сильного курса национальной валюты могут существенно корректировать как решения о вложениях в финансовые активы и сберегательное поведение населения, так и влиять на общие решения о потреблении – в особенности в экономиках, ориентированных на экспорт сырья;
- уровень дохода: общий уровень дохода или заработной платы. Ожидания относительно будущей динамики доходов влияют как на потребление, так и на инвестиционные и сберегательные решения. Кроме того, они могут оказывать влияние на решения, касающиеся рынка труда (смена работы и др.);
- ситуации на рынке труда: уровень безработицы, привлекательность поиска работы в отдельный период времени. Ожидания относительно ситуации на рынке труда оказывают прямое воздействие на решения населения о выходе на работу или увольнении, поиске и смене работы, а также на формирование

требований к работодателю относительно повышения заработной платы;

- устойчивости финансового сектора: привлекательность отдельных периодов времени для размещения депозитов или инвестиций в финансовые активы и рынок недвижимости, опасения относительно устойчивости банковского сектора. Данная категория ожиданий напрямую воздействует на инвестиционные и сберегательные решения населения. Например, ожидания снижения устойчивости банков могут значительно повысить вероятность возникновения «паники вкладчиков».

### *1.1.2. Влияние экономических ожиданий на макроэкономическую конъюнктуру*

Исследование вопроса влияния ожиданий экономических агентов на макроэкономическую конъюнктуру развивается уже около ста лет. Одними из первых в этой области стали исследования Найта [Knight, 1921] и Кейнса [Keynes, 1936]. Эти исследования обозначили проблему неопределенности среды, в которой экономические агенты принимают решения и необходимость понимания процесса влияния ожиданий агентов на механизм принятия решений. Так, в соответствии с теорией Кейнса, в краткосрочном периоде ожидания были определены как главный фактор неопределенности и волатильности экономических циклов.

Работа Эзекиела [Ezekiel, 1938] стала одной из первых попыток оценки влияния ожиданий на устойчивость экономического равновесия (в рамках паутинообразной модели). При этом Метцлер [Metzler, 1941] разработал две макроэкономические модели циклов запасов, включающих ожидания как один из параметров.

В исследовании Харта [Hart, 1949] получила значительное развитие теория самих ожиданий. Автор акцентировал внимание на роли знаний, планов и представлений о будущем со стороны экономических агентов в условиях неопределенности в рамках теории инвестиций. В исследовании

впервые была явно обозначена потребность в измерении ожиданий на историческом периоде – в целях более точного позиционирования концепции ожиданий в теории инвестиций фирм. В качестве возможного решения было предложено проведение регулярных оценок состояния ожиданий бизнеса или других экономических агентов со стороны экономистов и «специалистов по исследованию отношения» к различным проблемам экономики на основе опросов. Кроме того, в работе был поднят вопрос о различных механизмах формирования ожиданий, включая адаптивность последних и скорость их изменений в зависимости от происходящих событий.

Следующим важным этапом для интеграции концепции экономических ожиданий в экономическую теорию стала работа Мута [Muth, 1961]. Последняя оказала значительное влияние на рост внимания к анализу механизмов формирования ожиданий, а также на интеграцию количественных и качественных факторов, связанных с ожиданиями, в теоретические и практические модели множества областей экономики. В частности, это динамические модели потребления, рынка труда со стороны предложения, алгоритмы формирования цен инструментов на финансовых рынках, теория денежно-кредитной политики при режиме инфляционного таргетирования, а также решения экономических агентов при осуществлении покупок и продаж товаров и услуг или инвестиционных вложениях и др. При этом в теории бизнес-циклов экономические ожидания являются ключевым трансмиссионным механизмом для действий экономической политики [Kowalski, 1998, p. 99].

Возможности влияния экономических ожиданий на будущую макроэкономическую динамику, во многом, определяются алгоритмами их формирования. В качестве ключевых алгоритмов здесь следует выделить адаптивные и рациональные ожидания.

### *1.1.3. Адаптивные экономические ожидания*

Гипотеза адаптивных ожиданий стала первой гипотезой о механизме формирования ожиданий экономических агентов, основанной на

предположении об их последовательном развитии во времени или зависимости текущих уровней ожиданий от предшествующих значений ожиданий и их объекта. Простая форма данной гипотезы была впервые озвучена в работах Койка [Койск, 1954] в применении к теории инвестиций и Кагана [Cagan, 1956] в контексте изучения спроса экономических агентов на деньги в условиях гиперинфляции. Затем, гипотеза адаптивных ожиданий получила развитие в исследовании Фридмана [Friedman, 1957] в рамках выдвижения гипотезы постоянного дохода и Нерлов [Nerlove, 1958] при идентификации функции предложения на рынке агропродукции.

Пусть  $y_{t-i}^e$  ожидания  $y_{t-i}$  со стороны экономического агента, сформированные в предшествующий период времени  $t - i - 1$ , с учетом всей доступной на тот момент информации, т.е. в границах информационного множества  $\Omega_{t-i-1}$ . В таком случае, простую форму концепции адаптивных экономических ожиданий можно представить следующим образом:

$$y_t^e - y_{t-1}^e = \theta(y_{t-1} - y_{t-1}^e) \quad 0 < \theta < 1 \quad (1)$$

Параметр  $\theta$  здесь отвечает за адаптивность или преемственность ожиданий от предшествующего периода времени  $t$ . Таким образом, ожидания текущего периода пропорциональны ошибке ожиданий предшествующего периода. Как утверждает Песаран [Песаран, 2002, стр. 194], такая форма гипотезы адаптивных ожиданий называется моделью коррекции ошибки первого порядка.

Согласно выводам Мута [Muth, 1960], результатом выполнения формулы (1) может являться оптимальный с точки зрения статистики прогноз, если процесс формирования  $\{y_t\}$  будет представлен как интегрированный процесс первого порядка с параметром скользящей средней. Оптимальный прогноз  $y_t$ , сформированный на основе информации  $t - 1$  будет в таком случае задан следующим образом:

$$E(y_t | \Omega_{t-1}) = y_{t-1} - (1 - \theta)\varepsilon_{t-1}, \quad (2)$$

где  $\varepsilon_t$  – независимые одинаково распределенные случайные величины  $N(0, \sigma)$  или i.i.d.

В ряде работ гипотеза адаптивных ожиданий была расширена до предпосылки о формировании ожиданий на основе интегрированной авторегрессионной модели скользящего среднего – ARIMA. В обзоре Песарана [Песаран, 2002, стр. 195] такое расширение названо гипотезой экстраполяционных ожиданий. Процесс формирования ожиданий в контексте этой гипотезы может быть представлен классической моделью ARIMA (p, d, q):

$$\varphi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t, \quad (3)$$

где  $\varphi(L), \theta(L)$  – лаговые многочлены порядка p и q соответственно. При этом все абсолютные значения корней уравнения (3) больше единицы, т.е. авторегрессионный процесс формирования ожиданий не является стационарным. Тогда ожидания относительно  $y_t$  могут быть получены с помощью уравнения:

$$y_t^e = E(y_t | \Omega_{t-1}) = \sum_{i=0}^{\infty} w_i y_{t-i-1} = W(L)y_{t-1} \quad (4)$$

В соответствии с работой Песарана [Песаран, 2002, стр. 195], веса  $w_i$  в формуле (4) могут быть заданы в терминах параметрических оценок ARIMA. Однако подобный способ формирования ожиданий получил распространение лишь в небольшом круге экономических исследований [Trivedi, 1973; Nerlove et al., 1979; и др.]. В рамках гипотезы экстраполяционных ожиданий генерирующая функция для  $W(L)$  определяется априорно. Кроме того, экстраполяционный механизм формирования ожиданий включает предпосылку о возвращении ожиданий к нормальному уровню:

$$y_t^e = y_{t-1} - \lambda(y_{t-1} - y_{t-1}^*) \quad \lambda > 0 \quad (5)$$

Здесь  $y_{t-1}^*$  отвечает за средний уровень  $y_t$ , который также называют «нормальным».

Следует упомянуть, что при предположении об экстраполяционном механизме формирования ожиданий крупные блоки экономических исследований были посвящены, с одной стороны, поиску оптимального уравнения для  $y_{t-1}^*$  в уравнении (5) [Turnovsky, 1970 и др.], с другой – обобщению представленной выше модели (4) [Meiselman, 1962; Mincer, 1969; Frenkel, 1975].

В процессе своего развития гипотеза адаптивного механизма формирования ожиданий и ее расширения были интегрированы в различные экономические концепции, описывающие текущие макроэкономические процессы. Одна из основных причин распространения этой гипотезы – легкость в ее понимании и интерпретации [Pearce, 1986]. Так, в исследованиях Чоу [Chow, 1989; Chow, 2011] было показано строгое статистически значимое влияние ожиданий экономических агентов на приведенную стоимость цен на акции и дивиденды в период 1971-2010 гг. В работе Миллса [Mills, 1961] утверждалось, что результатом применения гипотезы адаптивных ожиданий к историческим данным является временной ряд ожиданий, соотносимый с траекторией фактических экономических показателей. При этом в случае, если исходные значения ожиданий и факта совпадают, то временные ряды ожиданий и факта также будут совпадать [Lovell, 1986, p. 113]. Это является верным и для ошибки прогноза. Среднее значение ошибки ожиданий во времени будет стремиться к нулю, если фактическая траектория показателя будет стремиться к равновесному состоянию. Важно отметить, что для гипотезы адаптивных ожиданий верно и обратное утверждение: в случае, если рынок не стремится к равновесию, различия ожиданий и факта становятся систематическими.

Несмотря на многочисленные экономические исследования, посвященные развитию гипотезы адаптивных ожиданий экономических агентов и ее расширений и исследованию их влияния на общую макроэкономическую динамику, данная концепция осталась в значительной степени уязвимой. Основной недостаток гипотезы адаптивных ожиданий базируется на слабости предпосылки об их зависимости исключительно от собственных предшествующих значений и соответственной концентрации гипотезы на прошедших событиях – без учета текущей экономической конъюнктуры и непосредственных прогнозов на будущее со стороны экономических агентов. В рамках такой парадигмы экономические агенты не способны учиться на своих ошибках – за исключением механической коррекции в соответствии с коэффициентом адаптации и ошибкой предшествующего периода. Таким образом при адаптивном механизме формирования ожиданий экономические агенты вынуждены ждать ускорения фактической инфляции, для того чтобы изменить свои ожидания относительно нее.

В то же время значительной критике подвергается и предположение о постоянном коэффициенте адаптации одновременно с эвристическим ограничением его масштаба в интервале от нуля до единицы [Gertchev, 2007]. Так, в работе Хана [Khan, 1977] была предложена модификация гипотезы адаптивных ожиданий, опирающаяся на изменяющийся коэффициент адаптации. В более поздних исследованиях было показано, что коэффициенты статистически значимо различаются как для разных групп экономических агентов [Mlambo, 2011], так и для отдельных людей, и в разные периоды времени [Figlewski and Watchtel, 1981]. Дополнением к критике постоянного коэффициента адаптации также является возможность скоррелированности ошибок прогноза во времени. Это может обусловить еще большее отставание изменения ожиданий от изменения тренда фактического показателя. Отсюда следует необходимость учета в теории формирования ожиданий не только предшествующих значений самих

ожиданий, но и предшествующих, настоящих и будущих (ожидаемых) значений различных экономических факторов.

Существенные недостатки гипотезы адаптивных ожиданий обуславливают необходимость рассмотрения другой важной в экономической теории гипотезы о механизме формирования ожиданий экономических агентов – предположения о рациональных ожиданиях. Последнее направлено на борьбу с недостатками гипотезы адаптивных ожиданий.

#### *1.1.4. Рациональные экономические ожидания*

В соответствии с Герчевым [Gertchev, 2007], отсутствие полноценного учета возможного рационального поведения экономических агентов в рамках гипотезы адаптивных ожиданий привело к долгосрочному лидерству гипотезы рациональных ожиданий в экономической литературе.

Гипотеза рациональных ожиданий была сформулирована в работе Мута [Muth, 1961] и в дальнейшем была значительно развита и расширена в исследованиях Сарджента и Уоллеса [Sargent, Wallace, 1975] и Лукаса [Lucas, 1976] и др.

В соответствии с гипотезой рациональных ожиданий, процесс формирования ожиданий экономических агентов оптимален и основан на объективной структурной модели экономики. В таком случае субъективные ожидания экономических агентов являются объективными [Pearce, 1986; Alejandro, 2008]. В общем случае рациональные ожидания находятся в зависимости от исторических значений набора факторов, влияющего на объект ожиданий и определяемого заданной моделью экономики.

Математически гипотеза рациональных ожиданий хорошо сформулирована в работе Мламбо [Mlambo, 2012, р. 6]. Пусть  $Y_t$  – объект ожиданий или целевая переменная, а  $X$  и  $Z$  – факторы, влияющие на объект ожиданий при объективной модели экономики.

Тогда гипотеза рациональных ожиданий может быть представлена следующим уравнением (6):



$$Y_t = a_0 + a_1 Y_{t-1} + a_2 X_{t-1} + a_3 Z_{t-1} + U_t, \quad (6)$$

где  $U_t$  – случайная величина или ошибка прогноза. Величины всех факторов периода  $t-1$  известны агенту в момент прогноза (формирования ожиданий). При этом ожидания экономического агента в момент времени  $t$  могут быть выражены как (7):

$$E_{t-1}(Y_t) = a_0 + a_1 Y_{t-1} + a_2 X_{t-1} + a_3 Z_{t-1}, \quad (7)$$

поскольку  $E_{t-1}(U_t) = 0$ . В случае, если модель экономики специфицирована верно,  $U_t$  является независимой одинаково распределенной случайной величиной. Таким образом,  $U_t$  отражает возможные аномалии в динамике объекта прогноза ( $Y_t$ ), вследствие событий, неожиданных для экономического агента.

По утверждению Песарана [Песаран, 2002, стр. 198], оптимальное решение в рамках гипотезы рациональных ожиданий возможно только при полной спецификации процесса, формирующего факторы, ключевые для определения ожиданий. В целях такой спецификации могут быть использованы авторегрессионные модели временных рядов. Однако следует учитывать, что сложность нахождения оптимального решения при гипотезе рациональных ожиданий напрямую зависит от сложности задаваемой модели временных рядов для факторов.

Гипотеза рациональных ожиданий получила широкую поддержку не только в теоретических работах, но и в эмпирических исследованиях. Согласно Муту [Muth, 1961, p. 316] для рассматриваемой гипотезы ожидания экономических агентов являются в то же время и предсказанием для экономических показателей – объектов ожиданий. Это подтверждается и математической формулировкой гипотезы, которая аналогична задаче регрессии или прогноза целевой переменной от набора релевантных факторов. В расширенном понимании гипотезы рациональных ожиданий динамику экономических показателей формируют именно субъективные

ожидания экономических агентов, поскольку агенты принимают решения именно на основе своих ожиданий [Levine, 2012 и др.]. На базе эмпирического анализа Мюэльбауэр [Muellbauer, 1981] заключил, что решения на рынке труда, во многом, обусловлены именно рациональными ожиданиями. В работе Гудвина и Шеффрина [Goodwin and Sheffrin, 1982] было показано влияние рациональных ожиданий экономических агентов на спрос и предложение на отдельных товарных рынках, на примере рынка куриных бройлеров. В исследовании Айгингера [Aiginger, 1981] было выявлено статистически значимое влияние рациональных ожиданий экономических агентов на потребление и инвестиции – на основе данных по продажам, инвестициям и ценам в США, Японии и Европе.

Основными направлениями критики гипотезы рациональных ожиданий являются:

- слабость предположения о том, что ошибка прогноза – это независимо распределенная случайная величина. С одной стороны, ряд эмпирических исследований показывает наличие систематической ошибки прогноза у различных категорий экономических агентов [Gertchev, 2007, p. 318]. С другой стороны, существование и развитие бизнеса профессиональных аналитиков указывает на сохранение постоянной потребности в точных прогнозах со стороны экономических агентов и подтверждает систематическое нарушение гипотезы рациональных ожиданий и наличие информационного арбитража;
- слабость предпосылки об одинаковой рациональности ожиданий для всех категорий экономических агентов [Gomes, 1982; Wible, 1982]. Данное предположение не учитывает различия в возрасте, уровне образования, опыте и навыках, исходную осведомленность об экономических процессах и доступ к информации о текущих процессах в экономике [Kozielecki, 1997]. Одинаковые

рациональные намерения не всегда означают одинаковые возможности для их реализации [Mlambo, 2012, p. 11];

- противоречие в формулировке самой гипотезы. Утверждение о том, что ожидания непосредственно влияют на экономическую реальность не в полной мере сопоставимо с тем, что объективная модель экономики, на основе которой агенты принимают решения, не учитывает ожиданий. Это противоречие привело к интеграции ожиданий экономических агентов в модели для различных областей экономики [Gertchev, 2007, pp. 324-325].

Таким образом, ключевыми теориями, описывающими процессы формирования ожиданий экономических агентов, являются гипотезы адаптивных и рациональных ожиданий.

Для обеих гипотез характерны как преимущества, так и недостатки. Важнейшим преимуществом гипотезы адаптивных ожиданий является легкость в ее понимании и интерпретации, недостатком – отсутствие учета факторов, влияющих на объект ожиданий, помимо предшествующей динамики целевой переменной. Устранение этого недостатка в гипотезе рациональных ожиданий привело к существенно большей ее популярности в экономической теории. Однако сама эта гипотеза также уязвима – в основном, за счет: слабости предположений об ошибке прогноза как о случайной величине и об одинаковой рациональности ожиданий для всех категорий экономических агентов, а также противоречий в формулировке самой гипотезы.

Реальный механизм формирования ожиданий представляет собой совокупность рациональных и адаптивных процессов, которая может отличаться в зависимости от особенностей рассматриваемой ситуации. На механизм формирования ожиданий могут оказывать влияние следующие особенности: категории экономических агентов (намерения, характеристики индивида, возможности и др.), экономическая среда (институциональная

инфраструктура, макроэкономическая динамика), объект ожиданий и его свойства.

## **1.2. Способы измерения экономических ожиданий**

В настоящее время существует два ключевых подхода к построению индикаторов экономических ожиданий: базирующийся на опросах экономических агентов и основанный на анализе высокочастотных биржевых параметров, учитывающих прогнозы участников финансового рынка относительно различных параметров экономики. Первый подход носит название *survey-based approach*, второй – *market-based approach*.

Кроме того, в последнее десятилетие на фоне активного развития методов машинного обучения, и в частности текстовой обработки, постепенно возрастает популярность построения экономических индикаторов на основе больших массивов данных. Этот подход может быть назван *big data approach*. В работах по теме экономических ожиданий населения также наблюдаются индикаторы, построенные с помощью элементов текстового анализа. В основном, это характерно для аппроксимации факторов, влияющих на формирование экономических ожиданий.

### *1.2.1. Измерение экономических ожиданий населения с помощью опросов*

Подход к измерению экономических ожиданий с помощью социологических опросов или *survey-based approach* широко применяется во множестве стран для измерения экономических ожиданий различных категорий экономических агентов. В особенности, это верно для центральных банков стран, реализующих режим инфляционного таргетирования в качестве режима денежно-кредитной политики. Ключевым объектом для измерения в данном случае служат инфляционные ожидания населения, бизнеса, а также профессиональных аналитиков и участников финансовых рынков [Голощапова, Андреев, 2017].

В общем, при измерении экономических ожиданий целевой аудиторией опросов являются домохозяйства (население), фирмы и профессиональные аналитики [Cunningham et al., 2010; Gali, 2008]. При этом наиболее распространенными и, соответственно, приоритетными для этого подхода являются опросы населения. Проведение масштабных исследований экономических ожиданий домохозяйств на уровне стран финансируется в США, Австралии, Швеции, ЕС, Канаде, России, и множестве других стран.

Эмпирические научные работы подчеркивают необходимость проведения опросов для оценки экономических ожиданий населения [Mlambo, 2012; Figlewski and Wachtel, 1981]. Индикаторы экономических ожиданий, полученные на основе опросов, являются ценными как для осуществления объективного прогноза будущей динамики ключевых макроэкономических параметров, так и для предсказания дальнейших финансовых решений населения.

Важными аспектами анализа результатов социологических опросов населения являются не только количественные оценки самих экономических ожиданий (доля респондентов, ожидающих определенной динамики целевого параметра, медианные оценки ожидаемых значений целевого параметра и др.) [Van der Klaauw et al., 2008], но и оценка неопределенности ожиданий в ответах респондентов (доля респондентов, которые не смогли определиться с ответом на вопрос, дисперсия ответов) [Sahm, Sockin, 2016; Coibion and Gorodnichenko, 2015].

Однако следует выделить также и ряд существенных недостатков метода измерения экономических ожиданий населения с помощью социологических опросов.

1. Значительные организационные и финансовые затраты на проведение опросов. К примеру, для ежемесячного проведения личных интервью с двумя тысячами домохозяйств по всей России информационному агентству «инФОМ», официальным заказчиком данных по экономическим ожиданиям населения для которого является Банк

России, требуется привлечение 42 организаций на территории страны. Корректное проведение опросов дополнительно осложняется тем, что пересечение респондентов возможно только раз в полгода. По завершении этапа проведения интервью, необходимо внести ответы респондентов в специальное программное обеспечение. При этом для того, чтобы минимизировать число возможных ошибок, осуществляется перекрестная проверка корректности внесенных данных с устранением допущенных неточностей, включая дополнительные обращения к респондентам при необходимости. Расчеты на основе полученных данных и их анализ становятся возможными только на завершающем этапе<sup>1</sup>.

2. Неустойчивость ответов респондентов в зависимости от формулировки вопроса, свойств выборки и расчетных особенностей работы с данными. Опросы населения в США осуществляются множеством организаций: New York Federal Reserve's Survey of Consumer Expectations, Reuters/University of Michigan Survey of Consumers, Conference Board's Consumer Confidence Survey и др. При этом агрегированные результаты опросов характеризуются систематическими расхождениями, несмотря на общие принципы проведения опросов [Armantier et al., 2013a]. Различия наблюдаются вследствие неодинаковых выборок для анализа, особенностей осуществления интервью [Bruine de Bruin et al., 2016], методик идентификации выбросов в собранных данных [Curtin, 1996], а также специфики округления агрегированных результатов [Binder, 2015].

При этом значительная часть эмпирических исследований посвящена детальному анализу реакции респондентов на разнообразные вариации формулировок вопросов относительно их экономических ожиданий [Armantier et al., 2016, Armantier et al., 2013b, van der Klaauw et al., 2008

---

<sup>1</sup> «Измерение инфляционных ожиданий и потребительских настроений на основе опросов населения». Развернутый отчет по результатам седьмого-двенадцатого опросов. ООО «инФОМ» по заказу Банка России. Июль-декабрь 2017.

и др.]. В целом, такая ситуация ведет к улучшению методологии опросов и анализа динамики и структуры экономических ожиданий. Однако существенным недостатком остается ухудшение сопоставимости агрегированных ответов респондентов во времени при пересмотрах методологии.

3. Низкая частотность публикации результатов опросов. Это является прямым следствием значительных временных и организационных издержек на проведение интервью, включая опросы по телефону, и обработку полученной информации. Как правило, максимальная частотность публикации результатов опросов населения – раз в месяц.

#### *1.2.2. Измерение экономических ожиданий населения на основе анализа биржевых данных*

Другой важный метод измерения экономических ожиданий населения базируется на анализе высокочастотных параметров финансовых рынков. Этот подход, как правило, применяется для выявления ожиданий участников финансовых рынков относительно будущей динамики инфляции, курса национальной валюты и иных показателей, касающихся условий и целей денежно-кредитной политики [Gurkaynak, Sack, Swanson, 2006].

Ожидания изменений ключевой ставки денежно-кредитной политики могут быть измерены с помощью текущих фьючерсных контрактов [Kuttner, 2001; Faust, Swanson and Wright, 2004], контрактов на 1 месяц вперед [Bomfim, 2003; Poole and Rasche, 2000], ставки по депозитам евродоллар на 1 месяц [Cochrane and Piazzessi, 2002], ставки по казначейским векселям (Treasury bills rate) на 3 месяца [Ellingsen, Soderstrom, 2004], фьючерсов на евродоллар на 3 месяца [Rigobon and Sack, 2002]. Для измерения инфляционных ожиданий могут быть использованы спреды суверенных облигаций с аналогичными облигациями, предполагающими защиту от инфляции [Barr, Campbell, 1997], а также инфляционные свопы.

Для метода измерения экономических ожиданий населения на основе биржевых данных не наблюдаются основных недостатков подхода,

базирующегося на опросах. Во-первых, данные по параметрам сделок на финансовых рынках, как правило, доступны в реальном времени. Во-вторых, интерпретация значений индикаторов, построенных на основе статистики финансовых рынков, как правило, не варьируется значительно, поскольку алгоритм их построения определен однозначно. Это совокупные данные о сделках, осуществленных на бирже, по конкретным финансовым инструментам, имеющим отношение к объекту ожиданий.

Вместе с тем рассматриваемый метод в значительной мере уязвим. В первую очередь, market based approach ориентирован на узкую группу населения – участников финансовых рынков, что значительно ограничивает репрезентативность выборки. Во вторую очередь, ожидания, вычисляемые на базе премии за риск по отдельным инструментам, как правило, не отражают конкретную область экономических ожиданий напрямую, так как в большинстве случаев показывают общую обеспокоенность участников финансовых рынков всеми возможными рисками, имеющими отношение к данным инструментам: риск ликвидности, инфляционные ожидания, кредитный риск и др. [Fleckenstein, Longstaff, Lustig, 2010; Millington, Pasaogullari, 2015]. В третью очередь, совокупные данные финансовых рынков могут отражать реальную траекторию ожиданий по отдельным параметрам только в случае относительно крупного объема торгов по рассматриваемым инструментам. Таким образом, вычисление индикаторов экономических ожиданий в большей степени релевантно для стран с развитыми финансовыми рынками.

### *1.2.3. Измерение экономических ожиданий населения с использованием методов анализа больших данных*

В последние десятилетия на фоне постепенного роста популярности методов анализа больших данных появляются эмпирические исследования по построению различных экономических индикаторов с помощью отдельных методов машинного обучения, включая подходы к работе с текстовой информацией.



В частности, среди исследований по теме моделирования экономических ожиданий населения следует выделить работы по оценке влияния новостей в экономических СМИ на динамику экономических ожиданий, где встречаются элементы текстовой обработки. В исследовании Бауэра [Bauer, 2014] показано значимое влияние новостей по теме макроэкономики на динамику инфляционных ожиданий экономических агентов. При этом работа Драгер, Ламла и Пфаджфара [Dräger, Lamla, and Pfajfar, 2016] также содержит оценку воздействия информационного фона на агрегированные результаты ежемесячных опросов населения относительно их экономических ожиданий. Информационный фон измеряется числом новостей на тему монетарной политики и ключевых макроэкономических событий, и трендов.

Следует также выделить блок эмпирических исследований, направленных на использование продвинутых методов текстового анализа и машинного обучения для создания опережающих экономических индикаторов.

Одной из первых наиболее существенных работ в этой области стало исследование Бейкера, Блума и Девиса [Baker, Bloom и Davis, 2016], где авторы применили частотный метод текстовой обработки на базе экономических статей ряда крупных СМИ с целью построения индексов неопределенности общей экономической конъюнктуры и политической ситуации для различных стран. Подход к построению индекса базировался на подсчете частоты появления слова «неопределенность» в собранной базе текстов экономических новостей. Итоговые индикаторы оказались значимо взаимосвязаны с основными макроэкономическими показателями рассматриваемых стран. Например, индекс неопределенности возрастает в периоды экономических кризисов и существенных политических событий (приближение выборов и др.). Более того, на микроуровне наблюдается значимая взаимосвязь между динамикой построенных индикаторов и цен на

акции предприятий, их инвестиционными решениями и политикой на рынке труда.

Помимо частотного подхода относительно активное развитие в эмпирических исследованиях с целью построения опережающих экономических индикаторов получили методы оценки эмоциональной окраски текстовой информации и тематическое моделирование.

В работе Нимана и соавторов [Nyman et al., 2014] был представлен подход к построению прокси для индекса потребительской уверенности Мичиганского университета как одного из ключевых индикаторов экономической конъюнктуры в США – на базе анализа эмоциональной окраски новостных статей и доступной текстовой информации финансовых рынков. В дальнейшем методология построения индикаторов была усовершенствована Ниманом и соавторами [Nyman et al., 2018] в части составления словарей для идентификации эмоциональной окраски текстов. Показатели, построенные по новой методологии, оказались весомыми для предсказания системных рисков финансовой стабильности. При этом составление словарей для измерения эмоциональной окраски текстов в большей степени базировалось не на методах машинного обучения, а на выводах моделей поведенческой экономики и психологии Шиллера [Shiller, 2017]. Последние представляют собой наборы слов, типичных для выражения людьми негативных и позитивных эмоций. Исследование Ардиа, Блюто и Боудта [Ardia, Bluteau, Boudt, 2017] также опирается на заранее подготовленные словари для составления оценок эмоциональной окраски новостных статей, которые затем комбинируются в высокочастотные общие тематические индексы, оптимизированные для предсказания ключевых макроэкономических параметров бизнес-циклов (ВВП, промышленное производство, PMI) с помощью одного из классических подходов машинного обучения – регрессии с регуляризацией (elastic net). В работах Тетлока [Tetlock, 2007] и Соо [Soo, 2013] авторы самостоятельно составляют словари на основе доступной текстовой информации для заданных экономических

тем и показывают значимость числа позитивно и негативно окрашенных слов в целях предсказания динамики финансового рынка и рынка недвижимости.

В работе Торсруда [Thorsrud, 2016] для предсказания параметров бизнес-циклов используются индексы вероятности появления основных тем экономических новостей, выявленных с помощью классического для машинного обучения подхода к вероятностному тематическому моделированию текстовой информации – Латентного размещения Дирихле (Latent Dirichlet Allocation, LDA). Исследование Ларсена и Торсруда [Larsen and Thorsrud, 2015] показывает, что индексы вероятности появления отдельных тем статистически значимы для улучшения предсказания параметров экономической конъюнктуры. Бейлк и соавторы [Balke et al., 2015] используют похожую методологию, опираясь на текстовые данные Beige Book. В работе Голощаповой и соавторов [Goloshchapova et al., 2018] на основе тематического моделирования осуществляется поиск ключевых тем в отчетности компаний, относящейся к их ответственности в области охраны окружающей среды, социальной сферы и корпоративного управления (Environment, social and governance, ESG), а также корпоративной социальной ответственности (Corporate social responsibility, CSR). Исследование произведено для компаний Великобритании и стран Евросоюза за период 1999-2017 гг. По его результатам удастся – в разрезе отраслей и с течением времени – проследить распространение наиболее важных для данного вида отчетностей тем, таких как обеспечение безопасности работников предприятий, соблюдения прав человека, пищевые отходы производства, эффективное использование энергоресурсов, контроль выбросов углекислого газа и др.

В российской практике также можно выделить исследование Яковлевой [Яковлева, 2017], посвященное данной теме, где автор применяет алгоритм Торсруда [Thorsrud, 2016] с элементами подхода Голощаповой и Андреева [Голощапова, Андреев, 2017] и собственными модификациями на новостных данных одного из крупных новостных СМИ в целях предсказания

индекса PMI для России. В частности, аналогично подходу Голощаповой и Андреева [Голощапова, Андреев, 2017], для определения эмоциональной окраски новостей автор использует метод обучения с учителем по заранее размеченной выборке текстов и дискретизацию по вероятности в целях выявления класса «нейтральных» новостей.

В части методов построения индикаторов, использующих элементы текстовой обработки для измерения экономических ожиданий населения, в качестве отдельного направления исследований следует выделить подход на основе анализа частоты запросов Интернет-пользователей в поисковой системе Google. Общедоступный сервис Google Trends в открытом доступе позволяет осуществить выгрузку нормализованных индексов частоты запросов Интернет-пользователей по заданным темам, включая экономические, на ежемесячной основе за любой период, начиная с января 2004 г. Фактически, обобщенная статистика запросов может являться прокси для экономических ожиданий населения, его финансовых и потребительских настроений [Choi, Varian, 2009; Schmidt, Vosen, 2010]. Ряд эмпирических исследований показал значительную эффективность использования подобных временных рядов для прогноза эпизодов финансовой нестабильности [Sarlin, 2014] и кризисных процессов в макроэкономике [D'Amuri, Marcucci, 2015]. В российской практике также встречаются исследования, использующие сервис Google Trends: в 2011 г. в работе Столбова [Столбов, 2011] был сконструирован сводный индекс – «барометр» финансовой конъюнктуры – и показана его статистически значимая роль в предсказании динамики депозитов населения.

Кроме того, следует также выделить широкую практику использования глубоких искусственных нейронных сетей или глубокого обучения (deep learning)<sup>2</sup> для работы с текстовыми данными [Рашка, 2017]. В частности, для

---

<sup>2</sup> Искусственные нейронные сети – группа алгоритмов машинного обучения, оперирующая концепцией комбинирования слоев искусственных нейронов (статистических моделей) в целях анализа больших данных. Глубокой или глубинной нейронной сетью называют нейронную сеть с большим количеством слоев нейронов [Рашка, 2017].

задач «разметки» или классификации новостных статей СМИ по темам, а также в целях предсказания временных рядов отдельных показателей финансовых рынков, популярными методами являются рекуррентные нейронные сети, включая сети с длительной краткосрочной памятью [Long Short-Term Memory, LSTM; Roondiwala, Patel, Varma, 2015]. Указанные подходы находятся на стадии активного развития и при корректном применении позволяют достичь точных прогнозов различных высокочастотных параметров на основе текстовых данных, а также произвести их классификацию по отдельным категориям. При этом анализ и прогноз высокочастотных параметров финансовых рынков стал одной из наиболее популярных сфер применения искусственных нейронных сетей в экономическом анализе в последнее десятилетие [Kutsurelis, 1998; Björklund, Uhlin, 2007].

Вместе с тем важно подчеркнуть то, что методы искусственных нейронных сетей обладают значительными требованиями к данным и задачам по их анализу, что ограничивает диапазон их применимости в анализе больших данных [Рассел, Норвиг, 2006], и в частности для построения экономических индикаторов. Корректное построение нейронной сети, включая оптимизацию всех необходимых гиперпараметров и достижение приемлемого качества работы алгоритма, в большинстве случаев предполагает наличие, во-первых, целевого параметра для прогноза, во-вторых – больших массивов данных с фактическими данными по целевому параметру. Это необходимо для реализации или «обучения» алгоритма в терминологии машинного обучения [Кафтанников, Парасич, 2016].

В экономическом анализе при построении экономических индикаторов на основе больших наборов данных целевой параметр может отсутствовать. Целью исследования, как и в данной работе, может являться само построение временного ряда целевого параметра, а не предсказание его будущих значений. В таких случаях применение искусственных нейронных сетей будет затруднено.

При этом, если в экономических исследованиях целью является прогноз категории каких-либо объектов или их классификация, целевой параметр, как правило, может быть создан экспертом – путем разметки данных вручную. Например, при задаче классификации новостных статей СМИ по категориям (тегирование) новости могут быть экспертно разделены по своему содержанию на нужные исследователю классы. Однако большой размер массива данных, необходимый для обучения искусственной нейронной сети и, соответственно, масштаб трудовых и временных затрат для ручной разметки данных, в большинстве случаев является значительным ограничением для использования нейронных сетей в таких случаях.

Ключевыми преимуществами подхода к построению экономических индикаторов с использованием методов анализа больших данных, включая методы текстовой обработки и машинного обучения, как правило, являются:

- высокая частотность обновления данных. К примеру, анализ новостного потока позволяет получать значения индикаторов по мере поступления новостей, а значит, обновление индикаторов на основе подобных данных возможно в реальном времени;
- широкий охват данных в релевантной области. Так, например, платформа Google Trends агрегирует статистику запросов, осуществляемых всеми интернет-пользователями поисковой системы Google, что несопоставимо больше выборки респондентов для осуществления любой формы опросов населения, или выборки участников финансовых рынков по отдельным инструментам;
- возможность коррекции исторических рядов индикаторов при изменении методики их построения. Подобное не представляется возможным для индикаторов, получаемых на основе опросов. В случае, если было принято решение о корректировке какого-либо из задаваемых респондентам вопросов, ответы респондентов на новый вопрос сформируют отдельный новый показатель, для которого не будут релевантными более ранние данные ответов на другие

вопросы. При использовании текстовой информации пересчет индикаторов возможен на всем историческом периоде – исследователь может самостоятельно формировать любые запросы к имеющемуся набору данных;

- низкие затраты на обеспечение расчетов. Издержки по настройке системы сбора, обработки и хранения больших данных для построения индикаторов несопоставимо ниже издержек на проведение корректной процедуры опросов населения.

Следует также выделить основные возможные недостатки подходов к построению экономических индикаторов с помощью методов анализа больших данных:

- различные аспекты репрезентативности данных, выбранных для анализа. В случае измерения экономических индикаторов на основе больших массивов новых данных, ранее не использовавшихся для этих целей, необходимо обоснование отношения выбранных данных к объекту анализа, а также возможности оценки конкретного индикатора на их основе. В настоящий момент, многие исследования не содержат достаточных обоснований репрезентативности используемых данных. Так, в работах Торсруда и Яковлевой [Thorsrud, 2016; Яковлева, 2017] для построения индексов и оценки их вклада в улучшение прогноза экономических трендов используется всего по одному локальному для страны источнику СМИ, тогда как результаты реализации предлагаемых алгоритмов обобщаются на уровень стран.

При этом в случае составления опросов информационные центры, как правило, аккуратно подходят к составлению выборки респондентов, обеспечивая ее репрезентативность для оценки конкретных экономических ожиданий на заданной территории по всем ключевым характеристикам респондентов (пол, возраст,

образование и др.) в соответствии с многолетней теоретической базой по данному вопросу.

- устойчивость результатов расчета индикаторов к правилам вычисления и обоснованность выбора алгоритма расчета. Многообразие методов анализа больших данных оставляет широкие возможности для необоснованного усложнения предлагаемых алгоритмов измерения экономических индикаторов. При этом чем больше объем исследуемых данных, тем проще найти определенную комбинацию параметров алгоритма, при которой результат окажется статистически значимым для объекта исследования [Мастицкий, Шитиков, 2015]. Однако, в соответствии с так называемым эффектом переобучения, полученный алгоритм не будет устойчивым даже к малому изменению параметров и с высокой вероятностью окажется далек от оптимального уже в ближайшем будущем – относительно периода, на котором происходила настройка параметров алгоритма.
- релевантность полученных индикаторов. Метод измерения экономических ожиданий с помощью больших данных является относительно новым для прикладной экономической науки. Соответственно, любой предлагаемый алгоритм оценки подобных индикаторов должен быть обоснован, в том числе, с точки зрения релевантности итоговых индикаторов, построенных на его основе, для объекта исследования.

### **1.3. Инфляционные ожидания населения**

#### *1.3.1. Воздействие инфляционных ожиданий на динамику экономики*

Инфляционные ожидания населения – один из ключевых объектов исследования для центральных банков при реализации денежно-кредитной политики. В особенности, если главной целью монетарной политики является стабильность цен. В долгосрочном периоде устойчивость ценовой



динамики оказывает значительное влияние как на общее благосостояние в экономике, так и на экономический рост [Loleyt, Gurov, 2011].

Стабильность цен является приоритетной целью режима денежно-кредитной политики, получившего название «инфляционное таргетирование». Этот режим впервые был официально объявлен Новой Зеландией в 1990 г. и с течением времени получил широкое распространение. Наиболее активный переход к данному режиму наблюдался, начиная с 2000 г. - идея поддержания ценовой стабильности в качестве основной цели монетарной политики пришла на смену режимам управляемого валютного курса и таргетирования денежных агрегатов, являвшихся основными в предшествующий период [Апокин и др., 2014; Голощапова, 2014].

Численный ориентир регулятора по инфляции при инфляционном таргетировании как режиме денежно-кредитной политики призван оказывать существенное влияние на инфляционные ожидания населения. Ожидания экономических агентов относительно темпов роста цен, в свою очередь, воздействуют на уровень цен в долгосрочном периоде в моделях экономики с рациональным механизмом формирования ожиданий. Это обусловлено тем, что экономические агенты учитывают уровень цен при принятии решений об инвестициях, потреблении и сбережениях. В работе Армантьера и соавторов [Armantier et al., 2016] подчеркивается существенное воздействие инфляционных ожиданий на будущее финансовое поведение населения.

Таким образом, режим инфляционного таргетирования предполагает критическую важность корректной и оперативной оценки инфляционных ожиданий различных категорий экономических агентов, а именно населения, предприятий и профессионального сообщества [Van der Klaauw et al., 2008].

С 2015 г. официальным режимом монетарной политики в России является инфляционное таргетирование. При этом Центральный банк Российской Федерации (ЦБ РФ) начал постепенный переход к этому режиму существенно раньше – с сентября 2013 г. – с момента введения такого инструмента денежно-кредитной политики, как ключевая процентная ставка.

Следовательно, в течение уже около пяти лет исследование динамики и структуры инфляционных ожиданий населения вносит существенный вклад в политику ЦБ РФ. Одним из основных источников информации с марта 2013 г. для регулятора являются опросы населения, проводимые агентством «инФОМ» по официальному заказу Банка России. Следует также отметить, что в официальных пресс-релизах ЦБ РФ регулярно выделяет фактор инфляционных ожиданий населения как один из определяющих при принятии решений об уровне ключевой процентной ставки. В настоящий момент количественный ориентир монетарной политики ЦБ РФ установлен на уровне 4% на период 2018-2020 гг., в соответствии с «Основными направлениями единой государственной денежно-кредитной политики на 2018 год и период 2019 и 2020 годов».

Важно отметить, что классическим определением инфляции в научной литературе является «устойчивое повышение общего уровня цен на товары и услуги в течение определенного периода времени» [Barro, 1997, p. 895; Blanchard and Johnson, 2013, p. 598; пер. автора с англ.]. При этом определением инфляционных ожиданий населения с учетом определения инфляции и общего определения экономических ожиданий, приведенного в разделе 1.1.1, являются «представления или прогнозы населения относительно будущей общей траектории темпов роста цен и о росте цен на отдельные товары и услуги или их группы». Такое определение соответствует определениям инфляционных ожиданий в современных научных работах [Койчужева, Богатырев, 2014, стр. 42; Ключев, 2010, стр. 16].

С точки зрения механизмов формирования инфляционных ожиданий в отечественной и зарубежной литературе выделяют две основных теории: адаптивных и рациональных ожиданий, как и для экономических ожиданий в целом. Содержание указанных теорий аналогично и было детально раскрыто в разделах 1.1.2 и 1.1.3.

Одной из первых работ в области адаптивного механизма формирования инфляционных ожиданий стало исследование Кагана [Cagan,

1956], посвященное изучению спроса экономических агентов на деньги в условиях гиперинфляции. В предложенной автором модели, ожидания будущей инфляции формировались на основе исторических значений фактической инфляции. В 1981 г. в работе Фиглевски и Уочтела [Figlewski and Wachtel, 1981] также было представлено эмпирическое подтверждение гипотезы адаптивных ожиданий для инфляционных ожиданий экономических агентов.

Несмотря на обнаруженные недостатки теории адаптивного механизма формирования ожиданий, указанная теория получила распространение в исследованиях и в дальнейшем [Carroll, 2003; Mankiw, Reis, Wolfers, 2003]. В работе Койчуевой и Богатырева [Койчуева, Богатырев, 2014] на основе эконометрического анализа показано, что прошлые значения инфляции являются существенным фактором для формирования инфляционных ожиданий населения в Киргизии.

Гипотеза рациональных инфляционных ожиданий также получила широкое распространение в зарубежных и отечественных исследованиях. На основе классической версии гипотезы было развито множество теорий о характере рациональности инфляционных ожиданий экономических агентов, таких как:

- концепции, где агенты не знают истинной модели экономики и формируют инфляционные ожидания, основываясь на доступной информации [Orphanides, Williams, 2005];
- модели, в которых агенты формируют прогнозы инфляции на основе набора правил или эвристик, описывающих их знания об экономике [Brazier et al., 2008];
- модель информационной игры, где агенты формируют ожидания на основе доступной информации с определенной долей недоверия или шума [Demertzis and Viegi, 2008];

- индивидуальных моделей экономических агентов, в которых последние формируют ожидания на основе собственной функции полезности [Gul, Pesendorfer, 2001] и др.

Что касается эмпирических исследований, в работе Форселлза и Кенни [Forsells, Kenny, 2002] на данных опросов населения Европейской комиссией было показано, что экономические агенты склонны к рационализации инфляционных ожиданий, включая устранение систематической ошибки прогноза инфляции с течением времени. Лайнз и Уэстерхофф [Lines, Westerhoff, 2009] на основе динамической модели показали, что совокупные рациональные ожидания экономических агентов оказывают существенное воздействие на уровень инфляции. В исследовании также была выдвинута гипотеза о смешанном характере ожиданий агентов (экстраполяционные и рациональные). Андольфатто, Хендри и Моран [Andolfatto, Hendry, Moran, 2005] показали, что эмпирическая критика гипотезы о рациональных инфляционных ожиданиях на базе классических эконометрических тестов часто бывает некорректна, вследствие особенностей работы с данными (малая выборка, спецификация модели и др.). По результатам модели, представленной в работе Кэролла [Carroll, 2003], в среднем инфляционные ожидания населения не являются рациональными, однако домохозяйства склонны формировать свои ожидания на основе медиа-информации и прогнозов профессионально аналитического сообщества, ожидания которых уже являются рациональными.

В российской практике следует выделить исследования Лолейт и Гурова [Лолейт, 2009; Loleyt, Gurov, 2011], где представлена модель, в которой механизм формирования инфляционных ожиданий экономических агентов определяется их уровнем восприятия информации и доверия к ней (в частности доверия к государству). В работах Соколовой [Соколова, 2012; Соколова, 2014] на базе сравнения кривых Филлипса с различными гипотезами о механизмах формирования инфляционных ожиданий заключается, что в России преобладают адаптивные инфляционные

ожидания. Перевышин, Рыкалин (2017) с помощью статистических тестов приводят эмпирическое доказательство смешанного характера инфляционных ожиданий российских домохозяйств, подчеркивая приоритет адаптивного механизма формирования ожиданий.

Таким образом, для исследования инфляционных ожиданий используются и находят подтверждение на практике гипотезы как об адаптивном, так и о рациональном характере их формирования [Silveira, Tadeu Lima, 2014]. При этом консенсусом в зарубежных и отечественных эмпирических и теоретических исследованиях является смешанный характер формирования ожиданий. Это означает, что с целью прогноза инфляции экономические агенты склонны использовать как значения инфляции предшествующих периодов, так и свои знания об экономике. Кроме того, значительное влияние на формирование ожиданий оказывает доступная экономическим агентам информация (включая медиа-ресурсы), степень достоверности этой информации, а также уровень ее восприятия агентами.

### *1.3.2. Подходы к измерению инфляционных ожиданий*

Ключевыми методами измерения инфляционных ожиданий населения, как и в общем экономических ожиданий, являются опросы и данные финансовых рынков.

Что касается первого подхода, исследования Хейфера и Хейна [Hafer, Hein, 1985], Томаса [Thomas, 1999], Бриана и Венкату [Bryan, Venkatu, 2001], Анга и соавторов [Ang et al., 2007] показали значимое влияние инфляционных ожиданий населения, измеренных с помощью опросов, на будущую траекторию инфляции.

Квантификация инфляционных ожиданий населения на основе опросов может быть произведена двумя основными способами: (1) с помощью прямого вопроса респондентам о численной оценке будущих значений инфляции; (2) с помощью вычисления прогноза инфляции из распределения ответов респондентов о будущих изменениях инфляции относительно текущих значений.

Как в зарубежной, так и в отечественной практике опросы о численных оценках ожидаемых значений инфляции являются классическим способом измерения инфляционных ожиданий профессионального аналитического сообщества (опросы Bloomberg, Reuters, НИУ ВШЭ, Центра развития и др.) и предприятий. Опросы фирм об их количественных оценках будущей инфляции часто осуществляются центральными банками различных стран (включая Банк России, ФРС, Банк Японии, Банк Канады и др.), а также отдельными аналитическими агентствами и научными институтами (NBER, GfK, «Российский экономический барометр» ИМЭМО РАН и др.).

При проведении опросов населения используются оба способа квантификации инфляционных ожиданий. Прямые опросы о численных оценках будущей инфляции (ФРС с помощью MSC Research Center, ЕЦБ, BER и др.) часто дополняются оценками на основе распределения ответов респондентов об их ожиданиях изменения инфляции относительно ее текущих значений. В российской практике ключевые количественные оценки инфляционных ожиданий населения осуществляются на основе опросов информационного агентства «инФОМ» для Банка России<sup>3</sup>. Итоговые оценки ожиданий, используемые Банком России для принятия решений о денежно-кредитной политике, формируются с помощью двух методов. Приоритетное значение имеют результаты оценки на основе прямых вопросов населению относительно их прогноза будущего значения инфляции (медианный индикатор численных ожиданий инфляции). Второй способ – расчет Банка России на базе предположения о нормальном и равномерном распределениях представлений респондентов ФОМ о будущем уровне инфляции<sup>4</sup>. Для расчетов Банк России использует данные ответов на вопрос: «Как, по Вашему мнению, в целом будут меняться цены в следующие 12 месяцев

---

<sup>3</sup> Банк России. О методике исследования инфляционных ожиданий и потребительских настроений населения.

[https://www.cbr.ru/Content/Document/File/17324/FOM\\_meth.pdf](https://www.cbr.ru/Content/Document/File/17324/FOM_meth.pdf)

<sup>4</sup> Банк России. Методика квантификации инфляционных ожиданий населения Банка России.

[https://www.cbr.ru/Content/Document/File/17325/Inflation\\_expectations\\_guide.pdf](https://www.cbr.ru/Content/Document/File/17325/Inflation_expectations_guide.pdf)

(год)?»<sup>5</sup>. Результатам этого способа оценки Банк России придает вторичное значение. В частности, в официальных пресс-релизах о денежно-кредитной политике регулятор, как правило, упоминает результаты именно первого способа оценки. Однако в ежемесячных отчетах «Инфляционные ожидания и потребительские настроения населения» Банк России публикует результаты всех вариантов оценки инфляционных ожиданий.

Комбинирование результатов двух способов квантификации инфляционных ожиданий населения позволяет достичь более точных оценок, поскольку для обоих методов характерны определенные недостатки.

Прямые численные оценки населения относительно как текущей, так и ожидаемой инфляции, как правило, систематически и значительно завышены [Detmeister, Lebow, Peneva, 2016]. Люди склонны не доверять уровню инфляции, согласно официальным фактическим данным, а также ориентироваться в своих оценках на отдельные категории, преобладающие в их потребительской корзине (хлеб, молоко и другие основные продукты питания, бензин и др.). Инфляция по отдельным, «чувствительным» для респондентов категориям, часто оказывается выше общей потребительской инфляции, что влияет и на агрегированные оценки на основе опросов [Murphy, Rohde, 2015]. Во втором случае респонденту не требуется давать точного ответа на вопрос о текущем уровне инфляции, а необходимо оценить ее изменение в будущем относительно любого текущего воспринимаемого уровня. Далее, исходя из различных предположений о распределении ответов респондентов и текущего фактического уровня инфляции, происходит квантификация инфляционных ожиданий – на основе расчета именно ожиданий ее изменения относительно текущего уровня [Arioli et al., 2017; Berk, 1999]. При этом следует отметить, что квантификация инфляционных ожиданий населения на основе распределения ответов респондентов также уязвима. Оценки по данному методу подвержены значительному влиянию

---

<sup>5</sup> Предполагается 5 вариантов ответа респондентов на вопрос: (1) будут расти быстрее, чем сейчас; (2) будут расти так же, как и сейчас; (3) будут расти медленнее, чем сейчас; (4) останутся на нынешнем уровне / не изменятся; (5) будут снижаться.

текущего фактического уровня инфляции и могут скрывать действительные отклонения инфляционных ожиданий населения от текущей инфляции.

Следует отметить, что перечисленные выше недостатки существующих методов измерения инфляционных ожиданий населения на основе опросов (не принимая во внимание общие недостатки метода для измерения экономических ожиданий населения), сконцентрированы на неточностях в определении уровня ожидаемой инфляции, но не его изменений. Это приводит к тому, что при мониторинге инфляционных ожиданий населения большее внимание, как правило, уделяется именно интерпретации их изменений, а не уровня.

Измерение инфляционных ожиданий на базе анализа высокочастотных параметров финансовых рынков основано на мониторинге двух типов индикаторов: спреда доходностей обыкновенных гособлигаций против гособлигаций с защитой от инфляции (TIPS) и инфляционных свопов [Голощапова, Андреев, 2017].

В работе Гуркайнака и соавторов [Gurkaynak et al., 2006] на основе подобных спредов авторами были рассчитаны инфляционные ожидания участников финансовых рынков США, Канады и Чили. В исследованиях Гуркайнака, Сака и Райта [Gurkaynak, Sack and Write, 2007] и Зёдерлинда [Söderlind, 2008] при осуществлении оценки инфляционных ожиданий по аналогичной методике подчеркивается необходимость исключения из спреда гособлигаций премии за риск ликвидности и за неопределенность инфляции для получения корректной итоговой оценки. В последующих работах авторы, как правило, приводят оценки инфляционных ожиданий с устранением указанных рисков [Beechey, Johannsen, Levin, 2011; Grishchenko, Huang, 2013].

Исследования, где расчет инфляционных ожиданий осуществляется на основе инфляционных свопов, менее распространены. Тем не менее, на основе этих биржевых инструментов также возможна квантификация инфляционных ожиданий. В исследовании Девлина и Пэтуордхан [Devlin,



Patwardhan, 2012] производится сравнение результатов оценок по методам на базе спреда суверенных облигаций и инфляционных свопов. Авторы делают вывод о том, что оценки на основе инфляционных свопов являются более корректными. В работе Герлач-Кристен, Моэсснер, Розенблатт-Уиш [Gerlach-Kristen, Moessner, Rosenblatt-Wisch, 2017] на базе сопоставления инструментов на мировых рынках были рассчитаны инфляционные ожидания для ряда стран, не имеющих собственных рынков инфляционных свопов.

Для России ярким примером оценки экономических ожиданий населения на основе биржевых данных является работа Жемкова и Кузнецовой [Жемков, Кузнецова, 2017]. В исследовании авторы рассчитывают инфляционные ожидания участников российских финансовых рынков в 2015-2016 гг. как разницу между номинальной и индексируемой на инфляцию доходностями гособлигаций. Эти показатели характеризуют прогнозы величины и волатильности инфляции со стороны участников финансовых рынков [Крепцев и др., 2016; Kliesen, 2015].

Несмотря на преимущества метода оценки инфляционных ожиданий на основе биржевых данных, связанных с возможностью высокочастотных расчетов, их относительной простотой и доступностью данных, применение данного метода как для оценки инфляционных ожиданий, так и экономических ожиданий в целом, ограничено значительными недостатками. Эти недостатки включают:

- ориентированность на узкую группу населения – участников финансовых рынков, что значительно ограничивает репрезентативность выборки.
- необходимость исключения влияния премии за риск ликвидности, неопределенность инфляции и др.
- необходимость наличия относительно крупного объема торгов по рассматриваемым инструментам в заданной стране.

Что касается применения современных методов анализа данных для оценки инфляционных ожиданий, то следует выделить работы, указывающие на высокое влияние новостного фона на инфляционные ожидания домохозяйств.

В работе Бауэра [Bauer, 2014] статистически подтверждается необходимость учета новостей о макроэкономических событиях и трендах при анализе инфляционных ожиданий различных категорий экономических агентов. Это исследование во многом базировалось на результатах работы Кэролла [Carroll, 2003], в которой была разработана теоретическая модель построения инфляционных ожиданий населения, включающая экономические новости как один из ключевых факторов. Данная модель была основана на гипотезе о том, что новости экономических СМИ – это основной источник информации о будущей динамике цен для домохозяйств. В исследовании Гузман [Guzman, 2011] осуществляется построение высокочастотного индикатора инфляционных ожиданий населения на базе данных о частоте запроса «инфляция» со стороны интернет-пользователей – по данным платформы Google Trends. По утверждению автора, полученный индикатор оказывает значимое влияние на динамику инфляции в пределах 12 месяцев, а также превосходит аналогичные индикаторы, измеренные другими способами, по качеству прогноза инфляционной динамики.

В российской практике работа Голощаповой и Андреева [Голощапова, Андреев, 2017], где представлены индикаторы неопределенности и интенсивности инфляционных ожиданий населения для российских интернет-пользователей, является первой в области применения текстового анализа для оценки высокочастотных экономических индикаторов и инфляционных ожиданий населения.

## **Выводы главы 1**

1. В данной главе конкретизированы наиболее существенные области экономических ожиданий населения в экономической теории с учетом их определения. Экономическими ожиданиями являются прогнозы экономических агентов относительно будущих событий в экономике, которые значимо влияют на принятие решений [Mikolajek-Gosejna, 2014, p. 35]. При этом ключевые области экономических ожиданий населения определены как ожидания относительно: общей экономической конъюнктуры и фазы бизнес-цикла, инфляции, курса национальной валюты, уровня дохода, ситуации на рынке труда, устойчивости финансового сектора.

2. Рассмотрены и систематизированы результаты ключевых теоретических и эмпирических исследований о влиянии ожиданий экономических агентов на экономическую динамику. В частности, в данной главе проанализированы особенности этого влияния при двух ключевых гипотезах о механизме формирования ожиданий: адаптивном и рациональном. Первый механизм предполагает последовательность образования ожиданий во времени или зависимость текущих уровней ожиданий от предшествующих значений ожиданий и их объекта. В основе второго механизма лежит предположение об оптимальности логики формирования ожиданий и их опоре на набор факторов, способных оказывать воздействие на объект ожиданий в условиях объективной структурной модели экономики.

Различная логика появления ожиданий определяет особенности их влияния на экономическую динамику – в условиях рассматриваемой модели экономики. Адаптивные и рациональные ожидания экономических агентов в их исходной форме, а также углубления, расширения и комбинации первоначальных гипотез оказывают значительное прямое и косвенное влияние на экономические процессы. Это подтверждается многочисленными отечественными и зарубежными теоретическими и эмпирическими

научными исследованиями, несмотря на существенную долю критики относительно самих гипотез.

Для гипотезы адаптивных ожиданий одним из основных доказательств существенного влияния ожиданий агентов на экономическую динамику является точность прогноза фактической траектории показателей от их предшествующих значений. При этом в концепции рациональных ожиданий объективная модель экономики в основе их формирования обуславливает теоретическую объективность самих ожиданий и их прямое влияние на динамику объекта ожиданий (различных параметров экономики).

3. Рассмотрены основные способы измерения экономических ожиданий – включая современные подходы на базе больших данных – и систематизированы их ключевые преимущества и недостатки. Двумя основными подходами к построению индикаторов экономических ожиданий в современной практике экономических исследований являются: базирующийся на опросах экономических агентов и основанный на анализе биржевых параметров, учитывающих прогнозы участников финансового рынка относительно различных параметров экономики.

Однако эти подходы имеют ряд существенных недостатков. Для подхода, базирующегося на основе опросов экономических агентов, это низкая частотность индикаторов, узкий охват целевой аудитории, высокие организационные и финансовые издержки, а также неустойчивость ответов респондентов в зависимости от формулировки вопросов. Для подхода, основанного на анализе биржевых параметров, основными недостатками являются низкая релевантность отдельным видам рисков и малая распространенность отдельных биржевых инструментов на локальных финансовых рынках.

При этом активное развитие методов анализа больших данных и алгоритмов машинного обучения способствует росту популярности построения экономических индикаторов на базе этих методов. Основными преимуществами индикаторов экономических ожиданий российского

населения на основе больших данных относительно существующих подходов, являются: (1) широкий охват целевой аудитории; (2) высокая частотность анализа, предполагающая возможность расчета индикаторов в реальном времени; (3) опережение индикаторов, построенных на основе опросов; (4) устойчивость результатов построения индикаторов к правилам вычисления и возможность перерасчета при изменении правил; (5) относительно низкие издержки внедрения автоматизированных расчетов.

4. Осуществлена систематизация исследований об инфляционных ожиданиях экономических агентов. В частности, рассмотрены аспекты влияния инфляционных ожиданий населения на экономическую динамику, а также основные методы их измерения.

Определением инфляционных ожиданий населения являются представления или прогнозы населения относительно будущей общей траектории темпов роста цен и о росте цен на отдельные товары и услуги или их группы.

Для исследования инфляционных ожиданий подтверждаются гипотезы как об адаптивном, так и о рациональном характере их формирования. При этом консенсусом научного сообщества является смешанный характер формирования ожиданий. При этом отмечается, что значительное влияние на формирование ожиданий оказывает доступная экономическим агентам информация (включая медиа-ресурсы), степень достоверности этой информации, а также уровень ее восприятия агентами.

Ключевыми подходами к измерению инфляционных ожиданий населения, как и в общем для экономических ожиданий, являются опросы и данные финансовых рынков.

Для измерения инфляционных ожиданий населения на основе опросов, как правило, комбинируют два способа квантификации (прямые вопросы о численных ожиданиях и расчет ожиданий на основе распределения ответов о будущем изменении темпов роста цен). Это необходимо, поскольку для обоих методов характерны определенные недостатки (завышение населением

прямых оценок инфляции, существенное влияние фактической инфляции на результаты расчетов на основе распределений). Следует также отметить, что при мониторинге инфляционных ожиданий населения большее внимание, как правило, уделяется именно интерпретации их изменений, а не уровня.

Несмотря на преимущества метода оценки на основе биржевых данных, связанных в основном с возможностью высокочастотных измерений, применение данного метода как для оценки инфляционных ожиданий, так и экономических ожиданий в целом, ограничено значительными недостатками (репрезентативность только для участников финансовых рынков, необходимость наличия крупных объемов торгов по специфическим финансовым инструментам и исключения эффектов премий за другие виды рисков).

В области применения современных методов анализа данных для оценки инфляционных ожиданий, исследования подтверждают значительное влияние новостного фона на инфляционные ожидания домохозяйств, а также возможность их измерения с помощью анализа активности пользователей в сети Интернет.

## **Глава 2. Методика построения высокочастотных индикаторов экономических ожиданий населения на основе больших данных**

В данной работе предлагается авторская методика построения высокочастотных индикаторов интенсивности и структуры экономических ожиданий российского населения – с помощью методов анализа больших данных. Основой для проведения исследования являются комментарии пользователей сети Интернет к новостям в экономических СМИ.

Предлагаемый подход позволяет осуществить построение различных индикаторов экономических ожиданий населения, имеющих значение как для денежно-кредитной, так и для макроэкономической политики. Кроме того, в разработанной методике предусмотрена комплексная процедура выявления ключевых факторов изменения индикаторов или событий и трендов, с которыми население связывает изменение собственных ожиданий.

### **2.1. Обоснование и общее описание методики построения индикаторов**

Ключевой гипотезой, на которой базируется предлагаемая методика построения высокочастотных индикаторов экономических ожиданий, является то, что интенсивность обсуждения читателями экономических СМИ в сети Интернет различных экономических событий и явлений в контексте релевантных экономических новостных статей, является отражением интенсивности ожиданий или обеспокоенности населения относительно реализации этих событий в ближайшем будущем. Обсуждениями статей являются комментарии интернет-пользователей к ним в сети Интернет.

Целесообразность этой гипотезы для измерения экономических, и в частности инфляционных, ожиданий населения подтверждается зарубежными теоретическими и эмпирическими исследованиями. Так, в работе Кэролла [Carroll, 2003] была построена теоретическая модель инфляционных ожиданий населения, главным фактором формирования и источником информации для которых являлись новости экономических

СМИ. В исследовании Бауэра [Bauer, 2014] эта концепция была статистически подтверждена. Результаты многочисленных опросов населения также свидетельствуют в пользу того, что новости экономических СМИ критически значимы для формирования экономических ожиданий населения [Dräger, Lamla, and Pfajfar, 2016].

Вторая важная гипотеза в основе предлагаемой методики состоит в том, что население, интересующееся экономическими событиями и высказывающее собственное мнение, оказывает значительное влияние на формирование инфляционных ожиданий российского населения в целом. Другими словами, ожидания именно активной части населения, интересующейся происходящим в экономике, определяют динамику экономических ожиданий населения в целом. В частности, это означает и влияние на ожидания менее активной части населения, не интересующейся экономическими новостями. В работах Гузман [Guzman, 2011], Чои и Вэриана [Choi, Varian, 2009], Шмидта и Возена [Schmidt, Vosen, 2010] показана возможность построения релевантных индикаторов экономических, и в частности инфляционных, ожиданий населения с помощью анализа активности пользователей в сети Интернет.

Несмотря на то, что доступ домохозяйств в интернет в целом по России на текущий момент является высоким (около 70%, по данным Росстата), выборка интернет-пользователей, комментирующих экономические новости, не является полностью сбалансированной в целом по стране. Отдельные оценки для рассматриваемой в данной работе выборки источников СМИ и комментариев показывают, что среди интернет-пользователей, комментирующих экономические новости, преобладает население 25-35 лет, а также население столичных регионов [Голощапова, 2017]. Следует отметить, что индикаторы экономических ожиданий на основе больших данных не ставят перед собой цели достижения сбалансированной выборки читателей экономических СМИ, а направлены на выявление совокупного мнения наиболее активной их части в сети Интернет.



Период, выбранный для построения индикаторов: с начала января 2014 г. до конца февраля 2018 г. Выбор начальной даты для сбора данных осуществлен на основе двух факторов. Первый фактор – это динамика доли домохозяйств, имеющих доступ к сети Интернет, в целом по России. Именно с 2014 г. эта доля превысила 60%, по данным Росстата (см. рис. 3), и в дальнейшем замедлила темпы прироста, вероятно, достигнув некоторого устойчивого для страны уровня. Второй фактор заключается в отсутствии открытого доступа к существенной доле материалов СМИ, опубликованных ранее 2014 г. Последнее, вероятно, связано с внутренними особенностями политики публикации материалов медиа-источников. Конечная дата периода для сбора данных не является строго фиксированной. Предлагаемая методика построения индикаторов экономических ожиданий предполагает регулярное обновление данных для анализа.

Таким образом, с учетом выдвинутых выше гипотез и указанных потенциальных недостатков методов анализа больших данных для построения экономических индикаторов (см. раздел 1.2.3, стр. 39-41) следует выделить ряд важных задач разрабатываемой методики.

Во-первых, это формирование репрезентативной выборки российских экономических СМИ как ключевого источника информации об экономических новостях для российского населения и выборки наиболее популярных ресурсов, где читатели этих СМИ обсуждают новостные статьи. Это будет способствовать максимизации охвата целевой аудитории индикаторов – населения, читающего и обсуждающего экономические новости. Кроме того, необходимость контроля репрезентативности данных в разрабатываемой методике подчеркивается по результатам анализа потенциальных недостатков подходов к построению экономических индикаторов с помощью методов анализа больших данных (см. раздел 1.2.3).

Во-вторых, это учет влияния медиа-ресурсов на мнение интернет-пользователей, или влияния интенсивности обсуждений какого-либо экономического явления в новостях СМИ на интенсивность его обсуждений

со стороны населения. Данный шаг необходим с целью корректировки экономических ожиданий интернет-пользователей на медиа-ожидания или на возможную искусственную популяризацию темы со стороны СМИ. Такая корректировка будет способствовать выявлению корректной или «истинной» обеспокоенности населения конкретным вопросом или проблемой экономики.

В-третьих, это необходимость разработки принципов формирования правил отбора новостных статей СМИ и комментариев интернет-пользователей к ним – в соответствии с конкретными видами экономических ожиданий населения, задаваемых исследователем. Отбор статей должен быть направлен на выявление тех из них, в которых обсуждается объект ожиданий – для того, чтобы исключить учет нерелевантных для исследуемых экономических ожиданий новостных статей и повысить качество рассчитываемого индикатора. Отбор комментариев должен быть направлен на выявление только тех из них, которые свидетельствуют об обеспокоенности интернет-пользователями конкретным экономическим явлением, ожидания которого измеряет исследователь. Качество отбора как новостных статей, так и комментариев читателей к ним, для каждого индикатора, рассчитанного на базе сформированных выборок, будет определять соответствие заданному виду экономических ожиданий населения.

В-четвертых, наличие гипотез в разрабатываемой методике предопределяет необходимость включения в нее этапа оценки релевантности полученных индикаторов экономических ожиданий. Целесообразность наличия такого этапа была выявлена и по результатам анализа потенциальных недостатков подходов к построению экономических индикаторов с помощью методов анализа больших данных (см. раздел 1.2.3). Динамика рассчитанного индикатора интенсивности экономических ожиданий должна быть соотносима с динамикой целевого экономического параметра (с учетом экономической теории), а также с динамикой наиболее релевантных индикаторов-аналогов, измеренных другими способами.

В-пятых, поскольку одним из потенциальных недостатков использования больших данных для построения экономических индикаторов также является низкая устойчивость результата к выбору алгоритма расчета индикаторов, в методике должна быть предусмотрена соответствующая проверка. С одной стороны, результаты расчета индикаторов ожиданий должны быть устойчивы к небольшим изменениям алгоритма расчета, с другой – обусловлены именно задачами исследователя (или запросами исследователя к данным), а не особенностями самого набора данных.

Кроме того, важным этапом построения индикатора – после обоснования его релевантности искомому виду экономических ожиданий – является выделение ключевых факторов, объясняющих его динамику. Инструментарий работы с большими данными, которые представлены текстами новостей СМИ и сообщений интернет-пользователей, позволяет осуществить комплексный анализ возможных драйверов изменения интенсивности ожиданий населения. Это способствует как дополнительной верификации релевантности полученного индикатора (при проверке теоретических предпосылок о взаимосвязи конкретного вида экономических ожиданий с макроэкономическими трендами), так и углубленному пониманию поведения населения в отдельные эпизоды изменения экономических ожиданий.

С учетом приведенных выше гипотез и сформулированных задач, предлагаемая методика построения индикаторов экономических ожиданий населения содержит пять этапов (см. рис. 1):

- 1) Определение источников информации и оценка репрезентативности выборки для анализа. Данный этап предполагает два шага: (1) формирование списка источников экономических СМИ и (2) ресурсов для сбора информации о мнении интернет-пользователей в отношении публикуемых новостей.
- 2) Отбор новостных статей выделенной выборки экономических СМИ на заданную тему или область экономических ожиданий.

- 3) Отбор комментариев интернет-пользователей к выделенному набору новостных статей экономических СМИ на тему конкретной области экономических ожиданий. На втором и третьем этапах методики предполагается единый алгоритм составления правил отбора, основанный на комбинировании регулярных выражений.
- 4) Расчет индикатора интенсивности ожиданий российского населения для заданной области экономики. Данный этап включает три шага:
  1. Вычисление временного ряда интенсивности заданного вида экономических ожиданий населения на базе сформированных правил отбора новостных статей СМИ и комментариев интернет-пользователей;
  2. Оценку релевантности полученного индикатора в две стадии: с точки зрения теоретических предпосылок о его динамике и соответствии ключевым макроэкономическим трендам и показателям-аналогам, измеренным иным способом (при наличии);
  3. Проверку устойчивости результатов расчета индикатора. Этот шаг опирается на две стадии: оценку чувствительности результатов построения индикатора в зависимости от сформированных на втором и третьем этапах методики правил отбора новостных статей СМИ и комментариев интернет-пользователей, а также в зависимости от изменения области экономических ожиданий.
- 5) Выявление факторов изменения интенсивности экономических ожиданий населения во времени. Данная часть представляет собой комплексную процедуру определения ключевых драйверов изменения ожиданий населения, основанную на трех шагах, а именно: анализе содержания комментариев интернет-пользователей с помощью различных подходов, включая (1) статистический анализ зависимостей, (2) экспертную оценку и (3) методы машинного обучения и текстовой обработки. Процесс выявления факторов изменения ожиданий также предусматривает постоянную верификацию полученных результатов.

## Методика построения высокочастотных индикаторов экономических ожиданий российского населения на основе больших данных

### Этап 1. Определение источников информации и оценка репрезентативности выборки для анализа

- **Шаг 1.** Составление выборки экономических СМИ, включая оценку репрезентативности.
- **Шаг 2.** Составление выборки источников информации о мнении читателей относительно материалов СМИ, включая оценку репрезентативности.

### Этап 2. Отбор новостных статей выделенной выборки экономических СМИ на заданную тему или область экономических ожиданий

### Этап 3. Отбор комментариев интернет-пользователей к выделенному набору новостных статей экономических СМИ для заданного вида экономических ожиданий

*Единый алгоритм составления правил отбора, базирующийся на комбинировании регулярных выражений*

### Этап 4. Расчет индикатора интенсивности ожиданий российского населения для заданной области экономики

- **Шаг 1.** Вычисление временного ряда интенсивности заданного вида экономических ожиданий населения на базе сформированных правил отбора новостных статей СМИ и комментариев интернет-пользователей;
- **Шаг 2.** Оценка релевантности полученного индикатора целям исследователя:
  - **Стадия 1.** Оценка соответствия динамики индикатора ключевым макроэкономическим трендам с учетом теоретических предпосылок.
  - **Стадия 2.** Оценка релевантности динамики индикатора показателям-аналогам, измеренным иным способом;
- **Шаг 3.** Проверка устойчивости результатов расчета индикатора:
  - **Стадия 1.** Оценка чувствительности результатов расчета индикатора в зависимости от сформированных на втором и третьем этапах методики правил отбора новостных статей СМИ и комментариев интернет-пользователей;
  - **Стадия 2.** Оценка чувствительности результатов расчета индикатора в зависимости от изменения области экономических ожиданий.

### Этап 5. Выявление факторов изменения интенсивности экономических ожиданий населения во времени

- **Шаг 1.** Оценка зависимости построенного индикатора интенсивности экономических ожиданий от релевантных экономических параметров;
- **Шаг 2.** Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе экспертного анализа:
  - **Стадия 1.** Выделение обобщающих категорий для экономических событий, явлений или трендов, которые население упоминает в одних и тех же сообщениях с упоминанием целевого для исследователя вида экономических ожиданий;
  - **Стадия 2.** Расчет показателей вкладов упоминаний каждой выделенной на первой стадии категории в общем числе упоминаний всех результирующих категорий;
  - **Стадия 3.** Сравнение динамики полученных на второй стадии показателей вклада различных факторов в динамику индикатора интенсивности ожиданий с динамикой индикатора интенсивности экономических ожиданий.
- **Шаг 3.** Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе методов машинного обучения и текстовой обработки.
  - **Стадия 1.** Предобработка текста целевых комментариев;
  - **Стадия 2.** Выбор оптимального числа тем и других параметров моделирования;
  - **Стадия 3.** Тематическое моделирование целевых комментариев.

*Источники:* составлено автором.

Рисунок 1. Общая схема методики построения высокочастотных индикаторов экономических ожиданий российского населения на основе больших данных.

Далее каждый из предлагаемых этапов методики будет рассмотрен в деталях, включая контроль на корректность получаемых результатов. Процедура контроля учитывает перечисленные в предшествующей главе потенциальные недостатки метода построения экономических индикаторов с использованием подходов к анализу больших данных.

## **2.2. Определение источников информации и оценка репрезентативности выборки для анализа**

Первичным источником информации для предлагаемой методики построения высокочастотных индикаторов экономических ожиданий российского населения служат новостные статьи в крупнейших экономических СМИ страны. При этом основной базой для проведения исследования являются комментарии пользователей сети Интернет к новостным статьям по заданной области экономических ожиданий в определенном на первом шаге наборе экономических СМИ.

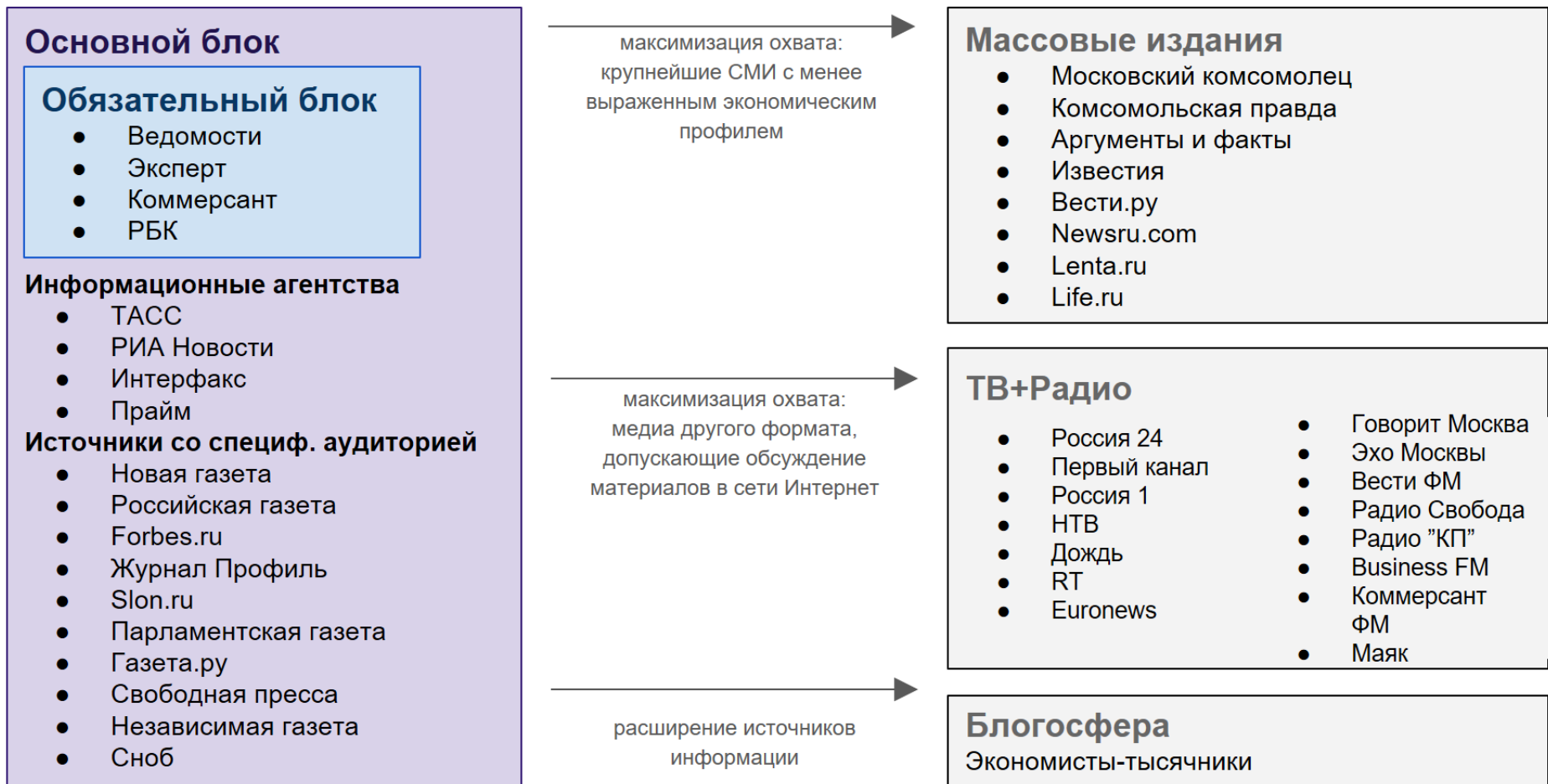
### *2.2.1. Составление выборки экономических СМИ, репрезентативной для построения индикаторов*

Выбор источников СМИ возможно осуществить с помощью анализа рейтингов медиа-источников. В качестве основы для выделения крупнейших и наиболее значимых экономических СМИ России были определены открытые рейтинги компании «Медиалогия». Эта компания представляет в открытый доступ ежемесячные отчеты о крупнейших и наиболее влиятельных по цитированию российских медиа-источниках, полученные на базе автоматической системы мониторинга СМИ в реальном времени. Охват медиа-источников для мониторинга крайне широк и составляет более 43 тыс. СМИ, цитируемость которых в том числе оценивается с помощью более чем

800 млн. социальных медиа. Разделы для составления рейтингов включают списки самых цитируемых газет, журналов и информационных агентств с учетом их представленности в интернете (анализ посещаемости официальных веб-страниц и официальных групп СМИ в социальных медиа). Кроме того, открытые рейтинги включают регулярное определение лидирующих ТВ-каналов и радиостанций, а также наиболее популярных блогов.

На базе анализа ежемесячных отчетов «Медиалогии» за период 2016-2017 гг. о рейтингах СМИ по всем вышеупомянутым разделам и с учетом профилей изданий, а именно релевантности их публикаций по экономической проблематике, была составлена авторская карта крупнейших российских экономических СМИ, релевантных для анализа экономических ожиданий российского населения (см. рис. 2).

Релевантность полученного набора СМИ была верифицирована и подтверждена с помощью опроса экспертов в области экономического анализа (ЦМАКП, МГУ, октябрь 2016 г.) и представления научному сообществу [Андреев, Голощапова, 2016]. Следует отметить, что большинство экспертов назвали в качестве ключевых крупных российских СМИ для проведения анализа экономических ожиданий российского населения следующие четыре источника: Ведомости, РБК, Эксперт, Коммерсант. Учитывая высокую цитируемость этих источников, согласно результатам анализа медиа-рейтингов «Медиалогии», они были определены как выборка крупнейших российских экономических СМИ, обладающая минимальной репрезентативностью для оценки экономических ожиданий населения [Голощапова, 2017]. При этом для расширения охвата релевантной аудитории – интересующейся внутренней экономической ситуацией, читающей и комментирующей экономические новости в сети Интернет – респонденты перечисляли дополнительные важные источники СМИ, которые были также учтены при составлении общей карты источников.



Источник: «Медиалогия», опросы экспертов ЦМАКП и МГУ в области экономики, расчеты автора.

Рисунок 2. Карта крупнейших российских экономических СМИ, релевантных для оценки экономических ожиданий российского населения.



Результирующая карта состоит из четырех блоков. Основной блок включает все крупные российские СМИ, публикующие значительное количество материалов, относящихся к проблематике российской экономики в целом, и одновременно имеющие высокую цитируемость. Последнее означает, что данные источники популярны среди аудитории, читающей и обсуждающей экономические новости в сети Интернет, являющейся целевой аудиторией для предлагаемой методики построения индикаторов экономических ожиданий.

Внутри основного блока выделено три категории. Первая категория, а именно «Обязательный блок», представляет собой список СМИ, обладающий минимальной репрезентативностью. Этот блок был сформирован на основе опросов экспертов и дискуссий в научном сообществе (см. стр. 63). Именно этот набор источников используется в настоящей работе для реализации и верификации результатов всех ключевых этапов разработанной методики.

Вторая категория основного блока содержит список крупнейших информационных агентств, обладающих высокой цитируемостью среди российских пользователей сети Интернет. Существенная доля Интернет-пользователей получает информацию об экономических событиях именно из этих источников. В рейтингах информационных агентств «Медиалогии» эти СМИ устойчиво занимают лидирующие позиции. Добавление данной категории в анализ способно существенно повысить его репрезентативность и качество.

Третья категория основного блока включает крупные и лидирующие по цитируемости среди российских пользователей источники, имеющие существенную долю публикаций экономических материалов среди всех материалов, но вместе с тем характеризующиеся достаточно специфической читательской аудиторией. Это может быть обусловлено как определенными политическими взглядами (Новая газета, Независимая газета и др.), так и уровнем дохода, и ориентированностью на отдельные категории бизнеса (Forbes.ru, Сноб и др.), или отношением к государственной службе и

юридической области деятельности (Российская газета) и др. Включение данной категории СМИ в базу для оценки экономических ожиданий населения также способно оказать существенное позитивное воздействие на качество и репрезентативность результатов.

Помимо основного блока карта содержит три дополнительных блока медиа-источников, способных существенно расширить базу данных для анализа. Перечисленные в дополнительных блоках источники содержат менее обширную относительно источников основного блока, но тем не менее значительную долю экономических материалов среди всех публикаций, а также занимают лидирующие места по цитируемости в целом среди российских СМИ.

Первые два дополнительных блока направлены на максимизацию охвата российской аудитории, получающей информацию о различных событиях, включая экономические, из сети Интернет. Блок «Массовые издания» направлен на охват крупнейших российских СМИ, включая печатные издания с менее выраженным экономическим профилем относительно источников основного блока. Второй блок – «ТВ+Радио» – направлен на включение в анализ форматов медиа-источников, отличных от СМИ, основной подачи информации для которых являются текстовые сообщения (газеты, журналы, информационные агентства), но тем не менее имеющие официальные веб-страницы и допускающие обсуждение собственных материалов в сети Интернет. Этот блок содержит ключевые телевизионные каналы и радиостанции.

Третий дополнительный блок «Блогосфера» направлен на расширение источников информации для анализа и включает относительно новый для интернета формат подачи информации – блоги отдельных авторов, как правило, экспертов в области экономики. При этом выбор авторов может быть основан как на экономической направленности подаваемых материалов и области профессиональной деятельности автора, так и на величине его

аудитории. Здесь порогом отсечения может быть одна тысяча постоянных читателей.

### *2.2.2. Составление выборки источников информации о мнении читателей относительно материалов СМИ*

В рамках разработанной методики базой для оценки экономических ожиданий российского населения является мнение активных интернет-пользователей относительно экономических событий, данных и трендов, освещаемых крупнейшими экономическими СМИ в статьях на тему рассматриваемых областей экономики.

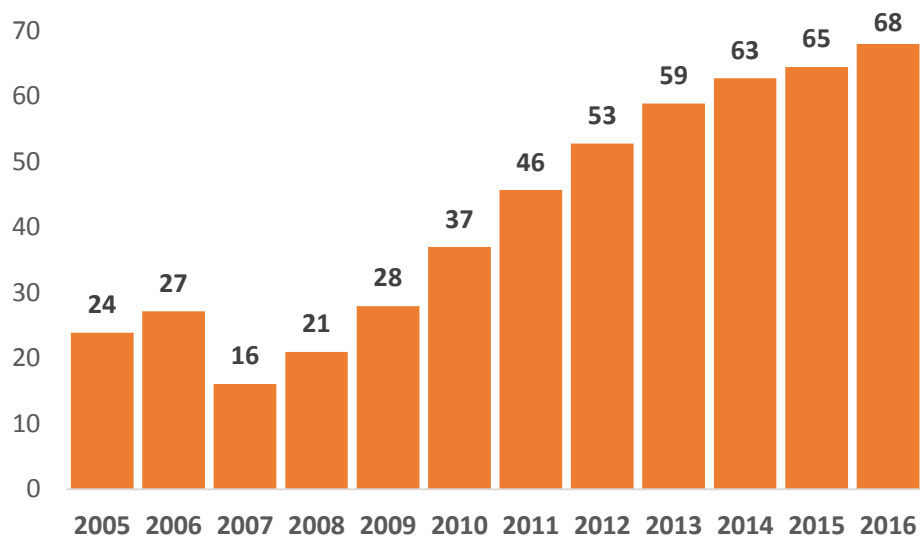
Предполагается, что ожидания активных интернет-пользователей, интересующихся экономическими событиями, способны определять динамику экономических ожиданий населения в целом. Так, по данным Росстата, доля домохозяйств, имеющих доступ к сети Интернет, в целом по России в 2016 г. составляла 68% (см. рис. 3). Тенденция устойчивого и стремительного роста показателя наблюдается в последние 10 лет. Так, еще в 2007 г. его значение составляло 16%. При этом, судя по последним имеющимся данным, тренд сохраняется и в настоящее время. По словам президента ПАО «Ростелеком» М. Осеевского от октября 2017 г.<sup>6</sup>, в настоящий момент доля домохозяйств, имеющих доступ к сети Интернет превысила 70%. В дальнейшем компания планирует увеличить этот показатель до 90%. Таким образом, агрегированные индикаторы экономических ожиданий, полученные на основе сбора и обработки мнений российских интернет-пользователей, могут являться репрезентативными для отражения совокупных экономических ожиданий российского населения.

Методика предполагает извлечение мнения интернет-пользователей по отдельным областям экономики из комментариев пользователей сети Интернет к новостным статьям по заданной области экономических ожиданий в определенном на первом шаге наборе экономических СМИ.

---

<sup>6</sup> Осеевский М. Цифровой мир – это реальность, в которой мы уже живем // Девятый Ежегодный Инвестиционный Форум ВТБ Капитал «Россия зовет!». – Октябрь 2017 г.

Поскольку комментарии пользователей представляют собой текстовые сообщения, получение агрегированных индикаторов экономических ожиданий производится с помощью методов текстовой обработки.



*Источник:* Росстат.

Рисунок 3. Доля российских домохозяйств, имеющих доступ к сети Интернет.

Для сбора комментариев интернет-пользователей к статьям в экономических СМИ используются три вида источников: официальные веб-страницы СМИ в сети Интернет, официальные группы СМИ в социальных сетях ВКонтакте и Facebook.

Первая группа источников – официальные веб-страницы СМИ – направлена на охват пользователей, ориентированных преимущественно на прочтение конкретных источников, и, предположительно, не пользующихся социальными сетями. Социальные сети выбраны на основе анализа количества комментариев пользователей среди интернет-пользователей. По мере распространения доступа к сети Интернет среди домохозяйств становятся все более популярными подписки на страницы СМИ в социальных сетях и обсуждение там публикуемых материалов. По результатам анализа на основе собранных по минимальной репрезентативной выборке СМИ данных (Коммерсант, РБК, Эксперт, Ведомости), соотношение

числа комментариев в социальных сетях к комментариям на официальных веб-страницах СМИ составляет двенадцать к одному.

Следует отметить, что в качестве существенного дополнения к базе социальных сетей может выступить другая распространенная в России сеть – «Одноклассники». Включение этого источника в анализ может в определенной мере повысить охват интернет-пользователей для оценки индикаторов экономических ожиданий.

Таким образом, в данном разделе были сформированы списки источников экономических СМИ и ресурсов для сбора информации о мнении интернет-пользователей в отношении публикуемых новостей, репрезентативные для оценки экономических ожиданий российского населения в рамках авторской методики.

### **2.3. Отбор новостных статей экономических СМИ и комментариев интернет-пользователей для измерения различных видов экономических ожиданий**

Вторым этапом построения индикаторов экономических ожиданий населения – после формирования выборки экономических СМИ и ресурсов для сбора информации о мнении интернет-пользователей – является отбор новостных статей СМИ. Цель отбора – получение выборки статей, наиболее релевантных какой-либо конкретной области экономики, измерение экономических ожиданий в которой является целью исследователя.

Затем, на следующем этапе, среди всех комментариев интернет-пользователей к отобранным статьям СМИ осуществляется отбор комментариев, имеющих непосредственное отношение к измеряемым экономическим ожиданиям.

Следует отметить, что предлагаемый для указанных двух этапов методики алгоритм отбора статей СМИ и комментариев интернет-пользователей является аналогичным. Все новостные статьи СМИ в сети Интернет представляют собой текстовые сообщения, поэтому к ним могут

быть применены методы текстовой обработки. Для этих этапов предполагается использование правил поиска информации в тексте, составленных экспертно на основе знаний экономической теории и теории ожиданий. Правила поиска формируются на базе комбинаций регулярных выражений. Последние представляют собой последовательности символов в рамках теории компьютерных наук и теории формальных языков, определяющие модель поиска, заданную исследователем [Friedl, 2006]. В упрощенной форме это означает, что статьи и комментарии отбираются по определенным экспертно наборам ключевых слов и словосочетаний, релевантным заданной области экономики и измеряемым экономическим ожиданиям.

Логика составления регулярных выражений для отбора новостных статей СМИ на тему конкретной области экономики, измерение экономических ожиданий в которой является целью исследователя, базируется на ряде принципов.

В первую очередь, в выборку статей должны быть включены все статьи, в которых происходит обсуждение объекта ожиданий, а именно упоминаются ключевые слова, определяющие этот объект в экономической теории. К примеру, в случае если цель исследования состоит в измерении ожиданий ухудшения ситуации на рынке труда, первичным словосочетанием для отбора будет «рынок труда». Для измерения ожиданий относительно ослабления курса национальной валюты – «курс валюты», «валютный курс», для инфляционных ожиданий – «инфляция», «темп роста цен».

Во вторую очередь, необходимо включить в правила отбора альтернативные определения исследуемого объекта экономических ожиданий. В некоторых случаях эти правила могут быть основаны на более сложных словосочетаниях. Так, правила определения статей на тему «рынок труда» могут быть дополнены словом «работа», на тему курса национальной валюты – словами «рубль», «доллар», «евро», на тему инфляции – «ИПЦ», «индекс потребительских цен», «темп снижения цен», «темп падения цен», а

также любыми словосочетаниями, связанными с динамикой цен (снижение, повышение, рост). В случае выявления статей, относящихся к динамике курса валюты, ключевые слова должны также содержать названия основных мировых валют, помимо рубля. Это является важным, поскольку при обсуждении новостных статей СМИ на тему динамики данных мировых валют, российское население часто дискутирует и на тему ожидаемой динамики курса рубля. Это может быть выявлено экспертно на основе итеративной корректировки правил (см. далее).

Предлагаемый алгоритм отбора углубляет и дополняет методы отбора новостных статей на определенные темы, представленные в экономической литературе и базирующиеся на регулярных выражениях [Голощапова, Андреев, 2017]. К примеру, выбор новостных статей СМИ на тему «инфляция» в целях исследования факторов изменения инфляционных ожиданий населения США в [Carroll, 2003; Dräger, Lamla, Pfajfar, 2016] был основан на поиске только слова «инфляция» (inflation) в заголовках первых страниц газет New York Times и Washington Post. Такая схема отбора оставляла возможность пропуска существенного числа статей СМИ, релевантных обсуждению динамики цен. Для предлагаемого в данной работе алгоритма характерны следующие преимущества: поиск осуществляется на основе полного текста каждой новости, а набор ключевых слов для поиска является значительно более широким.

Целью третьего этапа методики является отбор комментариев, имеющих непосредственное отношение к измеряемым экономическим ожиданиям – среди всех комментариев интернет-пользователей к отобранным на втором этапе методики статьям СМИ. Таким образом, алгоритм направлен на поиск упоминаний конкретных ожиданий населения – повышения или снижения определенного макроэкономического тренда или ухудшения, или улучшения ситуации в какой-либо конкретной области экономики. При этом для повышения качества такого поиска, он

производится только среди комментариев к статьям, имеющим отношение к теме или объекту ожиданий.

Логика отбора комментариев является сходной с логикой отбора новостных статей. Основное отличие состоит в том, что ключевые слова и словосочетания для формирования правил поиска должны иметь непосредственное отношение к исследуемым ожиданиям. Здесь, как показывает практика реализации методики, используемые модели поиска могут быть более сложными относительно моделей первого шага.

Для рассмотренных выше примеров областей экономических ожиданий правила отбора могут быть основаны на следующих ключевых словах и словосочетаниях:

- Для ожиданий ухудшения ситуации на рынке труда: слово «работа» в различных формах, встречающееся в одном комментарии совместно с различными формами слов «сложный», «плохой», «нет», «не будет», «ухудшение»;
- Для ожиданий ослабления курса национальной валюты: слова «рубль» и «курс» в различных формах, упоминаемые в одном комментарии совместно с различными формами слов «слабый», «девальвация», «падение», «снижение», «низкий»;
- Для инфляционных ожиданий: слова «инфляция» и «цены», встречающиеся в одном комментарии совместно с глаголами, относящимися к росту – «повышение», «рост».

Важно, что с помощью регулярных выражений поиск в тексте всех упомянутых слов производится для различных форм этих сочетаний (падежи, множественные и единственные числа, склонения). То есть, фактически, поиск осуществляется для основ слов в словосочетаниях. Регулярные выражения для приведенных в данном разделе примеров представлены в таблице 1.

Таблица 1. Примеры записи правил на основе регулярных выражений для отбора новостных статей СМИ и комментариев с целью построения



индикаторов отдельных видов экономических ожиданий российского населения

Область экономических ожиданий	Система правил для отбора новостных статей СМИ	Система правил для отбора комментариев к выбранным новостным статьям СМИ
Ожидания ухудшения ситуации на рынке труда	<p><i>Объект ожиданий:</i> рынок труда.</p> <p><i>Правила:</i></p> <p>(1) (работ[a e y ы ой] )</p> <p>(2) (рын[ок ка ку ки ков кам ком]) &amp; [«труда»]</p>	<p><i>Правила:</i></p> <p>[«работа », «работе », «работу », «работы », «работой »] &amp; [«нет», «не будет», «сложн», «плох», «худ», «хуж»]</p>
Ожидания ослабления курса национальной валюты	<p><i>Объект ожиданий:</i> курс национальной валюты.</p> <p><i>Правила:</i></p> <p>(1) (курс долл евро рубл)</p> <p>(2) [«курс»] &amp; [«валют»]</p>	<p><i>Правила:</i></p> <p>[«рубл», «курс»] &amp; [«слаб», «девальв», «пад», «ниж», «низ»]</p>
Инфляционные ожидания	<p><i>Объект ожиданий:</i> инфляция.</p> <p><i>Правила:</i></p> <p>(1) (инфляц ипц (темп.*цен)</p> <p>(2) ((рост повыс повыш ниж низ).*цен)</p> <p>(3) (индекс.*потребительских.*цен )</p>	<p><i>Правила:</i></p> <p>[«инфляц», «цен»] &amp; [«выс», «выш», «рост», «раст»]</p>

*Примечания:*

- (1) Все представленные регулярные выражения могут быть записаны в другой форме, в соответствии с правилами их составления;
- (2) Символ «&» в представленных регулярных выражениях обозначает поиск упоминаний в одном комментарии блоков слов, обозначенных в квадратных скобках. Такая модель поиска может быть записана при помощи различного синтаксиса в разных языках программирования, в которых используются регулярные выражения.
- (3) Перед поиском осуществляется базовая фильтрация текста (перевод символов в нижний регистр, удаление небуквенных символов).

*Источники:* составлено автором.

Следует подчеркнуть, что важной частью предлагаемого алгоритма отбора новостных статей СМИ и комментариев пользователей сети Интернет является итеративность процесса формирования правил и проведение тестов на устойчивость результата к изменению правил. Представленные в таблице 1 правила отбора были оптимизированы автором с учетом всех указанных критериев для базы данных СМИ за период 2014-2017 гг., обладающей минимальной репрезентативностью (см. раздел 2.2).

Итеративность процесса формирования правил предполагает последовательное улучшение комбинаций регулярных выражений на основе пошаговой коррекции правил. Коррекция осуществляется экспертно по результатам анализа качества полученных выборок новостных статей СМИ и комментариев после каждой итерации поиска. Процедура оценки основана на анализе случайных выборок новостных статей и комментариев, как из целевых наборов, так и из не включенных в них. Сама оценка качества и, соответственно, коррекция правил, базируются на минимизации доли нерелевантных полученным наборам примеров, а именно, снижении в итоговых наборах доли новостных статей СМИ, не относящихся к целевой для исследователя области экономики, и комментариев, не относящихся к целевым экономическим ожиданиям.

Представленные выше правила отбора новостных статей СМИ и комментариев не являются полностью оптимальными и могут быть скорректированы исследователем в зависимости как от характеристик базы исследуемых данных (СМИ-источников, источников получения информации о мнении интернет-пользователей), так и специфики целевых для исследователя экономических ожиданий населения.

С точки зрения оптимизации расчетов, следует отметить, что поиск более сложных, чем обыкновенные словосочетания из двух-трех слов, взаимосвязей слов значительно увеличивает расчетное время компьютера на поиск соответствующих моделей в тексте, однако при этом может повысить качество отбора. При увеличении объемов данных и частотности расчетов может возникать потребность в достижении компромисса между сложностью системы правил для поиска и производительностью расчетов. При необходимости проведения расчетов с очень высокой частотностью и на критически больших объемах данных оптимальным решением может стать отказ от поиска части словосочетаний в пользу повышения производительности. Примером здесь может служить необходимость расчета индикаторов экономических ожиданий в реальном времени на основе всех

перечисленных медиа-источников, указанных в соответствующей карте раздела 2.2.

Важно также отметить, что для составления правил отбора новостных статей СМИ и комментариев интернет-пользователей можно использовать современные алгоритмы текстового анализа, а именно тематическое моделирование, или искусственные нейронные сети.

Тематическое моделирование позволяет выделить ключевые сочетания слов в выделенных текстах (сообщений пользователей или новостей), обозначающие схожие темы – на основе анализа частотности встречаемых слов и ряда предположений, оптимизируемых в процессе работы алгоритма [Ramage et al., 2009]. Применение подобных алгоритмов могло бы способствовать снижению доли субъективной экспертной оценки при составлении правил отбора – при условии их формирования на основе математических расчетов (техник тематического моделирования).

Вместе с тем идея применения алгоритмов тематического моделирования к задаче составления правил не лишена недостатков. Этот способ текстового анализа характеризуется слабой устойчивостью к меняющимся наборам данных [Seiter, 2012]. Таким образом, автоматическое выделение конкретной темы для новостных статей СМИ и конкретной области экономических ожиданий среди комментариев интернет-пользователей на базе алгоритмов тематического моделирования является затруднительным. Результаты выделения тем будут различаться с поступлением новых данных. Фактически, выявление релевантных наборов новостных статей СМИ и комментариев не будет являться автоматической процедурой, поскольку регулярно при расширении выборки данных будет требовать новой и трудозатратной настройки параметров алгоритма для получения качественных результатов. При этом оценка качества получаемых наборов с точки зрения их релевантности поставленной задаче по-прежнему должна будет проводиться исследователем экспертно.

При этом тематическое моделирование в некоторых случаях может быть использовано для усовершенствования составленных экспертно правил. Среди тем, полученных на базе алгоритмов текстовой обработки, возможно выявление слов и словосочетаний, включение которых в процедуру отбора способно повысить качество правил. Это применимо для неоднородных наборов данных, где снижается качество экспертной оценки по случайным выборкам. Предположение о применении алгоритмов тематического моделирования для расширения наборов слов и словосочетаний в составленных экспертно правилах является одним из возможных направлений усовершенствования исследования.

Что касается применения искусственных нейронных сетей для задачи классификации новостных статей СМИ, то ключевым ограничением здесь является отсутствие большого массива новостных статей российских СМИ, размеченного по конкретным областям экономической науки, необходимым для анализа экономических ожиданий населения, таким как «инфляция», «рынок труда», «курс валюты», и т.п. Отдельные наборы данных новостей СМИ, имеющиеся в открытом доступе, оперируют более общими неспецифическими категориями, такими как «Россия», «Мир», «Экономика», «Политика», и т.п. [Юткин, корпус новостей Lenta.ru<sup>7</sup>].

Однако данное направление также можно рассматривать как одно из возможных направлений усовершенствования исследования – в случае решения проблемы наличия достаточного массива размеченных новостных данных для обучения алгоритмов. Проблема может быть решена с помощью привлечения внешних ресурсов для экспертной разметки и привлечения финансирования для оплаты труда экспертов, или при появлении в открытом доступе выборки с необходимыми для целей данной работы категориями новостных статей СМИ.

---

<sup>7</sup> Юткин Д., корпус новостей Lenta.ru с тегами за период с сентября 1999 г. по апрель 2017 г. <https://github.com/yutkin/lenta.ru-news-dataset>

## **2.4. Расчет индикатора интенсивности ожиданий населения для заданной области экономики**

Четвертым этапом методики является расчет индикаторов интенсивности ожиданий населения для различных областей экономики, заданных исследователем. Корректное выполнение этого этапа позволяет получить высокочастотную аппроксимацию ожиданий интернет-пользователей, представляющих значительную долю российского населения, относительно различных экономических явлений.

Корректная реализация данного этапа методики предполагает три шага: (1) вычисление самого индикатора интенсивности ожиданий; (2) оценку релевантности полученного индикатора заданному виду ожиданий; (3) проведение тестов на устойчивость результата к правилам расчета индикатора.

### *2.4.1. Расчет индикатора интенсивности экономических ожиданий и оценка его релевантности заданному виду ожиданий*

Индикатором, отвечающим за аппроксимацию экономических ожиданий населения, является показатель интенсивности ожиданий. Последний показывает, насколько высоки или «интенсивны» ожидания заданных изменений со стороны населения в каждый момент времени – например, ожидания роста уровня цен в экономике. Это позволяет осуществить аппроксимацию изменения ожиданий населения во времени.

Расчет индикатора базируется на определении числа комментариев, имеющих непосредственное отношение к измеряемым экономическим ожиданиям. Отбор комментариев осуществляется на предшествующем этапе методики – среди статей СМИ, релевантных заданной области экономики, измерение экономических ожиданий в которой является целью исследователя.

В каждый момент времени индикатор вычисляется как отношение числа релевантных комментариев к общему числу статей СМИ с упоминанием ожиданий, аналогичных искомым в комментариях (8).

$$\text{Интенсивность эк. ожиданий } I_T^{l^+} = \frac{\sum_{t=1}^T \sum_{l=1}^L \text{Число комментариев}_{tl}^{l^+}}{\sum_{t=1}^T \text{Число статей}_{t}^{l^+ \in L}}, \quad (8)$$

где  $T$  – период времени для расчета индикатора (день, месяц и др.),

$l^+$  – заданные исследователем экономические ожидания (например, ожидания роста цен),

$L$  – множество статей на тему конкретной области экономики, измерение экономических ожиданий в которой является целью исследователя (статьи СМИ, отобранные на втором этапе методики).

В соответствии с представленной формулой, числитель дроби включает суммарное для периода  $T$  число комментариев интернет-пользователей, упоминающих заданные экономические ожидания  $l^+$  в соответствии с системой правил отбора, к каждой из статей СМИ на тему  $L$ .

При этом знаменатель дроби включает суммарное для периода  $T$  число статей СМИ на тему  $L$ , упоминающих заданные экономические ожидания  $l^+$ , согласно системе правил, разработанной для отбора комментариев. Другими словами, отбор статей СМИ для расчета знаменателя дроби формулы (8) осуществляется в соответствии с правилами или регулярными выражениями, идентичными используемым для отбора комментариев интернет-пользователей. Это позволяет получить выборку статей СМИ, упоминающих заданные исследователем экономические ожидания. Данный шаг необходим с целью корректировки экономических ожиданий интернет-пользователей на медиа-ожидания или на возможную искусственную популяризацию темы со стороны СМИ. В результате, индикатор интенсивности экономических ожиданий будет стремиться отразить интенсивность «истинной» обеспокоенности населения конкретным вопросом или проблемой экономики – с исключением влияния медиа-ресурсов.

Корректировка на ожидания СМИ осуществляется путем деления (числа комментариев на число статей СМИ с упоминаниями аналогичных ожиданий). Таким образом, рост числа статей с обсуждением заданного вида экономических ожиданий способствует снижению итогового значения индикатора, тем самым корректируя на влияние СМИ соответствующий рост обсуждений этих статей интернет-пользователями. Фактически, индикатор представляет собой среднее число комментариев или интенсивность реакции читателей на одну статью, упоминающую целевые экономические ожидания.

Следует также отметить, что выбор деления для расчета индикатора осуществлен, вследствие характеристик рассматриваемых категорий данных. Среднее число статей и стандартное отклонение их временного ряда на порядки ниже среднего числа комментариев и стандартного отклонения их временного ряда. На одну статью, как правило, приходится более одного комментария. Так, в рассматриваемой в работе выборке данных среднее число статей в месяц составило 5131 шт. со средним отклонением 2793 шт. при среднем числе комментариев 31610 шт. со средним отклонением 27339 шт. При таких характеристиках выбор операции вычитания не привел бы к пропорциональным изменениям итогового значения индикатора при изменениях числа статей с упоминаниями целевых ожиданий. Таким образом, выбор операции деления представляется обоснованным.

Вторым шагом расчета индикатора интенсивности экономических ожиданий является оценка его релевантности заданной области ожиданий. В предлагаемом алгоритме расчета подобная оценка осуществляется в две стадии.

Во-первых, с целью оценки релевантности необходимо сравнить динамику полученного индикатора с существующими индикаторами аналогичных экономических ожиданий населения, измеренными на основе других методов – опросов населения или высокочастотных биржевых данных. Выбор индикаторов для оценки релевантности может быть осуществлен с помощью анализа международных эмпирических

исследований для заданной области экономических ожиданий. К примеру, для оценки релевантности индикатора интенсивности инфляционных ожиданий может быть использован медианный индикатор численных ожиданий инфляции, рассчитываемый информационным агентством «инФОМ» для Банка России на основе личных интервью домохозяйств в регионах России. Следует отметить, что в отдельных случаях показатели-аналоги для индикатора интенсивности ожиданий, измеренные другими способами – с помощью опросов или биржевых данных – могут отсутствовать. В таких случаях оценка релевантности осуществляется на основе только второй стадии.

Во-вторых, динамику полученного индикатора интенсивности ожиданий необходимо сравнить с динамикой целевого экономического параметра, ожидания в области которого измеряет исследователь. В частности, для инфляционных ожиданий подобным параметром выступает инфляция, для ожиданий ухудшения ситуации на рынке труда – уровень занятости или безработицы, а также темпы роста реальной заработной платы, для индикатора обеспокоенности кризисными процессами в экономике – темпы роста реального ВВП, для индикатора ожиданий обесценения курса национальной валюты – темпы роста курса доллара к рублю или курса бивалютной корзины к рублю и т.п.

Следует отметить, что сравнение динамики индикаторов может осуществляться в разных срезах частотности – в зависимости от доступной частотности данных по целевым экономическим параметрам или аналогичным индикаторам. К примеру, данные Росстата по темпам роста реального ВВП доступны только в поквартальном размере. Соответственно, полученный индикатор интенсивности обеспокоенности населения кризисными процессами может быть агрегирован до ежеквартального среза в целях соотнесения его динамики с динамикой экономического роста. Кроме того, в целях дополнительной проверки релевантности данного индикатора на высокочастотном уровне его динамика может быть соотнесена с



динамикой цен на нефть как одного из ключевых и признанных в экономической литературе факторов кризисов для экспортно-ориентированных экономик [Kitous et al., 2016]. В то же время в случае оценки релевантности индикатора интенсивности ожиданий обесценения курса национальной валюты, сравнение его динамики с темпами роста курса доллара к рублю или курса бивалютной корзины к рублю должно быть осуществлено в том числе и на ежедневной основе – вследствие доступности высокочастотных данных по валютным курсам.

Сравнение динамики построенного индикатора интенсивности конкретного вида экономических ожиданий с целевыми экономическими параметрами или аналогичными индикаторами, построенными на основе других методов, осуществляется, с одной стороны, с помощью визуального соотнесения динамики временных рядов показателей. С другой – с помощью расчета коэффициентов корреляции или критерия взаимной информации для сравниваемых рядов. При расчете критериев должна быть учтена статистическая значимость получаемых коэффициентов.

Критерий взаимной информации позволяет оценить наличие как значимой линейной, так и нелинейной взаимосвязи между сравниваемыми рядами, тогда как вычисление корреляции характеризует только силу линейной зависимости между рядами [Murphy, 2012]. В теории информации данный критерий представляет собой оценку количества информации, которое одна случайная величина содержит в отношении другой и вычисляется на основе условной энтропии рассматриваемых рядов данных [Cover, Thomas, 2006]. В общем виде для двух рядов непрерывных случайных величин этот критерий может быть рассчитан как:

$$I(X; Y) = \int_Y \int_X p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) dx dy, \quad (9)$$

где  $p(x, y)$  – это совместная плотность распределения случайных величин  $X$  и  $Y$ , а  $p(x)$  и  $p(y)$  – их частные функции распределения.

При этом при формировании итогового вывода о скоррелированности рядов и релевантности индикатора важно также учитывать возможные лаги зависимости целевых экономических параметров и оцениваемых экономических ожиданий – в соответствии с экономической теорией. Изменение экономических ожиданий населения может отражаться в динамике целевых экономических переменных по истечении определенного периода времени. В особенности, в случае если для данного вида ожиданий в большей степени характерен рациональный механизм формирования. Кроме того, в некоторых случаях ожидания, напротив, могут иметь запаздывающую взаимосвязь с динамикой экономических параметров. Это возможно в условиях, когда преобладает адаптивный механизм их формирования.

#### *2.4.2. Оценка устойчивости результатов расчета индикатора интенсивности экономических ожиданий*

При принятии решения о финальном наборе правил для расчета индикатора интенсивности ожиданий для заданной исследователем области экономики необходимо проведение ряда тестов на устойчивость результата к правилам расчета – в рамках рассматриваемого набора данных.

Оценка устойчивости включает две стадии.

Первая стадия представляет собой измерение чувствительности результатов расчета индикатора в зависимости от сформированных на втором и третьем этапах методики правил отбора новостных статей СМИ и комментариев интернет-пользователей. В частности, высокая чувствительность индикатора к небольшим изменениям правил отбора новостных статей и(или) комментариев интернет-пользователей указывает на недостаточное качество сформированных правил. Напротив, низкая чувствительность индикатора к небольшим изменениям правил свидетельствует в пользу устойчивости разработанной для данной области экономических ожиданий системы правил отбора новостных статей СМИ и комментариев интернет-пользователей. Результаты оценки первой стадии должны быть верифицированы и дополнены второй стадией оценки.

Небольшие изменения правил могут быть основаны на удалении их отдельных частей, но без потери содержательной нагрузки. При этом каждая вариация должна включать только одно изменение. Например, если финальное правило отбора основано на комбинации нескольких вариантов существительных с набором глаголов, вариации правил могут быть основаны на сочетании каждого из вариантов существительных с этим набором глаголов. Другой вариант построения вариации – сформировать их на базе последовательного исключения отдельных глаголов из набора, но с сохранением всех вариантов существительных. При этом в случае, если правило является комплексным и построено на нескольких отдельных комбинациях существительных и глаголов, вариации правил могут быть составлены с помощью последовательного удаления каждой из комбинаций. Важно подчеркнуть, что измененные правила отбора для проведения теста не должны терять своего содержательного значения для рассматриваемой области экономических ожиданий. К примеру, тестирование правила отбора для случая построения индикатора интенсивности ожиданий ослабления курса национальной валюты, не содержащего глагола, относящегося по смыслу к обесценению валюты, или существительного, обозначающего сам объект ожиданий (курс валюты), представляется некорректным.

В качестве примера проведения данного шага оценки устойчивости результатов расчета к правилам отбора новостных статей СМИ и комментариев интернет-пользователей в таблице 2 приведены возможные вариации правил отбора для случая построения индикатора интенсивности инфляционных ожиданий. Принципы построения вариаций правил в этом случае включают в себя все виды вариаций, рассмотренные выше. Предложенный тест на устойчивость представляет собой комплексную оценку устойчивости результата к правилам отбора как новостных статей СМИ, так и комментариев интернет-пользователей к ним, поэтому проверка устойчивости результата к правилам отбора комментариев должна также учитывать и изменения правил отбора новостных статей СМИ. Таким

образом, корректное проведение теста подразумевает расчет всех вариантов отбора комментариев интернет-пользователей для каждого варианта отбора новостных статей СМИ. Например, в случае проверки трех вариантов отбора новостных статей СМИ и трех вариантов отбора комментариев интернет-пользователей, общее число вариантов расчета индикатора интенсивности составит девять: по три варианта отбора комментариев для каждого из трех вариантов отбора статей. Следует отметить, что данное число представляется минимально возможным для получения выводов о чувствительности результатов построения индикатора в зависимости от сформированных правил отбора.

Таблица 2. Примеры вариации правил отбора новостных статей СМИ и комментариев интернет-пользователей для оценки устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий российского населения

Правила отбора новостных статей СМИ	Правила отбора комментариев интернет-пользователей к выбранным новостным статьям СМИ
<b>Исходный набор правил</b>	
<p><i>Объект ожиданий:</i> инфляция.</p> <p><i>Правила:</i></p> <p>(1) (инфляц ипц (темп.*цен)</p> <p>(2) ((рост повыс повыш ниж низ).*цен)</p> <p>(3) (индекс.*потребительских.*цен )</p>	<p><i>Целевые ожидания:</i> инфляционные ожидания.</p> <p><i>Правила:</i></p> <p>[«инфляц», «цен»] &amp; [«выс», «выш», «рост», «раст»]</p>
<b>Возможные вариации правил отбора</b>	
<p><i>Первая вариация:</i></p> <p>(1) (инфляц ипц (темп.*цен)</p> <p>(2) ((рост повыс повыш ниж низ).*цен)</p> <p><i>Вторая вариация:</i></p> <p>(1) (инфляц ипц (темп.*цен)</p> <p>(2) (индекс.*потребительских.*цен )</p> <p><i>Третья вариация:</i></p> <p>(1) ((рост повыс повыш ниж низ).*цен)</p> <p>(2) (индекс.*потребительских.*цен )</p>	<p><i>Первая вариация:</i></p> <p>[«инфляц»] &amp; [«выс», «выш», «рост», «раст»]</p> <p><i>Вторая вариация:</i></p> <p>[«цен»] &amp; [«выс», «выш», «рост», «раст»]</p> <p><i>Третья вариация:</i></p> <p>[«инфляц», «цен»] &amp; [«рост», «раст»]</p>

*Примечания:*

- (1) Все представленные регулярные выражения могут быть записаны в другой форме, в соответствии с правилами их составления;
- (2) Символ «&» в представленных регулярных выражениях обозначает поиск упоминаний в одном комментарии блоков слов, обозначенных в квадратных скобках. Такая модель поиска может быть записана при помощи различного синтаксиса в разных языках программирования, в которых используются регулярные выражения.
- (3) Перед поиском осуществляется базовая фильтрация текста (перевод символов в нижний регистр, удаление небуквенных символов).

*Источники:* составлено автором.

Вторая стадия заключается в оценке чувствительности результатов расчета индикатора в зависимости от изменения области экономических ожиданий. Данная стадия позволяет понять насколько в рассматриваемом наборе данных и с помощью применения разработанного алгоритма возможно выделение различных областей экономических ожиданий населения. Крайне высокая и стабильная взаимозависимость индикаторов интенсивности различных областей экономических ожиданий может, с одной стороны, означать неустойчивость разработанной системы правил отбора статей и комментариев, включая их низкую содержательную релевантность заданной области экономики. С другой стороны, это может свидетельствовать о специфике рассматриваемого набора данных, не позволяющей выделить различные по тематике экономические ожидания населения. К примеру, в каком-либо наборе данных могут преобладать статьи СМИ исключительно об одной теме экономики, в малой степени релевантной целевой области ожиданий. Вывод о причинах высокой и устойчивой взаимозависимости, в случае ее выявления, может быть получен с помощью результатов первой стадии оценки. В частности, если индикатор высокочувствителен к изменению правил отбора статей СМИ и комментариев интернет-пользователей, то высокая взаимозависимость индикаторов интенсивности для различных областей экономических ожиданий, скорее всего, обусловлена недостаточным качеством именно правил отбора новостных статей и комментариев, а не особенностями рассматриваемого набора данных. И наоборот, низкая чувствительность

индикатора к изменению правил отбора одновременно с высокой взаимозависимостью индикаторов для различных видов экономических ожиданий будет свидетельствовать о наличии негативной специфики в рассматриваемом наборе данных.

С точки зрения формальной процедуры проведения теста на устойчивость, на второй стадии оценки он может быть осуществлен с помощью расчета и сравнения динамики индикаторов интенсивности для различных областей экономических ожиданий, к примеру, таких как: ожидания ухудшения ситуации на рынке труда, ослабления курса национальной валюты, инфляционные ожидания и др. Правила отбора новостных статей СМИ и комментариев интернет-пользователей для этих областей экономических ожиданий были рассмотрены в разделе 2.3 в таблице 1. При этом по аналогии с логикой сравнения динамики рядов, предложенной в разделе 2.4.1, сравнение различных индикаторов интенсивности ожиданий может базироваться на визуальном соотнесении их динамики и статистической оценке их взаимозависимости на базе корреляции и критерия взаимной информации. Стоит отметить, что взаимосвязанность индикаторов для указанных областей экономических ожиданий не должна стремиться к нулю – экономические параметры, целевые для этих областей ожиданий, являются в значительной степени связанными в экономической теории. К примеру, историческая корреляция темпов роста курса доллара к рублю и индекса потребительских цен стремится к 90%<sup>8</sup>. Поэтому итоговая оценка наличия крайне высокой и устойчивой взаимозависимости различных индикаторов интенсивности должна производиться с учетом исторических – практических и теоретических – зависимостей экономических параметров, целевых для этих областей ожиданий.

---

<sup>8</sup> Рассчитано автором для показателей курса доллара к рублю и цен на нефть Brent в терминах темпов роста в каждом месяце к соответствующему месяцу предшествующего года за период с января 2014 г. по февраль 2018 г.

Положительный финальный вывод об устойчивости результатов расчета индикатора интенсивности может быть сформирован только на основе результатов оценки двух представленных выше стадий. Выводы об устойчивости правил отбора новостных статей СМИ и комментариев интернет-пользователей, полученные на первой стадии, должны быть подтверждены отсутствием высокой и устойчивой взаимозависимости в индикаторах интенсивности для различных областей экономических ожиданий. Получение отрицательного ответа по результатам хотя бы одной из стадий свидетельствует о необходимости доработки правил отбора новостных статей СМИ и(или) комментариев интернет-пользователей, или о необходимости расширения или изменения набора данных для анализа.

## **2.5. Выявление факторов изменения интенсивности экономических ожиданий населения во времени**

Алгоритм выявления факторов изменения интенсивности экономических ожиданий населения во времени представляет собой комплексную процедуру, базирующуюся на трех шагах. Результаты каждого из шагов должны быть верифицированы выводами остальных шагов.

Первый шаг заключается в оценке зависимости построенного индикатора интенсивности экономических ожиданий от различных экономических параметров, способных в экономической теории выступать факторами формирования ожиданий населения в заданной области экономики.

Второй шаг основан на экспертной оценке содержания комментариев с целью выявления факторов, с которыми население связывает изменение собственных ожиданий. Шаг предполагает детальное изучение текстового содержания финальной выборки комментариев, отнесенных к заданной области ожиданий и выделение категорий экономических параметров или тем, которые интернет-пользователи упоминают в комментариях одновременно с упоминанием целевых ожиданий. При этом процедура

выявления таких категорий осуществляется с учетом выделенных по результатам первого шага факторов изменения экономических ожиданий.

Третий шаг также направлен на выявление факторов, с которыми население связывает изменение собственных ожиданий, однако основан на методах машинного обучения и текстовой обработки – без вовлечения экспертной оценки. Этот шаг предполагает проведение тематического моделирования содержания комментариев, отнесенных к заданной области экономических ожиданий. Результаты моделирования позволяют выделить ключевые сочетания слов в сообщениях интернет-пользователей, обозначающих темы, которые интернет-пользователи упоминают в комментариях одновременно с упоминанием целевых ожиданий. Результаты данного шага используются для верификации и дополнения выводов первого и второго шагов. Такой подход способствует снижению доли субъективной экспертной оценки при экспертном выявлении факторов, осуществляемом на втором шаге, а также позволяет проверить появление новых категорий или тем, существенных для изменения ожиданий.

#### *2.5.1. Оценка зависимости индикатора интенсивности экономических ожиданий от релевантных экономических параметров*

Оценка зависимости рассчитанного индикатора интенсивности экономических ожиданий от различных экономических параметров, способных в экономической теории выступать факторами формирования ожиданий населения в заданной области экономики осуществляется с помощью, с одной стороны, проверки простой взаимозависимости рядов, и, с другой стороны, анализа причинно-следственных связей на основе статистических тестов.

Проверка простой взаимозависимости рядов включает в себя оценку наличия линейной взаимозависимости с помощью вычисления коэффициента корреляции и верификацию наличия значимых нелинейных взаимосвязей – на основе критерия взаимной информации (см. раздел 2.4.1).



Оценку наличия значимой причинно-следственной связи предполагается осуществлять с помощью теста Гренджера.

Тест Гренджера базируется на построении регрессии предполагаемой зависимой переменной на собственные предшествующие значения и предшествующие значения предполагаемого фактора и верификации нулевой гипотезы об одновременной незначимости или равенстве нулю коэффициентов при последних [Granger, 1969]. В случае отвержения нулевой гипотезы, можно сделать вывод о наличии причинности по Гренджеру между рассматриваемыми рядами. Тест также включает в себя проверку обратной взаимосвязи и проверку устойчивости выводов относительно числа используемых лагов для вычисления предшествующих значений переменной. Результаты последнего могут отличаться при изменении числа включаемых в анализ лагов. При этом важно отметить, что корректное проведение теста возможно только для стационарных временных рядов.

Следует подчеркнуть, что положительный результат теста на причинность по Гренджеру не обязательно означает наличие истинной причинно-следственной связи. Последняя должна быть обязательно подтверждена теоретическими предпосылками, а также может быть дополнительно проверена с помощью другого статистического инструментария, включая различные расширения и модификации теста Гренджера, тест на мгновенную причинно-следственную зависимость (*instantaneous causality*), тесты с использованием теории графов и др. [Jin-Lung Lin, 2008; Schwert, 1979].

Следует отметить, что углубленный причинно-следственный анализ не является приоритетной целью разработки алгоритма. Поэтому методологическая база предполагаемой оценки причинно-следственной связи является ограниченной. Осуществляемый на данном этапе методики анализ в большей степени ориентирован на получение первичного представления о возможных факторах изменения интенсивности ожиданий и на проверку ключевых теоретических предпосылок с точки зрения экономики.

Проведение углубленного анализа причинно-следственных цепочек для рассчитанных индикаторов интенсивности экономических ожиданий и усиление аналитического инструментария для него является одним из направлений развития данного исследования.

При этом новизной подхода, основанного на исследовании больших массивов текстовых данных является изучение событий и явлений, с которыми население связывает изменение собственных ожиданий, а также того, как население определяет важность каждого из событий и явлений относительно остальных в реальном времени – или за любой выбранный исследователем период. Кроме того, предлагаемая методика также позволяет проанализировать роль каждого фактора во времени, а также его влияние на повышение или снижение интенсивности ожиданий во время различных экономических шоков.

#### *2.5.2. Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе экспертного анализа*

Шаг выявления факторов, с которыми население связывает изменение собственных ожиданий на основе экспертного анализа включает три последовательных стадии.

Первая стадия заключается в детальном экспертном анализе содержания целевых комментариев – тех, в которых интернет-пользователи упоминают заданный вид экономических ожиданий. На основе данного анализа исследователь выделяет содержательные категории факторов или тем, с которыми население связывает изменение интенсивности собственных ожиданий, их усиление и снижение. Другими словами, целью этой стадии является выявление обобщающих категорий для экономических событий, явлений или трендов, которые население упоминает в одних и тех же сообщениях с упоминанием целевого для исследователя вида экономических ожиданий. Анализ содержания комментариев может быть осуществлен как с помощью исследования случайных выборок комментариев (в случае их большого для ручной обработки количества), так и с помощью изучения всех

комментариев интернет-пользователей, отмеченных как целевые (упоминающие заданный вид экономических ожиданий).

Важно, что процесс выделения содержательных категорий, или факторов изменения индикатора интенсивности заданной области экономических ожиданий, должен осуществляться с учетом выводов первого шага о макроэкономических параметрах, способных оказывать значимое влияние на изменение рассчитанного индикатора.

Для каждой из экспертно выделенных содержательных категорий должны быть составлены правила для их поиска в тексте комментариев интернет-пользователей – на основе регулярных выражений. Ключевые принципы составления таких правил отражены в разделе 2.3. Составление правил отбора является итеративным процессом, предполагающим их постепенное улучшение на основе экспертной оценки качества отбора комментариев и минимизации доли нерелевантных примеров.

Вторая стадия выявления факторов, с которыми население связывает изменение собственных ожиданий, на основе экспертного анализа состоит в расчете показателей вкладов упоминаний каждой выделенной на первой стадии категории в общем числе упоминаний всех результирующих категорий.

Показатель вклада для каждого выделенного фактора рассчитывается как отношение суммарного числа комментариев с упоминанием этого фактора в целевых комментариях в заданный период времени к общему числу упоминаний всех факторов в целевых комментариях в этот период (10). Следует отметить, что в одном комментарии интернет-пользователем может быть упомянуто более одного фактора. В таком случае все упоминания учитываются при расчете вкладов каждого из факторов как отдельные, без каких-либо корректировок.

$$\text{Вклад фактора}_{f,T} = \frac{\sum_{t=1}^T \text{Число упоминаний фактора}_{ft}}{\sum_{f=1}^F \sum_{t=1}^T \text{Число упоминаний фактора}_{ft}}, \quad (10)$$

где Вклад фактора  $f_{f,T}$  обозначает относительную важность фактора  $f$  из множества выделенных экспертом факторов  $F$  изменения интенсивности конкретного вида экономических ожиданий населения за период  $T$ .

Следует отметить, что для вычисления вкладов факторов используются только целевые для заданной области экономических ожиданий комментарии интернет-пользователей.

Результатом выполнения данной стадии являются временные ряды относительного вклада выделенных экспертно значимых факторов в изменение индикатора интенсивности заданного исследователем вида экономических ожиданий населения.

Третья стадия выявления факторов, с которыми население связывает изменение собственных ожиданий, включает сравнение динамики полученных на второй стадии показателей вклада различных факторов в динамику индикатора интенсивности ожиданий непосредственно с динамикой самого индикатора. В частности, данная стадия содержит проверку причинно-следственной зависимости между выделенными факторами и индикатором интенсивности ожиданий. Это осуществляется на основе визуального анализа, а также вычисления коэффициентов корреляции, критериев взаимной информации и проведения тестов Гренджера – в соответствии с алгоритмом, описанным в разделе 2.4.1.

### *2.5.3. Выявление факторов, с которыми население связывает изменение собственных ожиданий, на основе методов машинного обучения и текстовой обработки*

В основе третьего шага лежит процедура тематического моделирования целевых комментариев или комментариев интернет-пользователей, отнесенных к заданной категории экономических ожиданий при реализации методики. Это позволяет выделить ключевые сочетания слов в целевых для каждого вида ожиданий комментариях интернет-пользователей, обозначающих темы – экономические события, явления и тренды – которые

интернет-пользователи упоминают в комментариях одновременно с упоминанием целевых ожиданий. Цель данного шага состоит в верификации и дополнении выводов второго и первого шагов.

Это, во-первых, способствует снижению доли субъективной экспертной оценки относительно ключевых факторов и их роли в динамике индикатора интенсивности экономических ожиданий. В теории машинного обучения тематическое моделирование является так называемым методом обучения без учителя (unsupervised learning) [Bishop, 2006], который предполагает отсутствие целевой переменной для обучения модели и относительно низкую возможность влияния эксперта на процесс реализации алгоритма. Выделение тем происходит с помощью оптимизации параметров распределений, относящихся к словам и их сочетаниям в комментариях интернет-пользователей. Некоторые из этих параметров могут быть выбраны экспертно, однако содержательные корректировки тем в классическом тематическом моделировании не предусмотрены.

Во-вторых, проведение тематического моделирования в каждый рассматриваемый период времени позволяет проверить появление новых категорий или тем, существенных для изменения ожиданий. Кроме того, возможна коррекция содержательного состава уже выделенных на втором шаге значимых тем. В соответствии с процедурой тематического моделирования для целевых комментариев, относящихся к заданному периоду времени, выделяется оптимальное число тем, определенное для данного конкретного случая. Оптимальный набор тем будет представлять собой наиболее существенные в рамках обсуждения заданной области экономических ожиданий темы. При этом поскольку параметры алгоритма не фиксированы, появление новой существенной для динамики ожиданий темы в обсуждениях интернет-пользователей будет отражено в увеличении оптимального числа тем, и появлении словаря слов и словосочетаний для новой темы по результатам работы алгоритма. Кроме того, постоянно обновляемый и корректируемый алгоритмом состав слов и словосочетаний

ключевых для каждого периода времени тем также может быть учтен при улучшении правил отбора для экспертно выделенных на втором этапе содержательных категорий.

### **Тематическое моделирование**

Тематическое моделирование – это один из классических подходов машинного обучения, используемый в таком направлении исследований, как обработка естественного языка (natural language processing, NLP). Этот подход используется для автоматической работы с большими объемами текстовых данных – с целью выделения ключевых слов и словосочетаний, встречающихся в сходном контексте в рассматриваемом наборе данных. Другими словами, тематическое моделирование направлено на выявление кластеров слов и словосочетаний в определенном наборе данных [Jurafsky, Martin, 2018].

Одним из наиболее популярных и развитых алгоритмов тематического моделирования является метод Латентного размещения Дирихле (Latent Dirichlet Allocation, LDA). В соответствии с Хуанг [Huang et al., 2017] у алгоритма LDA можно выделить три ключевых преимущества над экспертным составлением правил для определения тем в тексте. Во-первых, автоматическая реализация алгоритма позволяет применить его к большим коллекциям документов, объем информации в которых будет представлять значительное затруднение для обработки экспертом. Во-вторых, результатом алгоритма LDA является надежная и воспроизводимая классификация тем для текстовой информации. Финальная модель LDA, обученная на определенном наборе текстов, может быть применена с целью выявления полученного набора тем для любой коллекции документов, включая новые тексты. В-третьих, алгоритм LDA не требует от исследователя предварительной спецификации правил поиска или ключевых слов для определения тем или других категорий в тексте. Темы и их вероятностные взаимосвязи с ключевыми словами определяются алгоритмом с помощью

поиска оптимальной статистической модели в заданном корпусе текстовой информации.

Тематическое моделирование в целом и метод LDA в частности начали получать активное распространение и в экономических исследованиях. К примеру, в работе Хуанг [Huang et al. 2017], используется алгоритм LDA для сравнения тематического содержания большой выборки отчетов аналитиков, опубликованных сразу после телефонных конференций с инвесторами со стороны публичных компаний по поводу финансовых результатов их деятельности (earnings conference calls) с тематическим содержанием самих звонков. В исследовании Дайер [Dyer et al., 2016] алгоритм LDA был применен с целью выявления трендов в раскрытии информации в детальной годовой отчетности компаний за период 1996-2013 гг., в работе Бейлка [Balke et al., 2015] – для текстовых данных Бежевой книги ФРС США (Beige Book). В работе Голощаповой [Goloshchapova et al., 2018] алгоритм LDA, дополненный методикой валидации результатов и настройки параметров, используется для выделения ключевых тем в отчетности компаний, относящейся к ее ответственности в области охраны окружающей среды, социальной сферы и корпоративного управления (Environment, social and governance, ESG), а также корпоративной социальной ответственности (Corporate social responsibility, CSR). Анализ осуществлен для компаний Великобритании и стран Евросоюза за период 1999-2017 гг. Важно отметить, что комплексная процедура выявления тем в больших массивах текстовой информации, представленная в [Goloshchapova et al., 2018], основана на методике, предложенной в этом разделе.

Исследование Торсруда [Thorsrud, 2016], как упоминалось в разделе 1.3.3, использует тематическое моделирование с целью прогноза различных характеристик бизнес-циклов. В данной работе одними из ключевых факторов для прогноза выступают индексы вероятностей появления определенного набора тем экономических новостей, полученных с помощью алгоритма LDA. При этом в исследовании Ларсена и Торсруда [Larsen and

Thorsrud, 2015] авторы показывают статистическую значимость этих индексов для предсказания не только характеристик бизнес-циклов, но и других важных параметров экономической конъюнктуры.

В российской практике одними из первых исследований в области экономической науки с использованием тематического моделирования стали работы Голощаповой и Андреева [Goloshcharova, Andreev, 2017] и Яковлевой [Яковлева, 2017]. В исследовании Голощаповой и Андреева [Goloshcharova, Andreev, 2017] тематическое моделирование осуществляется на базе комментариев интернет-пользователей, упоминающих ожидания роста цен. Комплексная процедура выявления факторов на основе методов текстовой обработки и машинного обучения, представленная в этом разделе, учитывает ключевые выводы Голощаповой и Андреева [Goloshcharova, Andreev, 2017]. В работе Яковлевой [Яковлева, 2017] осуществлена реализация алгоритма Торсруда [Thorsrud, 2016] на новостных данных одного из крупных новостных СМИ в целях предсказания индекса PMI для России.

### **Алгоритм Латентного размещения Дирихле (LDA)**

Алгоритм Латентного размещения Дирихле впервые был предложен в 2003 г. в работе Блэй [Blei et al., 2003]. В его основе находится предположение о том, что каждый рассматриваемый документ является набором или смесью различных латентных тем, а каждая тема – набором или смесью слов и словосочетаний. По результатам работы алгоритма темы в каждом документе представлены вероятностью появления в них каждого слова и (в зависимости от настройки параметров работы алгоритма) словосочетания в рассматриваемом корпусе текстовой информации. При этом полученный набор тем для всех документов также определен вероятностью появления каждой темы в каждом документе [Jelodar et al, 2017]. Априорным распределением для тем в корпусе документов является распределение Дирихле, что определяет название алгоритма.



Алгоритм LDA направлен на итеративный поиск оптимального распределения тем в каждом документе и оптимального распределения слов в каждой теме. Процесс оптимизации заключается в обучении двухшаговой генеративной модели на базе наблюдаемых в каждом документе слов – до момента выявления наилучшего набора параметров, описывающих как распределение тем в документах, так и распределение слов в темах. Таким образом, процесс оптимизации в определенной мере аналогичен кластерному анализу или вычислению главных компонент в применении к количественным данным.

Следует уточнить, что генеративными моделями классификации, в терминологии теории машинного обучения и обработки естественного языка, называются модели, обучение которых основано на предпосылке о совместном распределении факторов  $x$  и целевых параметров  $y$  –  $p(x, y)$ . Прогноз в таких моделях осуществляется на основе вычисления с помощью правила Байеса условного распределения вероятности  $y$  при  $x$  –  $p(y/x)$ <sup>9</sup> – и, затем, расчета наиболее вероятного значения  $y$  [Ng, Jordan, 2002]. Альтернативой генеративным моделям являются дискриминативные модели, в которых апостериорное условное распределение  $p(x, y)$ , или зависимость целевого параметра  $y$  от факторов  $x$ , моделируется напрямую на базе каждой отдельной величины параметра  $y$  [Bishop, Lasserre, 2007].

В финальном, по результатам работы алгоритма LDA, наборе тем слова, в наибольшей степени относящиеся к каждой теме, будут иметь наибольшую вероятность, в наименьшей – близкую к нулю. При этом вероятность наименее релевантных для всех тем слов будет примерно одинаковой для всех тем, или будет размещена алгоритмом в отдельную категорию (тему).

Перед началом работы алгоритма ни одна из тем не является строго определенной, ни семантически, ни эпистемиологически. Темы

---

<sup>9</sup> В соответствии с теоремой, названной в честь английского математика Т. Байеса (1702-1761), условная вероятность события  $A$  при условии реализации события  $B$  равна:  $p(A|B) = \frac{P(B|A)P(A)}{P(B)}$  [Murphy, 2012].

определяются автоматически в процессе работы алгоритма – на базе оптимизации параметров совместных распределений вероятностей появления слов и словосочетаний в темах и тем в документах. При этом одно и то же слово может встречаться в различных темах с разной вероятностью – однако с неодинаковым контекстом других слов и словосочетаний в одной теме. Технически алгоритм LDA осуществляет декомпозицию матрицы «документ-слово», обозначающей частоту появления каждого слова в каждом документе в рассматриваемом наборе данных, на две матрицы более низкого ранга: матрицу «документ-тема», обозначающую вероятность появления каждой темы в каждом документе, и матрицу «тема-слово», обозначающую вероятность появления каждого слова в каждой теме. Финальные вероятности вычисляются с учетом априорного распределения для тем в корпусе текстов, заданного самим алгоритмом как распределение Дирихле, а также ряда других параметров работы алгоритма, задаваемых исследователем.

В число параметров работы алгоритма, определяемых исследователем, входят:

- число тем в рассматриваемом наборе документов. Этот параметр может быть определен как экспертно на основе знаний предметной области, так и с помощью автоматических методов валидации результатов моделирования;
- гиперпараметр  $\alpha$  для распределения «документ-тема», принимающий значения от 1 до 0 и отвечающий за степень разреженности распределения. К примеру, значение 0.1 будет задавать относительно разреженную матрицу «документ-тема», или матрицу с преимущественно нулевыми элементами, предполагающую всего несколько существенных тем для каждого документа. Данный параметр определяется экспертно, в зависимости от целей исследователя;

- гиперпараметр  $\beta$  для распределения «тема-слово», значения которого также лежат в интервале от 0 до 1 и который аналогичным образом отвечает за степень разреженности распределения. Здесь, значение параметра близкое к 0, например, 0.001, будет означать в высокой степени разреженное распределение «тема-слово», ограничивающее число значимых для каждой темы слов – как правило, всего по несколько слов на тему (в зависимости от объема данных и числа заданных тем).

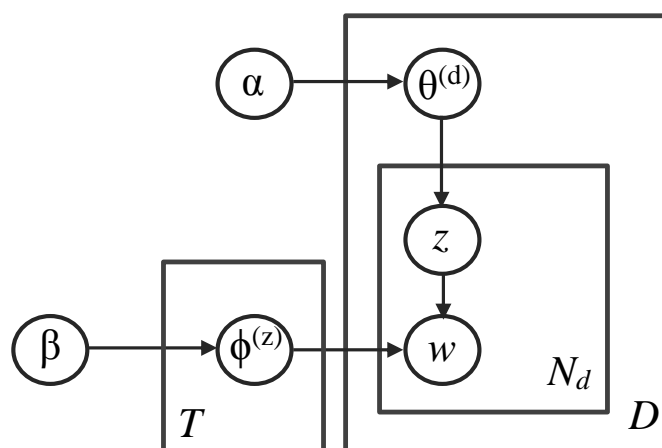
Пусть  $P(z)$  – функция плотности распределения тем  $z$  в каком-либо документе, тогда  $P(w|z)$  – функция плотности распределения слов в какой-либо теме  $z$ . Каждое слово  $w_i$  в документе, где  $i$  отвечает за номер слова в словаре всех слов в коллекции документов, выбирается или генерируется алгоритмом с помощью семплирования из распределения тем на первом шаге, и затем – из распределения слов в теме.  $P(z_i = j)$  обозначает вероятность того, что тема  $j$  окажется в выборке для слова  $i$ , а  $P(w_i|z_i = j)$  – вероятность того, что слово  $i$  окажется в выборке для темы  $j$ . Таким образом, в соответствии со Стейверсом и Гриффитсом [Steyvers, Griffiths, 2006], распределение слов в документе можно показать следующим образом (11):

$$P(w_i) = \sum_{j=1}^T P(w_i|z_i = j)P(z_i = j), \quad (11)$$

где  $T$  отвечает за общее число тем в коллекции документов.

Общая концепция алгоритма Латентного размещения Дирихле и его ключевые параметры представлены на рис. 4. Мультиномиальное распределение слов для темы  $j$  –  $P(w|z = j)$  – можно обозначить как  $\phi^{(j)}$ , а мультиномиальное распределение тем для документа  $d$  –  $P(z)$  – как  $\theta^{(d)}$ . Предположим, что общее число документов в коллекции равно  $D$ , причем каждый документ  $d$  состоит из  $N_d$  слов. При этом  $N$  является общим числом слов во всей коллекции документов ( $N = \sum N_d$ ). Параметры  $\phi$  и  $\theta$ , фактически, обозначают важность слов для различных тем и важность тем

для различных документов, соответственно. Подобная структура характерна для большинства вероятностных моделей тематического моделирования. Как уже упоминалось выше, основным отличием LDA алгоритма, зарекомендовавшим себя на практике, является предположение об априорном распределении Дирихле для параметра  $\theta$ . В рамках алгоритма LDA также следует обозначить  $\alpha$  и  $\beta$  – гиперпараметры распределений «документ-тема» и «тема-слово», регулирующие их разреженность (см. выше).



*Источник:* [Steyvers, Griffiths, 2006].

Рис. 4. Графическое представление для алгоритма LDA.

Выбор оптимального числа тем для алгоритма LDA, как правило, осуществляется экспертно – в зависимости от представлений о возможном количестве тем в рассматриваемой коллекции документов, а также от целей исследователя. Процесс определения оптимального числа тем при этом может быть итеративным: на каждом шаге автор исследования определяет качество полученных тем и их релевантность заданному набору текстов. На основе сделанных выводов автор меняет число тем и производит очередной пересчет алгоритма – до достижения оптимального с точки зрения содержания набора тем, и в частности его релевантности рассматриваемой коллекции документов. Следует отметить, что слишком малое число тем для заданного набора документов будет означать очень широкие темы, основной

содержательный смысл которых будет сложно сформулировать. В свою очередь, слишком большое число тем будет означать, что каждая тема будет являться неинтерпретируемой комбинацией слов, поскольку последних будет недостаточно для определения содержательного смысла темы.

Таким образом, принятие решения о точном количестве тем в документах во многих случаях может быть затруднительным. В особенности, в условиях больших наборов текстовой информации. В таких случаях – для помощи в принятии финального экспертного решения – могут быть использованы автоматические методы валидации результатов алгоритма LDA [Poon et al., 2017]. Последние, в соответствии с предлагаемой в данной работе методикой, состоят из вычисления и проверки совокупного результата для ряда статистических критериев, обоснованных научными исследованиями в качестве релевантных для алгоритма LDA.

Для определения оптимального числа тем предлагается использовать одни из наиболее популярных и признанных в научном сообществе статистических критериев, а именно:

- Критерий Гриффитса и Стейверса [Griffiths and Steyvers, 2004]. Расчет данного критерия основан на Байесовском подходе к отбору статистических моделей (Bayesian model selection). Общая идея состоит в оценке апостериорной вероятности модели с учетом всех возможных значений ее параметров (или вариантов принадлежности слов темам). Итоговое число тем по результатам вычислений равно максимальной апостериорной вероятности модели.
- Критерий Као Джуан [Cao Juan et al., 2009]. Этот показатель основан на адаптивном подходе к оптимизации числа тем в рамках алгоритма LDA на базе расстояний между распределениями слов в темах в соответствии с распределением Дирихле – в зависимости от задаваемого числа тем. В качестве меры расстояния выступает среднее косинусное расстояние между всеми возможными парами тем. Этот показатель, как показывают авторы, аппроксимирует

стабильность структуры тем для заданного числа тем по результатам работы алгоритма LDA: чем выше расстояние, тем ниже стабильность. Таким образом, итоговое число тем определяется путем минимизации заданного критерия (или меры расстояния между распределениями слов в темах).

- Критерий Аруна и соавторов [Arun et al., 2010]. Вычисление показателя базируется на применении теории информации к тематическому моделированию, а именно расчету меры симметричной дивергенции Кульбака-Лейблера между распределениями, полученными из результирующих матриц алгоритма LDA «документ-тема» и «тема-слово», для различного числа тем. Авторы доказывают, что в общем случае для неоптимального числа тем, указанная мера дивергенции выше. Таким образом, итоговое число тем определяется путем минимизации заданного критерия.
- Мера перплексии Жао [Zhao et al., 2015]. Оценка на основе вычисления перплексии представляет собой построение функции значений показателя перплексии от числа тем. Оптимальным диапазоном для выбора числа тем в рамках алгоритма LDA становится момент изменения в поведении функции – переход от резкого снижения к небольшим и малозначительным изменениям. Критерий перплексии является распространенным и популярным критерием оценки качества моделей тематического моделирования и обозначает степень несоответствия или «удивленности» модели появлению слов в каждом документе – в соответствии с итоговыми результатами работы алгоритма LDA. Другими словами, это величина, обратно связанная со способностью итоговой модели LDA при заданном количестве тем предсказать имеющийся в каждом документе набор слов. Таким образом, минимизация меры перплексии в общем случае ведет к более качественной модели LDA

[Воронцов, 2013]. Однако, оценка на основе критерия перплексии базируется именно на поиске момента изменения функции перплексии от числа тем, поскольку статистически глобальный минимум для данного критерия будет достигнут в точке, где число тем будет равно числу слов. В действительности, наилучшим образом можно предсказать текст, в каждой теме которого нет альтернатив для выбора слова.

Все перечисленные статистические критерии и функция перплексии рассчитываются для различного количества тем и изучаются экспертом в динамике – в целях сужения диапазона рассматриваемых для алгоритма LDA тем. У каждого из приведенных критериев есть недостатки, проявляющиеся в разных случаях и при различных особенностях рассматриваемых текстовых данных. Поэтому финальное решение о числе тем в соответствии с предлагаемым алгоритмом принимается экспертом с учетом всей доступной информации.

Следует также упомянуть, что для проведения тематического моделирования текстовые данные должны быть предварительно обработаны. Традиционно, этот процесс включает:

- Базовую фильтрацию текста. Эта ступень содержит перевод всех буквенных символов в нижний регистр и удаление из текста так называемых «лишних» символов, не относящихся к словам, важным для определения тем. В число лишних символов, как правило, включают: неязыковые символы, не буквенно-цифровые символы, знаки препинания, повторяющиеся пробелы, однобуквенные символы.
- Токенизацию. Эта ступень обозначает разделение сплошной текстовой информации документов на отдельные слова.
- Удаление стоп-слов или слов, не несущих существенной смысловой нагрузки для определения тем. Списки классических стоп-слов для каждого языка, как правило, определяются и регулярно обновляются лингвистами и являются общедоступными. В подобные списки чаще

всего входят предлоги, вводные конструкции, союзы, местоимения, оценочные определения и наречия и др. Списки стоп-слов для русского языка можно найти в библиотеке nltk – одной из наиболее распространенных программ по обработке естественного языка на языке программирования Python, или в библиотеке tm, исполняющей аналогичную роль для R – другого популярного языка программирования для анализа данных. При этом устоявшейся практикой текстовой обработки также является итеративное дополнение списка стоп-слов исследователем – на основе экспертного анализа качества результатов алгоритма LDA – в зависимости от особенностей исследуемых данных.

- **Стемминг.** Это означает приведение каждого слова к его основе или исходной форме. Например, слова «экономики», «кризисов» и «снижался» в результате стемминга будут приведены к формам «экономика», «кризис», «снижаться». Для проведения этой операции чаще всего используются готовые разработки для различных языков. К примеру, для русского языка классическим решением является программа mystem, разработанная в Yandex [Segalovich, 2003].

По результатам тематического моделирования ранжирование полученного набора тем или выбор тем, наиболее важных для данного индикатора интенсивности экономических ожиданий, осуществляется на базе не только итоговой вероятности появления каждой темы в различные периоды времени, но и с учетом экспертной оценки важности тем в контексте заданной области исследования. Выбор наиболее важных тем в каждый период времени, во-первых, может зависеть от факторов, оказавшихся наиболее существенными в этот период в рамках экспертной оценки, поскольку дополнение и верификация результатов экспертной оценки является одной из ключевых целей применения технологии тематического моделирования к выявлению факторов изменения интенсивности экономических ожиданий населения. Во-вторых,



определенный набор тем может приобретать особую значимость в определенные периоды времени, характеризующиеся какими-либо отдельными шоками в экономике. Например, в период введения новых санкций против России особую важность может приобрести оценка вклада этого фактора в интенсивность беспокойности российского населения кризисной ситуацией в экономике. В целях выявления подобных отдельных факторов тематическое моделирование может проводиться исследователем для различных промежутков времени: полугодий, кварталов и др.

Таким образом, итоговые выводы об основных факторах изменения интенсивности экономических ожиданий населения и о значимости этих факторов в контексте различных периодов времени, или финальные предположения относительно интерпретации динамики индикаторов интенсивности, осуществляются исследователем только на основе детального изучения результатов всех рассмотренных в данном разделе шагов выявления факторов во времени.

## **Выводы главы 2**

1. Разработана методика построения высокочастотных индикаторов экономических ожиданий населения на основе больших данных. В частности, в главе представлено общее описание предлагаемой методики, а также детальный состав каждого этапа, включая: (1) определение источников информации и оценку репрезентативности выборки для анализа; (2) отбор новостных статей выделенной выборки экономических СМИ на заданную тему или область экономических ожиданий; (3) отбор комментариев интернет-пользователей к выделенному набору новостных статей экономических СМИ на тему конкретной области экономических ожиданий; (4) расчет индикатора интенсивности ожиданий российского населения для заданной области экономики; (5) выявление факторов изменения интенсивности экономических ожиданий населения во времени.

2. Разработан алгоритм определения источников информации и правил отбора новостных статей СМИ и комментариев интернет-пользователей для

построения индикаторов экономических ожиданий. Источники информации определяются на основе результатов рейтингов медиа в области экономики и опросов экспертов с учетом российской специфики. Правила отбора новостных статей СМИ и комментариев интернет-пользователей основаны на комбинировании регулярных выражений с учетом итеративной верификации соответствия результатов отбора содержательным целям исследователя.

3. Разработан алгоритм расчета высокочастотных индикаторов экономических ожиданий российского населения. Алгоритм представляет собой комплексную процедуру, состоящую из трех шагов: (1) вычисление временного ряда интенсивности заданного вида экономических ожиданий населения на основе составленных правил отбора статей СМИ и комментариев интернет-пользователей, (2) оценку релевантности полученного индикатора с точки зрения теоретических предпосылок о его динамике и соответствии ключевым макроэкономическим трендам и индикаторам-аналогам, полученным другими способами; (3) проверку устойчивости результатов расчета индикатора интенсивности.

4. Разработан алгоритм выявления факторов изменения интенсивности экономических ожиданий населения во времени. Данный алгоритм также является комплексным процессом, объединяющим результаты анализа содержания комментариев с помощью трех шагов: (1) статистического анализа зависимостей; (2) экспертной оценки; (3) методов машинного обучения. При этом финальные выводы о факторах осуществляются только с учетом объединения и верификации результатов каждого шага.

### **Глава 3. Верификация методики построения высокочастотных индикаторов экономических ожиданий российского населения на примере инфляционных ожиданий**

В данной главе представлена реализация предложенной во второй главе методики построения высокочастотных индикаторов экономических ожиданий населения на основе больших данных на примере инфляционных ожиданий российского населения.

Реализация методики включает определение источников информации и данных для анализа, расчет индикатора интенсивности инфляционных ожиданий российского населения с учетом всех предусмотренных методикой оценок репрезентативности исходных данных и полученных результатов, а также тестов на устойчивость результатов расчета индикатора. Кроме того, в данной главе, в соответствии с предложенной методикой, выявлены факторы изменения интенсивности инфляционных ожиданий российского населения во времени. Предложена интерпретация динамики инфляционных ожиданий за период с января 2014 г. по февраль 2018 г.

#### **3.1. Определение источников информации и данных для анализа**

Для построения высокочастотного индикатора инфляционных ожиданий используется выборка крупнейших российских экономических СМИ, обладающая минимальной репрезентативностью для оценки экономических ожиданий населения – на основе авторской карты источников, составленной в главе 2. Данная выборка включает: Ведомости, РБК, Эксперт и Коммерсант. Этот блок источников был определен на основе рейтингов цитируемости СМИ, опросов экспертов и дискуссий в научном сообществе и получил название обязательного блока в рамках авторской карты источников (см. раздел 2.2.1).

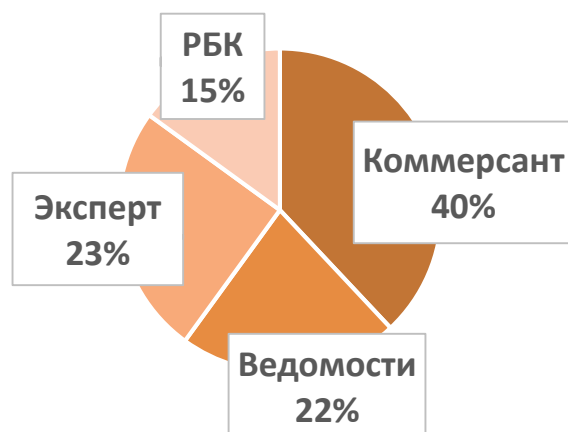
Источники информации о мнении читателей относительно материалов крупнейших экономических СМИ соответствуют источникам, определенным

в разделе 2.2.2: официальные страницы СМИ в сети Интернет, официальные группы СМИ в социальных сетях ВКонтакте и Facebook.

Все указанные данные собраны с помощью специальной программы-краулера, написанной на языке программирования Python при осуществлении проекта Big Data Indicators [Голощапова, Андреев, 2017; Андреев, 2017].

Период построения индикатора: с начала января 2014 г. до конца февраля 2018 г. Выбор начальной даты для сбора данных осуществлен на основе двух факторов (см. раздел 2.1). Первый фактор состоит в достижении с 2014 г. устойчивого для российской экономики уровня по показателю доли домохозяйств, имеющих доступ к сети Интернет в России (около 60%, см. рис. 3). Вторым фактором – закрытие доступа к существенной доле материалов СМИ с датой публикации до 2014 г. Важно отметить, что в соответствии с разработанной методикой конечная дата расчета не фиксирована, предусмотрено регулярное обновление данных и расчетов. Таким образом, индикатор инфляционных ожиданий в дальнейшем может быть построен для любого периода данных.

Рассматриваемый набор данных содержит 263 тыс. новостных статей СМИ. При этом около 40% из них принадлежат Коммерсанту, 23% опубликованы Экспертом, 22% размещены на сайте Ведомостей, и 15% статей приходится на РБК. Таким образом, с точки зрения источников информации о новостных статьях СМИ выборка представляется относительно сбалансированной (см. рис. 5).



*Источник:* составлено автором.

Рис. 5. Структура источников новостных статей СМИ в выборке данных за период с января 2014 г. по февраль 2018 г. для оценки инфляционных ожиданий российского населения.

Совокупное количество комментариев интернет-пользователей к материалам СМИ в перечисленных выше источниках составило 1.5 млн. сообщений. Число уникальных авторов комментариев составило около 140 тыс. человек. Такие объемы данных не могут быть полноценно вручную обработаны экспертом и требуют программных расчетов и существенных вычислительных мощностей. С точки зрения практики экономических исследований указанные объемы данных могут считаться большими.

### **3.2. Расчет индикатора интенсивности инфляционных ожиданий российского населения**

#### **Составление системы правил отбора новостных статей СМИ и комментариев интернет-пользователей**

Второй и третий этапы построения индикатора интенсивности инфляционных ожиданий российского населения – отбор новостных статей СМИ, наиболее релевантных области экономики, измерение экономических ожиданий в которой является целью исследователя, и отбор комментариев, имеющих непосредственное отношение к заданному виду экономических ожиданий, среди всех комментариев интернет-пользователей к отобранным

статьям СМИ – осуществляются на базе выводов и предположений, сделанных в разделе 2.3.

Для второго этапа построения индикатора интенсивности инфляционных ожиданий областью экономики, измерение экономических ожиданий в которой является целью, или объект ожиданий, следует определить как «инфляция».

В соответствии с определением инфляции, приведенным в разделе 1.3 как «устойчивое повышение общего уровня цен на товары и услуги в течение определенного периода времени», а также логикой составления регулярных выражений, представленной в разделе 2.3, ключевыми словами, определяющими объект ожиданий или инфляцию в экономической теории, будут являться «инфляция» и «темп роста цен». При этом альтернативными определениями исследуемого объекта можно считать «ИПЦ», «индекс потребительских цен», «темп снижения цен», «темп падения цен», а также любые словосочетания, связанные с динамикой цен (снижение, повышение, рост). Все выбранные сочетания умеренно упрощают определение инфляции, с учетом стиля изложения в СМИ, поскольку цель составления правил – отбор всех новостных статей СМИ, так или иначе посвященных теме инфляции.

Таким образом, для второго этапа методики итоговые правила отбора статей СМИ в терминах регулярных выражений можно представить следующим образом:

- (1) (инфляц|ипц|(темп.\*цен);
- (2) ((рост|повыс|повыш|ниж|низ).\*цен);
- (3) (индекс.\*потребительских.\*цен).

Третий этап состоит в отборе комментариев интернет-пользователей, имеющих непосредственное отношение к инфляционным ожиданиям населения, среди всех комментариев интернет-пользователей к отобранным статьям СМИ.

Для работы с комментариями интернет-пользователей и выявления инфляционных ожиданий определение инфляционных ожиданий, представленное в разделе 1.3, умеренно упрощено. Сообщениями с упоминанием инфляционных ожиданий признаются комментарии, где упоминается повышение как инфляции, так и цен, в том числе на отдельные товары и услуги. Это необходимо для учета стиля написания комментариев в СМИ. Цель отбора комментариев интернет-пользователей – выявление комментариев, так или иначе упоминающих рост цен.

Таким образом, ростом инфляционных ожиданий является рост числа комментариев интернет-пользователей, упоминающих рост цен, на одну статью СМИ с упоминанием роста цен – среди всех статей на тему инфляции.

С учетом указанных предпосылок, правила для поиска комментариев должны иметь непосредственное отношение к инфляционным ожиданиям или к ожиданиям роста цен. Для инфляционных ожиданий – в соответствии с правилами, представленными в разделе 2.3, и сформированной системой правил для отбора новостных статей СМИ – ключевыми словами для отбора комментариев будут слова «инфляция» и «цены», встречающиеся в одном комментарии совместно со словами, относящимися к росту – «повышение», «рост» в разных формах (отглагольные существительные и глаголы, включающие «выс», «выш», «рост», «раст»).

Итоговые правила отбора третьего этапа в терминах регулярных выражений можно представить следующим образом: [«инфляц», «цен»] & [«выс», «выш», «рост», «раст»].

Полученные правила отбора второго и третьего этапов были экспертно верифицированы и оптимизированы автором с помощью итеративного процесса анализа качества полученных выборок новостных статей СМИ и комментариев после каждой итерации поиска в рассматриваемом наборе данных с января 2014 г. по февраль 2018 г. Важно отметить, что для других

наборов данных, содержащих иные источники информации, правила могут быть скорректированы.

Следует также отметить, что полученная выборка новостных статей СМИ на тему «инфляция» – после реализации второго этапа в соответствии с итоговыми правилами – также является относительно сбалансированной с точки зрения источников информации. Так, рассматриваемый набор данных содержит около 35 тыс. новостных статей СМИ. При этом около 35% из них принадлежат Коммерсанту, 30% опубликованы Экспертом, 23% размещены на сайте Ведомостей, и 12% статей приходится на РБК.

Все этапы отбора были осуществлены из общей выборки новостных статей СМИ и комментариев проекта Big Data Indicators (см. раздел 3.1) за период с января 2014 г. по февраль 2018 г. с помощью языка программирования Python. Скрипт по отбору данных представлен в Приложении 1.

#### **Вычисление индикатора интенсивности инфляционных ожиданий**

Расчет индикатора интенсивности инфляционных ожиданий был произведен в соответствии с формулой, указанной в разделе 2.4, а именно, как отношение числа комментариев, отфильтрованных как упоминающие инфляционные ожидания, к общему числу статей СМИ с упоминанием инфляционных ожиданий, аналогичным искомым в комментариях – среди новостных статей СМИ на тему «инфляция» – для каждого периода времени.

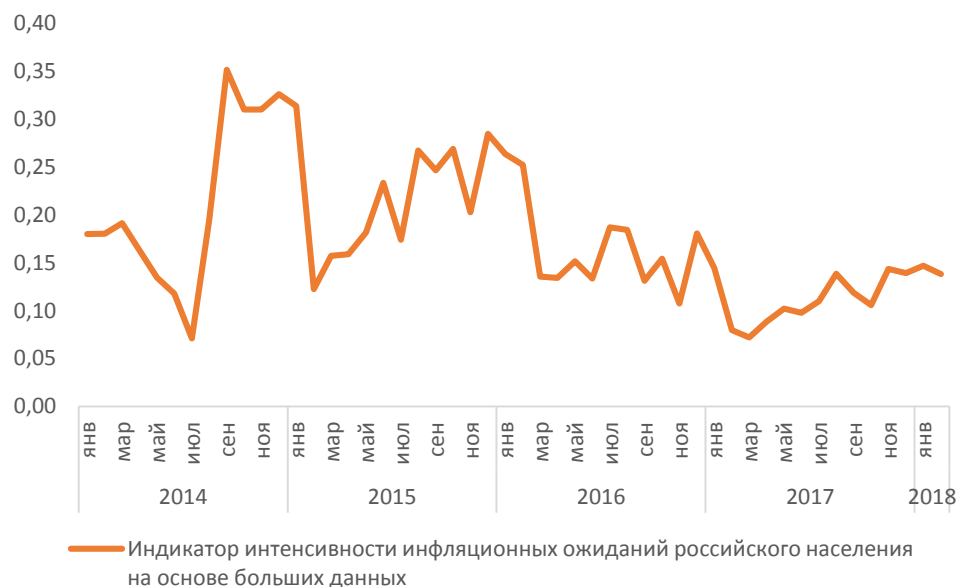
Для отбора статей СМИ и комментариев интернет-пользователей к ним были применены правила, составленные на втором и третьем этапах. В целях отбора статей СМИ с упоминанием инфляционных ожиданий были применены правила, идентичные составленным на третьем этапе для отбора комментариев интернет-пользователей, имеющих непосредственное отношение к инфляционным ожиданиям населения.

В качестве основной периодизации для построения динамики индикатора во времени используются ежемесячные и ежедневные данные.



При этом написанная программа-краулер для данных предполагает возможность построения индикатора с ежесекундной периодизацией.

Ежемесячный индикатор интенсивности инфляционных ожиданий российского населения представлен на рис. 6.



*Источник:* составлено автором.

Рис. 6. Индикатор интенсивности инфляционных ожиданий российского населения за период с января 2014 г. по февраль 2018 г., рассчитанный на основе больших наборов текстовых данных.

### **Оценка релевантности полученного индикатора**

В соответствии с предлагаемой в данной работе методикой построения высокочастотных индикаторов экономических ожиданий населения, оценка релевантности рассчитанного индикатора интенсивности инфляционных ожиданий осуществляется в две стадии.

На первой стадии необходимо провести сравнение динамики индикатора с существующими индикаторами инфляционных ожиданий населения, измеренными на основе других методов. Наиболее корректным представляется использование медианного индикатора численных ожиданий инфляции, рассчитываемого информационным агентством «инФОМ» (ФОМ) по официальному заказу Банка России на основе личных интервью домохозяйств в регионах России. Во-первых, индикаторы преследуют одну

цель: измерение инфляционных ожиданий российского населения. Указанный индикатор ФОМ является ключевым для регулятора показателем при вынесении суждений об инфляционных ожиданиях российского населения в рамках принятия решений о динамике ключевой ставки денежно-кредитной политики. Фактически, это единственный на текущий момент показатель, напрямую направленный на измерение инфляционных ожиданий российского населения на регулярной основе (ежемесячно). Во-вторых, аудитория, на основе мнений которой строятся индикаторы, также является аналогичной: это российское население, преимущественно не являющееся профессиональными участниками финансовых рынков.

Вместе с тем следует отметить и важные отличия в сравниваемых индикаторах. Аудиторией опросов для индикатора ФОМ является сбалансированная и репрезентативная для России выборка домохозяйств. Индикатор интенсивности инфляционных ожиданий на основе больших данных базируется на мнении интернет-пользователей, интересующихся экономическими новостями и готовыми самостоятельно высказывать собственные суждения о происходящем. Несмотря на то, что доступ домохозяйств в интернет в целом по России на текущий момент является высоким (около 70%, по данным Росстата), выборка интернет-пользователей, комментирующих экономические новости, не является полностью сбалансированной в целом по стране. Как указано в разделе 2.1, индикаторы экономических ожиданий на основе больших данных не ставят перед собой цели достижения полностью сбалансированной выборки по населению. Одна из гипотез для их построения состоит в том, что население, интересующееся экономическими событиями и высказывающее собственное мнение, оказывает значительное влияние на формирование инфляционных ожиданий российского населения в целом. Следовательно, динамика индикатора, разработанного в данном исследовании, и индикатора ФОМ должна быть сопоставимой.

Другое важное отличие состоит в периодах для формирования ожиданий. Медианный индикатор ФОМ направлен на измерение инфляционных ожиданий населения на срок 12 месяцев. При этом индикатор инфляционных ожиданий на основе больших данных не имеет четко заданного периода для измерения ожиданий. Однако, по оценкам автора на основе полученной выборки комментариев, это смешанные ожидания краткосрочного периода, примерно до 1 года. В действительности, поток медиа-информации относительно экономических событий и трендов в сети Интернет является масштабным и непрерывным, что способствует краткосрочному горизонту планирования при обсуждении экономических новостей. Более релевантным источником информации для прогнозов на долгосрочный период могут быть публикуемые прогнозы экспертного сообщества.

При выборе наиболее близкого индикатора-аналога следует также отметить расчетные оценки Банка России относительно динамики и уровня инфляционных ожиданий на основе данных отдельных вопросов ФОМ и предположениях о виде распределения ожиданий населения (равномерное или нормальное), а также индикатор инфляционных ожиданий профессиональных участников финансового рынка, рассчитанный для России [Жемков, Кузнецова, 2017].

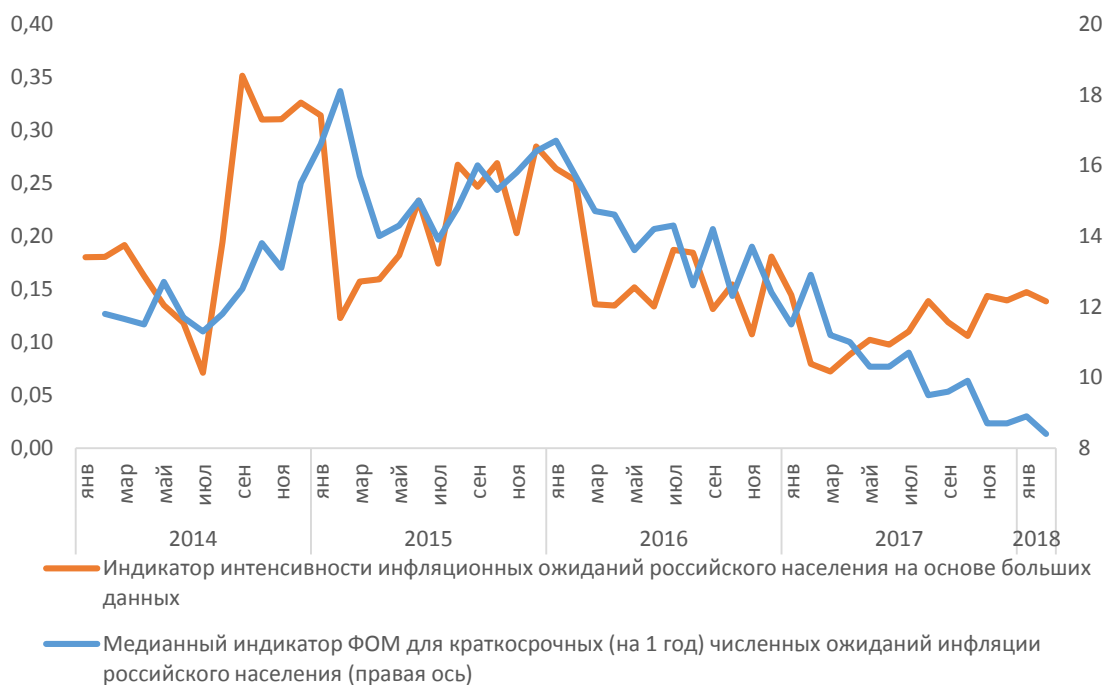
Оценки Банка России представляются менее релевантными для сравнения. Во-первых, гипотезы, лежащие в основе данных оценок одновременно с алгоритмом их расчета, предполагают высокое влияние текущей инфляции на итоговые индикаторы. В результате, динамика индикаторов практически полностью повторяет динамику текущей инфляции, что не представляется верным для понимания реальных инфляционных ожиданий населения и во многих случаях противоположно динамике медианного индикатора ФОМ. Высокая зависимость индикаторов инфляционных ожиданий, рассчитанных на основе предположений о распределении ответов домохозяйств о будущей динамике инфляции, от

фактических значений инфляции в текущий период, является одним из ключевых недостатков этого метода (см. раздел 1.3.2). Во-вторых, результаты данных оценок не используются самим Банком России для принятия решений о денежно-кредитной политике.

Что касается индикатора, вычисленного в работе Жемкова и Кузнецовой [Жемков, Кузнецова, 2017], то он также представляется в меньшей степени релевантным построенному индикатору интенсивности инфляционных ожиданий относительно индикатора ФОМ на основе опросов. Это обусловлено, во-первых, низкой развитостью биржевых инструментов с защитой от инфляции на российском финансовом рынке. В действительности, по данным авторов, в настоящий момент существует только один выпуск государственных облигаций с подобной опцией. Во-вторых, как было упомянуто выше, предполагается, что среди интернет-пользователей, читающих и комментирующих экономические новости, значительно более представленной является доля обычного населения, интересующегося экономическими событиями, относительно доли профессиональных участников финансового рынка. Для обсуждения ожиданий и стратегий, последние, как правило, используют специализированные сервисы.

Результаты сравнения динамики индикатора интенсивности инфляционных ожиданий на основе больших данных и медианного индикатора краткосрочных численных ожиданий инфляции ФОМ указывают на высокую и статистически значимую взаимосвязь индикаторов. Причем эта взаимосвязь оказывается наиболее существенной при предположении об опережающей динамике индикатора интенсивности инфляционных ожиданий над динамикой индикатора ФОМ на 2 месяца. При таком предположении значение корреляции возрастает с 52% до 72%, сохраняя статистическую значимость на уровне 1%, а критерий взаимной информации увеличивается с 0.24 до 0.36, повышая уровень статистической значимости с 20% до 5% (см. таблицу 3). В действительности, различия в методике

построения индикаторов также поддерживают гипотезу об опережающей динамике индикатора интенсивности инфляционных ожиданий на основе больших данных. Обработка текстовой информации в целях расчета последнего, а именно выгрузка и фильтрация новостных статей СМИ и комментариев интернет-пользователей и все расчеты, производится в реальном времени, в соответствии с датой и временем публикации комментариев в сети Интернет. Следовательно, получаемые значения индикатора за какой-либо период аппроксимируют интенсивность ожиданий населения именно за этот период, тогда как индикаторы на основе опросов могут отражать ожидания с определенным временным лагом, требующимся на сбор, верификацию и агрегацию ответов респондентов. По оценкам автора, на основе информации о методике проведения опросов ФОМ, размещенной на официальном сайте Банка России в разделе «Инфляционные ожидания и потребительские настроения населения», такой временной лаг может составлять от 2 недель и более.



Источник: составлено автором.

Рис. 7. Сравнение динамики индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных и динамики

медианного индикатора ФОМ для краткосрочных численных ожиданий инфляции российского населения.

Что касается визуального сопоставления индикаторов, то здесь следует отметить высокую сопоставимость их динамики в период с начала 2014 г. до середины 2017 г. (см. рис. 7). При этом с мая 2017 г. – момента временного разворота тренда инфляции к снижению из-за негативного эффекта аномально-холодных погодных условий на урожай сельскохозяйственной продукции – индикатор интенсивности роста цен существенно и устойчиво превышает индикатор ФОМ. Реакция интернет-пользователей на временный перелом тренда к замедлению инфляции оказалась масштабнее и не перешла к снижению до конца рассматриваемого периода, несмотря на возобновление тенденции к снижению индикатора ФОМ после его повышения в середине 2017 г. Таким образом, повышенная с мая 2017 г. интенсивность инфляционных ожиданий интернет-пользователей, читающих и комментирующих экономические новости, указывает на ожидания ускорения инфляции в 2018 г. Это соответствует консенсусу аналитического сообщества (VTV Capital, Bloomberg и др.) и самого регулятора относительно инфляционной динамики 2018 г., который заключается в ожиданиях ускорения ИПЦ во втором полугодии 2018 г.

Таблица 3. Статистические критерии взаимосвязи динамики индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных и динамики медианного индикатора ФОМ для краткосрочных численных ожиданий инфляции российского населения

Статистический критерий взаимосвязи		Корреляция с индикатором ФОМ	Взаимная информация с индикатором ФОМ
Интенсивность инфляционных ожиданий российского населения на основе больших данных	Лag = 0	0.52****	0.24*
	Лag = 1	0.67****	0.22
	Лag = 2	0.72****	0.36***
	Лag = 3	0.70****	0.27**
	Лag = 4	0.69****	0.22
	Лag = 5	0.61****	0.24*
	Лag = 6	0.48****	0.16

*Примечания:* использованы следующие обозначения для различных уровней значимости:  
\* – 20%, \*\* – 15%, \*\*\* – 5%, \*\*\*\* – 1%.

*Источник:* составлено автором.

Таким образом, в соответствии с первой стадией оценки релевантности индикаторов цели расчета, подтверждается релевантность полученного индикатора интенсивности инфляционных ожиданий российского населения заданной цели – измерению инфляционных ожиданий.

Вторая стадия оценки релевантности построенного индикатора предполагает сравнение его динамики с динамикой целевого экономического параметра, ожидания в области которого измеряет исследователь. Для инфляционных ожиданий таким индикатором является потребительская инфляция. В частности, представляется корректным выбрать для сравнения показатель, на который ориентируется Банк России при реализации режима инфляционного таргетирования, а именно общий базисный индекс потребительских цен (ИПЦ), по данным Росстата. Последний обозначает ежемесячные темпы роста цен к соответствующему месяцу предшествующего года.

Сравнение динамики индикатора интенсивности инфляционных ожиданий и динамики ИПЦ позволяет сделать вывод о значимой взаимосвязи показателей (см. рис. 8), однако не настолько существенной как в случае первого шага. При этом здесь также не отвергается гипотеза об опережающей динамике индикатора инфляционных ожиданий, что соответствует экономической теории, указывающей на значительное воздействие инфляционных ожиданий населения на значения и динамику будущей инфляции. В соответствии с различными статистическими критериями, временной лаг, при котором наблюдается наибольшая взаимосвязь показателей, является различным. Корреляция является максимальной при лаге в 3-4 месяца на 1% уровне значимости, а критерий взаимной информации предполагает опережение ожиданий над инфляцией в размере 8-9 месяцев на 10% уровне значимости (см. таблицу 4). Такая ситуация, с одной стороны, указывает на неопределенность в структуре

взаимосвязи между индикатором интенсивности инфляционных ожиданий и инфляцией, с другой – позволяет сделать вывод о наличии статистически значимой взаимосвязи между показателями.

Кроме того, дополнительно был проведен тест Гренджера для проверки причинно-следственной зависимости между индикатором интенсивности инфляционных ожиданий и инфляцией. Результаты теста не позволяют отвергнуть гипотезу о наличии отложенного во времени эффекта изменения инфляционных ожиданий на уровень инфляции для временных лагов воздействия 10-12 месяцев на 10% (для 10 месяцев) и 5% уровне значимости (для 11-12 месяцев). Это, с одной стороны, подтверждает наличие значимой взаимосвязи между рассматриваемыми показателями, с другой – соответствует теоретическим предпосылкам о влиянии инфляционных ожиданий населения на будущий уровень потребительской инфляции в стране. При этом дополнительно подтверждается авторская оценка периода инфляционных ожиданий в рамках рассматриваемого набора данных, как смешанных ожиданий краткосрочного периода до 1 года (см. стр. 115).

Совокупные результаты проведенных расчетов подтверждают положительную оценку релевантности индикатора интенсивности инфляционных ожиданий поставленной цели измерения инфляционных ожиданий населения в рамках разработанной методики их построения.

Таким образом, положительный результат двух стадий оценки позволяет сделать вывод о релевантности полученного индикатора интенсивности инфляционных ожиданий российского населения заданной исследователем цели – измерению инфляционных ожиданий.





Источник: составлено автором.

Рис. 8. Сравнение динамики индикатора интенсивности инфляционных ожиданий российского населения и динамики потребительской инфляции.

Таблица 4. Статистические критерии взаимосвязи динамики индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных и динамики базисного индекса потребительских цен на основе данных Росстата

Статистический критерий взаимосвязи		Корреляция с базисным ИПЦ	Взаимная информация с базисным ИПЦ	Тест Гренджера для эффекта на базисный ИПЦ
Интенсивность инфляционных ожиданий российского населения на основе больших данных	Лар = 0	0.51****	0.26*	0.54
	Лар = 1	0.55****	0.28*	0.88
	Лар = 2	0.59****	0.27*	0.87
	Лар = 3	0.62****	0.24	0.87
	Лар = 4	0.63****	0.22	0.19
	Лар = 5	0.58****	0.22	0.25
	Лар = 6	0.52****	0.21	0.52
	Лар = 7	0.45****	0.18	0.37
	Лар = 8	0.40****	0.40**	0.54
	Лар = 9	0.37***	0.27*	0.69
	Лар = 10	0.37***	0.22	0.06**
	Лар = 11	0.33***	0.12	0.04****
	Лар = 12	0.24*	0.18	0.05****

*Примечания:*

(а) использованы следующие обозначения для различных уровней значимости: \* – 15%, \*\* – 10%, \*\*\* – 5%, \*\*\*\* – 1%;

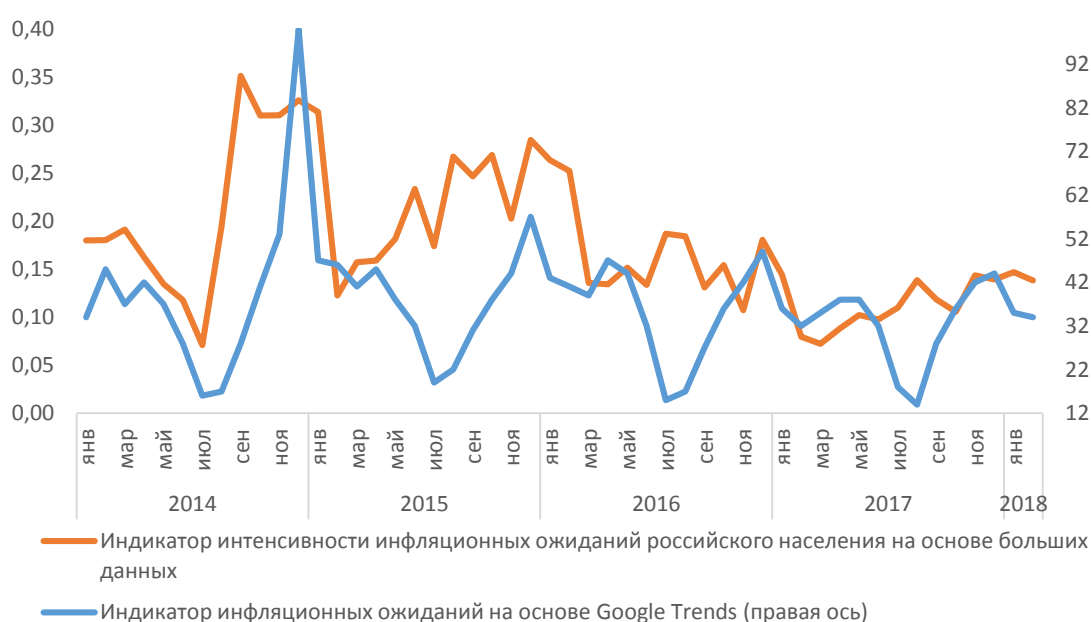
(б) В качестве результатов теста Гренджера приведены значения вероятностей принятия нулевой гипотезы об отсутствии причинно-следственной связи интенсивности инфляционных ожиданий и инфляции. Ряды приведены к стационарной форме с помощью вычисления первых разностей. Для проверки стационарности использованы тесты Дики-Фуллера и тест KPSS.

*Источник:* составлено автором.

В качестве дополнительного контроля релевантности динамики результирующего индикатора интенсивности инфляционных ожиданий может быть осуществлено сравнение с аналогичным по тематике индикатором, сконструированным с помощью платформы Google Trends. Возможность подобного сопоставления была рассмотрена в разделе 1.2.3. Однако в результате отсутствия на текущий момент общепризнанных индикаторов экономических ожиданий, построенных на основе этой платформы, как в зарубежной, так и в российской практике, сопоставление с индикаторами-аналогами, сконструированными с помощью Google Trends, не было включено отдельным этапом в методику построения индикаторов интенсивности экономических ожиданий населения.

Вместе с тем в целях достижения максимально полной оценки релевантности полученного индикатора интенсивности ожиданий в данной работе с помощью платформы Google Trends был сконструирован ежемесячный индикатор, отражающий частоту русскоязычных запросов «инфляция» в поисковой системе Google на территории России (см. рис. 9). Сопоставление динамики индикатора интенсивности инфляционных ожиданий, рассчитанного в данном разделе, и динамики индикатора на основе Google Trends позволяет сделать вывод о преимуществах построения индикатора инфляционных ожиданий на базе больших данных, представленного в данной работе. С одной стороны, динамика индикаторов является сходной на протяжении рассматриваемого временного периода – с января 2014 г. по февраль 2018 г. С другой стороны, явно прослеживается опережающая динамика индикатора интенсивности ожиданий: при

предположении о периоде опережения, равном 3 месяцам, коэффициент корреляции возрастает с 40% до 60%. В то же время визуальный анализ показывает, что кризисная ситуация в экономике конца 2014 г., которая привела к резкому росту потребительской инфляции, была отмечена индикатором интенсивности примерно на 3 месяца раньше – еще в сентябре 2014 г. Кроме того, следует отметить более волатильную динамику индикатора на основе Google Trends относительно индикатора интенсивности и его ярко выраженную сезонность. Другой локальный рост инфляционных ожиданий, происходивший на фоне очередного падения цен на нефть во второй половине 2015 г. не прослеживается в явном виде для индикатора Google Trends, вследствие одновременного сезонного роста этого показателя, тогда как для индикатора интенсивности пик роста инфляционных ожиданий в конце 2015 г. оказался близок к максимальным значениям конца 2014 г.



Источник: составлено автором.

Рис. 9. Сравнение динамики индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных и динамики индикатора, построенного с помощью платформы Google Trends.

## **Оценка устойчивости результатов расчета индикатора**

Проведение тестов на устойчивость результатов расчета индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных для рассматриваемого набора данных включает две стадии.

Первая стадия – это оценка чувствительности результатов построения индикатора в зависимости от сформированных на втором и третьем этапах правил отбора новостных статей СМИ и комментариев интернет-пользователей. Таблица вариаций правил отбора для индикатора инфляционных ожиданий в качестве примера была сформирована в разделе 2.4.2 – при описании методики построения индикаторов экономических ожиданий (см. таблицу 2). В соответствии со всеми вариациями правил отбора, представленными в этой таблице, были рассчитаны индикаторы инфляционных ожиданий (см рис. 10). Их визуальное сопоставление позволяет сделать вывод о сходной динамике индикаторов, устойчивости и сопоставимости трендов, представленных в полученных рядах. Средний коэффициент попарной корреляции всех рассматриваемых рядов составляет 70%. При этом подавляющее большинство корреляционных зависимостей (около 90%) является статистически значимым на уровне 1%. Таким образом, результаты первой стадии свидетельствуют об устойчивости результата к правилам отбора новостных статей СМИ и комментариев интернет-пользователей для расчета индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных.

Вторая стадия оценки устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий в рассматриваемом наборе данных заключается в оценке чувствительности данного индикатора к изменению области экономических ожиданий. Для проведения этого теста были сконструированы два альтернативных индикатора интенсивности для таких областей экономических ожиданий, как ожидания ухудшения ситуации на рынке труда и ожидания ослабления курса национальной валюты. Правила отбора новостных статей СМИ и комментариев интернет-пользователей для

этих областей экономических ожиданий были рассмотрены в разделе 2.3 в таблице 1. Результаты построения индикаторов и их сопоставление с динамикой индикатора интенсивности инфляционных ожиданий представлены на рис. 11.

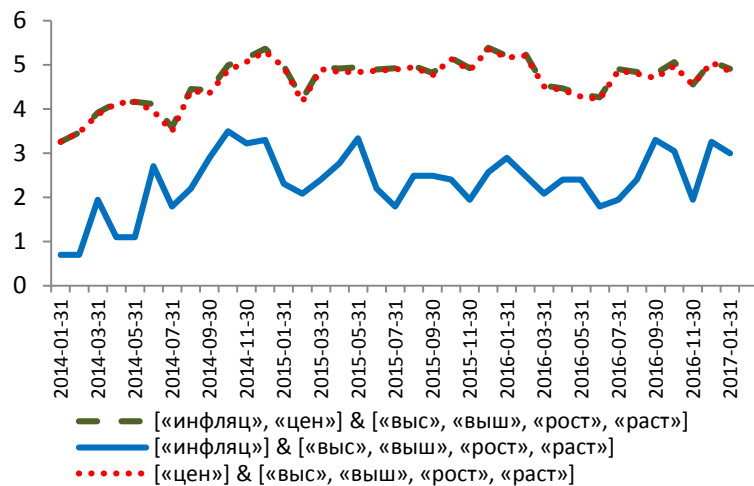
Сопоставление динамики полученных индикаторов позволяет сделать вывод о достаточном уровне чувствительности результатов построения индикаторов интенсивности к выбору различных областей экономических ожиданий. Полученные индикаторы характеризуются сопоставимой динамикой, но не идентичной и не близкой к идентичной. Между всеми построенными индикаторами не наблюдается крайне высокой и устойчивой взаимозависимости. Другими словами, ненулевая и значительная взаимосвязь построенных индикаторов интенсивности является адекватной с учетом существенной взаимосвязи динамики инфляции, курса валюты и состояния рынка труда в экономической теории. Значения корреляции варьируются от 30% до 73%, показывая наиболее значительную взаимосвязь для ожиданий обесценения курса национальной валюты и инфляционных ожиданий. Это соответствует фактическим макроэкономическим зависимостям в российской экономике: корреляция темпов изменения курса рубля к доллару и потребительской инфляции за аналогичный период времени (с января 2014 г. по февраль 2018 г.) приближается к 90%<sup>10</sup>. При этом также следует отметить, что в пользу отсутствия чрезмерно высокой и устойчивой взаимозависимости свидетельствует отсутствие статистической значимости попарных критериев взаимной информации между построенными индикаторами интенсивности (см. таблицу 5).

---

<sup>10</sup> Рассчитано автором для показателей курса доллара к рублю и цен на нефть Brent в терминах темпов роста в месяц к соответствующему месяцу предшествующего года.

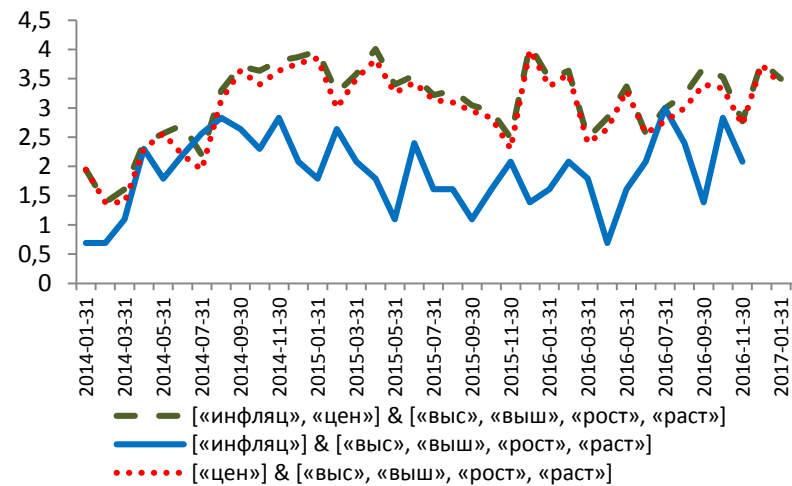
1) Правила отбора новостных статей СМИ на тему «инфляция»: (1) темп.\*цен;

(2) (рост|повыс|повыш|ниж|низ).\*цен.



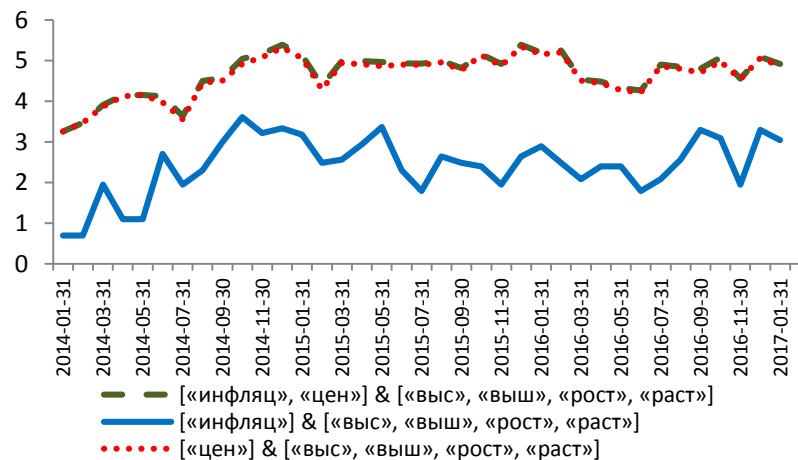
2) Правила отбора новостных статей СМИ на тему «инфляция»: (1) инфляц|ипц; (2)

(индекс.\*потребительских.\*цен )



3) Правила отбора новостных статей СМИ на тему «инфляция»: (1) инфляц|ипц|(темп.\*цен);

(2) (рост|повыс|повыш|ниж|низ).\*цен; (3) (индекс.\*потребительских.\*цен )



*Примечание:*

(1) Каждый рисунок представляет собой одну из возможных вариаций правил отбора новостных статей СМИ для оценки устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий, включая итоговый вариант, в соответствии с таблицей 2;

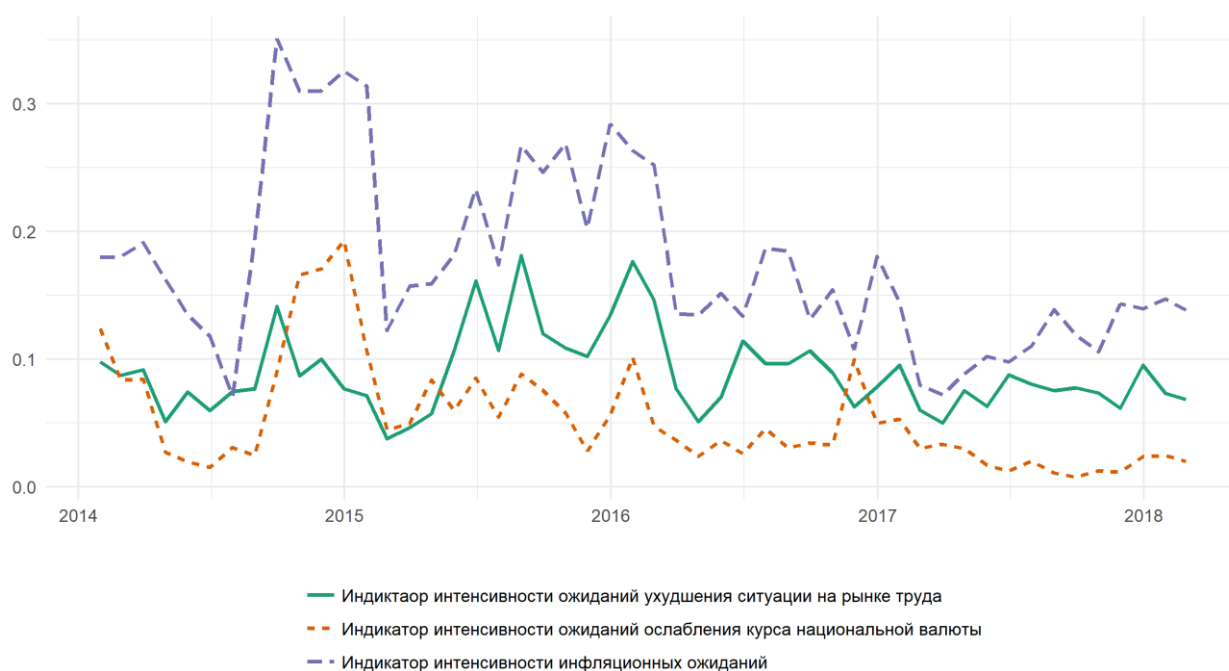
(2) Каждая линия на рисунке обозначает одну из возможных вариаций правил отбора комментариев интернет-пользователей, упоминающих ожидания роста цен, к выбранным новостным статьям СМИ, в соответствии с таблицей 2.

(3) В целях более удобного сопоставления динамики результирующих индикаторов, ось ординат на каждом рисунке представляет собой логарифмическую шкалу:  $\log(1 + \text{Число комментариев интернет-пользователей, упоминающих ожидания инфляции})$ .

(4) Тест на устойчивость результатов расчета индикатора интенсивности инфляционных ожиданий для различных вариаций правил отбора проводится для ежемесячных значений индикаторов – в целях повышения эффективности расчетов.

Рис. 10. Результаты первой стадии оценки устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных.

Это, с одной стороны, свидетельствует об устойчивости результатов расчета к выбранным правилам отбора статей и комментариев, включая их содержательную релевантность для измерения инфляционных ожиданий населения. С другой стороны, это показывает отсутствие негативной специфики в рассматриваемом наборе данных новостей СМИ и комментариев пользователей, которая могла бы ограничивать возможности расчетов индикаторов интенсивности для различных областей экономических ожиданий.



*Источник:* составлено автором.

Рис. 11. Результаты второй стадии оценки устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий: сопоставление динамики индикатора с альтернативными индикаторами для различных областей экономических ожиданий российского населения.

Таблица 5. Статистические критерии взаимосвязи динамики индикаторов интенсивности ожиданий российского населения на основе больших данных в области инфляции, обесценения курса национальной валюты и ухудшения ситуации на рынке труда

	Инфляционные ожидания	Ожидания обесценения курса рубля	Ожидания ухудшения ситуации на рынке труда
Инфляционные ожидания	-	К: 0.73**** ВИ: 0.18	К: 0.58**** ВИ: 0.18
Ожидания обесценения курса рубля	К: 0.73**** ВИ: 0.18	-	К: 0.30*** ВИ: 0.10
Ожидания ухудшения ситуации на рынке труда	К: 0.58**** ВИ: 0.18	К: 0.30*** ВИ: 0.10	-

*Примечания:* использованы следующие обозначения:

(а) для различных уровней значимости: \*\*\* – 5%, \*\*\*\* – 1%;

(б) для различных статистических критериев взаимосвязи: К – коэффициент корреляции, ВИ – критерий взаимной информации.

*Источник:* составлено автором.



Таким образом, на основе результатов двух стадий оценки можно сделать положительный вывод об устойчивости результатов расчета индикатора интенсивности инфляционных ожиданий – в рамках рассматриваемого набора данных за период с января 2014 г. по февраль 2018 г. Следует отметить, что обе стадии оценки дали положительные результаты, следовательно, разработанные правила отбора новостных статей СМИ и комментариев интернет-пользователей, а также составленный набор данных могут быть использованы для построения индикатора инфляционных ожиданий российского населения – в рамках предлагаемой в данной работе методики.

### **3.3. Выявление факторов изменения интенсивности инфляционных ожиданий населения во времени**

Алгоритм выявления факторов изменения интенсивности экономических ожиданий населения предполагает получение комплексных выводов на основе трех шагов. В случае инфляционных ожиданий такими шагами являются:

- оценка зависимости построенного индикатора интенсивности инфляционных ожиданий от различных экономических параметров, способных выступать факторами изменения инфляционных ожиданий в экономической теории;
- экспертное выявление факторов, с которыми население связывает изменение собственных инфляционных ожиданий, на основе анализа содержания комментариев интернет-пользователей;
- выявление факторов, с которыми население связывает изменение собственных инфляционных ожиданий на основе методов машинного обучения и текстовой обработки.

## **Оценка зависимости интенсивности инфляционных ожиданий российского населения от различных экономических параметров**

Полученный индикатор интенсивности инфляционных ожиданий в большей степени направлен на измерение именно ожиданий инфляции и, соответственно, может быть связан не только с динамикой самой потребительской инфляции, но и с ее ключевыми факторами. Кроме того, среди возможных факторов, способных оказать влияние на изменение инфляционных ожиданий населения, следует также выделить показатели, имеющие определяющее значение для проведения денежно-кредитной политики, вследствие тесной связи между инфляционными ожиданиями и характеристиками реализации политики инфляционного таргетирования.

На основе указанных предположений был сформирован следующий список показателей: индекс потребительских цен, цены на нефть, курс доллара к рублю, ключевая ставка Банка России, индекс промышленного производства, уровень безработицы, реальная заработная плата и реальные располагаемые доходы населения. Для выбранных показателей предполагаются следующие механизмы влияния на инфляционные ожидания населения:

- Индекс потребительских цен оказывает влияние на инфляционные ожидания населения при предположении об адаптивном механизме их формирования или о его значительном вкладе в принятие решений домохозяйствами (см. раздел 1.2). Существенная чувствительность ожиданий населения по инфляции к текущим значениям инфляции и их изменениям регулярно упоминается Банком России в официальных пресс-релизах о денежно-кредитной политике;
- Цены на нефть и курс национальной валюты сильно взаимосвязаны в условиях высокой экспортно-сырьевой ориентированности российской экономики [Blokhina, Karpenko, Guirinskiy, 2016] и оказывают значительное воздействие на большинство внутренних

экономических процессов, включая планирование потребительского и финансового поведения со стороны домохозяйств [Sibanda, Hove, Murwirapachena, 2015], а также динамику самой инфляции [Siok Kun Seka, Xue Qi Teoa, Yen Nee Wonga, 2015; ALsaedi, 2015]. В связи с этим, предположение о существенном влиянии цен на нефть на интенсивность инфляционных ожиданий населения представляется логичным;

- Ключевая ставка Банка России является основным инструментом денежно-кредитной политики при инфляционном таргетировании. При этом при реализации данного режима корректное управление инфляционными ожиданиями населения с помощью этого инструмента является одной из главных задач монетарной политики, направленной на достижение цели по инфляции и макроэкономической стабильности [Lewis, McDermott, Richardson, 2016];
- Индекс промышленного производства является одним из главных индикаторов экономической активности страны и учитывается центральным банком при принятии решений о денежно-кредитной политике. Уровень безработицы, в свою очередь, отражает как ситуацию на рынке труда, так и общее состояние экономики. Эти индикаторы также могут приниматься населением во внимание при вынесении суждений о будущей инфляции. В особенности при предположении о рациональном механизме формирования ожиданий;
- Реальная заработная плата и реальные располагаемые доходы населения отражают совокупное финансовое состояние домохозяйств, которое может оказывать существенное воздействие на внимание последних к инфляционным процессам в экономике. К примеру, в условиях падения доходов население может более остро воспринимать инфляционное давление, реагируя даже на

кратковременные шоки повышением инфляционных ожиданий, как это происходило в мае-июле 2017 г. Кратковременное и небольшое по масштабу ускорение инфляции с 4.1% до 4.4% г/г, вследствие аномально-холодных погодных условий апреля-июня 2017 г. и на фоне продолжительного падения реальных располагаемых доходов населения (с середины 2014 г.), привело к резкому изменению инфляционных ожиданий населения в середине 2017 г. как в соответствии с индикатором ФОМ, так и с индикатором интенсивности инфляционных ожиданий, рассматриваемом в данной работе. В дополнение, практические наблюдения для российской экономики показывают, что одним из факторов изменения ожиданий населения по инфляции может быть индексация заработных плат и пенсий государством [Карпов, 2017].

Построение на помесечном уровне корреляций и критериев взаимной информации для индикатора интенсивности инфляционных ожиданий и сформированного набора макроэкономических индикаторов позволяет сделать выводы о ключевых взаимодействиях между этими показателями (см. таблицу 6). В целях проведения более полноценного исследования взаимосвязей и проверки возможности отложенного во времени воздействия факторов на инфляционные ожидания статистические критерии были рассчитаны как для исходных рядов факторов, так и для временных лагов от одного до двенадцати месяцев. Выводы о степени взаимосвязанности рядов были сделаны на основе максимизации критериев корреляции и взаимной информации с учетом их статистической значимости.

Наиболее заметными и статистически подтверждёнными по результатам анализа являются взаимозависимости индикатора интенсивности инфляционных ожиданий с текущими значениями инфляции, темпами роста цен на нефть и курса рубля к доллару, а также изменением ключевой ставки Банка России с примерно полугодовым отложенным воздействием. Причем для первых трех факторов знаки зависимостей соответствуют

предполагаемым механизмам воздействия на инфляционные ожидания. Рост текущих значений инфляции связан с повышением инфляционных ожиданий так же, как и снижение цен на нефть и ослабление курса доллара к рублю, вследствие общего ухудшения экономической ситуации. Вместе с тем для эффекта изменений ключевой ставки положительный знак взаимосвязи является спорным. Реализация режима инфляционного таргетирования в общем случае предполагает отрицательную взаимосвязь изменений ключевой ставки и инфляционных ожиданий населения. Рост ключевой ставки по истечении определенного периода времени – при условии высокого доверия к проводимой регулятором денежно-кредитной политике – приводит к снижению и нормализации уровня инфляции, что при прочих равных должно вести к снижению ожиданий ускорения роста цен [Апокин и др., 2014]. Однако на рассматриваемом периоде наблюдается устойчивая положительная корреляция между изменениями ключевой ставки Банка России и интенсивностью инфляционных ожиданий. В случае, если это не является ложной взаимосвязью, обусловленной, к примеру, воздействием других общих для этих показателей внешних факторов, можно предположить следующий механизм воздействия: повышение ключевой ставки сигнализирует населению об ожиданиях роста инфляционного давления со стороны регулятора. Соответственно, в условиях пока еще недостаточно высокого уровня доверия к действиям центрального банка и его способностям управлять инфляцией население может повышать собственные инфляционные ожидания вслед за регулятором.

Второстепенными по значимости и корректными по знаку в рамках ожиданий являются взаимосвязи интенсивности инфляционных ожиданий и индекса промышленного производства и изменения уровня безработицы – с возможным отложенным эффектом в пределах одного квартала. Ухудшение общей экономической ситуации, отражающееся в повышении уровня безработицы и падении индекса промышленного производства, связано с повышением инфляционных ожиданий населения. Следует также отметить

значимую отрицательную взаимосвязь темпов роста реальной заработной платы и инфляционных ожиданий. С точки зрения простых критериев взаимосвязи – без проверки причинно-следственных связей и возможного эффекта других внешних факторов – снижение доходов является для населения негативным сигналом, способствующим повышению инфляционных ожиданий. Снижение доходов может способствовать повышению внимания к инфляционным процессам.

В целях проверки гипотез о причинно-следственных связях между выделенными факторами и интенсивностью инфляционных ожиданий, в соответствии с алгоритмом, описанным в разделе 2.5.1, был проведен тест Гренджера. Указанный тест был рассчитан для рядов, определенных как взаимосвязанные с индикатором интенсивности инфляционных ожиданий по результатам статистического анализа. При этом все ряды были приведены к стационарной форме с помощью вычисления первых разностей. Стационарность рядов и порядок для вычисления разностей верифицировались при помощи тестов Дики-Фуллера и KPSS [Banerjee et al., 1993; Kwiatkowski et al., 1992]. Проверка наличия отложенных во времени причинно-следственных зависимостей была осуществлена для временных лагов от 1 до 12 месяцев, обозначающих порядок проведения теста. Более расширенный период времени представляется малоинформативным, вследствие малого числа доступных данных: общее максимальное число наблюдений для рядов в первых разностях составило 49, с февраля 2014 г. по февраль 2018 г.

Таблица 6. Статистические критерии взаимосвязи индикатора интенсивности инфляционных ожиданий  
российского населения на основе больших данных с предполагаемыми факторами его изменений по ежемесячным  
данным

Временной лаг для факторов	Индекс потребитель ских цен (г/г, %)	Цены на нефть Brent (долл./бар.)	Курс рубля к доллару (рублей за доллар)	Темп роста цен на нефть Brent	Темп роста курса рубля к доллару	Изменение ключевой ставки Банка России (п.п.)	Индекс промышленного производства (г/г, %)	Уровень безработицы (% к ЭАН)	Изменение уровня безработицы (г/г, п.п.)	Темп роста реальной зарплатной платы (% г/г)	Темп роста реальных располагаемых доходов (%, г/г)
Исходный ряд	<b>K:0.51****</b> <b>V:0.29</b>	K:0.02 V:0.12	K:0.03 V:0.14	<b>K:-0.57****</b> <b>V:0.25</b>	<b>K:0.60****</b> <b>V:0.27*</b>	K:0.4**** V:0.14	K:-0.18 V:0.08	K:0.09 V:0.11	K:0.30*** V:0.18	<b>K:-0.47****</b> <b>V:0.26*</b>	K:-0.05 V:0.22
1 мес.	K:0.48**** V:0.28	K:0.15 V:0.09	K:-0.14 V:0.13	K:-0.55**** V:0.29*	K:0.51**** V:0.26	K:0.35*** V:0.12	K:-0.28*** V:0.15	K:-0.04 V:0.09	<b>K:0.32***</b> <b>V:0.28*</b>	K:-0.46**** V:0.23	K:0.01 V:0.12
2 мес.	K:0.48**** V:0.21	K:0.26** V:0.09	K:-0.26** V:0.12	K:-0.5**** V:0.23	K:0.45**** V:0.22	K:0.39*** V:0.17**	<b>K:-0.35****</b> <b>V:0.12</b>	K:-0.12 V:0.14	K:0.30*** V:0.25	K:-0.44**** V:0.16	K:0.13 V:0.13
3 мес.	K:0.48**** V:0.25	K:0.33*** V:0.11	K:-0.31*** V:0.13	K:-0.43**** V:0.17	K:0.42**** V:0.25	K:0.46**** V:0.29***	K:-0.3*** V:0.15	K:-0.24** V:0.20	K:0.27** V:0.27*	K:-0.42**** V:0.23	K:0.13 V:0.11
4 мес.	K:0.46**** V:0.27	K:0.35*** V:0.15	K:-0.32*** V:0.16	K:-0.36*** V:0.18	K:0.41**** V:0.25	K:0.53**** V:0.22***	K:-0.3** V:0.13	K:-0.29*** V:0.10	K:0.24* V:0.23	K:-0.35**** V:0.16	K:0.23* V:0.10
5 мес.	K:0.44**** V:0.23	K:0.35*** V:0.14	K:-0.31*** V:0.13	K:-0.33*** V:0.20	K:0.42**** V:0.27*	K:0.6**** V:0.24***	K:-0.24 V:0.09	K:-0.23* V:0.09	K:0.19 V:0.20	K:-0.32*** V:0.19	K:0.19 V:0.13
6 мес.	K:0.40**** V:0.18	K:0.36**** V:0.19	K:-0.31*** V:0.14	K:-0.28*** V:0.16	K:0.41**** V:0.27*	K:0.62**** V:0.21**	K:-0.14 V:0.10	K:-0.10 V:0.08	K:0.17 V:0.15	K:-0.24** V:0.20	K:0.08 V:0.10
7 мес.	K:0.35**** V:0.14	K:0.37**** V:0.14	K:-0.33*** V:0.14	K:-0.24** V:0.15	K:0.39**** V:0.21	K:0.67**** V:0.22**	K:-0.11 V:0.07	K:0 V:0.03	K:0.12 V:0.17	K:-0.17 V:0.15	K:0.14 V:0.12
8 мес.	K:0.29*** V:0.17	K:0.38**** V:0.18	K:-0.35*** V:0.14	K:-0.22* V:0.11	K:0.36*** V:0.19	<b>K:0.73****</b> <b>V:0.24**</b>	K:-0.03 V:0.06	K:0.10 V:0.12	K:0.03 V:0.11	K:-0.13 V:0.13	K:0.15 V:0.09
9 мес.	K:0.22*** V:0.16	K:0.39**** V:0.14	K:-0.38**** V:0.13	K:-0.17 V:0.10	K:0.29*** V:0.17	K:0.73**** V:0.22**	K:0 V:0.05	K:0.16 V:0.06	K:0 V:0.08	K:-0.09 V:0.14	K:0.05 V:0.12
10 мес.	K:0.14* V:0.13	K:0.41**** V:0.19	K:-0.39**** V:0.15	K:-0.13 V:0.15	K:0.24** V:0.16	K:0.71**** V:0.19*	K:0.13 V:0.05	K:0.06 V:0.08	K:-0.12 V:0.11	K:-0.03 V:0.12	K:0.22* V:0.10
11 мес.	K:0.06 V:0.14	K:0.44**** V:0.15	K:-0.41**** V:0.11	K:-0.05 V:0.10	K:0.18 V:0.18	K:0.67**** V:0.21**	K:0.21* V:0.09	K:-0.07 V:0.11	K:-0.22 V:0.20	K:0.06 V:0.09	K:0.29*** V:0.14
12 мес.	K:-0.03 V:0.14	K:0.49**** V:0.14	K:-0.47**** V:0.16	K:0.04 V:0.10	K:0.08 V:0.12	K:0.65**** V:0.19**	K:0.22* V:0.11	K:-0.26** V:0.10	K:-0.37*** V:0.20	K:0.13 V:0.10	K:0.32*** V:0.13

*Примечания:* использованы следующие обозначения:

- (а) для различных уровней значимости: \* – 15%, \*\* – 10%, \*\*\* – 5%, \*\*\*\* – 1%;
- (б) для различных статистических критериев взаимосвязи: К – коэффициент корреляции, В – критерий взаимной информации;
- (в) значение каждого критерия получено для рядов индикатора интенсивности инфляционных ожиданий и каждого предполагаемого фактора его изменений в период с января 2014 г. по февраль 2018 г. Каждый фактор обозначен в названиях столбцов таблицы. В каждой строке таблицы факторы используются с определенным, указанным в первом столбце временным лагом;
- (г) наиболее значимые взаимосвязи выделены жирным шрифтом и цветом, второстепенные взаимосвязи – курсивом и цветом.

*Источник:* составлено автором.

В то же время для дополнения анализа и верификации устойчивости полученных результатов для стационарных рядов был также рассчитан модифицированный для векторных авторегрессионных моделей (VAR) робастный (устойчивый) тест Гренджера. Данный тест был проведен для простых VAR моделей попарных зависимостей каждого фактора и индикатора интенсивности инфляционных ожиданий. При осуществлении модифицированного теста Гренджера использовалась робастная, устойчивая к гетероскедастичности, оценка матрицы ковариаций коэффициентов регрессий [Cribari-Neto, Da Silva, 2011]. При этом указанный тест осуществлялся только для оптимального лага каждой VAR-модели попарной зависимости, определенного с помощью информационного критерия Акаике [Akaike, 1974]. Таким образом, для каждой предполагаемой причинно-следственной зависимости интенсивности инфляционных ожиданий от экономических параметров была оценена одна векторная авторегрессия с оптимальным лагом и на основе нее рассчитан модифицированный робастный тест Гренджера.



Таблица 7. Результаты теста Гренджера на причинно-следственную зависимость индикатора интенсивности инфляционных ожиданий от различных экономических параметров по ежемесячным данным

Порядок для проведения теста Гренджера (временной лаг)	Индекс потребительских цен (г/г, %)	Темп роста цен на нефть Brent	Темп роста курса рубля к доллару	Изменение ключевой ставки Банка России (п.п.)	Индекс промышленного производства (г/г, %)	Изменение уровня безработицы (г/г, п.п.)	Темп роста реальной заработной платы (% г/г)
1 мес.	<b>0.20*</b>	0.32 <b>VAR: 0.19*</b>	0.48 VAR: 0.62	<b>0.04***</b> <b>VAR: 0.20*</b>	0.74 VAR: 0.97	0.89 VAR: 0.78	0.96 VAR: 0.93
2 мес.	0.25 VAR: 0.63	0.64	0.56	0.15**	0.85	0.99	0.29
3 мес.	0.53	0.81	0.80	0.36	0.93	0.80	0.57
4 мес.	0.72	0.83	0.79	0.56	0.98	0.83	0.66
5 мес.	0.81	0.72	0.92	0.75	0.98	0.30	0.76
6 мес.	0.92	0.78	0.91	0.54	0.94	0.49	0.80
7 мес.	0.74	0.89	0.96	0.70	0.79	0.39	0.86
8 мес.	0.19*	0.65	0.59	0.11**	0.96	0.52	0.57
9 мес.	0.32	0.82	0.74	0.20*	0.71	0.57	0.70
10 мес.	0.62	0.78	0.72	0.28	0.88	0.69	0.70
11 мес.	0.37	0.29	<i>0.19*</i>	0.14**	<i>0.17*</i>	0.46	0.47
12 мес.	0.26	0.15**	0.40	0.20*	0.16*	0.65	0.51

*Примечания:*

- (а) для различных уровней значимости использованы следующие обозначения: \* – 20%, \*\* – 15%, \*\*\* – 5%;
- (б) в качестве результатов теста Гренджера приведены значения вероятностей принятия нулевой гипотезы об отсутствии причинно-следственной связи интенсивности инфляционных ожиданий и предполагаемых факторов. Ряды приведены к стационарной форме с помощью вычисления первых разностей. Для проверки стационарности использованы тесты Дики-Фуллера и тест KPSS.
- (в) значение каждого критерия получено для предположения о зависимости рядов каждого предполагаемого фактора и индикатора интенсивности инфляционных ожиданий в период с февраля 2014 г. по февраль 2018 г. Период уменьшен за счет приведения рядов к стационарной форме. Каждый фактор обозначен в названиях столбцов таблицы. В каждой строке таблицы факторы используются с определенным, указанным в первом столбце временным лагом;
- (г) VAR обозначает результаты модифицированного робастного теста Гренджера для простых попарных векторных авторегрессий с оптимальным лагом, определенным на основе критерия Акаике. Результаты приведены только в ячейках таблицы, соответствующих оптимальным лагам модели.
- (д) наиболее значимые взаимосвязи выделены жирным шрифтом и цветом, второстепенные взаимосвязи – курсивом и цветом.

*Источник:* составлено автором.

Следует также отметить, что с учетом малого количества данных для оценки в качестве границы отсечения для статистической значимости теста Гренджера использовался уровень в 20%. При этом достижению приемлемого уровня значимости с большим временным лагом – в пределах 10-12 месяцев – придавалось второстепенное значение при вынесении суждений об итоговых причинно-следственных взаимосвязях. Это также обусловлено малым числом данных для анализа и возможным эффектом переобучения модели и снижения качества итоговой оценки при высоком порядке проведения теста. Так, увеличение порядка проведения теста на 1 единицу обозначает включение в анализ дополнительного временного лага для фактора и целевой переменной в дополнение к уже используемым на более низком порядке параметрам.

Результаты проведения теста Гренджера для ежемесячных данных не позволяют отвергнуть гипотезу о значимом влиянии на интенсивность инфляционных ожиданий со стороны текущих значений индекса потребительских цен, темпов роста цен на нефть и изменения ключевой ставки денежно-кредитной политики (см. таблицу 7). При этом второстепенную роль для изменения инфляционных ожиданий играет влияние факторов изменения индекса промышленного производства и темпов роста курса доллара к рублю. Вероятность значимого влияния немного возрастает при предположении о большем лаге взаимодействия.

В то же время по результатам анализа изменение доходов населения и индекса промышленного производства не являются причиной изменения индикатора интенсивности инфляционных ожиданий, несмотря на обнаруженную на предшествующем шаге статистически значимую взаимосвязь между этими показателями.

Среди рассмотренных выше факторов динамика цен на нефть, курса доллара к рублю и ключевой ставки Банка России может быть представлена и соотнесена с динамикой индикатора интенсивности инфляционных ожиданий не только на ежемесячном, но и на ежедневном уровне.

Визуальный анализ ежедневных рядов показателей в период с января 2014 г. по февраль 2018 г. подтверждает выводы о значительной взаимосвязи между перечисленными показателями и интенсивностью инфляционных ожиданий (см. рис. 12).

Изучение высокочастотной динамики индикаторов с учетом результатов теста Гренджера (см. Приложение 4, таблицы 1-2)<sup>11</sup> позволяет предположить, что первоочередным фактором для изменений как инфляционных ожиданий, так и курса национальной валюты<sup>12</sup> в 2014-2016 гг. являлись цены на нефть. Значительное влияние динамики цен на сырье на инфляционные ожидания подтверждается как на ежемесячных, так и на ежедневных данных. Во второй половине 2014 г. интенсивность инфляционных ожиданий фактически являлась отражением для динамики цен на нефть – до декабря 2014 г., момента не предполагаемого рынком решения регулятора о значительном повышении ключевой ставки. Затем, даже в условиях дальнейшего снижения цен на нефть – на фоне рекордно высокой ставки денежно-кредитной политики – инфляционные ожидания продолжали снижение до уровней середины 2014 г. Другое существенное повышение интенсивности ожиданий роста цен одновременно со значительным обесценением рубля произошло только после очередного «витка» снижения сырьевых цен в конце 2015 г. за пределы вероятного психологического для населения порога – значений цен на нефть в первые месяцы 2015 г.

В то же время результаты высокочастотного анализа указывают на то, что в краткосрочном периоде динамика курса рубля также оказывает влияние на инфляционные ожидания населения. Статистические критерии

---

<sup>11</sup> Следует отметить, что при осуществлении высокочастотного анализа для теста Гренджера используются более строгие критерии значимости, вследствие более существенного объема данных. Вывод о невозможности отклонить нулевую гипотезу об отсутствии причинно-следственной зависимости осуществляется при её вероятности (p-value) порога в 10%.

<sup>12</sup> Высокая значимость цен на нефть для предсказания курса национальной валюты при анализе высокочастотной статистики была продемонстрирована во множестве исследований [Ferraro, Rogoff, Rossi, 2015 и др.].

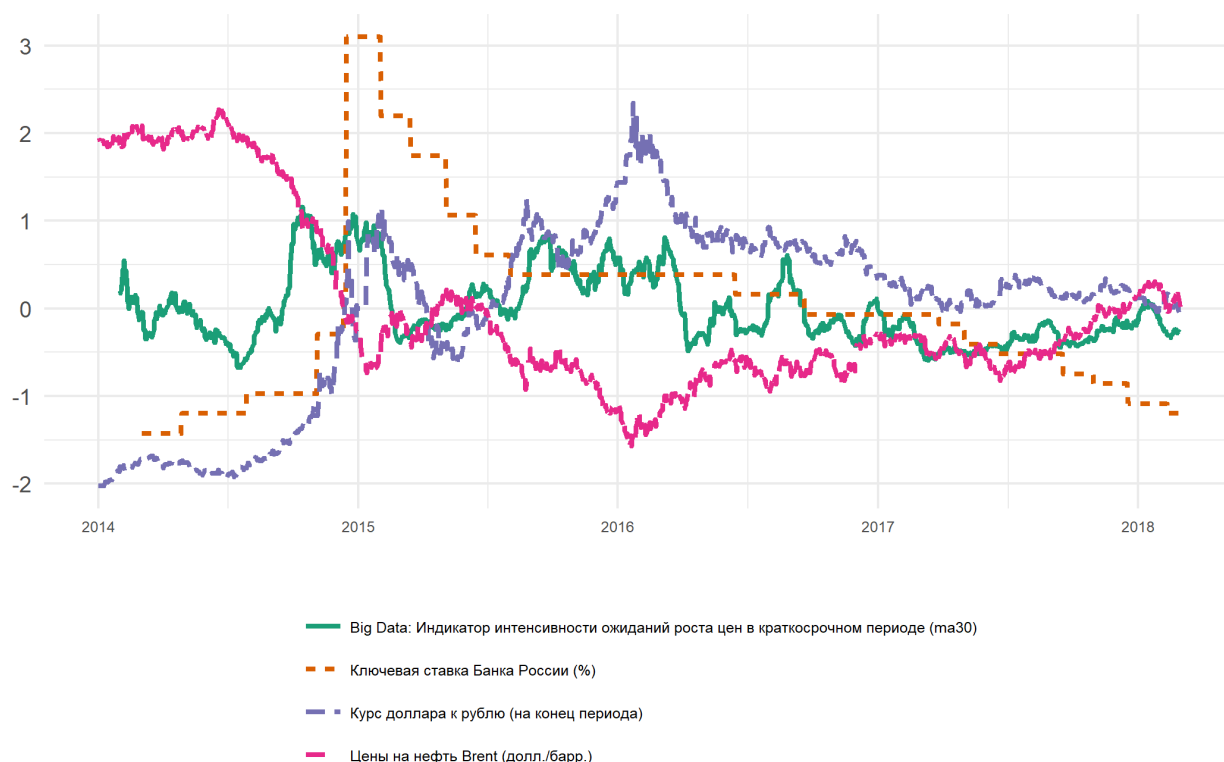
подтверждают значимую нелинейную и причинно-следственную взаимосвязь показателей по Гренджеру в пределах 2 недель.

Что касается влияния изменений ключевой ставки на инфляционные ожидания, то результаты высокочастотного анализа подтверждают наличие положительной взаимосвязи и причинно-следственного влияния ключевой ставки на инфляционные ожидания с лагом примерно в 1 месяц, как и по результатам анализа на ежемесячных данных. Такая ситуация в целом соответствует заявлениям руководства Банка России о том, что инфляционные ожидания населения пока не являются «полностью «заякоренными»<sup>13</sup>.

При этом с середины 2017 г. заметным становится усиление интенсивности инфляционных ожиданий в окрестности решений Банка России о снижении ключевой ставки. Это может свидетельствовать о постепенном повышении доверия населения к действиям регулятора и корректировке механизма формирования инфляционных ожиданий в сторону необходимого для успешной реализации режима инфляционного таргетирования.

---

<sup>13</sup> Заявление Э.С. Набиуллиной на пресс-конференции ЦБ РФ от 15.09.2017: <https://ria.ru/economy/20170915/1504859849.html>



*Примечания:*

(а) для удобства сопоставления индикаторов все ряды нормализованы – приведены к  $N(0, 1)$ ;

(б) в целях сглаживания неоднородной динамики в рамках месяцев индикатор интенсивности инфляционных ожиданий показан в виде скользящей средней за последние 30 дней (средняя продолжительность 1 месяца).

*Источник:* составлено автором.

Рис. 12. Сопоставление динамики индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных и различных экономических параметров на ежедневном уровне.

**Экспертное выявление факторов, с которыми население связывает изменение собственных инфляционных ожиданий**

Второй шаг выявления факторов интенсивности инфляционных ожиданий базируется на детальном изучении содержания комментариев интернет-пользователей, упоминающих ожидания роста цен (целевых комментариев).

В соответствии с данным шагом методики построения индикаторов интенсивности экономических ожиданий, описанным в разделе 2.5.2, для индикатора интенсивности инфляционных ожиданий в рассматриваемом

наборе данных был осуществлен детальный анализ содержания целевых комментариев.

По результатам проведенного анализа – с учетом выводов первого шага выявления факторов – было выделено восемь содержательных категорий факторов или тем, с которыми население в значительной доле комментариев связывало изменение интенсивности собственных инфляционных ожиданий на протяжении всего периода, с января 2014 г. по февраль 2018 г. [Голощапова, Андреев, 2017; Голощапова, 2017]. Эти категории интернет-пользователи, интересующиеся экономическими новостями СМИ, упоминали в одних и тех же сообщениях, где говорили о собственных инфляционных ожиданиях. Выделенные категории были обозначены как:

- Цены на сырье. Эта категория состоит в обсуждении в комментариях интернет-пользователей как цен на основные энергоресурсы (нефть, газ, уголь) и их производные (бензин, топливо и др.), так и объемов их предложения на рынке. Выбор этой категории подтвержден с точки зрения результатов анализа первого этапа, показавшего значительную взаимосвязь и наличие причинно-следственного влияния цен на нефть на изменение интенсивности инфляционных ожиданий населения;
- Денежно-кредитная политика. Данная категория учитывает любую связь инфляционных процессов со стороны населения и действий Банка России. Выбор данной категории был также подтвержден результатами анализа первого этапа, показавшими значительную корреляцию и причинно-следственную взаимосвязь изменений ставки денежно-кредитной политики и интенсивности инфляционных ожиданий;
- Курс национальной валюты. Категория включает упоминание курса национальной валюты, девальвации и спекулятивных атак при одновременном обсуждении инфляционного фона. Результаты первого этапа показали значительную взаимосвязь инфляционных

ожиданий и курса доллара к рублю без подтверждения причинно-следственного эффекта. Вместе с тем содержательный анализ комментариев показывает, что население склонно связывать динамику курса валюты и потребительских цен;

- Доходы. Категория определяет взаимосвязь между динамикой доходов населения в различных формах (заработная плата, пенсии и др.) и их внимания к инфляционным процессам. Наличие подобной взаимосвязи, но без подтверждения причинно-следственного эффекта, было подтверждено выводами первого этапа;
- Экономический кризис. Категория содержит упоминания кризисных процессов в экономике с изменением инфляционного давления. Результаты первого этапа показали существенную взаимосвязь инфляционных ожиданий и общей экономической динамики, аппроксимируемой с помощью уровня безработицы и индекса промышленного производства;
- Неопределенность экономической ситуации. Категория обозначает любые упоминания населением отсутствия понимания экономических трендов, связываемых с ожиданиями роста цен. Идея выявления подобного показателя на основе текстовой информации впервые была предложена [Baker, Bloom, Davis, 2016]. Аналогичная категория оказалась релевантной обсуждениям инфляционных ожиданий;
- Внутренняя политика. Категория содержит различные упоминания действий государства, которые население связывает с собственными ожиданиями роста цен. Это может иметь отношение как к тарифной, налоговой или бюджетной политике и изменениям в законодательстве, так и к менее определенным высказываниям о коррупции, бюрократии, действиях Министерств, Президента, Правительства или государства в целом. Данная категория не рассматривалась на предшествующем этапе, однако была выделена

как существенная по результатам анализа содержания комментариев;

- **Мировая политика.** Категория включает связь со стороны населения действий других государств и инфляционных процессов в российской экономике. Данная категория была выделена как значимая по результатам анализа содержания комментариев.

Для автоматизации поиска выделенных категорий в тексте комментариев интернет-пользователей были составлены правила на основе регулярных выражений в соответствии с принципами, отраженными в разделе 2.3. Составление правил осуществлялось итеративно, с постепенным улучшением качества отбора и минимизации доли нерелевантных примеров, определяемых с помощью экспертной оценки. Результаты составления правил представлены в таблице 8. Скрипт для отбора комментариев интернет-пользователей в соответствии со сформированными правилами представлен в Приложении 2.

На основе составленных правил для рассматриваемого набора данных был осуществлен поиск выделенных категорий среди комментариев интернет-пользователей, упоминающих ожидания роста цен. Доля сообщений, отнесенных к какой-либо из выделенных категорий, в общем числе целевых сообщений составила 93%. Это дополнительно свидетельствует об относительно высоком качестве выявления возможных факторов, с которыми население связывает изменение собственных инфляционных ожиданий, для рассматриваемого набора данных.

Для выделенных целевых комментариев, отнесенных к какой-либо категории, были рассчитаны показатели вкладов упоминаний каждой категории в общем числе упоминаний всех категорий – в соответствии с формулой (10), описанной в разделе 2.5.2. При этом из расчетов были исключены комментарии, не отнесенные ни к одной из выделенных категорий (7% всех целевых комментариев). В результате вычислений были получены временные ряды факторов, с которыми российское население



связывало изменение своих инфляционных ожиданий в период с января 2014 г. по февраль 2017 г., на ежемесячном и ежедневном уровнях.

Таблица 8. Правила отбора комментариев интернет-пользователей для факторов, с которыми российское население в существенной доле комментариев связывало изменение интенсивности инфляционных ожиданий на протяжении периода с января 2014 г. по февраль 2018 г.

Название категории	Правила отбора на основе регулярных выражений
Цены на сырье	(топлив бенз нефт барр brent брен опек энерг сырье brent urals wti добыч  газ ресурс уголь)
Денежно-кредитная политика	(процент цб банк россии центробанк цбр ставк ставок рефинанс репо регулят набиул юдаев дкп денежно-кредитн)
Курс национальной валюты	(долл евро \\\$ курс валют рубл девальв обесцен спекул керри кэрри сagгy)
Доходы	(пенс индекс зарп зараб зп з/п)
Экономический кризис	(дефицит банкрот дефолт производ спрос убыт кризис рецесс потрясение застой спад ввп стагн крах продаж расход промышл)
Неопределенность экономической ситуации	(неопред неясн непонятн не понятн неоднозначн нестабильн противореч неустойч не уверен неуверен запутан двусмысл опас непредсказ не предсказ волатил)
Внутренняя политика	(жкx минпром минфин медведев  бюрократ закон минист минэк конкур фас путин президент импортозамещ инициатив бюджет коррупц воров откат институт налог власт чиновн тариф государств правительств)
Мировая политика	( мир[ae]  внешн санкц сша  ес  евросоюз еврозон европ запад обама меркель германи фрс украин крым  кита великобр америк турц)

*Примечания:*

- (1) Все представленные регулярные выражения могут быть записаны в другой форме, в соответствии с правилами их составления;
- (2) Символ «&» в представленных регулярных выражениях обозначает поиск упоминаний в одном комментарии блоков слов, обозначенных в квадратных скобках. Такая модель поиска может быть записана при помощи различного синтаксиса в разных языках программирования, в которых используются регулярные выражения;
- (3) Перед поиском осуществляется базовая фильтрация текста (перевод символов в нижний регистр, удаление небуквенных символов за исключением «\$»).

*Источники:* составлено автором.

В течение всего рассматриваемого периода российское население связывало изменение собственных ожиданий роста цен преимущественно с внутренней политикой (21% сообщений из общего числа сообщений, отнесенных к выделенным категориям) и состоянием мировой конъюнктуры (16% сообщений). Кроме того, в подтверждение результатов первого шага выявления факторов инфляционных ожиданий, формирование инфляционных ожиданий оказалось в высокой степени связано с обсуждением курса национальной валюты и цен на сырье (по 14% сообщений для каждой категории), а также с обеспокоенностью кризисными процессами в экономике (12% сообщений), условиями денежно-кредитной политики (11% сообщений) и изменением доходов (8% сообщений). При этом российское население на протяжении всего рассматриваемого периода в наименьшей степени (всего в 4% сообщений) связывало изменение собственных ожиданий роста цен с неопределенностью экономической ситуации.

При этом детальный визуальный анализ полученных временных рядов вкладов различных категорий в изменение индикатора интенсивности инфляционных ожиданий позволяет сделать выводы о факторах, с которыми российское население связывало отдельные эпизоды повышения ожиданий роста цен (см. рис. 13):

- Период повышения ожиданий роста цен во второй половине 2014 г.:
  - Старт резкого повышения инфляционных ожиданий в 3 кв. 2014 г. происходил одновременно с очередным эпизодом ухудшения внешней конъюнктуры и с формированием вызвавшей обеспокоенность населения тенденции к повышению ключевой ставки Банком России. Ухудшение мировой конъюнктуры было связано с санкциями ряда стран против Российской Федерации на фоне обвинений в причинах катастрофы Boeing 777 в июле 2014 г. Что касается второго фактора, а именно обеспокоенности населения монетарной политикой регулятора, то в сентябре 2014

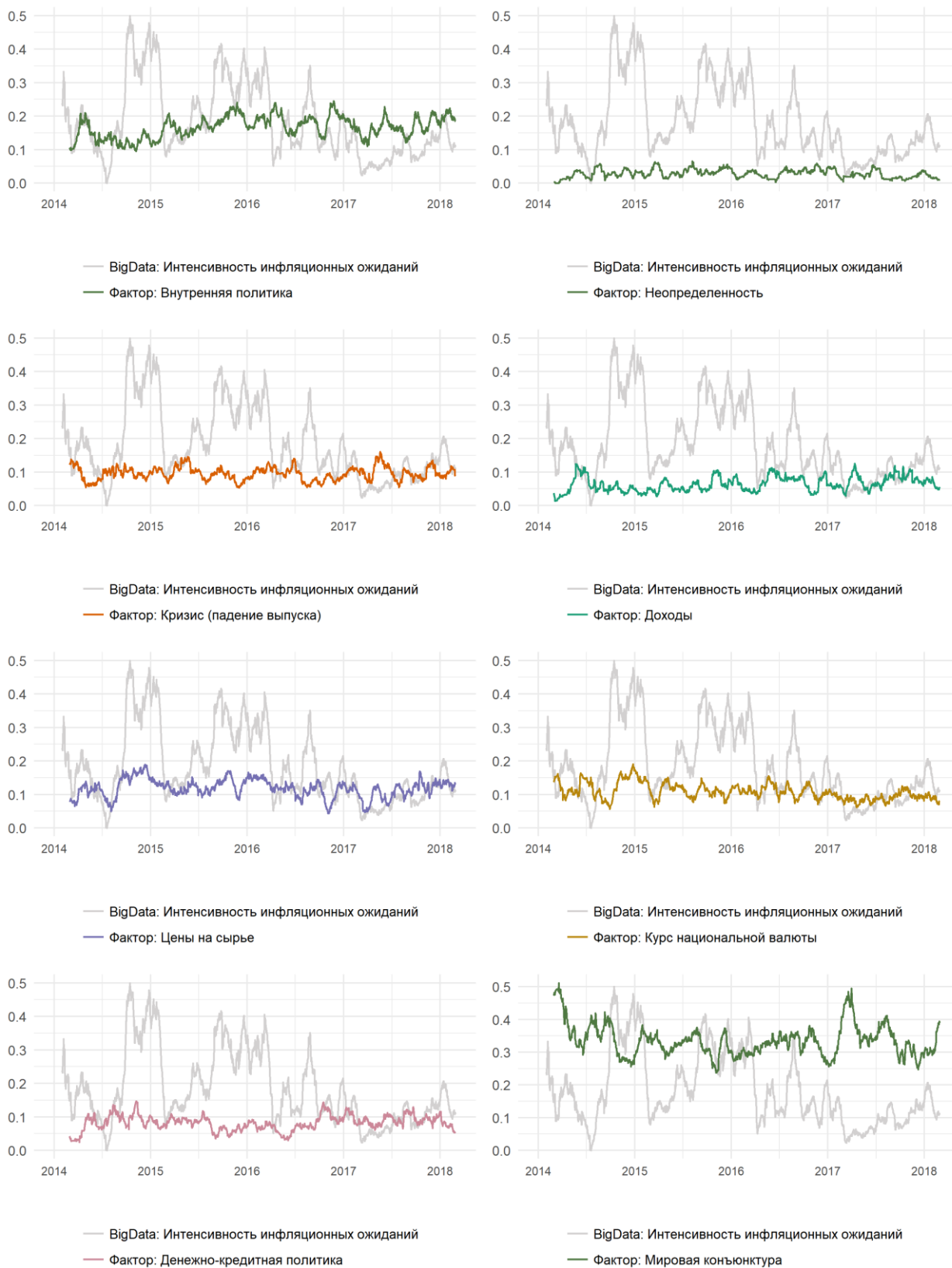
г. в сравнении с июнем 2014 г. повышение вклада категории «денежно-кредитная политика» составило 76%. По всей вероятности, это было связано с третьим повышением ключевой ставки (до 8% с конца июля 2014 г.), подтвердившим для населения намерения регулятора по ужесточению условий монетарной политики.

В то же время у существенной доли интернет-пользователей явный переход Банка России к ужесточению денежно-кредитной политики ассоциировался с ростом неопределенности экономической ситуации. При изучении ежедневных рядов вкладов категорий заметен резкий рост соответствующего фактора сразу после принятия регулятором решения об увеличении ключевой ставки до 8% (28.07.2014). При этом также следует отметить значительное повышение обеспокоенности населения возможностью реализации экономического кризиса через некоторое время после повышения опасений относительно ужесточения монетарной политики. Возможность такой взаимосвязи подтверждалась отдельными научными работами [Апокин и др., 2014; Апокин и др., 2015]. Детальное изучение содержания комментариев в категории «экономический кризис» в этот период позволяет заключить, что наибольшая доля опасений относительно возможного кризиса была связана с угрозой падения выпуска и роста кредитной нагрузки.

- Наблюдавшийся в 4 кв. 2014 г. дальнейший рост инфляционных ожиданий населения был в значительной степени связан с падением цен на нефть одновременно с существенным ослаблением курса национальной валюты. В этот период вклад категорий, помимо цен на сырье и курса рубля, в изменение инфляционных ожиданий был невысоким.

- Период повышения ожиданий роста цен в 4 кв. 2015 г. и 1 кв. 2016 г. был преимущественно соотнесен населением с очередным снижением сырьевых цен. Последние упали до значений 1 кв. 2015 г., локального минимума всего рассматриваемого периода. Второстепенными, но значимыми факторами выступали опасения населения относительно состояния и перспектив внутренней политики и развития экономического кризиса.
- Другой, менее заметный относительно других период повышения ожиданий роста цен со стороны населения во второй половине 2016 г., судя по результатам визуального анализа, был связан населением с двумя основными трендами: падением сырьевых цен и ухудшением состояния мировой конъюнктуры. Следует отметить, что на протяжении второго полугодия 2016 г. наблюдались существенные колебания цен на сырье в области 45-50 долларов за баррель – локального минимального значения трехлетнего периода и вероятной «психологической отметки» для населения, способной спровоцировать панические настроения. Наиболее значительный рост вклада соответствующей категории в повышение инфляционных ожиданий был зафиксирован в 4 кв. 2016 г. Вместе с тем в этот период вблизи локального максимума увеличения интенсивности ожиданий роста цен основным фактором их изменений являлась обеспокоенность населения состоянием мировой конъюнктуры на фоне итогового этапа выборов президента США.
- Последний и не заверченный по состоянию на февраль 2018 г. эпизод повышения инфляционных ожиданий наблюдается с мая 2017 г. – момента временного разворота тренда инфляции к снижению из-за временного негативного эффекта аномально-холодных погодных условий на урожай сельскохозяйственной продукции. С этого момента, несмотря на продолжение снижения

потребительской инфляции до рекордных значений в соответствии с официальными статистическими данными, ожидания роста цен со стороны населения остаются на повышенном уровне. Изучение временных рядов вкладов факторов на этом периоде позволяет заключить, что основными факторами, с которыми население связывает повышение ожиданий роста цен в этот период, являются снижение собственных доходов и опасения относительно развития кризисных процессов в экономике. Вклады соответствующих категорий значительно повысились в 4 кв. 2017 г. – на фоне продолжающегося с середины 2014 г. падения реальных располагаемых доходов населения и слабой и неустойчивой динамики производственной и инвестиционной активности в экономике. Другими значимыми факторами остаются цены на сырье, повышение которых в 4 кв. 2017 г. способствует росту цен на бензин, и обеспокоенность как внутренней, так и внешней экономико-политической ситуацией. Однако, несмотря на это, динамика вклада категории «мировая конъюнктура» не характеризовалась ни повышением, ни волатильностью в январе-феврале 2018 г. – в период возможных опасений относительно возможного введения новых санкций против России со стороны США при публикации доклада Минфина США.



*Примечания:*

(а) для удобства сопоставления индикаторов все ряды нормализованы – приведены к  $N(0, 1)$ ;

(б) в целях сглаживания неоднородной динамики в рамках месяцев индикатор интенсивности инфляционных ожиданий и временные ряды факторов его динамики

показаны в виде скользящей средней за последние 30 дней (средняя продолжительность 1 месяца).

*Источник:* составлено автором.

Рис. 13. Сопоставление динамики индикатора интенсивности инфляционных ожиданий населения на основе больших данных и различных экономических параметров на ежедневном уровне.

Результаты статистического анализа взаимосвязи рассчитанных на ежедневном уровне временных рядов вкладов категорий и изменения индикатора интенсивности инфляционных ожиданий, позволяют сделать следующие выводы.

С одной стороны, не наблюдается значительной взаимосвязи между динамикой выделенных категорий и динамикой индикатора интенсивности ожиданий на протяжении всего периода – за исключением факторов цен на сырье и курса национальной валюты. Корреляции для этих факторов равны 60% и 40% соответственно и являются значимыми на 1% уровне, однако с точки зрения критериев взаимной информации значимой взаимосвязи не выявлено. Такая ситуация объясняется высокой волатильностью полученных временных рядов факторов, обусловленной способом их построения. Структура факторов, связанных с изменением инфляционных ожиданий, постоянно меняется – каждый отдельный эпизод их изменения обусловлен разными причинами. Это определяет дополнительную изменчивость временных рядов, поскольку вклады факторов являются относительными величинами. Соответственно, при повышении относительного вклада одного из факторов, наиболее весомого для конкретного эпизода роста инфляционных ожиданий, снижается вклад других. В случае расчета указанных статистических критериев для отдельных однородных по структуре факторов периодов выводы о взаимосвязях могут оказаться значимыми для лидирующих в этом периоде категорий.

С другой стороны, результаты теста Гренджера в целом согласуются с выводами первого шага выявления факторов. Гипотеза о причинно-

следственном воздействии на изменение инфляционных ожиданий на высокочастотном уровне не может быть отвергнута на 20% уровне значимости для всех факторов, за исключением курса национальной валюты и внутренней политики (см. таблицу 9). Это подтверждает результаты тестов Гренджера предшествующего шага, показавшие значимый причинно-следственный эффект на ожидания со стороны темпов роста цен на нефть, изменения условий денежно-кредитной политики и общей экономической конъюнктуры, на первом шаге аппроксимированной индексом промышленного производства, а на текущем – категорией «экономический кризис». В то же время устойчивый причинно-следственный эффект на высокочастотном уровне был зафиксирован для категории «доход», несмотря на результаты предшествующего шага, где подобная гипотеза была отвергнута. Судя по содержательному анализу комментариев, снижение собственных доходов российское население в существенной мере связывает с повышением внимания к инфляционным процессам. Кроме того, для новых относительно предшествующего шага категорий, выделенных по результатам изучения текстов целевых сообщений, значимыми факторами для изменения инфляционных ожиданий оказались «неопределенность экономической ситуации» и «мировая конъюнктура».

Вместе с тем категория «внутренняя политика» по результатам теста Гренджера не оказалась устойчиво значимым фактором изменения ожиданий роста цен, несмотря на ее высокую распространенность среди целевых комментариев интернет-пользователей. Важно отметить, что категория «внутренняя политика» не всегда может содержать упоминания конкретных событий и трендов, оказывающих воздействие на ожидания роста цен со стороны населения в заданный момент времени (по построению правил отбора, за исключением роста стоимости жилищно-коммунальных услуг). Эта категория в большей степени отражает долгосрочные опасения населения относительно общей политико-экономической ситуации в стране, что может включать в себя обсуждение решений и законопроектов



министерств и Правительства в различных сферах, в том числе и в области налоговой политики, а также обеспокоенность уровнем коррупции. В некоторой степени эта категория может характеризовать даже неопределенность экономической ситуации. Это становится вероятным в случае, когда население испытывает трудности в нахождении конкретных факторов ожидаемого повышения цен, переходя к обобщениям [Голощапова, Андреев, 2017].

Процедура проведения теста Гренджера была аналогична реализованной на предшествующем шаге выявления факторов изменения инфляционных ожиданий, включая робастный тест Гренджера для простых попарных VAR моделей.

Таблица 9. Результаты теста Гренджера на причинно-следственную зависимость индикатора интенсивности инфляционных ожиданий от выделенных по результатам содержательного анализа целевых комментариев категорий

Порядок для проведения теста Гренджера (временной лаг)	Цены на сырье	Курс национальной валюты	Денежно-кредитная политика	Экономический кризис	Неопределенность экономической ситуации	Внутренняя политика	Мировая конъюнктура	Доход
1 день	0.96 <b>VAR: 0.07**</b>	0.32 VAR: 0.25	0.36	0.40 VAR: 0.13*	<b>0.09**</b> <b>VAR: 0.01****</b>	0.44 VAR: 0.15	0.14* <b>VAR: 0.02****</b>	0.79 VAR: 0.13*
2 дня	0.75	0.32	0.56	0.59	0.08**	0.30	0.57	0.25
3 дня	0.28	0.53	0.49 <b>VAR: 0.10*</b>	0.74	0.14*	0.46	0.72	0.11*
4 дня	0.12*	0.75	0.58	0.90	0.27	0.57	0.82	0.05**
5 дней	0.16	0.87	0.63	0.41	0.47	0.60	0.68	<b>0.04****</b>
6 дней	0.20	0.60	0.38	0.34	0.57	0.59	0.81	0.08**
7 дней	0.20	0.70	0.50	0.07**	0.67	0.56	0.39	0.13*
8 дней	0.26	0.57	0.53	<b>0.06**</b>	0.77	0.27	0.48	0.18
9 дней	0.31	0.50	0.56	0.08**	0.81	0.21	0.49	0.22
10 дней	0.42	0.38	0.66	0.10*	0.82	0.28	0.48	0.31
11 дней	0.44	0.50	0.68	0.09**	0.90	0.38	0.56	0.39
12 дней	0.34	0.34	0.66	0.12*	0.89	0.50	0.45	0.43
13 дней	0.38	0.38	0.74	0.10*	0.88	0.40	0.52	0.50
14 дней	0.46	0.38	0.71	0.12*	0.87	0.28	0.54	0.62

*Примечания:*

- (а) для различных уровней значимости использованы следующие обозначения: \* – 15%, \*\* – 10%, \*\*\* – 5%, \*\*\*\* – 1%;
- (б) в целях сглаживания неоднородной динамики в рамках месяцев индикатор интенсивности инфляционных ожиданий и временные ряды факторов его динамики используются для расчетов в виде скользящей средней за последние 30 дней (средняя продолжительность 1 месяца);
- (в) в качестве результатов теста Гренджера приведены значения вероятностей принятия нулевой гипотезы об отсутствии причинно-следственной связи интенсивности инфляционных ожиданий и предполагаемых факторов. Стационарность исходных рядов была верифицирована с помощью тестов Дики-Фуллера и KPSS.
- (г) значение каждого критерия получено для предположения о зависимости рядов каждого предполагаемого фактора и индикатора интенсивности инфляционных ожиданий в период с февраля 2014 г. по февраль 2018 г. Период уменьшен за счет приведения рядов к стационарной форме. Каждый фактор обозначен в названиях столбцов таблицы. В каждой строке таблицы факторы используются с определенным, указанным в первом столбце временным лагом;
- (д) VAR обозначает результаты модифицированного робастного теста Гренджера для простых попарных векторных авторегрессий с оптимальным лагом, определенным на основе критерия Акаике. Результаты приведены только в ячейках, соответствующих оптимальным лагам модели.
- (е) наиболее значимые взаимосвязи выделены жирным шрифтом и цветом.

*Источник:* составлено автором.

## **Выявление факторов, с которыми население связывает изменение инфляционных ожиданий, на основе методов машинного обучения и текстовой обработки**

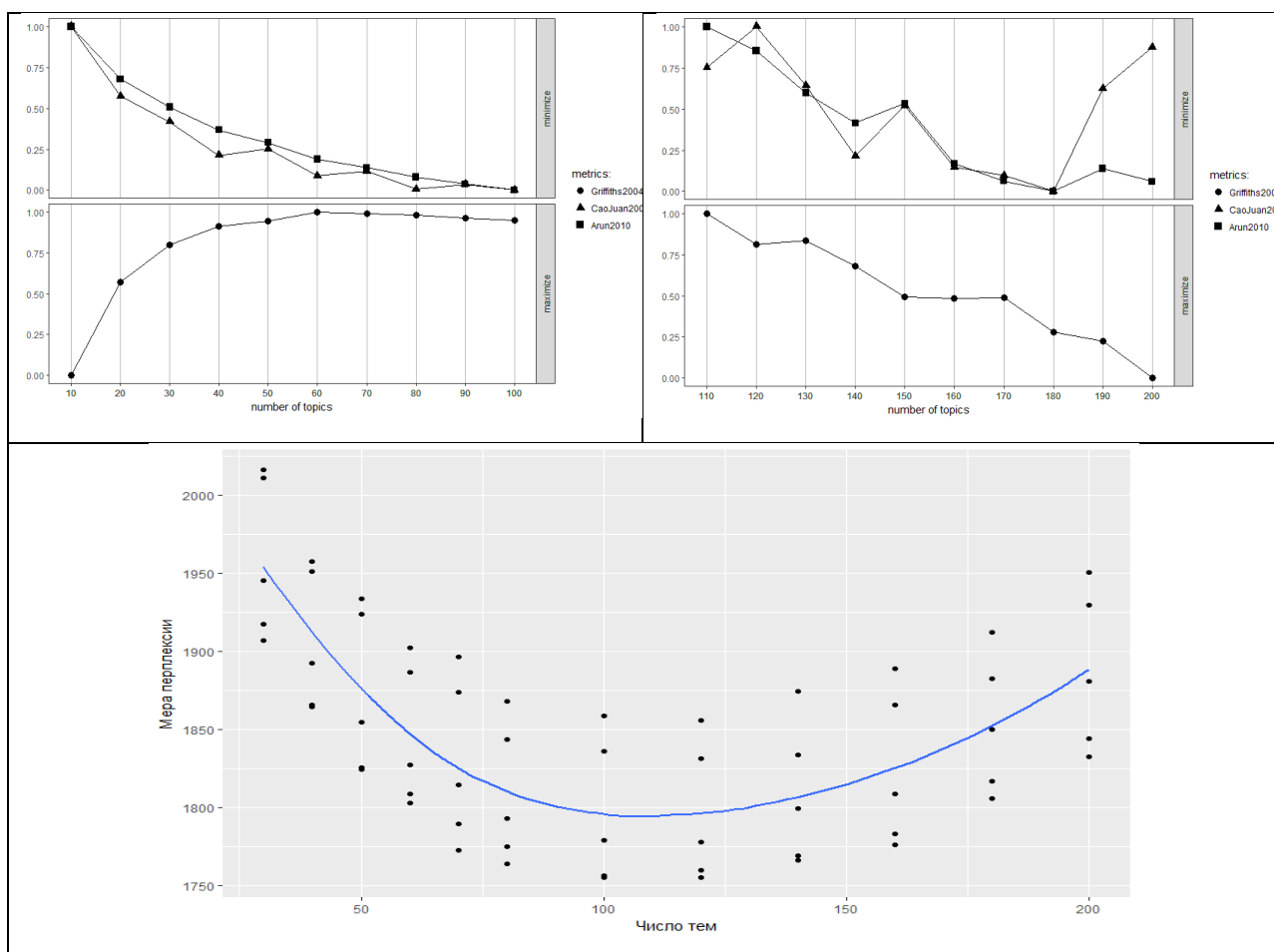
На данном шаге выявления факторов изменения инфляционных ожиданий российского населения, были последовательно реализованы все стадии, указанные в разделе 2.5.3 в рамках методики построения индикаторов.

Первая стадия заключалась в предобработке текстов целевых комментариев на всем рассматриваемом периоде, включившей базовую фильтрацию текста, токенизацию, удаление стоп-слов и стемминг. Все расчеты и операции по обработке текста были реализованы на языке программирования R, с использованием современных и эффективных библиотек для обработки естественного языка (`text2vec`, `tm`, `LDavis`, `topicmodels`, `ldatuning` и др.). В то же время стемминг был осуществлен с помощью программы `mystem` от Yandex, классического инструмента для выполнения данной задачи. При этом список стоп-слов для русского языка был загружен из международной коллекции стоп-слов для различных языков `Stopwords ISO`. Скрипт проведения тематического моделирования представлен в Приложении 3.

Затем, с помощью расчета указанных в разделе 2.5.3 статистических критериев было определено оптимальное число тем для всего рассматриваемого набора данных. При этом для повышения качества результата мера перплексии была рассчитана как среднее по результатам кросс-валидации с разделением выборки на 5 частей.

В соответствии с указанными критериями оптимальным для целевых комментариев на всем рассматриваемом периоде оказалось 100 тем (см. рис. 14). Именно в окрестности этого числа наблюдается переход от резкого снижения к небольшим и малозначительным изменениям для функции меры перплексии от числа тем. При этом критерии Као Джуан [Cao Juan et al., 2009] и Аруна и соавторов [Arun et al., 2010] стремятся к минимуму, а

критерий Гриффитса и Стейверса [Griffiths and Steyvers, 2004] – к максимуму, без существенных расхождений.



Источник: составлено автором.

Рис. 14. Статистические критерии выбора оптимального числа тем для целевых комментариев в период с января 2014 г. по февраль 2018 г.

На следующей стадии для оптимального числа тем был рассчитан алгоритм LDA. Параметры алгоритма помимо числа тем были определены в соответствии с одним из классических методов [Steyvers, Griffiths, 2006; Niraula et al, 2013]: гиперпараметр  $\alpha$  для распределения «документ-тема» выбран равным 0.5 (отношение 50 к числу тем), а гиперпараметр  $\beta$  для распределения «тема-слово» – равным 0.01. По результатам многочисленных эмпирических экспериментов при прочих равных данный метод по выбору параметров приводит к устойчиво качественному выявлению тем в различных текстах.

По результатам тематического моделирования целевых комментариев в соответствии с заданными параметрами алгоритма LDA на всем рассматриваемом временном периоде были получены списки ключевых слов для 100 тем. Изучение последних позволяет дополнить понимание факторов, с которыми население связывало изменение собственных инфляционных ожиданий в 2014-2018 гг. По результатам экспертного анализа среди выделенных тем были определены десять категорий, являющихся наиболее структурированными, логичными по составу слов и содержательными с точки зрения экономических процессов. Списки ключевых слов для каждой из выделенных категорий представлены в таблице 10.

Важно, что шесть из десяти выделенных категорий подтверждают результаты анализа предшествующих двух шагов выявления факторов. Так, значимыми категориями для изменения инфляционных ожиданий населения являются цены на сырье, курс национальной валюты, внутренние политические и экономические процессы, мировая конъюнктура, денежно-кредитная политика и состояние собственных доходов. При этом тематическое моделирование позволяет выделить дополнительные аспекты взаимосвязи инфляционных ожиданий и выделенных факторов. В категории «Доходы» отдельное значение имеет изменение пенсионных выплат и заработной платы, соотношение общего дохода и совокупных расходов. Для категории «Мировая конъюнктура» отдельными темами являются политические процессы в США, взаимодействие с Китаем, а также взаимодействие с различными странами по сырьевым вопросам и вопросы введения санкций. В категории «Цены на сырье» отдельное значение имеют вопросы цен на нефть в рамках внешней и внутренней политики, а также ценовая динамика производных от энергоресурсов. В категории «Денежно-кредитная политика» население акцентирует внимание как на процентной и валютной политике регулятора, так и на действиях в области финансового сектора. В категории «Внутренняя политика» отдельными темами выступает обсуждение общих внутренних политических процессов, включающих

обсуждение законодательства и ситуации с коррупцией, так и вопросов экономической политики, включая импортозамещение, состояние внутреннего спроса и перспектив экономического роста.

При этом тематическое моделирование также позволяет выделить новые с точки зрения результатов предшествующих шагов категории:

- Тарифы, налоги и платежи. Тема предполагает связь ожиданий повышения цен и наблюдаемого или ожидаемого повышения различных тарифов и платежей, включая жилищно-коммунальные услуги, бензин, а также повышение налогов и акцизов, включая, налоги на транспорт, НДС и др.;
- Недвижимость. Данная тема учитывает обеспокоенность стоимостью ипотечных кредитов и динамикой цен на недвижимость и их соотношением с собственными доходами;
- Торговая политика в розничной торговле. Данная категория включает обсуждение динамики цен на продовольственные и непродовольственные товары (в целом и на отдельные товары), а также представления населения относительно условий торговли для фирм и их будущей ценовой политики в связи с экономическими процессами.

На всем рассматриваемом периоде наиболее распространенными – по вероятности появления тем в комментариях на основе результатов алгоритма LDA – является категория цен на сырье. Совокупная вероятность обсуждения тем, связанных с ценами на сырье, в комментариях интернет-пользователей, упоминающих ожидания роста цен, равна 21% – относительно других выделенных категорий. Вторыми по популярности категориями, с которыми население связывает изменение инфляционных ожиданий стали темы внутренней конъюнктуры и торговых условий с вероятностями появления по 14%. Обсуждение вопросов мировой конъюнктуры и денежно-кредитной политики также в высокой степени распространено – вероятность появления тем данных категорий в текстах составила 12% и 11% соответственно. При

этом вероятность обсуждения проблем повышения тарифов, платежей и налогов, а также состояния собственных доходов в связи с ожиданиями роста цен равна 10% для каждой категории. Обсуждение курса валют (отдельно от других категорий) и ситуации с недвижимостью оказалось наименее распространенным на рассматриваемом периоде времени – вероятность появления каждой из этих категорий в целевых комментариях составила 3.5%. Такие результаты дополнительно подтверждают значимость фактора цен на сырье, внутренней и мировой конъюнктуры, денежно-кредитной политики и динамики собственных доходов и указывают на относительно низкую вероятность причинно-следственной связи «курс национальной валюты – инфляционные ожидания». Курс рубля редко упоминается населением отдельно от других категорий, более весомых по результатам тематического моделирования. Кроме того, результаты тематического моделирования свидетельствуют о необходимости учета категорий торговой политики розничных сетей и динамики тарифов, платежей и налогов при изучении изменения инфляционных ожиданий населения.

Следует отметить, что проведение полноценной процедуры тематического моделирования на отдельных временных промежутках может позволить выделить специфические категории – наиболее важные с точки зрения населения в заданный временной период. Результирующие категории могут отличаться от наиболее значимых на всем временном периоде, полученных в данном разделе, поскольку будут отражать специфику экономических процессов и их восприятия населением на более узком промежутке времени. Периодами для реализации алгоритма LDA и выделения категорий и тем, могут выступать отдельные годы, полугодия, кварталы, месяцы. При этом дезагрегация ниже уровня месяцев должна учитывать общее число сообщений в группе целевых комментариев. В случае малого количества комментариев реализация полноценного алгоритма LDA со стадией валидации числа тем может становиться нецелесообразной.

Таблица 10. Списки ключевых слов для отдельных тем в категориях, выделенных по результатам тематического моделирования целевых комментариев интернет-пользователей в период с января 2014 г. по февраль 2018 г.

Категория	Название темы	Список ключевых слов
Цены на сырье	Цены на нефть и международная конъюнктура	нефть, цена, доллар, добыча, рынок, нефтяной, баррель, сланцевый, экономика, увелич, сша, объем, стоимость, рост, месторождение, внутренний, зависеть, россия, упасть, крупный, сырье, самый, себестоимость, добывать, сектор, низко, поставка, норвегия, бюджет, опек, экспорт, обеспечивать, идти, запас, низкий, простой, спрос, становиться, скважина, расход, затрата, компенсировать, плюс, продажа, далеко, мировой
	Цены на сырье и внутренняя конъюнктура	цена, нефть, бензин, рост, повышение, повышать, рубль, вырастить, подымать, экономика, продавать, налог, упасть, правительство, газ, страна, повышаться, падать, дешеветь, дорожать, спад, ждать, просто, сырье, курс, стабильно, разница, сокращение, баррель, зарабатывать, тариф, литр, остальное, потеря, овощ, дешево, спрос, руб, энергоноситель, фас, мало, транспортный, перевозка, предложение, говорить, компенсировать
	Цены на сырье и производные	цена, расти, нефть, высокий, бензин, высоко, повышение, падать, повышать, низко, завышать, рынок, хотеть, низкий, увеличивать, акциз, темп, рано, упасть, покупать, спрос, пока, производитель, условие, быстро, просто, либо, приводить, бумажка, конкурент, запас, доля, скоро, падение, товар, реклама, элитный, недовольный, сравнивать
Курс национальной валюты	Курс национальной валюты	рубль, доллар, валюта, курс, страна, рф, валютный, девальвация, бюджет, евро, национальный, экономика, экспорт, падение, снижение, правительство, нефть, резерв, рейтинг, биржа, финансовый, санкция, цена, агентство, стоить, вводить, кризис, резервный, фонд, дефицит, близкий, продажа, триллион, повышение, фон, ввп, внутри, составлять, должный, декабрь, зависеть, иностранный, международный, момент, резко
Доходы	Пенсия	пенсия, получать, деньги, год, значить, пенсионер, свой, зарплата, процент, минимум, услуга, работать, простой, чиновник, государство, платить, инфляция, страховой, мало, делать, должный, рубль, становиться, высокий, месяц, давать, врач, стоимость, забывать, достойный, равно, либо, прожито, правило, предлагать, руб, сбербанк, уменьшаться, страхование, сказать, страховка, свободный, миллион, вырастить, заплатить
	Оплата труда	зп, труд, деньги, платить, повышать, налог, работник, нужно, производительность, работать, предприниматель, брать, процент, давать, купить, плата, огромный, водитель, сумма, зарплата, написать, услуга, высокий, затрата, число, старый, добавлять, образ, сколько, сотрудник
	Соотношение доходов и расходов	ребенок, человек, процент, работать, просто, расход, возраст, жить, доход, посмотреть, высокий, пенсия, сокращение, медицина, нормальный, школа, метод, размер, здравоохранение, статистика, деньги, порядок, благо, россиянин, жизнь, тыс, реальный, больной, сильно, тратить, выходить, пациент, необходимо, пойти, общество, бюджет, политика, наоборот, представлять
Мировая конъюнктура	Мировая конъюнктура в целом	сша, приводить, количество, отношение, влияние, результат, проводить, экономика, эксперт, американский, например, создавать, москва, большой, оценка, специалист, изменение, использовать, являться, исследование, область, рейтинг, мнение, экономический, проблема, продолжать, сотрудничать, прогноз, век, стол, доля, происходить, капитализм, рамка, представлять, внимание, вполне, вести, автор, новый, базовый, аналитик, выглядеть
	Политика США	сша, рынок, рост, политика, актив, инфляция, экономика, один, реальный, производство, стоимость, текущий, приводить, потребление, финансовый, менее, многих, функция, проблема, демократ, модель, трамп, процент, оказываться, денежный, фрс, американский, ем, продолжать, пространство, взять, являться, вернуться, иной, вводить, нужный, политический, валюта, весьма, развитый, позиция, увеличивать, конгресс, предполагать, переоценивать



	Взаимодействие с Китаем	китай, страна, кризис, экономика, мировой, китайский, рынок, значить, процент, показывать, приходить, один, оставлять, рука, свой, развитый, сделать, мир, нефть, полностью, поскольку, нужно, другой, попытка, выборы, полный, уходить, решение
	Международное взаимодействие по сырьевым товарам	россия, год, сша, китай, газ, мировой, страна, экономический, российский, америка, газпром, американский, план, европейский, конец, составлять, ресурс, инфраструктура, политический, сср, капитал, газовый, союз, центральный, дорога, база, европа, миллиард, дешевый, объем, стратеги, румыния, период, данные, газопровод, война, цель, общий, вооружение, длинный, разрушение, активный, оказываться, восток
	Санкции	год, ес, отношение, санкция, случай, являться, другой, вашингтон, против, один, самый, принимать, общий, предлагать, великобритания, франция, следовать, польша, германия, сша, позволять, создавать, угроза, совершенно, возможность, действовать, статус, выступать, понимать, кроме, восстановление, индия, устойчивый, следующий, британский, индийский, напоминать, структура, противостояние, рассматривать, смысл
Денежно-кредитная политика	Процентная политика	ставка, банк, кредит, инфляция, цб, процентный, банка, экономика, снижать, деньги, давать, рост, кредитный, риск, валюта, брать, рынок, долг, население, инвестор, денежный, масса, нужно, заем, фрс, компания, годовой, мера, финансовый, средство, бизнес, спрос, увеличивать, ипотека, мало, должный, депозит, кредитование, повышать, издержки, сектор, облигация, обязательство, рублевый, покупка
	Валютная политика и риски	деньги, инфляция, цб, рубль, видеть, высокий, печатать, говорить, создавать, валюта, банк, доллар, условие, высоко, покупка, продавать, курс, пойти, рынок, зависеть, снижение, рост, падение, вывод, уровень, эмиссия, центробанк, предложение, кроме, вкладывать, продажа, инфляционный, официальный, сколько, платить, потребительский, ценный, отдавать, думать
	Политика в области финансового сектора	банк, вклад, российский, совет, цель, вопрос, международный, лицензия, центральный, акция, организация, право, год, необходимо, асв, вводить, сделка, федерация, отзывать, денежный, целое, актив, мера, скорость, кредитный, принятие, возврат, законодательство, соглашение, круг, требование, признавать, полагать, план, автор, финансовый, всероссийский, привлекать, последствие, передавать, отделение, временной
Внутренняя конъюнктура	Внутренняя политика	чиновник, страна, другой, бизнес, система, работа, государство, компания, один, коррупция, пока, проблема, решать, производство, пример, хороший, просить, ехать, риск, помогать, частный, никакой, никто, ответственность, уровень, лекарство, немного, местный, стоять, снимать, работать, ростовщик, сокращать, приезжать, упасть, заниматься, и, уголовный, писать, капиталист, закон, министр, чиновник, контроль, свобода, фсб
	Внутренняя экономическая политика	россия, уровень, рост, политика, кредит, другой, позволять, правительство, некоторый, повышение, рынок, например, нужно, действие, оценка, долг, финансовый, пока, инфляция, ситуация, механизм, доход, время, резкий, экономист, существовать, отрасль, кредитование, биржа, норма, валютный, спекулятивный, институт, финансирование, глава, финансы, критерий, потенциал, экономика, падать, курс, деньги, спрос, реальный, экспорт, производитель, купить, население, инвестиция, импорт, потребительский, росстат, импортный, благосостояние
Тарифы, налоги и платежи	Тарифы и их соотношение с доходами	год, рост, тариф, бюджет, зарплата, инфляция, налог, знать, плохой, команда, мало, сумма, доллар, рубль, сколько, самый, расти, многих, последний, думать, цена, жкх, доход, увеличивать, компания, система, процент, завышать, топливо, миллиард, работа, на, стоять, спокойно, акциз, бедность, житель, обычный, услуга, высокий, выплачивать, любой, говорить, предупреждать, день, бакс, мера
	Тарифы	руб, месяц, нужный, рубль, человек, деятельность, руководитель, приватизация, решать, стоять, жкх, довольно, должность, тариф, май, заместитель, литр, основание, мнение, решение, ноябрь, вебсайт, имущество, главное, коммерсант, февраль, проводить, аи92, полномочия, роснефть, поступать, апрель, скоро, бензин, пообещать, контролировать, отменять
	Налоги	налог, уровень, страна, население, авто, мало, один, зарплата, золотой, проблема, низкий, врач, свой, увеличивать, взнос,

		автомобиль, сказать, доход, огромный, покупать, при, современный, прибыль, сделать, умирать, пустой, иностранный, предлагать, поборы, обслуживание, маленький, пара, возникать, реально, вопрос, явно, всякий, оценка, локализация, НДС, твой, больница, знание, менять, транспорт, газпром, обеспечивать
Недвижимость	Недвижимость	деньги, получать, квартира, проблема, хотеть, зарплата, жить, просто, ипотека, цена, платить, человек, аренда, общий, скоро, дом, труба, время, стоить, быстро, низкий, взять, нужно, раздавать, переставать, новость, многих, повышать, тратить, касаться, перевозка, недвижимость, земля, лишний, менять
Торговая политика	Торговая политика в целом	цена, товар, рынок, производство, рост, производить, повышать, спрос, импорт, продукт, магазин, стоимость, импортный, снижать, препарат, продавец, внутренний, повышаться, конкуренция, вырастать, пошлина, питание, средний, падение, налог, оплата, сбыт, ценовой, способность, поставка, дефицит, забывать, огранич, ассортимент, покупатель
	Продовольственные товары	цена, продукт, просто, рост, вырастать, магазин, стоять, покупать, высокий, производитель, разница, покупатель, доллар, деньги, греча, выгодно, картошка, фас, устанавливать, дорогой, второй, вещь, дешевый, услуга, фактор, идти, торговать, конкуренция, спекулянт, мешать, интересный, закон, розница, ценообразование, неделя, вести, закупочный
	Непродовольственные товары	новый, время, стоимость, автомобиль, каждый, бизнес, продукт, рынок, продукция, платить, разный, день, вид, вода, платеж, примерно, брат, простой, телефон, производить, месяц, некоторый, готовый, кризис, сила, повышение, старый, клиент, уходить, вводить, импортозамещение, сеть, увидеть, условие, цена, коне, проходить, большой

*Примечания:*

- (а) списки ключевых слов представляют собой слова с максимальной вероятностью появления для каждой выделенной темы по результатам применения алгоритма LDA на всем рассматриваемом временном периоде. При этом из списков исключены немногочисленные слова, не несущие смысловой нагрузки в рамках проводимого анализа (цензура, специфические выражения);
- (б) названия отдельных тем определены экспертно – в соответствии со списками ключевых слов для каждой темы;
- (в) имена собственные и аббревиатуры приведены в таблице строчными буквами, вследствие перевода всех буквенных символов в нижний регистр на стадии предварительной обработки текста (при базовой фильтрации текстовых данных) для корректной реализации тематического моделирования.

*Источник:* составлено автором.

### **Выводы главы 3**

1. Проведена верификация разработанной методики на примере построения высокочастотного индикатора инфляционных ожиданий российского населения.

Разработанная методика была реализована на основе текстовых данных новостных статей четырех крупнейших экономических СМИ России (Ведомости, Коммерсант, РБК, Эксперт) и комментариев интернет-пользователей к ним на официальных веб-страницах СМИ и в социальных сетях Facebook и ВКонтакте.

Индикатор интенсивности инфляционных ожиданий российского населения был построен на ежемесячном и ежедневном уровнях в период с января 2014 г. по февраль 2018 г. Оценка репрезентативности исходных данных и полученных результатов, а также тесты на устойчивость результатов расчета к выбранным правилам отбора статей и комментариев, предусмотренные методикой, дали положительный результат.

По результатам сравнения построенный индикатор значимо взаимосвязан с динамикой ближайшего индикатора-аналога: ключевого на текущий момент показателя инфляционных ожиданий российского населения, рассчитываемого ФОМ на основе опросов. При этом статистические критерии указывают на опережающую динамику полученного в работе индикатора (на 2 месяца). Подобные различия могут объясняться тем, что расчет индикатора на основе больших данных возможен в реальном времени, тогда как индикатор на основе опросов может отражать ожидания населения с временным лагом, требующимся на сбор, верификацию и агрегацию ответов респондентов.

В то же время построенный индикатор оказался релевантным объекту ожиданий, а именно потребительской инфляции. Статистические критерии указывают как на значимую взаимосвязь показателей, так и на возможность опережающего воздействия инфляционных ожиданий населения на будущую динамику инфляции.

Тесты на устойчивость результатов расчета к выбранным правилам отбора статей подтвердили их содержательную релевантность для измерения инфляционных ожиданий населения и отсутствие негативной специфики в рассматриваемом наборе данных новостей СМИ и комментариев интернет-пользователей, ограничивающих потенциал вычисления индикаторов интенсивности для различных областей экономических ожиданий.

2. В соответствии с разработанной методикой выявлены и проанализированы в динамике факторы изменения интенсивности инфляционных ожиданий российского населения во времени.

Совокупные результаты выявления факторов показали, что наиболее значимыми факторами (экономическими событиями и трендами), с которыми население связывало изменение собственных инфляционных ожиданий в период с января 2014 г. по февраль 2018 г., являлись: (1) текущая инфляция; (2) цены на сырье; (3) мировая конъюнктура (4) денежно-кредитная политика; (5) состояние доходов; (6) неопределенность экономической ситуации. Важно, что для этих факторов не была отвергнута гипотеза о причинно-следственной взаимосвязи с инфляционными ожиданиями.

Второстепенными, но значимыми и в высокой степени взаимосвязанными с инфляционными ожиданиями населения факторами стали: (1) внутренняя конъюнктура; (2) курс национальной валюты; (3) состояние тарифов, платежей и налогов; (4) торговая политика розничных сетей; (5) состояние и перспективы рынка недвижимости. Результаты теста Гренджера для этих факторов были неоднозначны, однако население в обсуждении новостных статей СМИ на тему «инфляция» со значимой вероятностью связывало с ними изменение собственных инфляционных ожиданий.

Значительное влияние текущей инфляции на динамику инфляционных ожиданий в данной работе дополнительно подтверждает гипотезу о существенном вкладе адаптивного механизма в формирование ожиданий российского населения. Ранее это также подтверждалось в эмпирических

исследованиях на российских данных с помощью других методов квантификации инфляционных ожиданий [Соколова, 2014; Перевышин, Рыкалин, 2017].

На высокочастотном уровне (по ежемесячным и ежедневным данным):

- показано, что общим движущим фактором для изменения как инфляционных ожиданий, так и курса национальной валюты в 2014-2016 гг. являлись цены на нефть. Именно этот оказывает причинно-следственное влияние на инфляционные ожидания по результатам всех шагов оценивания. При этом в краткосрочном периоде динамика курса доллара к рублю может оказывать причинно-следственное влияние на интенсивность инфляционных ожиданий.
- подтверждено влияние условий денежно-кредитной политики на инфляционные ожидания населения. При этом, по результатам тематического моделирования, население акцентирует внимание как на процентной и валютной политике регулятора, так и на действиях в области финансового сектора. Что касается процентной политики, то в среднем в период с января 2014 г. по февраль 2018 г. отмечается положительное воздействие изменения ключевой ставки на интенсивность инфляционных ожиданий населения, что противоречит ожидаемой в условиях режима инфляционного таргетирования взаимосвязи [Апокин и др., 2014]. Следует отметить, что подобная взаимозависимость возможна в условиях пока еще невысокого, формирующегося доверия населения к проводимой денежно-кредитной политике, когда население может повышать собственные инфляционные ожидания вслед за регулятором.

При этом с середины 2017 г. на высокочастотном уровне отмечается повышение инфляционных ожиданий в окрестности решений Банка России о снижении ключевой ставки. Это может свидетельствовать о постепенном повышении доверия населения к действиям регулятора и корректировке механизма формирования инфляционных ожиданий в

сторону необходимого для успешной реализации режима инфляционного таргетирования.

- подтверждено значимое влияние изменения доходов населения на инфляционные ожидания по результатам экспертного выделения факторов и тематического моделирования. Причинно-следственная зависимость подтверждается на высокочастотном уровне, несмотря на ее отсутствие между соответствующими показателями на ежемесячном уровне. Российское население в существенной мере связывает снижение собственных доходов с повышением внимания к инфляционным процессам. При этом тематическое моделирование свидетельствует о том, что в категории «Доходы» отдельное значение имеет изменение пенсионных выплат и заработной платы, соотношение общего дохода и совокупных расходов.

При этом экспертный анализ содержания комментариев интернет-пользователей позволил выделить новые в контексте исследования инфляционных ожиданий российского населения факторы, с которыми население связывает изменение ожиданий. Это факторы неопределенности экономической ситуации, внутренней политики и мировой конъюнктуры за исключением воздействия динамики курса доллара и цен на нефть. Причем статистические критерии указывают на то, что неопределенность экономической ситуации и мировая конъюнктура оказывают причинно-следственное воздействие на инфляционные ожидания, тогда как внутренняя политика, отражающая общие опасения населения относительно политико-экономической ситуации в стране, скорее, значительно взаимосвязана с ростом инфляционных ожиданий, но не оказывает на них значимого влияния. При этом в категории «мировой конъюнктуры», по результатам тематического моделирования, можно выделить обеспокоенность населения вопросами санкций, политическими процессами в США, взаимодействием с Китаем и другими странами по сырьевым вопросам.

Кроме того, тематическое моделирование также позволило выделить новые в контексте исследования инфляционных ожиданий российского населения факторы, с которыми население связывает изменение ожиданий, а именно: тарифы, налоги и платежи; ситуация в секторе недвижимости; торговая политика в розничной торговле.

3. С учетом высокочастотных текстовых данных выявлены факторы, с которыми население связывало повышение инфляционных ожиданий в окрестности отдельных эпизодов их резкого роста (конец 2014 г., середина 2015 г. - начало 2016 г.) или структурного изменения уровня (вторая половина 2017 г. - начало 2018 г.).

- первое повышение ожиданий в 2014 г. было связано не только с ухудшением внешней конъюнктуры, но и с обеспокоенностью ужесточением политики Банка России (формирование тенденции к повышению ключевой ставки к середине 2017 г.). Причем последнее население также связало с ростом неопределенности экономической ситуации и, позднее, с кризисной ситуацией в экономике. В конце 2014 г. развитие роста инфляционных ожиданий было в большей степени связано с динамикой цен на нефть и ослаблением курса рубля.
- рост инфляционных ожиданий в середине 2015 г. начале 2016 г. был преимущественно соотнесен населением с очередным снижением сырьевых цен.
- повышенный уровень инфляционных ожиданий с мая 2017 г., в основном, обусловлен обеспокоенностью населения относительно состояния собственных доходов и возможного развития кризисных процессов в экономике, а также динамикой цен на нефть, способствующей повышению цен на бензин. При этом влияние фактора сырьевых цен усиливается к концу периода – на фоне закрепления цен на нефть на повышенных уровнях.

## Заключение

В процессе проведения диссертационного исследования были получены следующие результаты:

1. Разработана авторская методика построения высокочастотных индикаторов экономических ожиданий населения, основанная на пяти основных этапах: (1) определение источников информации и оценка репрезентативности выборки для анализа; (2) отбор новостных статей выделенного набора российских экономических СМИ; (3) отбор комментариев интернет-пользователей к выделенному набору новостных статей экономических СМИ на тему конкретной области экономических ожиданий; (4) расчет индикатора интенсивности заданного вида экономических ожиданий, включая оценку релевантности полученного индикатора целям исследования и проверку устойчивости результатов к правилам расчета; (5) выявление факторов изменения интенсивности заданного вида экономических ожиданий во времени.

2. Составлена авторская карта крупнейших российских экономических СМИ, релевантных для анализа экономических ожиданий населения. Карта верифицирована с помощью опросов экономического экспертного сообщества. Карта состоит из четырех блоков: основной блок, блок массовых изданий, блок телевидения и радио, блок блогосферы.

В рамках основного блока карты определен обязательный блок, а именно, набор СМИ, обладающий достаточной репрезентативностью для исследования. В состав этого блока были включены четыре крупнейших российских источника экономической информации: Ведомости, РБК, Эксперт, Коммерсант.

3. В качестве базы для построения индикаторов экономических ожиданий использованы комментарии интернет-пользователей к новостным статьям экономических СМИ. При этом установлены и обоснованы источники для сбора информации о сообщениях интернет-пользователей,



включающие: официальные веб-страницы СМИ в сети Интернет, а также официальные группы СМИ в социальных сетях ВКонтакте и Facebook.

4. Разработан алгоритм составления правил отбора новостных статей СМИ для заданной области экономики, а также отбора комментариев интернет-пользователей, упоминающих заданный вид ожиданий. Составление правил основано на экспертном подборе комбинаций регулярных выражений, релевантных заданной области экономики и виду экономических ожиданий. Алгоритм предполагает итеративную верификацию соответствия результатов отбора содержательным целям исследователя.

5. Показано, что в рамках разработанной методики ключевым индикатором для аппроксимации экономических ожиданий населения является индикатор интенсивности ожиданий. Показатель отражает степень обеспокоенности населения относительно заданного вида экономических ожиданий во времени.

Расчет показателя основан на числе комментариев интернет-пользователей, упоминающих заданный вид ожиданий, к новостным статьям СМИ по заданной области экономики, скорректированный на интенсивность исследуемых ожиданий со стороны СМИ.

6. Предусмотрена оценка релевантности рассчитанных индикаторов интенсивности экономических ожиданий на основе сравнения с индикаторами-аналогами, вычисленными другими способами, а также с динамикой целевого экономического параметра, ожидания в области которого измеряет исследователь.

7. Разработан алгоритм оценки устойчивости результатов построения индикатора интенсивности экономических ожиданий, предполагающий две стадии: измерение чувствительности результатов построения индикаторов в зависимости от сформированных правил отбора новостных статей СМИ и комментариев интернет-пользователей, а также от изменения области экономических ожиданий.

8. Разработан трехшаговый алгоритм выявления факторов изменения интенсивности экономических ожиданий населения во времени. Шагами алгоритма являются: (1) анализ зависимости индикаторов интенсивности экономических ожиданий от различных экономических параметров, способных в экономической теории выступать факторами формирования ожиданий населения в заданной области экономики; (2) экспертное исследование содержания комментариев с целью выявления факторов, с которыми население связывает изменение собственных ожиданий; (3) анализ содержания комментариев интернет-пользователей с помощью современных методов текстовой обработки и алгоритмов тематического моделирования.

9. Разработанная методика была последовательно применена для построения высокочастотных индикаторов инфляционных ожиданий российского населения на основе определенной ранее минимально репрезентативной выборки экономических СМИ в период с января 2014 г. по февраль 2018 г.

Рассматриваемый для анализа набор данных включил около 260 тыс. новостных статей СМИ, 1.5 миллиона комментариев от 140 тыс. интернет-пользователей. При этом выборка является сбалансированной по медиа-источникам.

10. На основе рассматриваемого набора данных в соответствии с разработанной методикой рассчитан индикатор интенсивности инфляционных ожиданий. Показатель представлен на ежемесячном и ежедневном уровне. При этом предусмотрена возможность расчета индикатора на ежесекундной основе. Оценка репрезентативности исходных данных и полученных результатов, а также тесты на устойчивость результатов расчета к выбранным правилам отбора статей и комментариев, предусмотренные методикой, дали положительный результат.

По результатам сравнения построенный индикатор значимо взаимосвязан с динамикой ближайшего индикатора-аналога: ключевого на

текущий момент индикатора инфляционных ожиданий российского населения, рассчитываемого ФОМ на основе опросов. При этом статистические критерии указывают на опережающую динамику полученного в работе индикатора (на 2 месяца).

В то же время построенный индикатор оказался релевантным объекту ожиданий, а именно потребительской инфляции. Статистические критерии указывают как на значимую взаимосвязь показателей, так и на возможность опережающего воздействия инфляционных ожиданий населения на будущую динамику инфляции.

11. На основе реализации трехшагового алгоритма выявления факторов (экономических событий и трендов), с которыми население связывало изменение собственных инфляционных ожиданий, в период с января 2014 г. по февраль 2018 г. в качестве первоочередных были определены следующие категории факторов: (1) текущая динамика потребительской инфляции; (2) цены на сырье; (3) мировая политико-экономическая конъюнктура; (4) денежно-кредитная политика; (5) состояние доходов населения; (6) неопределенность экономической ситуации. Для этих категорий, по совокупным результатам проведения теста Гренджера, не может быть отвергнута гипотеза о наличии причинно-следственной взаимосвязи с индикатором интенсивности инфляционных ожиданий российского населения.

При этом в качестве второстепенных факторов, с которыми население в значимой степени связывало изменение собственных инфляционных ожиданий, но для которых не было показано наличие причинно-следственной связи с инфляционными ожиданиями, были определены: (1) внутренняя политико-экономическая конъюнктура; (2) курс национальной валюты; (3) состояние тарифов, платежей и налогов; (4) торговая политика в розничной торговле; (5) состояние рынка недвижимости.

По результатам анализа выявлено значительное влияние текущей инфляции на динамику инфляционных ожиданий, что дополнительно

подтверждает гипотезу о существенном вкладе адаптивного механизма в формирование ожиданий российского населения – в соответствии с предшествующими исследованиями по данной теме [Соколова, 2014; Перевышин, Рыкалин, 2017].

По результатам выявления факторов для инфляционных ожиданий российского населения на высокочастотном уровне (по ежемесячным и ежедневным данным) в период с января 2014 г. по февраль 2018 г.:

- показано, что причинно-следственное воздействие на инфляционные ожидания было характерно именно для цен на нефть, тогда как динамика курса национальной валюты может оказывать влияние на ожидания в краткосрочном периоде (в течение около 2 недель).
- подтверждено влияние условий денежно-кредитной политики на инфляционные ожидания населения. При этом по результатам тематического моделирования население акцентирует внимание как на процентной и валютной политике регулятора, так и на действиях в области финансового сектора. Кроме того, отмечается положительная взаимосвязь изменения ключевой ставки на интенсивность инфляционных ожиданий населения, что может свидетельствовать о пока еще недостаточно высоком уровне доверия населения к политике регулятора.

Тем не менее, с середины 2017 г. проявилась обратная зависимость – инфляционные ожидания возрастают после решений Банка России о снижении ключевой ставки. Это может свидетельствовать о постепенном повышении доверия населения к действиям регулятора.

- на высокочастотном уровне подтверждено значимое влияние изменения доходов населения на инфляционные ожидания. Российское население в существенной мере связывает снижение собственных доходов с повышением внимания к инфляционным процессам. По результатам тематического моделирования обеспокоенность состоянием доходов включает опасения относительно изменения

пенсионных выплат и заработной платы, а также соотношения общего дохода и совокупных расходов.

12. Экспертный анализ содержания комментариев интернет-пользователей позволил выделить новые в контексте исследования инфляционных ожиданий российского населения факторы, с которыми население связывает изменение ожиданий, а именно: неопределенность экономической ситуации, внутренняя политика и мировая конъюнктура за исключением воздействия динамики курса доллара и цен на нефть. Среди них факторами, оказывающими причинно-следственное воздействие на инфляционные ожидания российского населения, являются неопределенность экономической ситуации и мировая конъюнктура. Обеспокоенность состоянием внутренней политики взаимосвязана с ростом инфляционных ожиданий, но не оказывает на них значимого влияния.

13. Тематическое моделирование также позволило выделить новые в контексте исследования инфляционных ожиданий российского населения факторы, с которыми население связывает изменение ожиданий, а именно: тарифы, налоги и платежи; ситуация в секторе недвижимости; торговая политика в розничной торговле.

14. С учетом высокочастотных текстовых данных выявлены факторы, с которыми население связывало повышение инфляционных ожиданий в окрестности отдельных эпизодов их резкого роста (конец 2014 г., середина 2015 г. - начало 2016 г.) или структурного изменения уровня (вторая половина 2017 г. - начало 2018 г.).

- первое повышение ожиданий в 2014 г. было связано не только с ухудшением внешней конъюнктуры, но и с обеспокоенностью ужесточением политики Банка России (формирование тенденции к повышению ключевой ставки к середине 2017 г.). Причем последнее население также связало с ростом неопределенности экономической ситуации и позднее с кризисной ситуацией в экономике. В конце 2014

г. развитие роста инфляционных ожиданий было в большей степени связано с динамикой цен на нефть и ослаблением курса рубля.

- рост инфляционных ожиданий в середине 2015 г. начале 2016 г. был преимущественно соотнесен населением с очередным снижением сырьевых цен.
- повышенный уровень инфляционных ожиданий с мая 2017 г. в основном обусловлен обеспокоенностью населения относительно состояния собственных доходов и возможного развития кризисных процессов в экономике, а также динамикой цен на нефть, способствующей повышению цен на бензин. При этом влияние фактора сырьевых цен усиливается к концу периода – на фоне закрепления цен на нефть на повышенных уровнях.

### **Возможные направления продолжения исследования**

В качестве ключевого возможного направления продолжения исследования следует выделить применение разработанной методики построения высокочастотных индикаторов экономических ожиданий населения на развитых и развивающихся рынках других стран. Потенциал такого применения может быть оценен как высокий с учетом специфики языков, используемых для общения в различных странах.

Одним из ключевых и сложных этапов в случае применения методики к данным других стран будет определение релевантных для конкретной страны источников информации и оценка репрезентативности выборки для анализа. При выборе источников новостных статей исследователю в соответствии с разработанной методикой необходимо провести анализ рейтингов популярности медиа-источников для отдельной страны или группы стран на основе рейтингов СМИ и учесть экономический профиль популярных источников с помощью опросов локального экспертного сообщества. При этом выбор источников для изучения комментариев интернет-пользователей будет включать сообщения к статьям на страницах официальных веб-сайтов полученного набора экономических СМИ и

комментарии к ним в социальных сетях, популярных в данной стране. Следует учитывать, что используемая для российской экономики социальная сеть «ВКонтакте» является популярной, прежде всего, в России и странах СНГ. Следовательно, ее включение в набор источников для других стран может оказаться нецелесообразным.

Составление правил отбора новостных статей выделенной выборки экономических СМИ и затем комментариев интернет-пользователей на заданную тему или область экономических ожиданий может полностью следовать логике, представленной в методике. В некоторых случаях, возможны небольшие модификации регулярных выражений, вследствие специфических особенностей отдельных языков. При этом оценка устойчивости результатов отбора также может быть беспрепятственно осуществлена для текстовых данных на любом языке.

Этап расчета индикаторов интенсивности ожиданий населения для заданной области экономики, включая оценку его релевантности с точки зрения теоретических предпосылок о его динамике и соответствии ключевым макроэкономическим трендам и показателям-аналогам, измеренным иным способом (в случае их наличия) также может быть реализован в полном соответствии с разработанной методикой. При этом выбор показателей для оценки релевантности должен учитывать особенности экономики выбранной страны. Так, к примеру, для экономик с невысокой экспортно-сырьевой зависимостью показатель цен на нефть может иметь ограниченный потенциал для сравнения с индикаторами интенсивности экономических ожиданий.

Выявление факторов изменения интенсивности экономических ожиданий населения во времени и в особенности тематическое моделирование комментариев интернет-пользователей должно учитывать специфику языка выбранной для исследования страны. Предполагаемые макроэкономические факторы изменения интенсивности ожиданий так же, как и на предшествующем этапе, должны учитывать специфику

экономических процессов выбранных для исследования стран. При этом тематическое моделирование, включая этапы предобработки текста, будет основываться на локальных программах или библиотеках, ориентированных на конкретный язык, используемый в данной стране.

Следует отметить, что в случае преобладания нескольких различных языков в выбранной для исследования стране, построение индикатора может представлять собой совокупный результат реализации разработанной методики для каждого отдельного языка. При этом автоматическая идентификация языка может быть осуществлена через подключение к API (автоматический программный интерфейс) Google. Другим способом построения индикатора может быть автоматический перевод всех сообщений интернет-пользователей и новостных статей СМИ на один язык, преобладающий в данной стране. Это также может быть осуществлено с помощью подключения к API соответствующего сервиса Google (Google Translate) или других сервисов автоматического перевода. Последний указанный способ является оптимальным для построения с точки зрения организации вычислений и соответствия разработанной методике, однако характеризуется рядом недостатков, связанных с качеством автоматического перевода и вероятным искажением смысла некоторых комментариев. При этом важно отметить, что в последнее десятилетие на фоне развития алгоритмов работы с текстом на основе глубоких нейронных сетей качество автоматического перевода значительно возросло. Это приводит к высокому качеству тематического моделирования для текстовых данных на различных языках при использовании описанного подхода [Goloshchapova et al., 2018].



## Библиография

### Книги

1. Мостицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R / М.: ДМК Пресс, 2015. – 496 с.: цв. ил. ISBN 978-5-97060-301-7.
2. Песаран Х. Ожидания в экономической теории / Панорама экономической мысли конца XX столетия; под ред. Д. Гринэуэя, М. Блини, И. Стюарта; пер. с англ. под ред. В. С. Автономова, С.А. Афонцева. – Библиотека «Экономической школы». Санкт-Петербург: Экономическая школа. – Т.1-XVI. – 670 с. – Т.2. – 352 с. – 2002.
3. Рашка С. Python и машинное обучение. / пер. с англ. А.В. Логунова. – М. ДМК Пресс, 2017. – 418 с. ISBN 978-5-97060-409-0.
4. Рассел С., Норвиг П. Искусственный интеллект. Современный подход / Второе издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с. ISBN 5-8459-0887-6.

### Препринты

5. Соколова А. Являются ли инфляционные ожидания в России впередсмотрящими? // Препринт WP12/2012/02 Серия WP12, Научные доклады лаборатории макроэкономического анализа. – 2012.

### Авторефераты диссертаций

6. Ключосов М.Н. Специфика инфляционных процессов в современной российской экономике. – Автореферат диссертации канд. эк. наук. – Волгоград. – 2010.

### Материалы конференций, научных семинаров

7. Андреев М. Новый подход к определению отношения авторов коротких текстов к обсуждаемым темам на примере оценки инфляционных ожиданий: Труды XIX Международной конференции

- «Аналитика и управление данными в областях с интенсивным использованием данных» (DAMDID/RCDL'2017). – Москва, 10-13 октября 2017 г.
8. Андреев М.Л., Голощапова И.О. Индексы финансовой паники: Научный семинар программы «Большие данные» ВМК МГУ, Платформы и аналитика больших данных. – Москва, 30 ноября 2016 г.
  9. Голощапова И. Оценка экономических ожиданий населения на основе big data (на примере инфляционных ожиданий населения России): Научный семинар ЦМАКП. – Москва, 30 ноября 2017 г.
  10. Карпов А. Измерение инфляционных ожиданий населения с помощью опросов: возможности и вызовы: Научный семинар ЦМАКП. – Москва, 30 ноября 2017 г.
  11. Осеевский М. Цифровой мир – это реальность, в которой мы уже живем: Девятый Ежегодный Инвестиционный Форум ВТБ Капитал «Россия зовет!». – Москва, октябрь 2017 г.

#### Электронные ресурсы

12. Банк России. Методика квантификации инфляционных ожиданий населения Банка России [Электронный ресурс]. – Материалы официального сайта Банка России. – 2017. – Режим доступа: [https://www.cbr.ru/Content/Document/File/17325/Inflation\\_expectations\\_guide.pdf](https://www.cbr.ru/Content/Document/File/17325/Inflation_expectations_guide.pdf).
13. Банк России. О методике исследования инфляционных ожиданий и потребительских настроений населения [Электронный ресурс]. – Материалы официального сайта Банка России. – Режим доступа: [https://www.cbr.ru/Content/Document/File/17324/FOM\\_meth.pdf](https://www.cbr.ru/Content/Document/File/17324/FOM_meth.pdf).
14. Воронцов К.В. Вероятностное тематическое моделирование [Электронный ресурс]. – 2013. – Курс лекций. – Режим доступа: <http://www.machinelearning.ru/wiki/images/2/22/Voron-2013-ptm.pdf>.

15. Крепцев Д., Поршаков А., Селезнев С., Синяков А. Равновесная процентная ставка: оценки для России [Электронный ресурс]. – Банк России, Серия докладов об экономических исследованиях. – № 13. – Июль 2016 г. – Режим доступа: [https://www.cbr.ru/Content/Document/File/16727/wps\\_13.pdf](https://www.cbr.ru/Content/Document/File/16727/wps_13.pdf).
16. Перевышин Ю.Н., Рыкалин А.С. Моделирование инфляционных ожиданий в российской экономике [Электронный ресурс]. – РАНХиГС. – 2018. – Москва. – Режим доступа: <ftp://w82.ranepa.ru/rnp/wpaper/031816.pdf>.
17. ФОМ. Измерение инфляционных ожиданий и потребительских настроений на основе опросов населения. Развернутый отчет по результатам седьмого-двенадцатого опросов. Июль-декабрь 2017 [Электронный ресурс]. – ООО «инФОМ» по заказу Банка России. – 2017. – Режим доступа: [https://www.cbr.ru/Collection/Collection/File/3739/FOM\\_17\\_12.pdf](https://www.cbr.ru/Collection/Collection/File/3739/FOM_17_12.pdf).
18. Яковлева К. Оценка экономической активности на основе текстового анализа [Электронный ресурс]. – Банк России, Серия докладов об экономических исследованиях. – № 25. – Москва, Октябрь 2017 г. – Режим доступа: [https://www.cbr.ru/Content/Document/File/27714/wp\\_25.pdf](https://www.cbr.ru/Content/Document/File/27714/wp_25.pdf).
19. Официальный сайт Центрального Банка Российской Федерации ([www.cbr.ru](http://www.cbr.ru)).
20. Официальный сайт компании «Финам» (данные по финансовым рынкам, ценам на нефть) ([www.finam.ru](http://www.finam.ru)).
21. Официальный сайт Министерства экономического развития РФ (<http://economy.gov.ru>).
22. Официальный сайт Федеральной службы государственной статистики (<http://www.gks.ru/>).
23. Официальный сайт Единой межведомственной информационно-статистической системы (<https://fedstat.ru/>).

24. Официальный сайт газеты «Ведомости» (<https://www.vedomosti.ru/>).
25. Официальный сайт журнала «Эксперт» (<http://expert.ru/>).
26. Официальный сайт РБК (<https://www.rbc.ru/>).
27. Официальный сайт материалов «Коммерсантъ» (<https://www.kommersant.ru/>).
28. Официальный сайт социальной сети ВКонтакте (<https://vk.com/>).
29. Официальный сайт социальной сети Facebook (<https://www.facebook.com/>).
30. Официальный сайт новостного ресурса «РИА Новости» (<https://ria.ru/>).

#### Статьи

31. Апокин А., Галимов Д., Голощапова И., Сальников В., Солнцев О. Денежно-кредитная политика: работа над ошибками // Вопросы экономики. – 2015. – №9. – С. 136-151.
32. Апокин А., Белоусов Д., Голощапова И., Ипатов И., Солнцев О. О фундаментальных недостатках современной денежно-кредитной политики // Вопросы экономики. – 2014. – №12. – С. 80-100.
33. Пестова А.А., Панкова В.А., Ахметов Р.Р., Голощапова И.О. Разработка системы индикаторов финансовой нестабильности на основе высокочастотных данных // Деньги и кредит. – 2017. – №6. – С. 49-58.
34. Голощапова И., Андреев А. Оценка инфляционных ожиданий российского населения методами машинного обучения // Вопросы экономики. – 2017. – №6. – С. 71-93.
35. Голощапова И. Как регулировать регулятора // Эксперт. – 2014. – №45 (922). – С. 38-42. – М.
36. Солнцев О., Сухарева И., Дешко А., Пенухина Е. В плену старой парадигмы // Эксперт, №3 (882). – С. 30-33. – М. – 2014.

37. О. Солнцев, И. Сухарева, А. Дешко. Как распутать проблемные узлы. О вызовах, с которыми столкнется новое руководство Банка России // Эксперт. – №14 (846). – С. 18-20. – М.
38. Жемков М., Кузнецова О. Измерение инфляционных ожиданий участников финансового рынка в России // Вопросы экономики. – 2017. – №10.
39. Кафтанников И.Л., Парасич А.В. Проблемы формирования обучающей выборки в задачах машинного обучения // Вестник Южно-Уральского государственного университета. Серия: компьютерные технологии, управление, радиоэлектроника. – 2016. – Т. 16. – №3. – С. 15-24.
40. Койчужева М.Т., Богатырёв М.М. Инфляционные ожидания и методы их изучения // Научный журнал НИУ ИТМО. Серия «Экономика и экологический менеджмент». – 2014. – №4. – С. 41-49.
41. Лолейт А. Инфляционные ожидания экономических агентов в России // Экономическая политика. – 2009. – № 6. – С. 34-59.
42. Соколова А. Инфляционные ожидания и кривая Филлипса: оценка на российских данных // Деньги и кредит. – 2014. – № 11. – С. 61-67.
43. Столбов М. Статистика поиска в Google как индикатор финансовой конъюнктуры // Вопросы экономики. – 2011. – № 11.

#### Книги на иностранных языках

44. Arun R., Suresh V., Veni Madhavan C.E., Narasimha Murthy M. N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations / In Advances in knowledge discovery and data mining. Mohammed J. Zaki, Jeffrey Xu Yu, Balaraman Ravindran and Vikram Pudi (eds.). Springer Berlin Heidelberg, 391–402. ISBN: 0-262-56097-6.
45. Banerjee A., Dolado J. J., Galbraith J. W., Hendry D. F. Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data / Oxford University Press, Oxford. – 1993. – 352 p. ISBN-10: 0198288107.

46. Barro R.J. *Macroeconomics* / Cambridge, Massachusetts: MIT Press. – 1997. ISBN 0-262-02436-5.
47. Bishop, C.M. *Pattern Recognition and Machine Learning* / New York: Springer. – 2006. ISBN 978-0-387-31073-2.
48. Blanchard O., Johnson D.R. *Macroeconomics* (6th ed.) / Englewood Cliffs, N.J: Prentice Hall. – 2013. ISBN-13: 978-0-13-306163-5.
49. Cagan P. The monetary dynamics of hyperinflation / In M. Friedman (ed.). *Studies in the Quantity Theory of Money*. Chicago, IL: University of Chicago Press. – 1956.
50. Cover T.M., Thomas J.A. *Elements of Information Theory* / Wiley-Interscience, 2nd edition. – 2006. – 776 p. ISBN 978-0-471-24195-9.
51. Friedl J.E.F. *Mastering Regular Expressions*. Third Edition / O'Reilly Media, Inc., USA. – 2006. – 544 p.
52. Friedman M. *A Theory of the Consumption Function* / Princeton, NJ, Princeton, University Press. – 1957. – 259 p. ISBN: 0-691-04182-2.
53. Gali J. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework* / Princeton University Press. – 2008. – 224 p. ISBN-10: 0691133166.
54. Keynes J. M. *The General Theory of Employment, Interest and Money* / London: Macmillan. – 1936.
55. Knight F. K. *Risk, Uncertainty and Profit* / London: Frank Cass. – 1921.
56. Koyck L. M. *Distributed Lags and Investment Analysis* / Amsterdam. North Holland. – 1954.
57. Meiselman D. *The Term Structure of Interest Rates* / Englewood Cliffs, NJ, Prentice-Hall. – 1962.
58. Mikolajek-Gocejna M. *Investor Expectations in Value Based Management* / Springer International Publishing Switzerland. – 2014. DOI 10.1007/978-3-319-06847-3\_2.
59. Mincer J. Models of adaptive forecasting / In J. Mincer (ed.). *Economic Forecasts and Expectations: Analysis of Forecasting Behaviour and*

- Performance. New York: Columbia University Press for the National Bureau of Economic Research. – 1969.
60. Murphy K.P. Machine Learning: A Probabilistic Perspective / The MIT Press, Cambridge, Massachusetts, London, England. – 2012. ISBN 978-0-262-01802-9.
  61. Nerlove M., Grether D.M., Carvalho J. Analysis of Economic Time Series: A Synthesis / New York: Academic Press. – 1979. – 468 p. ISBN-10: 0125157517.
  62. Pearce D. The Dictionary of Modern Economics. Third Edition / London: ELBS/MacMillan. – 1986.
  63. Sarlin P. Mapping Financial Stability / Springer Science & Business Media. – 2014. ISIN: 978-3-642-54955-7.
  64. Steyvers M., Griffiths T.L. Probabilistic topic models. / In T. Landauer, D McNamara, S. Dennis, and W. Kintsch (Eds.), Latent Semantic Analysis: A Road to Meaning. Mahwah, NJ: Erlbaum. – 2006.

#### Препринты

65. Ardia D., Bluteau K., Boudt K. Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. Preprint submitted to SSRN. – July 21, 2017.

#### Монографии

66. Balke, N. S., M. Fulmer, and R. Zhang. Incorporating the Beige Book into a Quantitative Index of Economic Activity. – Mimeo, Southern Methodist University. – 2015.
67. Binder C. Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations. – Manuscript. – 2015.
68. Choi H., Varian H. Predicting the Present with Google trends. – Mimeo. – April 2009.

69. Curtin, R. Procedure to estimate price expectations. – Manuscript, University of Michigan Survey Research Center. – 1996.
70. Ellingsen T., Soderstrom U. Monetary Policy and the Bond Market. – Unpublished manuscript. – IRIER. – 2004. – Referenced by Gurkaynak, Sack, Swanson (2006).
71. Koziielecki J. Psychologiczna teoria decyzji. PWN. Warsaw. – 1997.

#### Диссертации

72. Björklund S., Uhlin T. Artificial neural networks for financial time series prediction and portfolio optimization. Master of Science Thesis in Industrial Engineering and Management Department of Management and Engineering. – Linköping University. – June 2017.
73. Kutsurelis J.E. Forecasting financial markets using neural networks: an analysis of methods and accuracy. Theses and Dissertations. – Calhoun, The NPS Institutional Archive DSpace Repository. – September 1998.

#### Материалы конференций, научных семинаров на иностранных языках

74. Niraula N., Banjade R., Stefanescu D., Rus V. (2013). Experiments with Semantic Similarity Measures based on LDA and LSA // Statistical Language and Speech Processing, First International Conference, SLSP 2013. – Tarragona, Spain. – July 2013. – Proceedings.
75. Goloshchapova I., Andreev M. Big Data approach to measure inflation expectations: the case of the Russian economy // IFABS 2017. – Oxford. – July 2017.
76. Griffiths T.L., Steyvers M. Finding scientific topics // Proceedings of the National Academy of Sciences 101. – 2004. – Suppl. 1. – P. 5228–5235.
77. Hart A.G. Assets, Liquidity, and Investment // The American Economic Review. – 1949. – Vol. 39, No. 3. – Papers and Proceedings of the Sixty-first Annual Meeting of the American Economic Association (May, 1949). – P. 171-181.



78. Lucas R. Econometric policy evaluation: A critique // Carnegie-Rochester conference series on public policy North-Holland. – 1976. – P. 19-46.
79. Ng A.Y., Jordan M.I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes // Advances in neural information processing systems. – 2002.
80. Nyman R., Ormerod P., Smith R., Tuckett D. Big Data and Economic Forecasting: A Top-Down Approach Using Directed Algorithmic Text Analysis // Workshop on using big data for forecasting and statistics. European Central Bank. – 7-8 April, 2014.
81. Ramage D., Rosen E., Chuang J., Manning C.D., McFarland D.A. Topic Modeling for the Social Sciences // Workshop on Applications for Topic Models, NIPS. – 2009.
82. Schwert G. Tests of causality: The message in the innovations // Carnegie-Rochester Conference Series on Public Policy. – 1979. – Vol. 10, Issue 1. – P. 55-96.
83. Segalovich I. A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine // Proc. of MLMTA-2003. – 2003.
84. Seiter J., Amft O., Troster G. Assessing Topic Models: How to Obtain Robustness? // Proceedings of the AwareCast 2012: workshop on recent advances in behaviour prediction and pro-active pervasive computing. – 2012.
85. Ser-Huang Poon, Goloshchapova Irina, Matthew Pritchard and Phil Reed. Corporate Social Responsibility Reports: Topic Analysis Big Data Approach // Workshop on “Recent Developments in Econometrics and Financial Data Science” at ICMA Centre, Reading University. – November 2017.
86. Siok Kun Seka, Xue Qi Teoa, Yen Nee Wonga. A Comparative Study on the Effects of Oil Price Changes on Inflation // 4th World Conference on Business, Economics and Management (WCBEM). – Procedia Economics and Finance. – 2015. – Vol. 26. – P. 630 – 636.

87. Alejandro R. The Stock Market, the Theory of Rational Expectations and the Efficient Market Hypothesis [Электронный ресурс]. – Lecture notes. – ECON 351. Pennsylvania: Penn State University. – 2008. Режим доступа: <https://www.economicsnetwork.ac.uk/sites/default/files/Kent%20Matthews/MBF2.ppt>.
88. Detmeister, A., Lebow D., Peneva E. Inflation Perceptions and Inflation Expectations [Электронный ресурс]. – FEDS Notes. Washington: Board of Governors of the Federal Reserve System. – December 5, 2016. Режим доступа: <https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/inflation-perceptions-and-inflation-expectations-20161205.html>.
89. Jelodar H., Wang Y., Yuan C., Feng X. Latent Dirichlet Allocation (LDA) and Topic modeling - models, applications, a survey [Электронный ресурс]. – arXiv. – 2017. Режим доступа: <https://arxiv.org/abs/1711.04305>.
90. Jin-Lung Lin. Notes on Testing Causality [Электронный ресурс]. – Institute of Economics, Academia Sinica, National Chengchi University. – 2008. Режим доступа: <http://www.sciencedirect.com/reference/117051>.
91. Jurafsky D., Martin J.H. (2018). Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition [Электронный ресурс]. – Third Edition Draft. Stanford University. – September 2018. Режим доступа: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.
92. Levine D.K. Why Economists Are Right? Rational Expectations and the Uncertainty Principle in Economics – Part I [Электронный ресурс]. – 2012. Режим доступа: [https://www.huffingtonpost.com/david-k-levine/uncertainty-principle-economics\\_b\\_1220796.html](https://www.huffingtonpost.com/david-k-levine/uncertainty-principle-economics_b_1220796.html).
93. Millington S., Pasaogullari M. Survey and Market Measures of Inflation Expectations. Economic Trends [Электронный ресурс]. – Federal Reserve Bank of Cleveland. – 2015, September 25. Режим доступа: <https://www.clevelandfed.org/newsroom-and-events/publications/economic->

trends/2015-economic-trends/et-20150924-survey-and-market-measures-of-inflation-expectations.aspx.

94. Sahm C., Sockin J. 'Limited Attention' and Inflation Expectations of Households [Электронный ресурс]. – FEDS Notes. Washington: Board of Governors of the Federal Reserve System. – October 19, 2016. Режим доступа: <https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/limited-attention-and-inflation-expectations-of-households-20161019.html>.

#### Статьи на иностранных языках

95. Aiginger K. Empirical Evidence on the Rational Expectations Hypothesis Using Reported Expectations // *Empirica*. – 1981. – Volume 8, Issue 1. – P. 25-72. <https://doi.org/10.1007/BF00924901>.
96. Akaike H. A new look at the statistical model identification // *IEEE Transactions on Automatic Control*. – 1974. – AC-19: 716-723.
97. ALsaedi R. N. R. M. The Relationship between Oil Prices, Inflation, Exchange Rate and Economic Activities: Cases GCC between 2010 to 2014 // *International Journal of Sciences: Basic and Applied Research (IJSBAR)*. – 2015. – Vol. 24, No. 1. – P. 117-137.
98. Andolfatto D., Hendry S., Moran K. Are Inflation Expectations Rational? // *Journal of Monetary Economics*, Elsevier. – 2005. – Vol. 55(2). – P. 406-422.
99. Ang A., Bekaert G., Min Wei. Do macro variables, asset markets, or surveys forecast inflation better? // *Journal of Monetary Economics*. 2007. – No 54. – P. 1163-1212.
100. Arioli R., Colm Bates, Dieden H., Duca I., Friz R., Gayer Ch., Kenny G., Meyler A., Pavlova I. EU Consumers' Quantitative inflation perceptions and expectations: an evaluation // *ECB, Occasional Paper Series*. – 2017, April. – No 186.

101. Armantier O., Bruine de Bruin W., Potter S., Topa G., van der Klaauw W., Zafar B. Measuring inflation expectations // *Annual Review of Economics*. – 2013. – No 5. – P. 273-301.
102. Armantier O., Nelson S., Topa G., van der Klaauw W., Zafar B. The Price is Right: Updating Inflation Expectations in a Randomized Price Information Experiment // *The Review of Economics and Statistics*. – 2016. – No 98(3). – P. 503-523.
103. Baker S.R., Bloom N., Davis S.J. Measuring Economic Policy Uncertainty // *The Quarterly Journal of Economics*. – 2016. – Vol 131 (4). – P. 1593-1636.
104. Barr, David F., and John Y. Campbell. Inflation, real interest rates, and the bond market: a study of UK nominal and index-linked government bond prices // *Journal of Monetary Economics*. – 1997. – Vol. 39, No. 3. – P. 361-383.
105. Bauer M.D. Inflation Expectations and the News // *Working Paper Series, Federal Reserve Bank of San Francisco*. – Working Paper. – 2014. – Vol 9.
106. Berk J.M. Measuring inflation expectations: a survey data approach // *Applied Economics*. – 1999. – Vol. 31. – P. 1467–1480.
107. Beechey M., Johansson B.K., Levin A. Are Long-Run Inflation Expectations Anchored More Firmly in the Euro Area than in the United States? // *CEPR Discussion Papers from C.E.P.R. Discussion Papers*. – 2011. – No 6536.
108. Bishop C.M., Lasserre J. Generative or discriminative? getting the best of both worlds // *Bayesian Statistics*. – 2007. – Vol. 8. – P. 3-24.
109. Blei, D. M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation // *Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 993–1022.
110. Blokhina T.K., Karpenko O.A., Guirinskiy A.V. The relationship between oil prices and exchange rate in Russia // *International Journal of Energy Economics and Policy*. – 2016. – Vol. 6, No. 4.
111. Bomfim A. Pre-Announcement Effects, News Effects, and Volatility: Monetary Policy and the Stock Market // *Journal of Banking and Finance*. – 2003. – Vol. 27. – P. 133-151.

112. Brazier A., Harrison R., King M., Yates T. The Danger of Inflating Expectations of Macroeconomic Stability: Heuristic Switching in an Overlapping-Generations Monetary Model // *International Journal of Central Banking*. – 2008, June.
113. Bruine de Bruin W., van der Klaauw W., van Rooij M., Teppa F., Klaas de Vos. Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise // *De Nederlandsche Bank NV*. – 2016. – Working Paper No. 506.
114. Bryan M., Venkatu G. The curiously different inflation perspectives of men and women // *Federal Reserve Bank of Cleveland Economic Commentary Series*. – 2001.
115. Cao Juan, Xia Tian, Li Jintao, Zhang Yongdong, Tang Sheng. A density-based method for adaptive IDA model selection // *Neurocomputing*. – 2009. – Vol. 72, Issues 7–9. – P. 1775–1781.
116. Carroll C.D. Macroeconomic Expectations of Households and Professional Forecasters // *Quarterly Journal of Economics*. – 2003. – Vol. 118, No 1.
117. Chow G. C. Rational versus Adaptive Expectations in Present Value Models // *Review of Economics and Statistics*. – 1989. – 71(3). – P. 376-384.
118. Chow G. C. Usefulness of Adaptive and Rational Expectations // *CEPS Working Paper No. 221*. – September 2011.
119. Cochrane J.H., Piazzessi M. The Fed and Interest Rates: A High Frequency Identification // *American Economic Review Papers and Proceedings*. – 2002. – Vol. 92. – P. 90-101.
120. Coibion O., Gorodnichenko Y. Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts // *American Economic Review*. – 2015. – No 105(8). – P. 2644-2678.
121. Cribari-Neto F., Da Silva W.B. A New Heteroskedasticity-Consistent Covariance Matrix Estimator for the Linear Regression Model // *Advances in Statistical Analysis*. – 2011. – 95(2). – P. 129–146.

122. Cunningham R., Desroches B., Santor E. Inflation Expectations and the Conduct of Monetary Policy: A Review of Recent Evidence and Experience // *Bank of Canada Review* (spring). – 2010. – P. 13-25.
123. D'Amuri F., Marcucci J. The Predictive Power of Google Searches in Forecasting Unemployment // *Bank of Italy, Economic Research and International Relations Area. – Economic working papers series.* – 2015. – No. 891.
124. Demertzis M., Viegi N. Inflation Targets as Focal Points // *International Journal of Central Banking.* – 2008. – Vol. 4, No. 1. – P. 55-87.
125. Devlin W., Patwardhan D. Measuring Market Inflation Expectations // *Economic Roundup.* – 2012. – Issue 2. – P. 5-17.
126. Dräger L., Lamla M.J., Pfajfar D. Are Survey Expectations Theory-Consistent? The Role of Central Bank Communication and News // *European Economic Review.* – 2016. – Vol. 85. – P. 84-111.
127. Dyer Travis, Mark Lang, Lorien Stice-Lawrence. The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation // *Journal of Accounting and Economics.* – 2016.
128. Ezekiel, M. The cobweb theorem // *The Quarterly Journal of Economics.* – 1938. – Vol. 52, No. 2. – P. 255–280.
129. Faust J., Swanson E., Wright J.H. Identifying VARs Based on High Frequency Futures Data // *Journal of Monetary Economics.* – 2004. – 51(6). – P. 1107-1131.
130. Ferraro D., Rogoff K., Rossi B. Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates // *Journal of International Money and Finance.* – Vol. 54. – June 2015. – P. 116-141.
131. Figlewski S., Wachtel P. The Formation of Inflationary Expectations // *The Review of Economics and Statistics.* – 1981. – Vol. 63(1). – P. 1-10.
132. Fleckenstein M., Longstaff F.A., Lustig H. Why Does The Treasury Issue TIPS? The TIPS-Treasury Bond Puzzle // *NBER Working Paper No 16358.* – 2010.

133. Forsells M., Kenny G. The Rationality of Consumers' Inflation Expectations: Survey-Based Evidence for the Euro-Area // European Central Bank, Working Paper № 163. – August 2002.
134. Gerlach-Kristen P., Moessner R., Rosenblatt-Wisch R. Computing long-term market inflation expectations for countries without inflation expectation markets // SNB Working Papers. – 2017. – Vol. 9.
135. Gertchev N. A Critique of Adaptive and Rational Expectations // Quarterly Journal of Austrian Economics. – 2007. – Vol. 10. – P. 313-329.
136. Grishchenko O.V., Jing-Zhi Huang. Inflation Risk Premium: Evidence from the TIPS Market // Journal of Fixed Income. – 2013. – Vol. 22, No. 4.
137. Gul F., Pesendorfer W. Temptation and Self-Control // Econometrica. – 2001. – Vol. 69, issue 6. – P. 1403-1435.
138. Guzman G. Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations // The Journal of Economic and Social Measurement. – 2011. – Vol. 36, No. 3.
139. Goloshchapova Irina, Ser-Huang Poon, Matthew Pritchard and Phil Reed. Corporate Social Responsibility Reports: Topic Analysis Big Data Approach // European Journal of Finance. – 2018. – Forthcoming.
140. Gomes G.M. Irrationality of «rational expectations» // Journal of Post Keynesian Economics. – 1982. – Vol. 5, No. 1. – P. 51-65.
141. Goodwin T.H., Sheffrin S.M. Testing the Rational Expectations Hypothesis in an Agricultural Market // The Review of Economics and Statistics. – 1982. – Vol. 64(4). – P. 658-667.
142. Granger C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods // Econometrica. – 1969. – Vol. 37, No. 3. – P. 424-438.
143. Gürkaynak R.S., Levin A.T., Marder A.N., Swanson E.T. Inflation Targeting and the Anchoring of Inflation Expectations in The Western Hemisphere // Working Papers Central Bank of Chile 400, Central Bank of Chile. – 2006.

144. Gurkaynak R.S., Sack B., Swanson E. Market-Based Measures of Monetary Policy Expectations // Federal Reserve Bank of San-Francisco, Working Paper Series. – WP 2006-04. – 2006.
145. Hafer R.W., Hein S.E. On the accuracy of time-series, interest rate, and survey forecasts of inflation // The Journal of Business. – 1985. – Vol 58. – P. 377-398.
146. Hartley, J. E. Retrospectives: The Origins of the Representative Agent // Journal of Economic Perspectives. – 1997. – Vol. 10 (2). – P. 169–177.
147. Huang Allen, Reuven Lehavy, Amy Zang, Rong Zheng. Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach // Management Science. – 2017.
148. Khan M.S. The Variability of Expectations in Hyperinflations // Journal of Political Economy. 1977. – Vol. 85(4). – P. 817-827.
149. Kitous A., Saveyn B., Keramidas K., Vandyck T., Luis Rey Los Santos, Wojtowicz K. Impact of low oil prices on oil exporting countries // JRC Science for Policy Report, European Commission, Joint Research Centre. – 2016. – EUR 27909 EN.
150. Kliesen K.L. How Accurate Are Measures of Long-Term Inflation Expectations? // Economic Synopses. – 2015. – No. 9, 2015-05-06.
151. Kowalski T. Wprowadzenie oczekiwan podmioto w gospodarczych do analizy polityki makroekonomicznej // Ruch Prawniczy, Ekonomiczny i Socjologiczny. – 1998. – Vol. 1.
152. Kuttner K. Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market // Journal of Monetary Economics. – 2001. – P. 523-544.
153. Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root // Journal of Econometrics. – 1992. – Vol. 54. – P. 159–178.
154. Larsen V., Thorsrud L. The Value of News // BI Norwegian Business School, Working Papers from Centre For Applied Macro- and Petroleum economics (CAMP). – No.6. – 2015.



155. Lewis M., McDermott J., Richardson A. Inflation expectations and the conduct of monetary policy in New Zealand // Reserve Bank of New Zealand. – 2016. – Bulletin Vol. 79, No. 4.
156. Lines M., Westerhoff F. Inflation Expectations and Macroeconomic Dynamics: The Case of Rational Versus Extrapolative Expectations // Journal of Economic Dynamics and Control. – 2009.
157. Loley A., Gurov I. The Process of Formation of Inflation Expectations in an Information Economy // IFC Bulletin, Vol. 34. – 2011.
158. Lovell M.C. Tests of the Rational Expectations Hypothesis // The American Economic Review. – 1986. – 76(1). – P. 110-124.
159. Mankiw G., Reis R., Wolfers J. Disagreement about inflation expectations // NBER Macroeconomics Annual 2003. – 2003. – Vol. 18. – P. 209-248.
160. Metzler L.A. The nature and stability of inventory cycles // The Review of Economics and Statistics. – 1941. – Vol. 23, No. 3. – P. 113-129.
161. Mills E.S. The Use of Adaptive Expectations in Stability Analysis: A Comment // The Quarterly Journal of Economics. – 1961. – 75(2). – P. 330-335, Oxford Journals.
162. Mlambo L. Adaptive Price Expectations and the Modelling of World Gold Price Dynamics // International Journal of Management, IT and Engineering. – 2011. – Vol. 2(9). – P. 1-23.
163. Mlambo L. Adaptive and Rational Expectations Hypotheses: Reviewing the Critiques // Journal of Economic Behavior. – 2012. – Vol. 2. – P. 3-15.
164. Muellbauer J. Are Employment Decisions Based on Rational Expectations? // Journal of Econometrics. – 1981. – Vol. 16(1). – P. 156.
165. Murphy R.G., Rohde A. Rational Bias in Inflation Expectations // Eastern Economic Journal. – 2015. DOI 10.1057/ej.2015.50.
166. Muth J. F. Rational expectations and the theory of price movements // Econometrica. – 1961. – Vol. 29. – P. 315-335.
167. Nerlove M. Adaptive expectations and cobweb phenomena // Quarterly Journal of Economics. – 1958. – Vol. 72. – P. 227-240.

168. Nyman R., Kapadia S., Tuckett D., Gregory D., Ormerod P., Smith R. News and narratives in financial systems: exploiting big data for systemic risk assessment // Bank of England, Staff Working Paper. – 2018. – No. 704.
169. Orphanides A., Williams J.C. Inflation Scares and Forecast-Based Monetary Policy // Review of Economic Dynamics. – 2005. – Vol. 8(2). – P. 498–527.
170. Poole W., Rasche R.H. Perfecting the Market’s Knowledge of Monetary Policy // Journal of Financial Services Research. – 2000. – P. 255-298.
171. Rigobon R., Sack B. The Impact of Monetary Policy on Asset Prices // National Bureau of Economic Research Working Paper 10547. – 2002.
172. Roondiwala M., Patel H., Varma S. Predicting Stock Prices Using LSTM // International Journal of Science and Research. – 2015. – Vol. 6, Issue 4.
173. Sargent T., Wallace N. Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule // Journal of political economy. – 1975. – Vol. 83, No. 2. – P. 241-254.
174. Schmidt T., Vosen S. A Monthly Consumption Indicator for Germany Based on Internet Search Query Data // Ruhr Economic Paper. – 2010. – No 208.
175. Shiller R. Narrative economics // Cowles Foundation Discussion Paper, No. 2069. – 2017.
176. Sibanda K., Hove P., Murwirapachena G. Oil Prices, Exchange Rates, And Inflation Expectations In South Africa // International Business & Economics Research Journal. – 2015. – Vol. 14, No. 4.
177. Silveira J. J., Tadeu Lima G. Heterogeneity in Inflation Expectations and Macroeconomic Stability under Satisficing Learning // Working Papers, Department of Economics 2014\_28, University of São Paulo (FEA-USP). – 2014.
178. Söderlind P. Inflation Risk Premia and Survey Evidence on Macroeconomic Uncertainty // University of St.Gallen, Department of Economics Discussion Paper No. 2008-11, International Journal of Central Banking. – 2008.
179. Soo C. K. Quantifying Animal Spirits: News Media and Sentiment in the Housing Market // Ross School of Business Working Paper, No. 1200. – 2013.

180. Tetlock, P. C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market // *The Journal of Finance*. – 2007. – Vol. 62(3). – P. 1139–1168.
181. Thomas, L.B. Jr. Survey measures of expected US inflation // *Journal of Economic Perspectives*. – 1999. – Vol 13. – P. 125–144.
182. Thorsrud A. Words are the new numbers: A newsy coincident index of business cycles // *Norges Bank Research. Working Paper*. – February, 2016.
183. Trivedi P.K. Retail inventory investment behaviour // *Journal of Econometrics*. – 1973. – Vol. 1. – P. 61-80.
184. Turnovsky S.J. Empirical evidence on the formation of price expectations // *Journal of the American Statistical Association*. – 1970. – Vol. 65. – P. 1441-1454.
185. Van der Klaauw W., Bruine de Bruin W., Topa G., Potter S.M., Bryan M.F. Rethinking the measurement of household inflation expectations: preliminary findings // *Staff Reports 359, Federal Reserve Bank of New York*. – 2008.
186. Wible R.J. The rational expectations tautologies // *Journal of Post Keynesian Economics*. – 1982. – Vol. 5, No. 2. – P. 199-207.
187. Tyszka T. *Psychologia zachowan ekonomicznych*. Warszawa: PWN. – 1997. – Roz.1. – S. 15-31.
188. Weizhong Zhao, James J Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, Wen Zou. A heuristic approach to determine an appropriate number of topics in topic modeling // *BMC Bioinformatics*. – 2015. – 16(Suppl. 13):S8.

## Приложение 1

Скрипт для отбора новостных статей СМИ и комментариев интернет-пользователей в целях расчета индикатора инфляционных ожиданий российского населения. Язык программирования: Python.

Набор данных: новостные статьи СМИ (Эксперт, Коммерсант, Ведомости, РБК) и комментарии интернет-пользователей к ним на официальных сайтах СМИ, в социальных сетях Facebook, Вконтакте.

Период: январь 2014 г. – февраль 2018 г.

```
from IPython import get_ipython
import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sns
from tqdm import tqdm, tqdm_notebook
import pandas as pd
import numpy as np

import re
import json
import os

# Загрузка данных. Необходимо предварительно определить параметр INPUT_FILENAME
data = pd.read_csv('/data/' + INPUT_FILENAME, encoding='utf-8')
data['date'] = pd.to_datetime(data['date'])
data.drop_duplicates(inplace=True)
data.index = data['date']

data.index.min(), data.index.max()
## (Timestamp('2014-01-01 00:06:43'), Timestamp('2018-02-28 23:49:02'))

# Отбор новостных статей СМИ
regexr = re.compile(r'(инфляц|ипц|((рост|повыс|повыш|ниж|низ).*цен)|(темп.*цен)|
(индекс.*потребительских.*цен ))', re.I+re.U)
data['I_patterns+'] = data['text'].apply(lambda text: regexr.search(str(text)) is not None)
relevant_data_articles = data[ (data['I_patterns+']==True) & ((data['contentType']=='ARTICLE') |
(data['contentType']=='PREVIEW'))]
relevant_data_articles = relevant_data_articles['parentId'].values

# Отбор комментариев интернет-пользователей к выбранным новостным статьям СМИ
relevant_data_comments = data[ (data['contentType']=='COMMENT') &
(data['parentId'].isin(relevant_data_articles)) ].copy()

key_words1 = ['инфляц','цен']
key_words2 = ['выс','выш','рост','раст']
relevant_data_comments['I_inflation+ (pattern marked)'] = relevant_data_comments['text'].apply(lambda text :
any(word in str(text) for word in key_words1)
and any(word in str(text) for word in key_words2))

target_data = relevant_data_comments[ relevant_data_comments['I_inflation+ (pattern marked)']==True]
```

```

# Отбор новостных статей с упоминаниями инфляционных ожиданий
relevant_data_articles_target = data[ (data['I_patterns+']==True) & ((data['contentType']=='ARTICLE') |
(data['contentType']=='PREVIEW'))]
relevant_data_articles_target['I_inflation+ (pattern marked)'] = relevant_data_articles_target['text'].apply(lambda
text : any(word in str(text) for word in key_words1)
and any(word in str(text) for word in key_words2))

target_data_articles = relevant_data_articles_target[ relevant_data_articles_target['I_inflation+ (pattern
marked)']==True]

# Получение временных рядов. Параметр FREQ может быть задан как '1D' (ежедневные данные) и '1M'
(ежемесячные данные)
ts_target = target_data.groupby(pd.TimeGrouper(freq=FREQ)).count()['I_inflation+ (pattern marked)']
ts_target_articles = target_data_articles.groupby(pd.TimeGrouper(freq=FREQ)).count()['I_inflation+ (pattern
marked)']

```

*Примечание:* # - комментарий к выполнению кода, ## - результат исполнения кода для отдельных операций.

## Приложение 2

Скрипт для отбора комментариев интернет-пользователей в целях экспертного выделения факторов, с которыми российское население в значительной доле комментариев связывало изменение интенсивности инфляционных ожиданий, в рамках построения индикатора инфляционных ожиданий российского населения. Язык программирования: Python.

Набор данных: новостные статьи СМИ (Эксперт, Коммерсант, Ведомости, РБК) и комментарии интернет-пользователей к ним на официальных сайтах СМИ, в социальных сетях Facebook, Вконтакте.

Период: январь 2014 г. – февраль 2018 г.

```

import pandas as pd
import numpy as np
from tqdm import tqdm_notebook

# Загрузка данных. Необходимо предварительно определить параметр INPUT_FILENAME
data = pd.read_csv('/data/' + INPUT_FILENAME, encoding='utf-8')
data['date'] = pd.to_datetime(data['date'])
data.drop_duplicates(inplace=True)
data.index = data['date']

data.index.min(), data.index.max()
## (Timestamp('2014-01-01 00:06:43'), Timestamp('2018-02-28 23:49:02'))

import re

regexp_infup = re.compile(r'(инфляц|ипц|((рост|повыс|повыш|ниж|низ).*цен)|(темп.*цен)|
(индекс.*потребительских.*цен ))', re.I+re.U)
data_actual['I_patterns+'] = data_actual['text'].apply(lambda text: regexp_infup.search(str(text)) is not None)

relevant_data_articles = data_actual[ (data_actual['I_patterns+']==True) &
((data_actual['contentType']=='ARTICLE') | (data_actual['contentType']=='PREVIEW'))]

```

```

relevant_data_articles = relevant_data_articles['parentId'].values
relevant_data_comments = data_actual[ (data_actual['contentType']=='COMMENT') &
(data_actual['parentId'].isin(relevant_data_articles)) ].copy()

key_words1 = ['инфляц','цен']
key_words2 = ['выс','выш','рост','раст']
relevant_data_comments['I_inflation+ (pattern marked)'] = relevant_data_comments['text'].apply(lambda text :
any(word in str(text) for word in key_words1)
and any(word in str(text) for word in key_words2))

target_data = relevant_data_comments[ relevant_data_comments['I_inflation+ (pattern marked)']==True ]

regex_inf_pattern_oil = re.compile(r'(топлив|бенз|нефт|барр|brent|брен|опек|энерг|сырье|brent|urals|wti|добыч|
газ|ресурс|уголь)', re.I+re.U)
regex_inf_pattern_currency = re.compile(r'(долл|евро
\\$|курс|валют|рубл|девальв|обесцен|спекул|кэрри|кэрри|carry)', re.I+re.U)
regex_inf_pattern_cbr = re.compile(r'(процент| цб|банк россия|центробанк|цбр|ставка|ставок|рефинанс|репо |
регулят|набиул|юдаев|дкп|денежно-кредитн)', re.I+re.U)
regex_inf_pattern_crisis =
re.compile(r'(дефицит|банкрот|дефолт|производ|спрос|убыт|кризис|рецесс|потрясение|застой|спад|ввп|стагн|кр
ах|продаж|расход|промышл)', re.I+re.U)
regex_inf_pattern_uncertainty = re.compile(r'(неопред|неясн|непонятн|не
понятн|неоднозначн|нестабильн|противореч|неустойч|не
уверен|неуверен|запутан|двусмысл|опас|непредсказ|не предсказ|волатил)', re.I+re.U)
regex_inf_pattern_macropolicy = re.compile(r'(жкх|минпром|минфин|медведев
|бюрократ|закон|минист|минэк|конкур|
фас|Путин|путин|президент|импортзамещ|инициатив|бюджет|коррупц|воров|откат|институт|налог|власт|чин
овн|тариф|государств|правительств)', re.I+re.U)
regex_inf_pattern_external = re.compile(r'( мир|ае] |внешн|санкц|сша| ес
|евросоюз|еврозон|европ|запад|обама|меркель|германи|фрс|украин|крым| кита|великобр|америк|турц)',
re.I+re.U)
regex_dohod = re.compile(r'(пенс|индекс|зарп|зараб|зп|з/п)', re.I+re.U)

target_data['inf_pattern_oil'] = target_data['text'].apply(lambda text : regex_inf_pattern_oil.search(text) is not None)
target_data['inf_pattern_currency'] = target_data['text'].apply(lambda text : regex_inf_pattern_currency.search(text)
is not None)
target_data['inf_pattern_cbr'] = target_data['text'].apply(lambda text : regex_inf_pattern_cbr.search(text) is not
None)
target_data['inf_pattern_crisis'] = target_data['text'].apply(lambda text : regex_inf_pattern_crisis.search(text) is not
None)
target_data['inf_pattern_uncertainty'] = target_data['text'].apply(lambda text :
regex_inf_pattern_uncertainty.search(text) is not None)
target_data['inf_pattern_macropolicy'] = target_data['text'].apply(lambda text :
regex_inf_pattern_macropolicy.search(text) is not None)
target_data['inf_pattern_external'] = target_data['text'].apply(lambda text : regex_inf_pattern_external.search(text) is
not None)
target_data['inf_pattern_dohod'] = target_data['text'].apply(lambda text : regex_dohod.search(text) is not None)

# Запись файла
target_data[['inf_pattern_oil','inf_pattern_currency', 'inf_pattern_cbr', 'inf_pattern_crisis', 'inf_pattern_uncertainty',
'inf_pattern_macropolicy', 'inf_pattern_external',
'inf_pattern_dohod']].to_csv('../data/processed/inf_pattern_all.csv')

```

*Примечание:* # - комментарий к выполнению кода, ## - результат исполнения кода для отдельных операций.

## Приложение 3

Скрипт тематического моделирования целевых комментариев интернет-пользователей в целях выявления факторов изменения интенсивности инфляционных ожиданий населения во времени в рамках построения индикатора инфляционных ожиданий российского населения. Язык программирования: R.

Набор данных: новостные статьи СМИ (Эксперт, Коммерсант, Ведомости, РБК) и комментарии интернет-пользователей к ним на официальных сайтах СМИ, в социальных сетях Facebook, Вконтакте.

Период: январь 2014 г. – февраль 2018 г.

```
# -----
# Загрузка библиотек для моделирования
# -----
packages <- c('tidyverse', 'data.table', 'lubridate', 'text2vec',
             'stringr', 'LDavis', 'tm', 'topicmodels', 'doParallel',
             'scales', 'RColorBrewer', 'wordcloud', 'ldatuning',
             'SnowballC')
lapply(packages, require, character.only = TRUE)

# -----
# Функции
# -----
clean_text <- function(doc) {
  stopwords <- c(tm::stopwords("russian"), c("в", "без", "до", "из", "к", "на", "по", "о", "от", "перед", "при",
"через", "за", "над", "об", "под", "про", "для", "вблизи", "вглубь", "вдоль", "возле", "около", "вокруг",
"впереди", "после", "посредством", "в роли", "в зависимости от", "путём", "насчёт", "по поводу", "ввиду", "по
случаю", "в течение", "благодаря", "несмотря на", "спустя", "с", "из-под", "из-за", "по-над", "в отличие от",
"в связи", "как", "словно", "так как", "для того чтобы", "тоже", "зато", "чтобы", "также", "потому что", "и", "а",
", "что", "или", "но", "однако", "когда", "лишь", "едва", "где", "куда", "откуда", "столько", "настолько", "так",
"до такой степени", "до того", "такой", "как будто", "будто", "точно", "как бы", "если", "если бы", "коли",
"ежели", "несмотря на то", "хотя", "хоть", "пускай", "дабы", "с тем чтобы", "так что", "ли", "не", "какой"))
  doc <- tolower(doc) # приводим все к одному регистру
  doc <- enc2utf8(doc) # меняем кодировку
  doc <- removeWords(doc, stopwords) # удаляем служебные слова
  doc <- removePunctuation(doc) # удаляем знаки пунктуации
  doc <- str_replace_all(doc, "\\s+", " ") # удаляем лишние пробелы
  doc <- str_trim(doc, side = "both") # удаляем пробелы в начале и в конце строки
  doc
}

stem_text <- function(x) {
  x <- enc2utf8(x %>% str_to_lower)
  res <- system('mystem -cl -e cp1251', intern = TRUE, input = x)
  res <- paste(res, sep = ", collapse = ")
  res <- str_replace_all(res, "[^\\w][\\d\\.]+", " ")
  res <- gsub("[{}]", "", res)
  res <- gsub("(\\[[^ ]+)", "", res)
  res <- gsub("\\?", "", res)
  res <- gsub("\\s+", " ", res)
  res
}
```

```

prep_fun <- function(x) {
  iconv(x, to = "cp1251")
}

# without stemming
tok_fun = function(x) {
  word_tokenizer(enc2native(x))
}

# -----
# Загрузка данных
# -----
df <- fread('data/target_comments_all.csv', sep = ';') %>% mutate(date = dmy_hms(date))

df <- df %>% select(date, text) %>% bind_cols(data.frame(comment_id = 1:nrow(df))) %>%
  mutate(text = as.character(text))

# -----
# Препроцессинг и стемминг
# -----
# стемминг
df_stemmed <- df %>% group_by(comment_id) %>% mutate(text = stem_text(clean_text(text))) %>% ungroup

# удаление идентификаторов пользователей
df_stemmed <- df_stemmed %>% mutate(text = gsub('id\\d+\\w+', "", text),
  text = gsub('id\\d+', "", text))
df_stemmed <- df_stemmed %>% mutate(text = gsub('quot', "", text))

# saveRDS(df_stemmed, 'data/df_stemmed.rds')

# -----
# Тематическое моделирование
# -----
textdf <- df_stemmed %>% select(comment_id, text)

# формирование data.frame со всеми текстами
setDT(textdf)
setkey(textdf, comment_id)
set.seed(2018L)

it_train = itoken(textdf$text,
  tokenizer = tok_fun,
  ids = textdf$comment_id,
  progressbar = TRUE)

# создание словаря
vocab = create_vocabulary(it_train)
# vocab <- vocab %>% mutate(term = iconv(term, to = "cp1251"))

# сокращение словаря для редких выражений
pruned_vocab = prune_vocabulary(vocab,
  term_count_min = 5,
  doc_proportion_max = 1,
  doc_proportion_min = 0)

# создание матрицы «документ-слово» (document term matrix)
vectorizer = vocab_vectorizer(pruned_vocab)

# document term matrix
dtm <- create_dtm(it_train, vectorizer)

```



```

saveRDS(dtm, 'data/dtm.rds')

# -----
# Validation
# -----
full_data <- as.matrix(dtm)

full_data <- full_data[rowSums(full_data) > 0, ]
full_data <- as.DocumentTermMatrix(full_data, weighting = weightTf)

cores <- detectCores() - 1
cl <- makeCluster(cores)

system.time({
  tunes <- FindTopicsNumber(
    full_data,
    topics = c(1:10 * 10 + 100),
    metrics = c("Griffiths2004", "CaoJuan2009", "Arun2010"),
    method = "Gibbs",
    control = list(seed = 77),
    mc.cores = cl,
    verbose = TRUE
  )
})

stopCluster(cl)

FindTopicsNumber_plot(tunes)

# Валидация по параметру перплексии
burnin = 1000
iter = 1000
keep = 50
n <- nrow(full_data)
folds <- 5
k <- 5
splitfolds <- sample(1:folds, n, replace = TRUE)

cluster <- makeCluster(detectCores(logical = TRUE) - 1)
registerDoParallel(cluster)

clusterEvalQ(cluster, {
  library(topicmodels)
  library(ldatuning)
  library(text2vec)
})

candidate_k <- c(30, 40, 50, 60, 70, 80, 100, 120, 140, 160, 180, 200)
clusterExport(cluster, c("full_data", "burnin", "iter", "keep", "splitfolds", "folds", "candidate_k"))

# Параллелизация LDA
system.time({
  results <- foreach(j = 1:length(candidate_k), .combine = rbind) %dopar% {
    k <- candidate_k[j]
    results_1k <- matrix(0, nrow = folds, ncol = 2)
    colnames(results_1k) <- c("k", "perplexity")
    for(i in 1:folds){
      train_set <- full_data[splitfolds != i, ]
      valid_set <- full_data[splitfolds == i, ]

      fitted <- LDA(train_set, k = k, method = "Gibbs",
        control = list(burnin = burnin, iter = iter, keep = keep) )
    }
  }
})

```

```

    results_1k[i,] <- c(k, topicmodels::perplexity(fitted, newdata = valid_set))
  }
  return(results_1k)
}
})
stopCluster(cluster)

results_df <- as.data.frame(results)

# визуализация результатов
ggplot(results_df, aes(x = k, y = perplexity)) +
  geom_point() +
  geom_smooth(se = FALSE) +
  labs(x = "Число тем", y = "Мера перплексии")
# -----
# Тематическое моделирование по алгоритму LDA
# -----
# 100 тем по результатам валидации
topics_number <- 100

# установка начальной случайной величины для достижения воспроизводимости результатов LDA
set.seed(2345)

lda_model <- text2vec::LDA$new(n_topics = topics_number, doc_topic_prior = 0.05, topic_word_prior = 0.01)

doc_topic_distr <- lda_model$fit_transform(dtm, n_iter = 1000, convergence_tol = 0.001,
  check_convergence_every_n = 10)

# -----
# Результаты LDA
# -----
word_vectors <- lda_model$get_top_words(n = 50, lambda = 1)

# вектора слов для каждой темы
word_vectors_to_write <- data.frame(word_vectors)
names(word_vectors_to_write) <- paste0('topic_', 1:topics_number)
word_vectors_to_write <- word_vectors_to_write %>% mutate_all(funs(iconv(., to = "cp1251")))

# вероятность появления тем в документах
doc_topic_distr_to_write <- data.frame(date = df$date) %>% bind_cols(data.frame(doc_topic_distr))
names(doc_topic_distr_to_write) <- c('date', paste0('topic_', 1:topics_number))
Примечание: # - комментарий к выполнению кода.

```

## Приложение 4

Таблица 1. Статистические критерии взаимосвязи индикатора интенсивности инфляционных ожиданий российского населения на основе больших данных с предполагаемыми факторами его изменений по ежедневным данным

Временной лаг для факторов	Ключевая ставка Банка России (% годовых)	Цены на нефть Brent (долл./бар.)	Курс рубля к доллару (рублей за доллар)
Исходный ряд	K:0.37*** V:0.54***	K:-0.12*** V:0.23**	K:0.18*** V:0.40***
1 день	K:0.36*** V:0.54***	K:-0.11*** V:0.25**	K:0.18*** V:0.41***
2 дня	K:0.36*** V:0.54***	K:-0.11*** V:0.25**	K:0.17*** V:0.41***
3 дня	K:0.35*** V:0.54***	K:-0.11*** V:0.25**	K:0.17*** V:0.43***
4 дня	K:0.34*** V:0.54***	K:-0.10*** V:0.24**	K:0.17*** V:0.42***
5 дней	K:0.34*** V:0.54***	K:-0.10*** V:0.26**	K:0.16*** V:0.42***
6 дней	K:0.33*** V:0.54***	K:-0.10*** V:0.26**	K:0.16*** V:0.42***
7 дней	K:0.32*** V:0.55***	K:-0.10*** V:0.26**	K:0.15*** V:0.43***
8 дней	K:0.32*** V:0.55***	K:-0.09*** V:0.25**	K:0.15*** V:0.43***
9 дней	K:0.31*** V:0.55***	K:-0.09*** V:0.25**	K:0.15*** V:0.43***
10 дней	K:0.30*** V:0.55***	K:-0.09*** V:0.25**	K:0.14*** V:0.43***
11 дней	K:0.29*** V:0.55***	K:-0.09*** V:0.26**	K:0.14*** V:0.42***
12 дней	K:0.29*** V:0.56***	K:-0.08*** V:0.26**	K:0.13*** V:0.42***
13 дней	K:0.28*** V:0.56***	K:-0.08*** V:0.27**	K:0.13*** V:0.41***
14 дней	K:0.27*** V:0.56***	K:-0.08*** V:0.27**	K:0.12*** V:0.41***
15 дней	K:0.26*** V:0.56***	K:-0.07** V:0.28**	K:0.12*** V:0.41***
16 дней	K:0.26*** V:0.56***	K:-0.07** V:0.28**	K:0.11*** V:0.41***
17 дней	K:0.25*** V:0.55***	K:-0.07** V:0.28**	K:0.11*** V:0.41***
18 дней	K:0.25*** V:0.55***	K:-0.06** V:0.29**	K:0.10*** V:0.41***
19 дней	K:0.24*** V:0.55***	K:-0.06** V:0.28**	K:0.10*** V:0.41***
20 дней	K:0.23*** V:0.55***	K:-0.06** V:0.29**	K:0.09*** V:0.41***
21 день	K:0.23*** V:0.54***	K:-0.05* V:0.29**	K:0.09*** V:0.40***
22 дня	K:0.22*** V:0.54***	K:-0.05 V:0.29***	K:0.08*** V:0.40***
23 дней	K:0.22*** V:0.54***	K:-0.04 V:0.29***	K:0.08*** V:0.40***

24 дней	K:0.21*** V:0.54***	K:-0.04 V:0.29**	K:0.07*** V:0.39***
25 дней	K:0.20*** V:0.54***	K:-0.04 V:0.28**	K:0.07*** V:0.38***
26 дней	K:0.20*** V:0.54***	K:-0.03 V:0.28**	K:0.06** V:0.38***
27 дней	K:0.19*** V:0.53***	K:-0.03 V:0.28**	K:0.06** V:0.39***
28 дней	K:0.18*** V:0.53***	K:-0.02 V:0.28**	K:0.05* V:0.38***
29 дней	K:0.17*** V:0.53***	K:-0.02 V:0.29***	K:0.04* V:0.38***
30 дней	K:0.17*** V:0.52***	K:-0.01 V:0.30***	K:0.04 V:0.38***

*Примечания:* использованы следующие обозначения:

(а) для различных уровней значимости: \* – 10%; \*\* – 5%; \*\*\* – 1%;

(б) для различных статистических критериев взаимосвязи: К – коэффициент корреляции, В – критерий взаимной информации;

(в) значение каждого критерия получено для рядов индикатора интенсивности инфляционных ожиданий и каждого предполагаемого фактора его изменений в период с января 2014 г. по февраль 2018 г. Каждый фактор обозначен в названиях столбцов таблицы. В каждой строке таблицы факторы используются с определенным, указанным в первом столбце временным лагом;

*Источник:* составлено автором.

Таблица 2. Результаты теста Гренджера на причинно-следственную зависимость индикатора интенсивности инфляционных ожиданий от различных экономических параметров по ежедневным данным

Порядок для проведения теста Гренджера (временной лаг)	Индекс потребительских цен (г/г, %)	Темп роста цен на нефть Brent	Темп роста курса рубля к доллару
1 день	0.45 VAR: 0.30	0.07* VAR: 0.06*	0.04** VAR: 0.20
2 дня	0.79	0.07*	0.02**
3 дня	0.91	0.14	0.04**
4 дня	0.30	0.26	0.01***
5 дней	0.46	0.03**	0.02**
6 дней	0.61	0.05**	0.05**
7 дней	0.67	0.08*	0.07*
8 дней	0.48	0.11	0.08*
9 дней	0.56	0.11	0.12
10 дней	0.25	0.16	0.18
11 дней	0.20	0.18	0.16
12 дней	0.24	0.10*	0.08
13 дней	0.21	0.09*	0.06*
14 дней	0.26	0.12	0.01***
15 дней	0.37	0.14	0
16 дней	0.29	0.17	0.01
17 дней	0.30	0.21	0
18 дней	0.32	0.22	0
19 дней	0.06*	0.26	0
20 дней	0.06*	0.32	0
21 день	0.04**	0.39	0.01
22 дня	0.05**	0.18	0.01
23 дней	0.05**	0.21	0.01
24 дней	0.07*	0.08	0.01

25 дней	0.09*	0.04	0
26 дней	0.09*	0.02	0.01
27 дней	0.10*	0.03	0.01
28 дней	0.12	0.04	0.01
29 дней	0.08*	0.05	0.02
30 дней	0.37	0.14	0

*Примечания:*

(а) для различных уровней значимости использованы следующие обозначения: \* – 10%, \*\* – 5%; \*\*\* – 1%.

(б) в качестве результатов теста Гренджера приведены значения вероятностей принятия нулевой гипотезы об отсутствии причинно-следственной связи интенсивности инфляционных ожиданий и предполагаемых факторов. Ряды являются стационарными. Для проверки стационарности использованы тесты Дики-Фуллера и тест KPSS.

(в) значение каждого критерия получено для предположения о зависимости рядов каждого предполагаемого фактора и индикатора интенсивности инфляционных ожиданий в период с февраля 2014 г. по февраль 2018 г. Период уменьшен за счет приведения рядов к стационарной форме. Каждый фактор обозначен в названиях столбцов таблицы. В каждой строке таблицы факторы используются с определенным, указанным в первом столбце временным лагом;

(г) VAR обозначает результаты модифицированного робастного теста Гренджера для простых попарных векторных авторегрессий с оптимальным лагом, определенным на основе критерия Акаике. Результаты приведены только в ячейках таблицы, соответствующих оптимальным лагам модели.

*Источник:* составлено автором.