

Прикладная эконометрика, 2026, т. 81, с. 26–45.

Applied Econometrics, 2026, v. 81, pp. 26–45.

DOI: 10.22394/1993-7601-2026-81-26-45

EDN: BAMTKX

Ф.И. Миничев¹

Потенциал использования текстовых данных для оценки инфляционных ожиданий в России

В статье предлагается методология измерения инфляционных ожиданий фирм в России на основе анализа текстовых данных крупнейших СМИ. Из социальной сети ВКонтакте выгружены более 2.3 млн новостных публикаций 13 российских изданий за период с января 2015 по октябрь 2024 г. Для выделения публикаций, посвященных именно инфляции, а не просто содержащих соответствующие ключевые слова, применяется модель латентного размещения Дирихле (LDA) — метод вероятностного тематического моделирования. Построенный индикатор интенсивности инфляционных ожиданий демонстрирует корреляцию 78% с ценовыми ожиданиями предприятий Банка России, что существенно выше корреляции с ожиданиями населения (48%). С помощью LASSO-регрессии установлено, что наибольший вклад в объяснение динамики ожиданий фирм обеспечивают три издания экономической направленности — «Коммерсантъ», «Ведомости» и «Эксперт», которые совместно объясняют 75% дисперсии ценовых ожиданий предприятий. Результаты подтверждают, что текстовые данные СМИ могут служить эффективным дополнением к опросным методам оценки инфляционных ожиданий.

Ключевые слова: инфляционные ожидания; обработка естественного языка; денежно-кредитная политика; средства массовой информации; инфляционное таргетирование.

JEL classification: C55; D84; E31; E37; E52.

1. Введение

В результате перехода к политике инфляционного таргетирования (ИТ), который окончательно завершился в 2015 г., основной целью Банка России (БР) стало поддержание постоянной годовой инфляции вблизи 4%. Классика литературы об ИТ свидетельствует о том, что управление инфляционными ожиданиями (ИО) — ключевой элемент при политике таргетирования инфляции (Svensson, 2010). Сотрудники Международного валютного фонда считают, что важность управления ИО сопоставима с важностью наличия номинального таргета при режиме ИТ (Adrian et al., 2018).

Инфляционные ожидания влияют на эффективность денежно-кредитной политики (ДКП) через несколько каналов. Они оказывают воздействие на потребительские решения домохозяйств и ценообразование фирм, отражают доверие экономических агентов к денежным властям. Их важность особенно высока при проведении ДКП в развивающихся странах

¹ Миничев Фёдор Ильич — ПАО «Сбербанк России», Москва; fedor.minichev@gmail.com.

и во время кризисов, когда незаякоренные ИО могут существенно ограничить возможности денежных властей и эффективность проводимой политики. Все это подчеркивает актуальность изучения ИО в России, и исследователи Банка России указывают, что «основным фокусом ДКП на предстоящем периоде должно быть заякоривание инфляционных ожиданий» (Смирнова, 2023, с. 5).

Исходя из важности ИО для достижения целевых ориентиров ДКП, возникает методологически сложная задача корректного измерения ИО. Для определения ИО пользуются двумя традиционными методами: проведение опросов и использование привязанных к ИО финансовых инструментов (например, государственных облигаций). В последнее время, благодаря увеличению объема больших данных, развитию методов работы с ними и росту вычислительных мощностей появляется новый способ оценки ИО на основе текстовых интернет-данных. Преимуществом такого способа, дополняющего традиционные методы, выступают низкие затраты оценки ИО и возможность отслеживать их в реальном времени, по дням или неделям. При этом, имея высокую частоту и легкость сбора, свойственные оценкам, полученным на основе анализа финансовых инструментов, исследователь может рассчитывать на точность, сравнимую с оценками по опросным данным.

Обычно под текстовыми данными понимаются посты и комментарии в социальных сетях. Текстовые данные, используемые при построении индикатора ИО в России, сужаются до новостных публикаций крупнейших СМИ в социальной сети Вконтакте (постов). Существует несколько причин, по которым новостные публикации считаются не только важным текстовым источником информации об ИО, но и достаточным для построения показателя интенсивности инфляционных ожиданий.

Во-первых, ИО формируются на основании воспринимаемой экономической картины мира, которую в большой степени формируют СМИ. Во-вторых, СМИ пишут о важных для читателя темах, и рост публикаций на заданную тематику может не только увеличивать интерес к заданной теме, но и быть индикатором этого интереса. Тем самым СМИ не только формируют картину мира, но и отражают ее. В-третьих, потенциальные дополнительные источники текстовых данных, комментарии и публикации пользователей социальных сетей не содержат много информации. В таких текстах большая доля ботов, бессмысленных сообщений и споров пользователей между собой. Расширение выборки такими данными внесет очень много шума и лишь исказит оценки ожиданий, а определение, что комментарий написан ботом, заслуживает отдельного исследования.

Исходя из потенциального качества используемых данных, текстовые данные трактуются в нашем исследовании довольно узко, и возникает вопрос, связанный с формированием ожиданий у разных групп населения. Чьи инфляционные ожидания будут отражать индикатор, основанный на текстовых данных крупнейших СМИ (в том числе экономических)? Предположительно, текстовые СМИ формируют экономическую картину мира у профессионалов и экспертов, заинтересованных в подобной публицистике. Население же, вероятно, придает больший вес при формировании ожиданий другим факторам, например, наблюдаемому росту цен на «товары-маркеры» (Грищенко и др., 2023).

Может возникнуть опасение, что текстовый индикатор ИО, по сути отражающий интенсивность обсуждения инфляции в СМИ, не является однозначным свидетельством в пользу роста ИО. СМИ по каким-то иным причинам могут уделять инфляции большее внимание. Например, к этому их может подталкивать прошлая высокая инфляция, изменение геополитической ситуации, налоговые поправки и прочее. Однако все подобные факторы, которые

могут привести к большему обсуждению инфляции в СМИ, приводят к росту фактической инфляции и ИО, поэтому гипотеза о связи интенсивности обсуждения инфляции и ИО является теоретически обоснованной.

Другим релевантным соображением является то, что разные СМИ с разной успешностью формируют и отражают экономическую картину мира населения и профессионалов. Вероятно, СМИ с большим акцентом на экономические публикации могут отражать динамику интенсивности инфляционных ожиданий лучше, чем остальные.

Исходя из этих рассуждений, формулируются *гипотезы* исследования.

1. Индикатор интенсивности инфляционных ожиданий, основанный на текстовых данных крупнейших СМИ, будет в большей степени отражать инфляционные ожидания фирм, чем населения.

2. Лучше всего отражают динамику инфляционных ожиданий публикации СМИ, акцент которых сделан на бизнесе и экономике.

Исследование построено следующим образом. Во втором разделе обсуждаются теории формирования инфляционных ожиданий, подчеркивается роль СМИ, что мотивирует использовать новостные публикации для более точного понимания динамики ожиданий. В третьем разделе представлены обзор и сравнение методов измерения инфляционных ожиданий на основе опросов, финансовых инструментов и больших текстовых данных. Описание методологии построения показателя интенсивности ИО, интерпретация результатов и проверка гипотез представлены в четвертом разделе. В заключении подводятся итоги и выделяются направления для дальнейших исследований.

2. Теории формирования инфляционных ожиданий

В данном разделе будут рассмотрены теории, объясняющие формирование инфляционных ожиданий. Это позволит обосновать использование текстовых данных для измерения ИО и соотнесет теорию ИО с гипотезами.

Одним из первых подходов к моделированию ИО считается теория адаптивных ожиданий Фридмана (Friedman, 1968). До этого инфляционные ожидания в макроэкономике практически игнорировались. Базовая макроэкономическая модель, дающая экономической политике выбирать между инфляцией и безработицей — кривая Филлипса — предполагала постоянные инфляционные ожидания (Phillips, 1958). Фридман показал, что в долгосрочном периоде выбор между инфляцией и безработицей недоступен политикам именно из-за того, что ИО меняются во времени, поскольку формируются на основании предыдущих значений инфляции. Если политики для достижения низкой безработицы допустят высокую инфляцию, то это приведет к росту ИО и, как следствие, сдвигу кривой Филлипса. Учет ИО усовершенствовал кривую Филлипса.

Теория адаптивных ожиданий подразумевает, что ИО зависят только от прошлых значений инфляции. Однако в данных наблюдается наличие других факторов формирования ИО, что породило альтернативные теории.

Одним из ключевых направлений в моделировании ИО является теория рациональных ожиданий. В 1970-е годы критика Лукаса (или критика рациональных ожиданий) показала, что существовавшие на тот момент макроэкономические модели не подходили для анализа последствий экономической политики, т.к. новые меры политики приводят к новым

зависимостям в экономике (Lucas, 1972). Это привело к существенному изменению методологии построения макроэкономических моделей и активному использованию микроэкономических обоснований решений агентов в экономике.

Наиболее стандартная теория полной информации и рациональных ожиданий (full information rational expectations, FIRE), с которой удобно сравнивать дальнейшие модификации и отклонения от предпосылок, подразумевает следующее. Агенты обладают полной информацией о макроэкономических переменных и их истории, и на этой основе делают оптимальные прогнозы. Иначе говоря, нет систематических отклонений прогнозов агентов от прогнозов модели. Это строгие, хотя и нереалистичные предпосылки, однако теория полной информации и рациональных ожиданий легла в основу активно используемой в Центральных Банках новокейнсианской кривой Филлипса (Galí, Gertler, 1999) — модели, основанной на микроуровневых обоснованиях решений и имеющей понятную интерпретацию, что обеспечило популярность модели FIRE.

Можно выделить два основных направления теорий, модифицирующих модель полной информации и рациональных ожиданий — ослабление предположения о полной информации и ослабление предпосылки о рациональных ожиданиях.

Диагностические ожидания — один из основных подходов к моделированию иррациональных ожиданий наряду с уже рассмотренными адаптивными ожиданиями. Одна из эвристик, которая была представлена в работе (Tversky, Kahneman, 1974), заключалась в том, что люди переоценивают вероятность «репрезентативных» событий. В теории нерациональных ожиданий это может означать следующее. Получая информацию о новостях, которые повышают вероятность каких-то возможных событий, индивиды начинают чересчур переоценивать вероятность этих событий, что приводит к неоптимальным прогнозам (Gennaioli, Shleifer, 2010; Bordalo et al., 2018). Теория диагностических ожиданий на индивидуальном уровне находит подтверждение в данных (Bordalo et al., 2020).

Другой пласт исследований рассматривает модели, ослабляющие предположение о наличии полной информации. Идея, что информация у агентов может не быть полной, высказывалась еще в 1977 г. (Lucas, 1977), но начала активно применяться в моделях только в начале 2000-х гг. Модели с информационными фрикциями (information frictions) содержат предпосылку, что в каждом периоде только доля агентов получает новую информацию об экономике (Mankiw, Reis, 2002).

Модификация модели с информационными фрикциями, актуальная в контексте обсуждения роли публикаций СМИ в формировании ИО — это эпидемиологическая модель (Cagroll, 2003). Основной вывод таких моделей следующий: люди формируют ИО, исходя из статей в СМИ, а не из новой макроэкономической статистики или правильной модели функционирования экономики. Процесс формирования ожиданий устроен таким образом. СМИ распространяют консенсусные ИО экспертов, которыми «заражаются» домохозяйства с определенной вероятностью, затем придерживаются таких ИО до следующего «заражения». Так в теоретической модели формирования ИО возникают СМИ. Этот подход используется во многих эмпирических работах. Lei et al. (2015), используя эпидемиологическую модель, показывают, что СМИ транслируют типичные ИО, и таким образом воздействуют на ИО в Китае. В Индии ИО формируются на основе информации, полученной из интернета (Saakshi et al., 2020). Работа экономистов Федеральной резервной системы США показывает, что ИО из социальных сетей значимо влияют на индивидуальные ИО (Garcia-Lembergman et al., 2023).

Важным фактором при формировании ИО оказывается персональный опыт и воспринимаемая инфляция, что сказывается на наблюдаемой гетерогенности ИО и систематических различий ожиданий в зависимости от пола, возраста, дохода и образования (D'Acunto et al., 2023; Malmendier, Nagel, 2016). В случае с ИО и воспринимаемой инфляцией могут возникнуть сложности в определении причины и следствия, но эксперименты демонстрируют, что воспринимаемая инфляция казуально влияет на ИО (Huber et al., 2023). Также, на воспринимаемую инфляцию и ИО могут иметь значительное влияние товары-маркеры (Грищенко и др., 2023).

Теоретически может существовать канал косвенного влияния СМИ на ИО через влияние на воспринимаемую инфляцию, которая по итогу влияет на ожидания. Исследование на опросных данных немецких домохозяйств подтверждает наличие этого канала (Conrad et al., 2022). К таким же выводам приходит Dräger (2015), исследующий связь воспринимаемой инфляции, ИО и СМИ в Швеции.

В российских условиях есть свидетельства, что новости об инфляции значимо влияют на воспринимаемую инфляцию и инфляционные ожидания населения (Евстигнеева, Карпов, 2023).

Обосновано ли использование новостей СМИ и текстовых данных для анализа инфляционных ожиданий фирм? Вероятно, ответ на этот вопрос варьируется в зависимости от экономической ситуации. Фирмы не обращают внимания на инфляцию и связанные с ней новости, если она их серьезно не затрагивает, это иллюстрирует концепцию рациональной невнимательности (Coibion et al., 2018). Как итог — в развитых странах с низкой и стабильной инфляцией фирмы не уделяют особого внимания инфляции и денежно-кредитной политике (Candia et al., 2023). Однако в странах, где денежные власти еще не закрепили успех в достижении целевых ориентиров политики, инфляция касается фирм, и они реагируют на новости о ней. Работа (Candia et al., 2023) показывает, что в Израиле в период высокой и волатильной инфляции новости о макроэкономике (в частности, инфляции) побуждали фирмы к обновлению своих инфляционных ожиданий.

На основе обзора приведенных исследований можно сказать, что современные теоретические и эмпирические работы рассматривают СМИ как существенный фактор, формирующий ИО. Поэтому можно предположить, что показатель интенсивности инфляционных ожиданий, основанный на публикациях СМИ, будет отражать динамику этих ожиданий.

3. Способы измерения инфляционных ожиданий

3.1. Традиционные методы измерения

Обычно определение ИО осуществляется с помощью двух стандартных методов — проведение опросов населения и фирм, а также использование финансовых инструментов, связанных к ИО (например, государственные облигации).

Анализ финансовых инструментов позволяет недорого и относительно просто получать высокочастотные оценки ИО, закладываемых в цену соответствующих финансовых инструментов. Ограничение данного подхода заключается в том, что исследователь получает оценку ожиданий исключительно участников финансового рынка. Кроме этого, оценка является зашумленной, т. к. на изменение цены влияют не только динамика ИО, но и изменения премий за риск и ликвидность (Haubrich et al., 2012; Gürkaynak et al., 2010).

Проведение опросов — другой традиционный метод для определения ИО населения и фирм. Важность ИО населения очень высока, поскольку именно они влияют на потребительские решения домохозяйств и являются одним из основных факторов инфляционной динамики. Центральные банки в большинстве стран мира оценивают ИО домохозяйств, проводя или заказывая регулярные опросы на больших репрезентативных выборках. В США ежемесячные опросы проводят New York Fed's Survey of Consumer Expectations и University of Michigan Survey of Consumers; в Великобритании исследовательская компания Ipsos по заказу Банка Англии ежеквартально проводит Inflation Attitudes Survey, а Банк Японии также ежеквартально предоставляет результаты проведенного Opinion Survey (Armantier et al., 2017). В России социологическая организация «инФОМ»² по заказу ЦБ на ежемесячной основе проводит опросы, определяя ИО и потребительские настроения российского населения. Таким же образом, с помощью опросов, оцениваются инфляционные ожидания предприятий.

Опросы для оценки инфляционных ожиданий населения в России проводятся следующим образом³. Проводится стандартизированный опрос населения по репрезентативной общероссийской выборке, где исследуемая генеральная совокупность — граждане РФ старше 18 лет. Количество респондентов — около 2000, число населенных пунктов — 105, технология опроса — интервью по месту жительства в режиме «face-to-face». В 2020 г. из-за пандемии COVID-19 вместо «face-to-face» были проведены телефонные опросы. Вопросы анкеты делятся на открытые и закрытые. Открытые вопросы предполагают, что респондент сам сформулирует вариант ответа, а закрытые дают выбрать из числа предложенных.

В закрытых вопросах для расчета медианных оценок происходит перекодировка ответов. Ответу «будут снижаться» присваивается значение –5. Ответу «останутся на нынешнем уровне / не изменятся» соответствует значение 0. Ответы «на 2% и менее» и «на 51% и более» кодируются как 1 и 75 соответственно. Для остальных ответов используются средние значения интервалов. Например, «на 31–40%» используется в расчетах как 35%. На основе этого рассчитывается медианное значение инфляционных ожиданий населения на год вперед.

В случае фирм методология измерения инфляционных ожиданий отличается. Данные о ценовых ожиданиях предприятий Банк России (БР) получает в ходе мониторинга предприятий — регулярного опроса предприятий нефинансового сектора экономики⁴. Опросы в каждом регионе проводят территориальные учреждения БР. В каждом регионе формируется многоступенчатая случайная выборка. Вся выборка включает в себя более 14 тыс. предприятий. В анкетах присутствуют вопросы закрытого типа, предполагающие качественную оценку. Для расчета ценовых ожиданий предприятий рассчитывается балансовый показатель ответов на вопрос «Как изменятся в следующие три месяца цены на готовую продукцию (услуги) предприятия?». Несмотря на прямую связь именно с ценами производителей, а не потребителей, полученная оценка ценовых ожиданий предприятий широко используется в литературе в качестве метрики инфляционных ожиданий, т.к. сильно коррелирована с динамикой индекса потребительских цен⁵. Корреляция между ценовыми ожиданиями предприятий

² <https://oirom.ru/infom>.

³ https://www.cbr.ru/Content/Document/File/59814/FOM_meth.pdf.

⁴ https://www.cbr.ru/Content/Document/File/130872/mm_br.pdf.

⁵ <https://www.cbr.ru/Collection/Collection/File/59322/1025.pdf>.

и динамикой ИПЦ (индекса, очищенного от сезонности, усредненного за три месяца) составляет 45% за период исследования (с января 2015 по октябрь 2024 г.).

При выборе способа оценки ИО, основанного на опросных данных, исследователь сталкивается с рядом трудностей. Во-первых, это наиболее дорогой способ оценки, подразумевающий достаточно высокие затраты на проведение опросов. Во-вторых, проведение корректного опроса, дающего ответы на вопросы исследователя, и составление репрезентативной выборки, отражающей мнение всего населения, являются сложными методологическими задачами. В-третьих, частотность таких опросов — ежеквартальная или ежемесячная, что не позволяет отслеживать ИО в реальном времени, в отличие от оценки, основанной на финансовых инструментах.

3.2. Метод измерения, основанный на текстовых данных

Альтернативой традиционным методам представляется методология, основанная на работе с текстовыми данными. Суть метода заключается в сборе текстов путем парсинга данных из социальных сетей или медиа, обработке этих текстовых данных и построении индикатора, отражающего ИО на основании того, что пишут в интернете.

Одной из первых таких работ на российских данных является статья (Голощапова, Андреев, 2017), в которой авторы использовали новости четырех крупнейших российских СМИ и комментарии пользователей в социальных сетях для оценки интенсивности ИО в краткосрочном периоде. Полученный индикатор показал высокую корреляцию с медианным индикатором численных ожиданий инфляции на 12 месяцев, который предоставляет «инФОМ», т. е. была продемонстрирована адекватность подхода. Однако построенная авторами система мониторинга не поддерживается начиная с 2019 г. Кроме этого, авторы отбирали посты и комментарии по словам и словосочетаниям «инфляция», «рост цен» и т. д., а затем для построения индикатора оценивали относительное число этих постов, что безусловно является достаточно зашумленной оценкой ожиданий общей инфляционной динамики. Теми же ограничениями обладает более современная работа (Петрова, 2022), где автор оценивает интенсивность ИО как долю публикаций со словами и словосочетаниями «инфляция», «ипц», «завышение цен», «цены на газ» и т. д. среди всех публикаций крупнейших новостных сообществ. Такой подход тоже оказывается адекватен — оцененные ИО являются причиной по Грейнджеру для инфляции. Из зарубежных работ схожую методологию и результаты можно встретить в (Aromí, Llada, 2020), где ИО, оцененные как доля постов в Твиттере⁶ Аргентины со словами «инфляция» или «инфляционные», вносят значимый вклад в прогнозирование инфляции наряду с другими макропоказателями.

Gabrielyan et al. (2020) предлагают более обоснованный и продвинутый метод оценки ИО на основе текстовых данных, что позволяет авторам, в отличие от представленных выше работ, разграничивать посты со словом «инфляция» и посты *про инфляцию*.

Более аккуратно и полно этот подход применяется в работе Банка Италии (Angelico et al., 2022), поэтому идеи метода проиллюстрируем именно на основе этой работы. Для нахождения публикаций про инфляцию авторы, собрав посты в Твиттере с подходящими словами и словосочетаниями, решают задачу вероятностного тематического моделирования,

⁶ Ограничен в доступе на территории РФ.

применяя модель латентного размещения Дирихле⁷. Такой метод позволяет определить, к какой тематике принадлежит документ на основе того, какие слова встречаются в нем чаще всего. Исследователи оставляют только те посты, тематика которых — именно инфляция и общая ценовая динамика. Затем авторы строят индекс ожиданий роста инфляции (доля постов с соответствующими словосочетаниями: «более дорогой», «высокие цены» и т.д.) и индекс ожиданий снижения инфляции (доля постов с соответствующими словосочетаниями: «низкие цены», «снижение цен», «дефляция» и т.д.). Итоговый индекс ИО есть разность двух индексов. Полученный индекс вносит значимый вклад в ожидания, заявляемые в рамках опросов. Такой метод оценки, так же как и прошлые методы, несложен в реализации и не несет финансовых затрат, но более обоснован и содержательно интерпретируем.

Схожей методологией пользуется Банк Франции при построении индикатора ИО на базе текстовых данных (Denes et al., 2022). Основное отличие от работы (Angelico et al., 2022) состоит в способе выявления постов *про инфляцию* среди собранных во французском Твиттере постов *со словом «инфляция»*. Здесь используется не латентное размещение Дирихле, а решается задача классификации. Исследователи отбирают 800 случайных постов и самостоятельно размечают, относятся ли они к общей ценовой динамике или нет. Затем алгоритм случайного леса обучается на этой выборке из 800 постов и используется для определения того, какие из остальных (неразмеченных) постов относятся к общей динамике цен, а какие — нет. Как и итальянские коллеги, авторы демонстрируют, что построенный индекс адекватен опросным ИО и фактической динамике инфляции.

Рассмотренные выше традиционные методы оценки ИО имеют ряд ограничений. Метод, основанный на финансовых инструментах, показывает только ожидания участников финансового рынка и не совсем точен из-за премий за ликвидность и риск. Метод, основанный на опросных данных, требует высоких затрат и не позволяет получить оценки ИО населения в реальном времени. Поэтому в данной работе будет применен приобретающий популярность метод, основанный на использовании текстовых данных из социальных сетей. Этот подход сочетает в себе низкие затраты, возможность следить за ИО в реальном времени и потенциальную точность оценок.

4. Методология построения и потенциал применения индикатора на текстовых данных

4.1. Методология построения

Для построения индикатора ИО из социальной сети Вконтакте выгружаются все новостные публикации 13 крупнейших российских СМИ за период с января 2015 по октябрь 2024 г., что значительно больше, чем в предыдущих российских исследованиях, использующих данные СМИ для решения такой же задачи. Так, Петрова (2022) использует 8 СМИ и не приводит размер выборки, а Голощапова, Андреев (2017) использовали 4 СМИ.

Суммарное количество загруженных публикаций, содержащих тексты, превышает 2.3 млн.

⁷ Решая задачу тематического моделирования, исследователь пытается определить, к каким темам относятся собранные документы. Модель латентного размещения Дирихле позволяет обнаружить тему документа, покрывающую большую часть слов в документе.

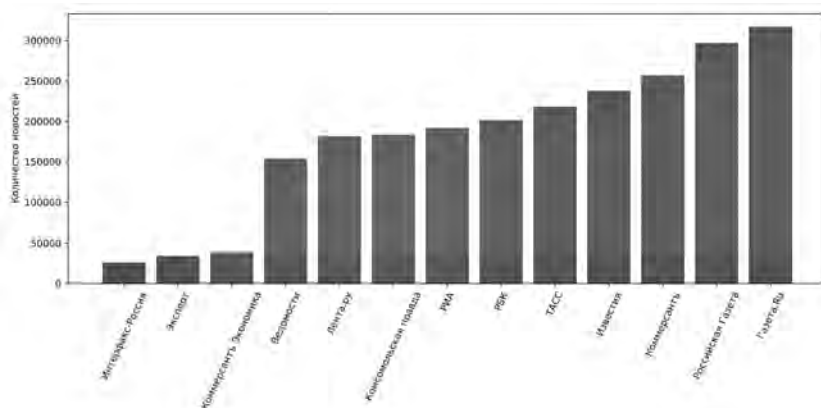


Рис. 1. Распределение СМИ по количеству постов за период 01.01.2015–31.10.2024

Распределение публикаций по СМИ ожидаемо неравномерно. Больше всего постов опубликовали журналисты *Газета.Ру*, а наименьшее количество публикаций содержится в журналах «Интерфакс-Россия», «Коммерсантъ Экономика» и «Эксперт» (рис. 1). На рисунке 2 представлена динамика количества загруженных постов.

Предобработка загруженных текстов состоит из нескольких этапов. Во-первых, удаляются пунктуация, переносы, интернет-ссылки, табуляции и символы из текстов, все заглавные буквы заменяются на строчные. Во-вторых, при дальнейшем анализе текстов будет важно, чтобы, например, «цена» и «цены» воспринимались как одно и то же слово. Поэтому проводится процедура лемматизации — приведение словоформы к лемме, нормальной (словарной) форме. Затем следует удаление «стоп-слов», которые очень часто встречаются в тексте, но не очень информативны: некоторые союзы, местоимения, предлоги и так далее. Пример изначальной новостной публикации и результата ее обработки приведен на рис. 3.

После обработки новостных постов следует решение задачи по нахождению постов про инфляцию. Сначала выборка сужается до постов, включающих хотя бы одно слово из следующего списка: «цена»; «инфляция»; «инфляционный»; «дефляция»; «дефляционный»; «ипц».

Итого за рассматриваемый период было получено почти 49 тыс. постов, т.е. в месяц выходит приблизительно 479 публикаций со словами из списка слов про инфляцию. Однако

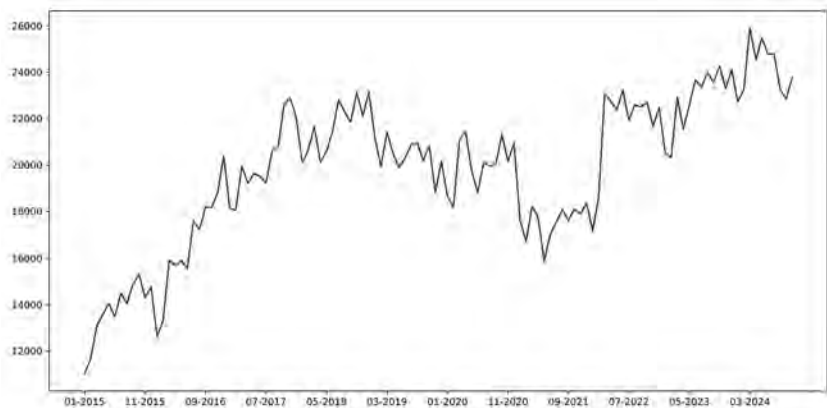


Рис. 2. Динамика количества постов за период 01.01.2015–31.10.2024

text	date	text_clean_lemma_wo_stop
Совет директоров российского Центробанка принял решение снизить с 2 мая ключевую ставку на 0,5 процентного пункта. Причина — снижение инфляционных ожиданий http://izvestia.ru/news/6952407 utm_source=VKutm_medium=ekonomika@izvestia	Fri, 28 Apr 2017 14:20:03 +0000	совет директор российский центробанк принять решение снизить 2 май ключевой ставка 0,5 процентный пункт причина снижение инфляционный ожидание

Рис. 3. Пример предварительной обработки текста

не все такие посты посвящены инфляции. Например, такие публикации, как «За год долларовые цены на вторичное жилье в Москве снизились на 47%» или «В России предложили установить минимальную цену на пиво» не несут информации об общей ценовой динамике. Это вынуждает прибегнуть к следующему методу, позволяющему выделить посты именно про инфляцию, а не просто со словом из вышеприведенного списка.

Для выделения текстов про инфляцию необходимо определить, к какой тематике относится каждый документ. Для решения этой задачи часто используется модель латентного размещения Дирихле (LDA) (Blei et al., 2003), позволяющая определить темы документов на основе слов, которые чаще всего в них встречаются. Идея модели заключается в следующем. Предполагается, что документ порождается смесью ключевых для него тем. Тема документа — та тема, к которой относятся слова, представленные в документе в наибольшей степени. Каждая тема — распределение слов. Исследователь сам выбирает, как интерпретировать получившиеся темы. Если наиболее вероятные слова в какой-то теме — «клюв», «птенец», «голубь», «ворона», то можно интерпретировать такую тему как тему о птицах, и, соответственно, классифицировать текст, в котором эта тема основная, как текст о птицах.

Число тем, порождающих документы в выборке, является гиперпараметром, который исследователь задает самостоятельно. На корпусе текстов строятся модели LDA с разным числом тем для того, чтобы выбрать их оптимальное количество. Критерием оптимальности выступает метрика согласованности — насколько схожи по смыслу слова, находящиеся внутри одной темы. Чем больше значение этой метрики, тем более интерпретируемы получившиеся темы с точки зрения человека. На основе этой метрики выбирается число тем, равное 7.

Из получившихся тем две охватывают документы, наиболее подходящие по содержанию на тексты про инфляцию. Наиболее вероятные слова в этих темах — «инфляция», «год», «рост», «россия», «цена», «вырасти». Остальные темы либо охватывают разнородные тексты, либо посвящены какой-то другой теме, а не инфляции. Так, ключевые слова одной из оставшихся тем — «нефть», «цена», «баррель», «газ», «бензин». Считаем постами про инфляцию посты, основная тема в которых — одна из двух выбранных нами.

Итоговый индикатор интенсивности инфляционных ожиданий — это отношение количества постов про инфляцию к общему количеству постов за рассматриваемый период⁸:

$$TextIndicator_t = \alpha \cdot \frac{numTextsInfl_t}{numTextsAll_t} + \beta, \quad (1)$$

где α и β — константы, вводимые для удобства сопоставимости индикатора с официальными показателями ИО Банка России и не влияющие на его динамику.

⁸ Если в заданный период времени, например месяц, было 1000 постов, из которых 100 про инфляцию, то отношение равно 0.1.

Нормировка производится, поскольку нас интересует относительная доля постов про инфляцию в общем информационном фоне. Нормировка именно на общее количество постов, а не количество постов про экономику, выбрана потому, что не добавляются модельные оценки там, где это может внести дополнительный шум и не является обязательным.

4.2. Связь индикатора с опросными показателями

Для оценки адекватности построенного индикатора оцениваются коэффициенты корреляции текстового индикатора месячной частоты с оценками инфляционных ожиданий населения и фирм на основе опросных данных. Медианные инфляционные ожидания населения на год вперед предоставляет ООО «инФОМ» по заказу БР, а ценовые ожидания предприятий оцениваются исходя из опроса, проводимого БР⁹.

С инфляционными ожиданиями населения корреляция индикатора равна 48%, а корреляция пятого лага — 58%.

Корреляция индикатора с инфляционными ожиданиями фирм равна 78%, что значимо больше корреляции с ИО населения¹⁰ (рис. 4). Тем самым подтверждена первая гипотеза о том, что тексты СМИ формируют (или отражают) картину мира скорее для профессионалов, чем для всего населения.

Насколько хорошо построенный индикатор объясняет динамику инфляционных ожиданий фирм? Для ответа на этот вопрос имеет смысл оценить, какую долю дисперсии ценовых ожиданий предприятий объясняет вариация в нашем текстовом индикаторе. Как известно, R^2 в парной регрессии равен квадрату коэффициента корреляции. Можно сделать вывод, что, используя только одну переменную (индикатор на основе частоты упоминания инфляции в постах СМИ ВКонтакте), можно объяснить почти две трети (61%) вариации ИО фирм.

Почти две трети объясненной вариации ценовых ожиданий фирм на фоне гораздо более низкого качества объяснения ИО населения — важный результат, предоставляющий

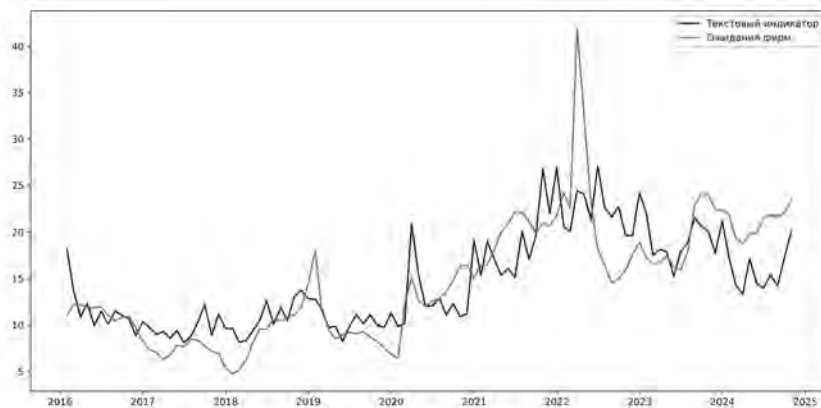


Рис. 4. Динамика текстового индикатора и показателя ценовых ожиданий фирм Банка России

⁹ https://www.cbr.ru/analytics/dkp/inflationary_expectations/.

¹⁰ Доверительные интервалы для двух полученных коэффициентов корреляции составляют (45; 68%) и (70; 84%).

новую информацию о том, как формируются ИО населения и фирм в России и что на них влияет (СМИ, собственный опыт или что-то еще). Насколько известно автору, это первый опыт применения текстовых данных для оценки инфляционных ожиданий фирм в России.

4.3. Разграничение СМИ

Для ответа на вопрос, какие СМИ в большей степени отражают динамику инфляционных ожиданий (проверка второй гипотезы), и построения более точного показателя имеет смысл разграничение изданий в выборке. На данном этапе исследования строится отдельный показатель интенсивности инфляционных ожиданий для каждого из 13 СМИ:

$$TextIndicator_{ij} = \alpha_j \cdot \frac{numTextsInfl_{ij}}{numTextsAll_{ij}} + \beta_j, \quad (2)$$

где j — индекс СМИ, остальные обозначения аналогичны соответствующим в формуле (1) для показателя на выборке всех интернет-изданий.

В качестве предварительного анализа полученных показателей рассчитываются коэффициенты корреляции каждого индикатора с показателем ценовых ожиданий предприятий БР (без лагов). Результаты представлены в табл. 1.

Издания, целевой аудиторией которых в первую очередь выступают профессионалы, эксперты и представители индустрий, демонстрируют большие коэффициенты корреляции по сравнению с более массовыми СМИ. Коэффициенты корреляции, рассчитанные для показателей изданий «Ведомости», «Эксперт» и «Коммерсантъ», превышают 0.65. Наименьшие коэффициенты корреляции получены для изданий «РБК», «Интерфакс-Россия», «Лента.ру». Единственный неожиданный результат — очень низкий коэффициент корреляции между показателем для «РБК» и ИО фирм. Интересно, что при этом «РБК» демонстрирует наибольшую связь с ИО населения среди всех СМИ — 0.54. Остальные СМИ имеют меньшую корреляцию с ИО населения и, преимущественно, коррелируют с ИО населения меньше, чем с ИО фирм. Остальные результаты соответствуют ожиданиям — наибольшие коэффициенты корреляции у изданий с бизнес-тематикой, а более массовые издания демонстрируют менее тесную связь с показателем ценовых ожиданий предприятий БР. Кроме этого, если учесть количество публикаций каждого СМИ, то можно предположить, что малое число публикаций затрудняет расчет адекватного показателя. Так, коэффициент корреляции у «Коммерсантъ» значительно выше, чем у «Коммерсантъ Экономика» — 0.79 и 0.62 соответственно. Вероятно, это можно объяснить относительно малым количеством публикаций у последнего (см. рис. 1). Кроме этого, один из наименьших коэффициентов у издания «Интерфакс-Россия», которое в наименьшей степени представлено в выборке.

На данном этапе на основании анализа коэффициентов корреляции можно говорить о подтверждении второй гипотезы.

Очевидно, показатель инфляционных ожиданий можно улучшить, если использовать одновременно несколько показателей СМИ (с их лагами), которые в наилучшей степени отражают ИО фирм. Кроме этого, такой подход позволит более обоснованно сделать выводы о СМИ, чья динамика публикаций об инфляции наиболее тесно связана с показателем ЦБ.

Таблица 1. Корреляция СМИ с ценовыми ожиданиями предприятий

Издание	Корреляция с ИО фирм
Коммерсантъ	0.79
Эксперт	0.71
Ведомости	0.66
Коммерсантъ Экономика	0.62
Комсомольская правда	0.60
Российская Газета	0.56
РИА	0.55
ТАСС	0.46
Известия	0.41
Газета.Ru	0.30
РБК	0.27
Интерфакс-Россия	0.25
Лента.ру	0.05

Построение регрессии с помощью обычного метода наименьших квадратов (МНК) в данной ситуации затруднительно. Во-первых, предстоит столкнуться с проблемой мультиколлинеарности: показатели, построенные для разных СМИ, должны быть тесно коррелированы между собой, что подтверждается расчетом коэффициентов VIF (variance inflation factor). Во-вторых, такое соотношение параметров и наблюдений приводит к «проклятию размерности», когда имеющейся «скромной» выборки явно недостаточно для корректной оценки большой модели.

Существует несколько потенциальных путей решения проблемы. Можно прибегнуть к методу главных компонент, позволяющему снизить число признаков и перейти к новым признакам, в наибольшей степени объясняющим дисперсию исходных переменных. Использование метода главных компонент не позволит ответить на вопрос о главных СМИ, но зато позволит увидеть, отражает ли общая компонента показателей разных СМИ динамику ИО фирм.

С помощью метода главных компонент число признаков, описывающих пространство, сокращается до двух. Две новые переменные описывают суммарно 81% дисперсии (73 и 8% по отдельности), а увеличение числа главных компонент не приводит к заметному росту объясненной дисперсии. В регрессии показателя ценовых ожиданий предприятий на эти две компоненты $R^2 = 71\%$, что говорит о тесной связи между официальным показателем БР и главными компонентами, характеризующими дисперсию показателей, построенных для разных СМИ:

$$\widehat{InflExpFirm}_i = 13.54 + 0.06 \underset{(0.8)}{Comp1}_i + 0.09 \underset{(0.005)}{Comp2}_i, \quad R^2 = 0.709. \quad (3)$$

Более подходящим в рамках задач выявления наиболее важных СМИ для динамики ИО и построения более точного показателя представляется семейство подходов, основанных на идее регуляризации. Регуляризация заключается в изменении функционала ошибки — к сумме квадратов остатков добавляется, например, сумма квадратов или модулей коэффициентов. Таким образом, к штрафу за ошибочные прогнозы добавляется штраф за чересчур большие параметры или их излишнее количество. Наиболее известные модели линейной

регрессии с регуляризацией — LASSO и RIDGE регрессии (L1 и L2 регуляризация соответственно). В первом случае в функции используется сумма модулей коэффициентов, во втором — сумма квадратов. Основное отличие между двумя моделями заключается в том, что LASSO регрессии гораздо более свойственно обнуление коэффициентов (оптимальные оценки коэффициентов при не важных переменных будут равны нулю). Это свойство делает ее более подходящей в контексте «проклятия размерности», которого желательно избежать, найдя только самые важные СМИ, наиболее тесно связанные с динамикой ИО фирм.

В LASSO регрессии решается оптимизационная задача:

$$\sum_{i=1}^n e_i^2 + \alpha \sum_{j=1}^k |\beta_j| \rightarrow \min, \quad (4)$$

где e_i — остатки уравнения, α — гиперпараметр, β_j — параметры регрессии (без константы), n — число наблюдений, k — число переменных.

Гиперпараметр α , определяющий относительный вес суммы модулей коэффициентов в штрафной функции, определяется исследователем самостоятельно. Рост этого гиперпараметра приводит к снижению R^2 и числа ненулевых оценок параметров. Модель становится менее точной, но более устойчивой (в терминах bias-variance tradeoff).

Оценивается LASSO регрессия, где в качестве объясняющих переменных используются индикаторы разных СМИ (13) и три их первых лага. Для того чтобы оценки коэффициентов корректно отражали веса переменных, все переменные предварительно стандартизированы (вычитается среднее и делится на стандартное отклонение). LASSO регрессия оценивается для разных значений α , для разных моделей рассчитываются R^2 и число ненулевых параметров. При $\alpha = 1.35$ достигается достаточно высокая объясняющая способность модели ($R^2 = 75\%$) при малом числе параметров (7 переменных с ненулевыми оценками). Выбранное значение гиперпараметра является «точкой перелома», поскольку меньшие значения приводят к росту числа ненулевых оценок параметров, что дает лишь незначительный рост R^2 . Так, при $\alpha = 0.3$ получается 15 ненулевых оценок параметров, а $R^2 = 80\%$. Результаты оцененной регрессии представлены в табл. 2.

Видно, что не обнуляются оценки параметров при показателях всего у трех СМИ из 13. Три четверти дисперсии инфляционных ожиданий фирм объясняется с помощью комбинации различных лагов и текущих значений показателей, построенных на данных изданий «Ведомости», «Эксперт» и «Коммерсантъ». Регрессионный анализ с использованием

Таблица 2. Ненулевые оценки коэффициентов в LASSO регрессии

Переменная	Оценка коэффициента
Коммерсантъ_lag_3	1.3
Коммерсантъ_lag_1	1.2
Коммерсантъ	1
Эксперт_lag_1	0.6
Коммерсантъ_lag_2	0.6
Эксперт	0.4
Ведомости_lag_3	0.07

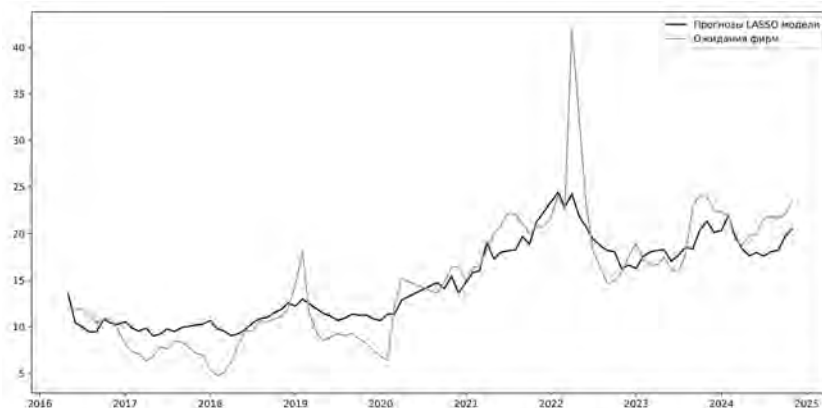


Рис. 5. Динамика прогнозов LASSO модели и показателя ценовых ожиданий фирм Банка России

регуляризации позволил подтвердить предварительные выводы: наиболее тесно с ИО фирм связаны показатели, основанные на СМИ, у которых акцент делается на публикациях о бизнесе и экономике.

Динамика прогнозов LASSO модели и фактического показателя Банка России представлена на рис. 5.

5. Заключение

Основным вопросом данного исследования является потенциал использования текстовых данных для оценки инфляционных ожиданий (ИО). Замена существующей общепринятой методики проведения опросов не кажется обоснованной. Опросные данные останутся наилучшим источником информации о ценовых ожиданиях. Они гораздо яснее интерпретируются, методика их получения уже отработана и применяется во множестве стран, они позволяют получить информацию не только об оценке ИО в среднем, но и об их разбросе. В конце концов, единственный способ в литературе проверить адекватность нового текстового индикатора — оценить его связь с опросным показателем.

Текстовый показатель вместо замены опросному показателю может пригодиться для его дополнения. Дело в том, что опросный показатель ИО доступен, как правило, в помесечной частоте, т.к. измерять ИО с помощью опросов достаточно дорого. Это является его главным недостатком, который частично может скомпенсировать предложенная методология. Текстовые данные могут пригодиться для интерполяции — оценки динамики ИО с большей частотой, например, еженедельной.

В настоящем исследовании предложена методология построения индикатора интенсивности ИО, который может объяснить три четверти вариации ИО предприятий в России, что является новым результатом. Кроме этого, показано, публикации каких СМИ наиболее тесно связаны с динамикой ценовых ожиданий предприятий.

В рамках исследования было собрано более 2 млн публикаций СМИ, что значительно больше, чем в аналогичных исследованиях, решающих такую же задачу. Построенный на их основе показатель продемонстрировал тесную связь с официальным показателем ИО,

подтвердив адекватность подхода. Предложена и опробована методология выделения новостных изданий, наиболее важных для динамики ожиданий. Также предложены потенциальные способы использования метода для анализа ИО фирм.

В исследовании собрана наибольшая база публикаций СМИ среди российских исследований, предлагающих оценки инфляционных ожиданий на основе текстовых данных. Использование модели LDA улучшило показатель. Кроме этого, насколько известно автору, впервые предложена методология выделения наиболее важных СМИ.

Теоретический результат заключается в подтверждении гипотезы, что профильные СМИ, пишущие про инфляцию, отражают или формируют картину мира скорее профессионалов (экспертов, предпринимателей, менеджмента фирм), чем всего населения. Более того, теснота связи с ИО фирм очень отличается между разными изданиями, и наиболее тесную связь демонстрируют медиа, ориентирующиеся на экономику: «Коммерсантъ», «Ведомости» и «Эксперт». В качестве практического применения построенного индикатора в исследовании предложено дополнение официального показателя БР с целью его интерполяции и получения дополнительной информации об ИО как одной из важнейших переменных для денежных властей.

Можно выделить несколько наиболее перспективных направлений дальнейших исследований. Во-первых, заслуживает внимания развитие методологии построения индикатора. Например, разметка публикаций про инфляцию и дальнейшее применение машинного обучения с учителем для нахождения публикаций про инфляцию. Во-вторых, применение предложенного подхода требует построения системы автоматического мониторинга, позволяющей в реальном времени наблюдать за ИО фирм. В-третьих, высокочастотные оценки ИО можно использовать для анализа характеристик ИО в России: заякоренность, адаптивность, рациональность и т. д. В-четвертых, интерес представляет построение показателей, учитывающих эмоциональный окрас новостей и оценка асимметрии в реакции ИО на новости в зависимости от их окраса.

Список литературы

- Голощапова И. О., Андреев М. Л. (2017). Оценка инфляционных ожиданий российского населения методами машинного обучения. *Вопросы экономики*, 6, 71–93. DOI: 10.32609/0042-8736-2017-6-71-93.
- Грищенко В., Гасанова Д., Фомин Е., Кореньяк Г. (2023). Идентификация товаров-маркеров и влияние их цен на инфляционные ожидания российских домохозяйств. *Серия докладов об экономических исследованиях* № 117. Банк России.
- Евстигнеева А., Карпов Д. (2023). Влияние негативных новостей на восприятие инфляции населением. *Серия докладов об экономических исследованиях* № 111. Банк России.
- Петрова Д. А. (2022). Оценка инфляционных ожиданий на основе интернет-данных. *Прикладная эконометрика*, 66, 25–38. DOI: 10.22394/1993-7601-2022-66-25-38.
- Смирнова Ж. И. (2023). Новые вызовы для денежно-кредитной политики. *Аналитическая записка*. Банк России.
- Adrian T., Laxton M. D., Obstfeld M. M. (2018). *Advancing the frontiers of monetary policy*. International Monetary Fund. DOI: 10.5089/9781484325940.071.
- Angelico C., Marcucci J., Miccoli M., Quarta F. (2022). Can we measure inflation expectations using Twitter? *Journal of Econometrics*, 228 (2), 259–277. DOI: 10.1016/j.jeconom.2021.12.008.

- Armantier O., Topa G., Van der Klaauw W., Zafar B. (2017). An overview of the Survey of Consumer Expectations. *Economic Policy Review*, 23 (2), 51–72.
- Aromí J. D., Llada M. (2020). Forecasting inflation with Twitter. *Economica*, 69, 64–85.
- Blei D. M., Ng A. Y., Jordan M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bordalo P., Gennaioli N., Shleifer A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, 73 (1), 199–227. DOI: 10.1111/jofi.12586.
- Bordalo P., Gennaioli N., Ma Y., Shleifer A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110 (9), 2748–2782. DOI: 10.1257/aer.20181219.
- Candia B., Coibion O., Gorodnichenko Y. (2023). The macroeconomic expectations of firms. In: *Handbook of Economic Expectations*, 321–353. Academic Press. DOI: 10.1016/B978-0-12-822927-9.00018-5.
- Carroll C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118 (1), 269–298. DOI: 10.1162/00335530360535207.
- Coibion O., Gorodnichenko Y., Kumar S. (2018). How do firms form their expectations? New survey evidence. *American Economic Review*, 108 (9), 2671–2713. DOI: 10.1257/aer.20151299.
- Conrad C., Enders Z., Glas A. (2022). The role of information and experience for households' inflation expectations. *European Economic Review*, 143. DOI: 10.1016/j.eurocorev.2021.104015.
- D'Acunto F., Malmendier U., Weber M. (2023). What do the data tell us about inflation expectations? In: *Handbook of Economic Expectations*, 133–161. Academic Press. DOI: 10.1016/B978-0-12-822927-9.00012-4.
- Denes J., Lestrade A., Richardet L. (2022). Using Twitter data to measure inflation perception. In: *Machine learning in central banking*, 57. Bank for International Settlements.
- Dräger L. (2015). Inflation perceptions and expectations in Sweden — Are media reports the missing link? *Oxford Bulletin of Economics and Statistics*, 77 (5), 681–700. DOI: 10.1111/obes.12078.
- Friedman M. (1968). The role of monetary policy. *American Economic Review*, 58 (1), 1–17.
- Gabrielyan D., Masso J., Uusküla L. (2020). Mining news data for the measurement and prediction of inflation expectations. In: *Theory and Applications of Time Series Analysis*, 253–271. Springer International Publishing.
- Galí J., Gertler M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44 (2), 195–222. DOI: 10.1016/S0304-3932(99)00023-9.
- Garcia-Lembergman E., Hajdini I., Leer J., Pedemonte M., Schoenle R. (2023). The expectations of others. *CEPR Discussion Papers*, No. 18243. DOI: 10.18235/0013191.
- Gennaioli N., Shleifer A. (2010). What comes to mind. *The Quarterly Journal of Economics*, 125 (4), 1399–1433. DOI: 10.1162/qjec.2010.125.4.1399.
- Gürkaynak R. S., Sack B., Wright J. H. (2010). The TIPS yield curve and inflation compensation. *American Economic Journal: Macroeconomics*, 2 (1), 70–92. DOI: 10.1257/mac.2.1.70.
- Haubrich J., Pennacchi G., Ritchken P. (2012). Inflation expectations, real rates, and risk premia: Evidence from inflation swaps. *Review of Financial Studies*, 25 (5), 1588–1629. DOI: 10.1093/rfs/hhs003.
- Huber S., Minina D., Schmidt T. (2023). The pass-through from inflation perceptions to inflation expectations. *Deutsche Bundesbank Discussion Paper*, No. 17/2023.
- Lei C., Lu Z., Zhang C. (2015). News on inflation and the epidemiology of inflation expectations in China. *Economic Systems*, 39 (4), 644–653. DOI: 10.1016/j.ecosys.2015.04.006.

- Lucas R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4 (2), 103–124.
- Lucas R. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy*, 5 (1), 7–29.
- Malmendier U., Nagel S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131 (1), 53–87. DOI: 10.1093/qje/qjv037.
- Mankiw N. G., Reis R. (2002). Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics*, 117 (4), 1295–1328. DOI: 10.1162/003355302320935034.
- Phillips A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, 25 (100), 283–299.
- Saakshi, Sahu S., Chattopadhyay S. (2020). Epidemiology of inflation expectations and internet search: An analysis for India. *Journal of Economic Interaction and Coordination*, 15, 649–671. DOI: 10.1007/s11403-019-00255-4.
- Svensson L. E. O. (2010). Inflation targeting. In: *Handbook of Monetary Economics*, 1237–1302. Elsevier.
- Tversky A., Kahneman D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124–1131. DOI: 10.1126/science.185.4157.1124.

Поступила в редакцию 04.12.2024;
принята в печать 20.11.2025

Minichev F. I. The potential of using textual data to estimate inflation expectations in Russia. *Applied Econometrics*, 2026, v. 81, pp. 26–45.

DOI: 10.22394/1993-7601-2026-81-26-45.

Fedor Minichev

Sberbank of Russia, Moscow, Russian Federation;
fedor.minichev@gmail.com

The potential of using textual data to estimate inflation expectations in Russia

The paper proposes a methodology for measuring inflation expectations of firms in Russia based on the analysis of textual data from major media outlets. Over 2.3 million news publications from 13 Russian media were collected from the social network VKontakte for the period from January 2015 to October 2024. To identify publications specifically devoted to inflation, rather than merely containing relevant keywords, Latent Dirichlet Allocation (LDA) — a probabilistic topic modeling method — is applied for the first time in the Russian literature. The constructed indicator of inflation expectation intensity exhibits a 78% correlation with the Bank of Russia's measure of firms' price expectations, which is substantially higher than the correlation with household expectations (48%). Using LASSO regression, it is established that the largest contribution to explaining the dynamics of firms' expectations comes from three business-oriented media outlets — Kommersant, Vedomosti, and Expert — which jointly account for 75% of the variance in firms' price expectations. The results confirm that media textual data can serve as an effective complement to survey-based methods of measuring inflation expectations.

Keywords: inflation expectations; natural language processing; monetary policy; mass media; inflation targeting.

JEL classification: C55; D84; E31; E37; E52.

References

- Goloshchapova I., Andreev M. (2017). Measuring inflation expectations of the Russian population with the help of machine learning. *Voprosy Ekonomiki*, 6, 71–93 (in Russian). DOI: 10.32609/0042-8736-2017-6-71-93.
- Grishchenko V., Gasanova D., Fomin E., Korenyak G. (2023). Identifikaciya tovarov-markeroi i vliyaniye ikh tsen na inflyatsionnye ozhidaniya rossiyskikh domokhozyaystv. *Seriya dokladov Banka Rossii ob ekonomicheskikh issledovaniyakh*, 117 (in Russian).
- Evstigneeva A., Karpov D. (2023). Vliyaniye negativnykh novostey na vospriyatiye inflyatsii naseleniem. *Seriya dokladov Banka Rossii ob ekonomicheskikh issledovaniyakh*, 111 (in Russian).
- Petrova D. A. (2022). Assessment of inflation expectations based on internet data. *Applied Econometrics*, 66, 25–38 (in Russian). DOI: 10.22394/1993-7601-2022-66-25-38.
- Smirnova Zh. I. (2023). Novye vyzovy dlya denezhno-kreditnoy politiki. *Analiticheskaya zapiska*. Bank Rossii (in Russian).
- Adrian T., Laxton M. D., Obstfeld M. M. (2018). *Advancing the frontiers of monetary policy*. International Monetary Fund. DOI: 10.5089/9781484325940.071.
- Angelico C., Marcucci J., Miccoli M., Quarta F. (2022). Can we measure inflation expectations using Twitter? *Journal of Econometrics*, 228 (2), 259–277. DOI: 10.1016/j.jeconom.2021.12.008.
- Armantier O., Topa G., Van der Klaauw W., Zafar B. (2017). An overview of the Survey of Consumer Expectations. *Economic Policy Review*, 23 (2), 51–72.
- Aromí J. D., Llada M. (2020). Forecasting inflation with Twitter. *Economica*, 69, 64–85.
- Blei D. M., Ng A. Y., Jordan M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bordalo P., Gennaioli N., Shleifer A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, 73 (1), 199–227. DOI: 10.1111/jofi.12586.
- Bordalo P., Gennaioli N., Ma Y., Shleifer A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110 (9), 2748–2782. DOI: 10.1257/aer.20181219.
- Candia B., Coibion O., Gorodnichenko Y. (2023). The macroeconomic expectations of firms. In: *Handbook of Economic Expectations*, 321–353. Academic Press. DOI: 10.1016/B978-0-12-822927-9.00018-5.
- Carroll C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118 (1), 269–298. DOI: 10.1162/00335530360535207.
- Coibion O., Gorodnichenko Y., Kumar S. (2018). How do firms form their expectations? New survey evidence. *American Economic Review*, 108 (9), 2671–2713. DOI: 10.1257/aer.20151299.
- Conrad C., Enders Z., Glas A. (2022). The role of information and experience for households' inflation expectations. *European Economic Review*, 143. DOI: 10.1016/j.euroecorev.2021.104015.
- D'Acunto F., Malmendier U., Weber M. (2023). What do the data tell us about inflation expectations? In: *Handbook of Economic Expectations*, 133–161. Academic Press. DOI: 10.1016/B978-0-12-822927-9.00012-4.
- Denes J., Lestrade A., Richardet L. (2022). Using Twitter data to measure inflation perception. In: *Machine learning in central banking*, 57. Bank for International Settlements.
- Dräger L. (2015). Inflation perceptions and expectations in Sweden — Are media reports the missing link? *Oxford Bulletin of Economics and Statistics*, 77 (5), 681–700. DOI: 10.1111/obes.12078.

- Friedman M. (1968). The role of monetary policy. *American Economic Review*, 58 (1), 1–17.
- Gabrielyan D., Masso J., Uusküla L. (2020). Mining news data for the measurement and prediction of inflation expectations. In: *Theory and Applications of Time Series Analysis*, 253–271. Springer International Publishing.
- Gali J., Gertler M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44 (2), 195–222. DOI: 10.1016/S0304-3932(99)00023-9.
- Garcia-Lembergman E., Hajdini I., Leer J., Pedemonte M., Schoenle R. (2023). The expectations of others. *CEPR Discussion Papers*, No. 18243. DOI: 10.18235/0013191.
- Gennaioli N., Shleifer A. (2010). What comes to mind. *The Quarterly Journal of Economics*, 125 (4), 1399–1433. DOI: 10.1162/qjec.2010.125.4.1399.
- Gürkaynak R. S., Sack B., Wright J. H. (2010). The TIPS yield curve and inflation compensation. *American Economic Journal: Macroeconomics*, 2 (1), 70–92. DOI: 10.1257/mac.2.1.70.
- Haubrich J., Pennacchi G., Ritchken P. (2012). Inflation expectations, real rates, and risk premia: Evidence from inflation swaps. *Review of Financial Studies*, 25 (5), 1588–1629. DOI: 10.1093/rfs/hhs003.
- Huber S., Minina D., Schmidt T. (2023). The pass-through from inflation perceptions to inflation expectations. *Deutsche Bundesbank Discussion Paper*, No. 17/2023.
- Lei C., Lu Z., Zhang C. (2015). News on inflation and the epidemiology of inflation expectations in China. *Economic Systems*, 39 (4), 644–653. DOI: 10.1016/j.ecosys.2015.04.006.
- Lucas R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4 (2), 103–124.
- Lucas R. (1977). Understanding business cycles. *Carnegie-Rochester Conference Series on Public Policy*, 5 (1), 7–29.
- Malmendier U., Nagel S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131 (1), 53–87. DOI: 10.1093/qje/qjv037.
- Mankiw N. G., Reis R. (2002). Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *The Quarterly Journal of Economics*, 117 (4), 1295–1328. DOI: 10.1162/003355302320935034.
- Phillips A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, 25 (100), 283–299.
- Saakshi, Sahu S., Chattopadhyay S. (2020). Epidemiology of inflation expectations and internet search: An analysis for India. *Journal of Economic Interaction and Coordination*, 15, 649–671. DOI: 10.1007/s11403-019-00255-4.
- Svensson L. E. O. (2010). Inflation targeting. In: *Handbook of Monetary Economics*, 1237–1302. Elsevier.
- Tversky A., Kahneman D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185 (4157), 1124–1131. DOI: 10.1126/science.185.4157.1124.

Received 04.12.2024; accepted 20.11.2025