

М.А. Габов

Отделение по Пермскому краю Уральского главного управления Центрального банка Российской Федерации; НИУ «Высшая школа экономики», Пермь

Т.В. Букина

НИУ «Высшая школа экономики», Пермь

Д.В. Кашин

НИУ «Высшая школа экономики», Пермь

Сравнительный анализ моделей прогнозирования региональной инфляции¹

Аннотация. Исследование посвящено сравнению подходов к прогнозированию уровня индекса потребительских цен (ИПЦ г/г) в регионах Приволжского федерального округа с использованием эконометрических моделей временных рядов и методов машинного обучения. В статье предпринята попытка выбрать наиболее подходящие и эффективные модели для прогнозирования регионального индекса общего уровня цен. Исследование включает использование комбинированного подхода, который основан на сочетании как эконометрических моделей, так и методов машинного обучения. Полученные результаты свидетельствуют о том, что модели машинного обучения обеспечивают более стабильные и точные прогнозы, чем эконометрические модели – особенно на длительных периодах прогнозирования (6 и более месяцев). Однако для ряда регионов мы нашли доказательства эффективности эконометрических моделей временных рядов на краткосрочном периоде – для нескольких регионов различные спецификации расширенных моделей авторегрессии работают лучше, чем подход, основанный на моделях машинного обучения, при прогнозировании на 1 и 3 месяца. Результаты комбинированного подхода сопоставимы с прогнозами моделей машинного обучения и чаще обеспечивают более точный прогноз на 12 и 24 месяца. Проведенное исследование показало, что не удалось обнаружить устойчивого эффекта региональных особенностей в результатах прогнозирования, вызванного спецификой региона, а именно волатильностью инфляции, структурой экономики региона.

Ключевые слова: *прогнозирование, региональная инфляция, машинное обучение, эконометрическое моделирование, комбинированный подход.*

Классификация JEL: E31, E37.

Для цитирования: **Габов М.А., Букина Т.В., Кашин Д.В.** (2025). Сравнительный анализ моделей прогнозирования региональной инфляции // *Журнал Новой экономической ассоциации*. № 4 (69). С. 87–117.

DOI: 10.31737/22212264_2025_4_87-117

EDN: WBWFAU

1. Введение

В период структурной перестройки экономики прогнозирование инфляции вновь обрело особый статус как для Банка России и правительства страны, так и для домохозяйств и фирм. Потребность в наиболее точном прогнозе одного из ключевых показателей состояния экономики обусловлена проведением соот-

¹ Настоящая статья отражает личную позицию авторов. Содержание и результаты данного исследования не следует рассматривать, в том числе цитировать в каких-либо изданиях, как официальную позицию Банка России или указание на официальную политику или решения регулятора. Любые ошибки в данном материале являются исключительно авторскими.

Авторы выражают благодарность анонимным рецензентам за полезные замечания, рекомендации и ценные советы, позволившие доработать рукопись.

ветствующей экономической политики, с одной стороны, принятием решений об инвестиционных стратегиях и разработкой бизнес-планов – с другой.

Вопрос прогнозирования региональной инфляции является менее исследованным по сравнению с общестрановым показателем. По данным Google Scholar², к маю 2024 г. число публикаций по запросу «прогнозирование инфляции» кратно превышает число работ, посвященных прогнозу регионального показателя (454 против 30). Тем не менее, прогноз инфляции в регионах остается не менее важной задачей.

Потребность в наиболее точном прогнозировании региональной инфляции продиктована тем, что макроэкономические агенты могут быть заинтересованы в разработке стратегий на основе будущих тенденций уровня цен, учитывая региональные особенности. Более того, комбинация региональных прогнозов может обеспечить более точный прогноз общероссийской инфляции (Tena, Espasa, Pino, 2010). Вдобавок, ввиду неоднородности структуры экономики регионов существуют различия в результатах прогнозирования, которые необходимо учитывать при моделировании.

Для прогнозирования макроэкономических показателей зачастую используются традиционные эконометрические модели благодаря своей относительной простоте в использовании и интерпретации. В то же время современные подходы к прогнозированию, основанные на методах машинного обучения, в последние несколько лет уже успели доказать свою практическую ценность в сравнении с наиболее распространенными в прошлом моделями авторегрессии (AR) и скользящего среднего (MA). Однако в литературе часто отмечается, что классические модели временных рядов обеспечивают более точный прогноз на краткосрочный период, в то время как модели машинного обучения – на среднесрочный и долгосрочный периоды. Кроме того, исследования (Семитуркин, Шевелев, 2022; Pratap, Sengupta, 2019) являются примерами того, что правильно подобранные спецификации эконометрических моделей временных рядов могут быть более точными при прогнозе на более чем три периода.

Из этого следует, что не существует однозначного вывода о том, какой подход является наиболее эффективным. Сравнительный анализ моделей прогнозирования позволит выявить наилучшие подходы, учитывая текущие экономические и геополитические условия. Так, наиболее точные прогнозы региональной инфляции могут обеспечить благоприятные условия для перехода на траекторию экономического роста.

Цель статьи – провести сравнительный анализ наиболее точных и эффективных подходов к прогнозированию регионального уровня цен с использованием различных эконометрических моделей временных рядов и методов машинного обучения.

В качестве объекта исследования рассматриваются уровни ИПЦ (индекса потребительских цен) по месяцам год к году (г/г) регионов Приволжского федерального округа (ПФО) в период с января 2003 г. по декабрь 2023 г. (всего 252 наблюдения) – источник данных – Единая межведомственная информационно-статистическая система (ЕМИСС). В состав ПФО входят шесть республик (Башкортостан, Марий-Эл, Мордовия, Татарстан, Удмуртия, Чувашия), семь

² Поиск система по научным публикациям. С помощью поисковых роботов портал индексирует метаданные и осуществляет полнотекстовый поиск по научной литературе, включая журнальные статьи, препринты, диссертации, книги и технические отчеты. Пользователи могут искать нужные работы по авторам, ключевым словам, названию журнала. Углубленный поиск позволяет ранжировать материалы по публикации, дате и предметной области.

областей (Кировская, Нижегородская, Оренбургская, Пензенская, Самарская, Саратовская, Ульяновская) и Пермский край. Центр округа – город Нижний Новгород.

Мы выявили, что модели машинного обучения обеспечивают более точные прогнозы в части среднеквадратичной ошибки прогноза, чем эконометрические модели – особенно на длительных периодах прогнозирования (6 и более месяцев). Для нескольких регионов ПФО мы подтвердили предположение о том, что классические модели временных рядов эффективны при прогнозе на краткосрочный период. Кроме того, простая наивная модель демонстрирует более точные результаты при прогнозировании уровня ИПЦ (г/г) в Республике Мордовия по сравнению с обоими подходами. Результаты комбинированного подхода сопоставимы с прогнозами моделей машинного обучения и чаще обеспечивают более точный прогноз на 12 и 24 месяца.

2. Теоретический обзор моделей прогнозирования инфляции

В литературе в основном уделяется внимание однофакторным моделям временных рядов, которые часто превосходят многофакторные модели с использованием показателей экономической активности и подходы, основанные на модели кривой Филлипса (Faust, Wright, 2013). В то же время, число публикаций в годовом выражении, в которых исследуется применимость моделей машинного обучения для прогнозирования, за последние десять лет выросло более чем в 20 раз. Динамика, представленная на рис. 1, подчеркивает возросшую популярность в работах, посвященных прогнозированию методами машинного обучения. Также часто рассматривается комбинированный подход, основанный на сочетании двух методов – с помощью однофакторных моделей (временные

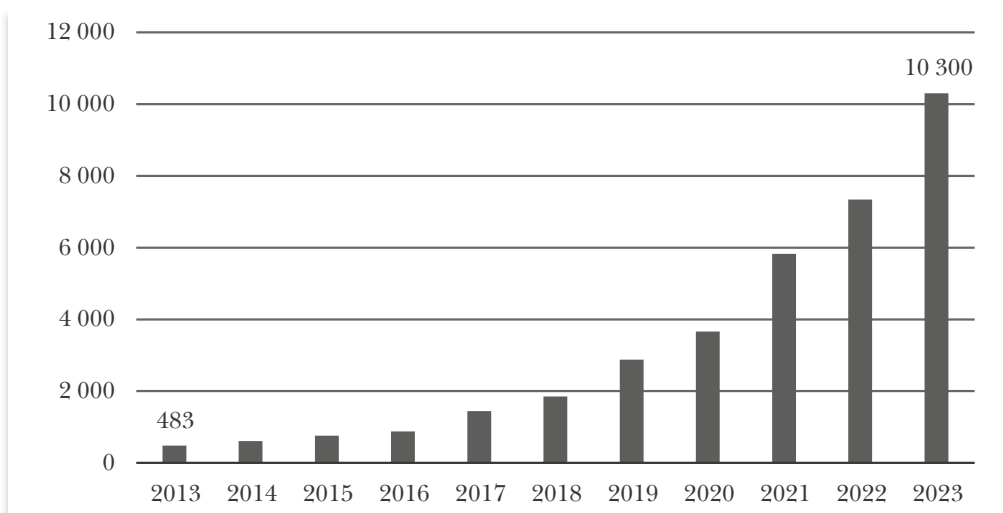


Рис. 1.

Ежегодное число публикаций, в которых исследуется применимость моделей машинного обучения для прогнозирования

Источник: составлено авторами на основе данных Google Scholar по запросу «inflation forecasting “machine learning”».

ряды) и моделей машинного обучения. Ряд авторов отмечают превосходство такого подхода с точки зрения устойчивости результатов (Pratar, Sengupta, 2019; Ögünç et al., 2013).

2.1. Эконометрические модели временных рядов

В настоящее время наиболее распространенным подходом к прогнозированию является применение классических моделей временных рядов – в первую очередь модели авторегрессии и скользящего среднего. Ключевое преимущество эконометрических моделей временных рядов для прогнозирования состоит в возможности интерпретации полученной модели с помощью экономической теории.

В эмпирических исследованиях зачастую отмечается превосходство модели авторегрессии скользящего среднего (autoregressive moving-average model, ARMA) для прогнозирования инфляции (Udom, Phumchusri, 2014; Mustapha, Kubalu, 2016; Popoola et al., 2017). Одним из наиболее цитируемых эмпирических исследований на тему прогнозирования инфляции является работа (Faust, Wright, 2013). В своем исследовании авторы рассмотрели около 20 различных спецификаций моделей прогнозирования инфляции в США на основе квартальных данных за период с 1985 по 2011 г. Основным выводом работы состоит в том, что, основываясь на метриках ошибок прогноза, многофакторные модели (DSGE, EWA, BMA и др.) не превосходят однофакторные – например, модель авторегрессии (autoregressive, AR) и модель случайного блуждания (random walk, RW) обладают более эффективной прогностической силой в сравнении с другими рассмотренными моделями и экспертными прогнозами.

В (Nyoni, Mutongi, 2019) авторы исследовали годовую инфляцию в Гамбии за период с 1962 по 2016 г. Для построения модели прогнозирования авторы использовали модель ARMA (autoregressive moving average). Наилучшей из рассмотренных спецификаций оказалась модель ARMA (1, 0, 0) – она обеспечивает наименьшее значение информационного критерия Акаике (AIC). С помощью полученной модели авторы спрогнозировали уровень инфляции в Гамбии на 2017 г. – по их расчетам, годовая инфляция в стране составит 7,56%³.

Для прогнозирования инфляции в Румынии (Baciu, 2015) использовал следующие модели – AR, MA (moving average), ARMA. Используя данные инфляции по месяцам в период с января 1997 г. по август 2013 г., автор пришел к выводу о том, что модели ARMA не подходят для среднесрочных и долгосрочных прогнозов. Однако для прогноза на более короткие периоды данная спецификация способна обеспечить точные результаты.

Исследуя годовую инфляцию в Нигерии, помимо модели ARMA (Nyoni, Nathaniel, 2018) также рассмотрели более продвинутые модели для прогнозирования – в частности модель ARMA совместно со спецификациями ARCH и GARCH (generalized autoregressive conditional heteroskedasticity). Особенность такого подхода состоит в том, что помимо точечного прогноза инфляции прогнозируется ее дисперсия. Авторы отмечают, что модель AR(3) – GARCH(1, 1) может

³ Прогноз авторов несколько отличается от фактического показателя. Согласно данным Всемирного банка, инфляция в Гамбии по итогам 2017 г. составила 8,03% в годовом выражении. Причины отклонения продиктованы тем, что в модели учтено только предыдущее значение инфляции, в то время как ряд других макроэкономических показателей, определяющих динамику инфляции, не рассмотрены. Тем не менее результаты прогноза верно предсказали динамику ускорения темпов годовой инфляции.

быть использована для прогнозирования волатильности инфляции в Нигерии, обеспечивая устойчивые и надежные результаты.

Авторы (Uwilingiyimana, Munga'tu, Naregimana, 2015) также отмечают практическую ценность использования моделей ARMA–ARCH / GARCH для прогноза волатильности инфляции. Авторы обращают внимание на то, что на основе данных с 2000 по 2014 г. полученная модель прогнозирования инфляции по месяцам в Кении ARMA(1, 1, 12) – GARCH(1, 2) обеспечивает оптимальные результаты и эффективно повышает точность оценки и прогнозирования в сравнении с базовыми моделями AR и ARMA. Более того, в той же работе утверждается, что использование данной спецификации позволяет нивелировать ограничения, связанные с типом линейной связи данных, что можно рассматривать в качестве потенциально нового подхода к анализу и прогнозированию инфляции.

Использование различных спецификаций модели ARMA также является распространенным в эмпирических исследованиях. Так, например, в работах (Lidiema, 2017; Otu et al., 2014) отмечается превосходство в учете сезонности в модели для прогнозирования инфляции – модели SARMA (seasonal ARMA). Практическая необходимость в таком подходе заключается в том, что инфляция зачастую характеризуется *сезонным характером* – особенно в тех странах, где динамика темпов роста экономики подвержена определенным временным закономерностям, т.е. сезонности. Одним из примеров такой экономики является Турция – страна, на которую приходится 4% международного туристического рынка. Исследование (Saz, 2011) является подтверждением того, что модель SARMA способна точно демонстрировать прогнозы ежемесячной инфляции в Турции.

Ранее было отмечено, что учет дополнительных (экзогенных или эндогенных) переменных для построения моделей прогноза инфляции обеспечивает более устойчивые результаты. Стохастическое моделирование инфляции с учетом других экзогенных переменных (например, обменный курс национальной валюты или денежная масса) сводится к модели ARMAX (autoregressive moving average eXtended). Авторы (Tamuke, Jackson, Sillah, 2018) в своей работе сравнивают два подхода к прогнозированию инфляции – с помощью модели ARMA и ARMAX (ковариата, т.е. экзогенная переменная – номинальный обменный курс). Результат работы состоит в том, что модель ARMAX для прогнозирования ИПЦ по месяцам в Сьерра-Леоне выглядит более предпочтительной, чем модель ARMA.

В (Doguwa, Alade 2013) сравнивали результаты прогнозирования инфляции по месяцам в Шри-Ланке на основе моделей SARMA и SARMAX (в качестве ковариат были рассмотрены государственные расходы, агрегатор денежной массы M2, номинальный обменный курс и др.). Однозначного вывода о результатах прогнозирования моделей сделать нельзя, отмечают авторы – однако для прогноза на краткосрочный период модель SARMAX выглядит более предпочтительной, основываясь на метриках ошибок прогноза.

Таким образом, для точечного прогнозирования инфляции среди эконометрических моделей временных рядов в основном используются спецификации ARMA. Стоит отметить, что данная модель чаще всего способна давать точные и устойчивые прогнозы лишь на краткосрочные периоды, от одного до трех рассматриваемых интервалов. Для более совершенных результатов принято рас-

сма́тривать более продвину́тые специфика́ции модели ARMA: SARMA – если в данных есть сезонность; ARMAX – если существует связь с другими макроэкономическими показателями; SARMAX – если существует сезонность и есть предпосылки в использовании дополнительных показателей, которые способны более полно описывать динамику темпов роста цен. Более того, для прогнозирования волатильности (дисперсии) инфляции используется комбинированный подход на основе моделей ARMA–ARCH / GARCH.

В вышеупомянутых исследованиях (см., например, (Saz, 2011; Lidiema, 2017; и др.)), посвященных прогнозированию инфляции классическими эконометрическими моделями, не так часто встречается обсуждение процесса выбора оптимальной спецификации моделей и подбора параметров авторегрессионных моделей. В данной работе мы подробно рассмотрим этот вопрос, именно осуществим подбор лагов на основе анализа графиков автокорреляционной функции (ACF) и частной автокорреляционной функции (PACF); с помощью данных функций подберем параметры для моделей ARIMA (p, d, q) и ARIMAX (p, d, q). По результатам проведенного анализа мы далее оценим, как наиболее распространенную в литературе модель авторегрессии с первым лагом и ее вариации, так и другие спецификации, и проверим, обеспечивают ли более сложные модели высокую точность прогнозов.

2.2. Модели машинного обучения

Модели машинного обучения уже были разработаны к началу XXI в. Однако в профессиональной литературе, посвященной прогнозированию инфляции, они долгое время оставались без должного внимания (Baybuza, 2018). В работе (Baybuza, 2018) отмечается, что ввиду своей специфики модели машинного обучения способны точно предсказывать числовую переменную на основе входных данных – этот метод лучше учитывает статистические выбросы в отличие от традиционных прогнозных моделей.

Исследование (Baybuza, 2018) является одной из первых попыток применить методы машинного обучения для прогнозирования инфляции в России – LASSO, Ridge, Elastic Net, Random Forest и Boosting. Основным результатом данной работы является подтверждение возможности более точного прогнозирования инфляции в России с использованием методов машинного обучения – модели Random Forest и Boosting также хороши в прогнозировании инфляции, как и традиционные эконометрические модели, отмечает автор.

Работа (Almosova, Andresen, 2023) фокусируется на прогнозировании инфляции в США. Авторы пришли к выводу о том, что использование методов машинного обучения позволяет улучшить прогнозы инфляции – такие модели, как Lasso, Bagging, Random Forest и др., способны давать более точные прогнозы, чем стандартные ориентиры (AR/ARMA). Эти результаты также подчеркивают преимущества методов машинного обучения для макроэкономического прогнозирования.

В (Chakraborty, Joseph 2017) исследован вопрос прогнозирования ИПЦ в Великобритании на среднесрочный период (24 месяца). В работе были использованы популярные подходы машинного обучения к моделированию –

KNN (K-Nearest Neighbors), SVM (Support Vector Machine), DT (Decision Trees), Random Forest и др. Более того, авторы также рассмотрели эконометрические модели временных рядов – AR, ARMA и др. Используя квартальные данные за период с 1988 по 2015 г. о ВВП, уровне безработицы, денежной массе и других макроэкономических показателей, выступавших в качестве экзогенных переменных, авторам удалось установить следующее: основываясь на метриках средних абсолютных ошибок прогноза (MAE, mean absolute error), модели машинного обучения превосходят эконометрические модели в прогнозировании ИПЦ.

Однако, учитывая качество и точность прогностической силы моделей машинного обучения, остаются открытые исследовательские вопросы в отношении свойств причинно-следственной связей рассмотренных моделей – отмечают (Chakraborty, Joseph, 2017). Это свидетельствует о том, что ключевым ограничением методов машинного обучения является интерпретация полученных результатов. Иными словами, существует ряд сложностей с возможностью количественной оценки влияния фактора, включенного в модель, на зависимую переменную – в отличие от эконометрических моделей зачастую не удается определить величину коэффициента экзогенной переменной, поскольку связь между условными X и Y может быть нелинейной, что как раз и учитывают модели машинного обучения.

Еще одним подтверждением превосходства методов машинного обучения для прогнозирования инфляции является работа (Pratap, Sengupta, 2019). Исследуя данные по инфляции в Индии за 2002–2018 гг., авторы определили, что модели машинного обучения обеспечивают более точные прогнозы инфляции, чем дает модель ARMA. В частности, методы XgBoost и Random Forest обеспечили одни из наименьших среднеквадратических ошибок прогноза (RMSE) – модели примерно в три раза точнее предсказали инфляцию в Индии, чем модель ARMA (0,67 и 0,62 против 1,62 соответственно).

Тем не менее стоит отметить, что авторами была обнаружена характерная годовая сезонность в данных. С помощью модели SARMA (Pratap, Sengupta, 2019) добились более точного прогноза инфляции – величина RMSE оказалась примерно в три раза меньше, в сравнении с результатами моделей XgBoost и Random Forest (0,67 и 0,62 против 0,24 соответственно). Данная работа также является одним из примеров того, что с помощью правильно подобранной спецификации эконометрических моделей временных рядов можно добиться более точного прогноза инфляции.

В литературе можно также встретить работы, в которых отмечается превосходство моделей авторегрессии над машинным обучением в прогнозе на короткий период. Подтверждением данного тезиса служит работа (Семитуркин, Шевелев, 2022). Оценивая качество прогнозов ИПЦ по Новосибирской области с помощью относительной величины RMSE (относительно модели ARMA), авторы установили, что модель ARMAX значительно лучше прогнозирует инфляцию на три месяца, чем модели машинного обучения. Более того, сравнивая метрики ошибок прогноза ИПЦ по всему макрорегиону Сибирь, авторы (Семитуркин, Шевелев, 2022) отмечают существенное превосходство модели ARMAX на 3, 6, 9 и 12 месяцев – однако прогноз на 24 месяца оказался менее точным в сравнении с моделями машинного обучения. Анализ результатов прогноза

ИПЦ остальных регионов Сибири в основном свидетельствует о превосходстве моделей машинного обучения на более длительный период.

Одна из недавних работ на тему сравнения качества прогнозов региональной инфляции эконометрическими моделями и моделями машинного обучения — работа (Букина, Кашин, 2024). В статье авторы строили прогнозы региональной инфляции на примере субъектов, входящих в Приволжский федеральный округ (ПФО). Авторы статьи выявили, что модели машинного обучения не хуже справляются с задачей прогнозирования региональной инфляции в регионах ПФО на длительных временных периодах, в то время как эконометрические модели достаточно точно прогнозируют на краткосрочных временных горизонтах. В работе использовались такие методы машинного обучения, как метод опорных векторов, градиентный бустинг и случайный лес. Следуя методологии, используемой в работе (Букина, Кашин, 2024), в данной статье, на примере регионов ПФО, мы сравним модели временных рядов с моделями машинного обучения, используя более широкий модельный аппарат в части двух классов моделей, а также представим результаты комбинированного подхода, сравнив устойчивость и качество такого подхода с традиционными методами прогнозирования.

Таким образом, можно заключить, что модели машинного обучения могут применяться для прогнозирования инфляции как на страновом (Baybuza, 2018), так и на региональном уровнях (Семитуркин, Шевелев, 2022; Букина, Кашин, 2024). Зачастую модели машинного обучения способны обеспечивать более точные прогнозы инфляции, чем традиционные модели временных рядов, — особенно на средне- и долгосрочный периоды. Однако важно учитывать специфику рассматриваемого объекта — сезонность и структуру экономики, зависимость от экспортно-импортных условий и т.д., что позволит определить целесообразность применения моделей машинного обучения.

Необходимо отметить, что эконометрические модели до сих пор могут выступать в качестве более предпочтительных подходов к прогнозированию инфляции на короткий период, а также в случаях с наличием сезонности в данных. В связи с этим имеет место использование комбинированного подхода к прогнозированию инфляции — сочетание эконометрических моделей с методами машинного обучения выступает в качестве одного из направлений развития исследования на эту тему.

В большинстве исследований, посвященных оценке применимости моделей машинного обучения для прогнозирования инфляции, рассматриваются подходы, которые методологически отличаются между собой. Например, ансамблевые модели на основе таких деревьев решений, как Random Forest и XgBoost, могут работать с неструктурированными данными, не требуя предварительной обработки (например, нормализации или стандартизации входных данных), поскольку отличаются от других моделей инвариантностью к масштабированию признаков. В противоположность им линейные модели (LASSO, RIDGE, SVM и Elastic Net) требуют нормализации признаков, поскольку они чувствительны к масштабированию данных. К примеру, в работах (Семитуркин, Шевелев, 2022; Baybuza, 2018) одновременно анализируются как модели деревьев решений, так и линейные регрессионные модели. В свою очередь, в исследованиях (Букина, Кашин, 2024; Pratap, Sengupta, 2019) также рассматривается модель SVM.

Учитывая эти особенности, далее в данном исследовании особое внимание уделяется только одному классу моделей машинного обучения – ансамблевым моделям деревьев решений (Random Forest и XgBoost) – как наиболее распространенным и зарекомендовавшим себя с точки зрения точности прогнозирования в литературе.

2.3. Комбинированный подход

Необходимость комбинации моделей прогнозирования обусловлена тем, что любой прогностический метод в лучшем случае является всего лишь приближением истинного процесса генерации данных (DGP), в результате чего для одной и той же целевой переменной можно получить несколько моделей с одинаковой точностью прогнозирования. Комбинированные прогнозы демонстрируют более высокую эффективность эмпирически, обеспечивая согласованность результатов и защиту от волатильности модели (Pratap, Sengupta, 2019; Elliott, Timmermann, 2005).

Анализируя точность прогнозирования инфляции с помощью эконометрических моделей временных рядов и методов машинного обучения, в (Pratap, Sengupta, 2019) рассмотрен комбинированный подход и построена модель, основанная на результатах предсказания наилучших из ранее рассмотренных (SARMA, XGBoost, NNAR и др.). В результате авторы оценили точность предсказания полученной модели и пришли к следующим выводам. Комбинированный подход превосходит все рассмотренные спецификации моделей машинного обучения, однако обеспечить лучший прогноз, чем модель SARMA, не удалось – величина RMSE комбинированного подхода составила 0,38 против 0,24. Применение двух методов учета весов в комбинированном подходе (усредненный и основанный на величине обратной RMSE каждой из рассмотренных моделей) обеспечивает более точные результаты, чем модели машинного обучения по отдельности. Комбинированный подход может применяться в случае, если нельзя однозначно определить, какая из моделей является более эффективной – сочетание всех методов в одном позволяет добиться точных результатов.

Одной из особенностей работы (Семитуркин, Шевелев, 2022) также является использование комбинированного подхода к прогнозированию региональной инфляции. Авторы использовали методику взвешенной модели, основываясь на относительных метриках ошибок. В результате им удалось построить модель, которая не уступает Random Forest и Gradient Boosting в точности прогноза на 12 и 24 месяца соответственно. Тем не менее, другие спецификации машинного обучения оказались более эффективными, чем комбинированный подход в прогнозе на период до 1 года.

Комбинированный подход также применяется и на основе сочетания эконометрических моделей временных рядов. Так, например, в (Ögünç et al., 2013) авторы использовали различные спецификации векторных моделей авторегрессии (VAR – авторегрессия с учетом эндогенных переменных) в комбинированном подходе для краткосрочного прогноза инфляции в Турции. Авторы утверждают, что объединение прогнозов приводит к меньшей ошибке прогнозирования по сравнению с большинством моделей, однако на некоторых периодах отдельные модели показывают

одинаковую точность. В статье (Karpetanios, Labhard, Price, 2008) были исследованы различные макроэкономические показатели Великобритании с помощью эконометрических моделей (RW, AR, VAR и др.) и различных способов их комбинирования. Один из основных результатов работы состоит в том, что комбинированные подходы обеспечили более точные прогнозы инфляции. Более того, комбинация теоретико-информационных прогнозов (WITMA) систематически превосходила AR на всех горизонтах — в некоторых случаях с большим отрывом.

Автор (Mihretu, 2023) использовал шесть различных моделей инфляции — RW, ARIMA, ECM, VECM, BVAR и Phillips Curve. Сравнивая оценки метрик ошибок моделей, автор отмечает превосходство ARMA для прогнозирования инфляции в Эфиопии. Однако в своей работе М. Мирету акцентирует внимание на важности использования комбинированного подхода для прогнозирования, ссылаясь на возможные структурные изменения и погрешность спецификации, присущие одной модели. Сочетание нескольких подходов позволяет нивелировать данную проблему — так, используя различные техники объединения моделей, автор обнаружил существенные преимущества в точности прогноза с помощью комбинированных моделей.

Таким образом, комбинированный подход к прогнозированию инфляции может обеспечить более точные результаты, чем одна отдельно взятая модель — эконометрическая или модель машинного обучения. Более того, одним из ключевых преимуществ такого подхода является объединение моделей, которые не отличаются друг от друга с точки зрения метрик ошибок прогноза. Это может быть особенно полезно в случае моделирования с помощью методов машинного обучения — зачастую различные подходы одинаково хорошо демонстрируют точность прогноза, поэтому выбрать среди них наилучшую кажется сложным. В связи с этим с помощью комбинированного подхода можно не только объединить множество комплексных моделей в одну, но и добиться не менее точных результатов.

В данной работе для комбинации моделей будут использованы только те модели, которые оказались наилучшими на каждый анализируемый период прогноза. Корректировка весов моделей будет проведена согласно числу таких моделей. Данный подход к построению комбинированного прогноза представляет собой объединение тех подходов, которые были использованы в работе (Pratap, Sengupta, 2019). В данном случае мы учитываем как простоту применения — установку одинаковых весов для моделей, так и относительную эффективность, что аналогично подбору весов на основе обратной величины RMSE. Более того, в следующем разделе статьи рассмотрен подход, при котором веса моделей в их комбинации рассчитываются на основе того, сколько раз та или иная модель оказалась наилучшей в прогнозе на определенный период.

3. Данные и методология исследования

В качестве объекта данного исследования рассматриваются уровни ИПЦ по месяцам год к году (г/г)⁴ регионов Приволжского федерального округа (ПФО) в период с января 2003 г. по декабрь 2023 г. (всего 252 наблю-

⁴ Исследование годовых данных по инфляции обусловлено тем, что Банк России таргетирует именно этот показатель. Тем не менее, в годовых данных уже учтена сезонность. В данной статье рассматриваются спецификации моделей временных рядов с сезонностью в 12 месяцев. Вместе с тем важно отметить, что для каждого региона ПФО была построена модель авторегрессии с помощью функции `auto.arima`, благодаря которой можно подобрать наилучшую спецификацию модели авторегрессии с точки зрения AIC. Так, для каждого региона, подбирая модели ARIMA и ARIMAX с помощью `auto.arima`, наилучшей оказалась спецификация SARIMA и SARIMAX с сезонностью в 12 месяцев соответственно.

дения, источник данных – ЕМИСС). В состав ПФО входит шесть республик (Башкортостан, Марий-Эл, Мордовия, Татарстан, Удмуртия, Чувашия), семь областей (Кировская, Нижегородская, Оренбургская, Пензенская, Самарская, Саратовская, Ульяновская) и Пермский край.

ПФО является развитым экономическим округом страны с большим объемом экспортно-импортных отношений. Так, основываясь на обзоре литературы и учитывая структуру экономики округа, рыночный обменный курс доллара США в рублях, устанавливаемый Банком России, будет рассмотрен в качестве экзогенной переменной⁵ (источник данных – Банк России).

В Приложении 1, п. 1 представлена описательная статистика ИПЦ год к году (г/г) регионов ПФО за 2023 г., в Приложении 1, п. 2 – по рыночному обменному курсу доллара США в руб. в период с 2003 г. по 2023 г.

Согласно обзору литературы, для прогнозирования инфляции в регионах в работе рассматриваются различные спецификации эконометрических моделей временных рядов – за исключением векторных моделей авторегрессии – а также наиболее распространенные методы машинного обучения, основанные на методе «деревья решений» (Decision Trees) – Random Forest и XgBoost. Вместе с тем исследуются подходы с использованием комбинаций прогнозов обоих методов.

Процесс построения моделей временных рядов осуществляется в соответствии с результатами анализа иллюстраций ACF и PACF. На всех графиках ACF наблюдался значимый коэффициент корреляции между Y_t и Y_{t-1} . Графики PACF, в свою очередь, позволили сделать вывод о том, что для процессов могут быть характерны модели AR(1) и AR(12). Таким образом, мы определили, что возможны следующие спецификации моделей временных рядов: AR(1), AR(12), MA(1), MA(12), ARIMA(1; 1), ARIMA (12; 1).

Спецификация ARIMA(12, 12) далее не рассматривается, так как не выполняется условие о соотношении числа наблюдений и переменных – $k+1 \leq \sqrt{N}$. В данном случае возможны лишь 14 переменных и 1 константа для 225 наблюдений из тренировочной выборки.

Таким образом, были получены следующие модели временных рядов для прогнозирования ИПЦ г/г на основе тренировочной выборки: AR(1), AR(12), ARIMA(1, 0, 1), ARIMA(12, 0, 1), SARIMA(3, 0, 2) (2, 0, 1) [12], ARIMAX(1, 0, 1), ARIMAX(12, 0, 1), SARIMAX(3, 0, 2) (2, 0, 1) [12].

Среди моделей машинного обучения рассматриваются следующие спецификации:

1) $cpi_t = f(cpi_{t-i}; Month)$, где cpi – ИПЦ год к году (г/г) в Пермском крае (первая разность и замена выбросов); i – номер лага (от 1 до 12); $Month$ – номер месяца (от 1 до 12);

2) $cpi_t = f(cpi_{t-i}; usd_{t-j}; Month)$, где cpi – ИПЦ год к году (г/г) в Пермском крае (первая разность и замена выбросов); usd – рыночный обменный курс доллара США в рублях (первая разность и замена выбросов); i – номер лага (1, 3, 6, 12); j – номер лага (0, 1, 3, 6, 12); $Month$ – номер месяца (от 1 до 12);

3) RF – Random Forest; XgB – eXtreme gradient Boosting. Номер в названии модели отвечает за спецификации, описанные выше.

⁵ Согласно обзору литературы (Семитуркин, Шевелев, 2022), обменный курс национальной валюты зачастую выступает в качестве экзогенной переменной при прогнозировании инфляции, в том числе региональной.

Параметры для моделей RF: *ntree* – число деревьев в лесу. Использовалось значение по умолчанию – 500; *mtry* – число предикторов, которые выбираются случайно для каждого дерева; $mtry = p/3$, где p – число предикторов; *nodesize* – минимальное число наблюдений в листе дерева (использовалось значение по умолчанию – пять наблюдений); *maxnodes* – максимальное число узлов в каждом дереве (по умолчанию – без ограничений, т.е. деревья росли до полной глубины); *replace* – указывает, следует ли выбирать наблюдения с возвращением (по умолчанию – TRUE).

Параметры для моделей XgB: *booster* – тип бустера (использовалось значение по умолчанию – *gbtree*, что означает использование деревьев решений); *eta* (или *learning_rate*) – скорость обучения (по умолчанию – 0,3); *max_depth* – максимальная глубина деревьев (по умолчанию – 6); *min_child_weight* – минимальный вес дочернего узла (по умолчанию – 1); *subsample* – доля обучающих данных, используемых для построения каждого дерева (по умолчанию – 1 (100%)); *gamma* – параметр, который контролирует минимальное уменьшение потерь, необходимое для разделения узла (по умолчанию – 0).

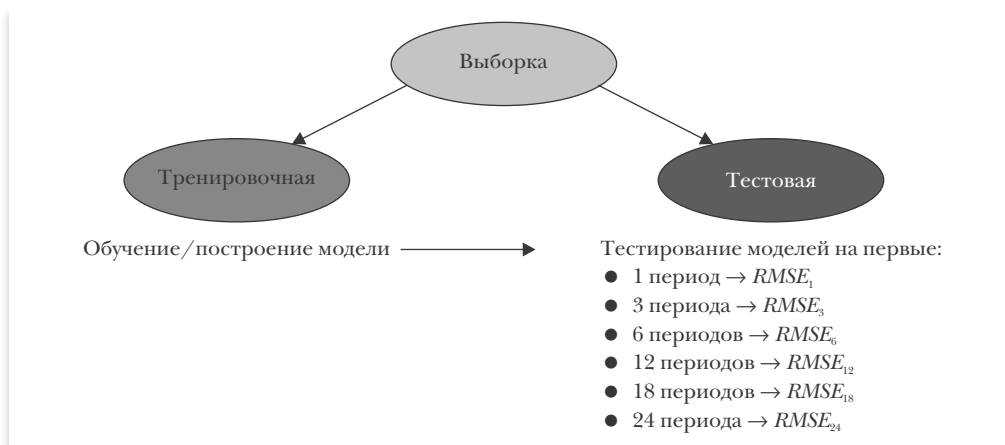
Первым этапом эмпирической части исследования будет приведение рассматриваемых данных к слабо стационарной форме. На их основе будут построены как эконометрические модели временных рядов, так и модели машинного обучения – несмотря на то что теоретических требований для данных в случае применения методов машинного обучения нет, построение моделей на одних и тех же наблюдениях необходимо для возможности справедливого сравнения эффективности и прогностической силы моделей.

Также важно упомянуть, что, решая задачу прогнозирования, необходимо разделить исходную выборку на тренировочную и тестовую. На тренировочной выборке мы будем обучать модели, на тестовой – оценивать точность их прогноза. В тренировочную выборку войдут данные с 01.01.2003 по 01.12.2021, в тестовую – 24 последних месяца (период с 01.01.2022 по 01.12.2023), что составляет примерно 20% всех наблюдений. Период, соответствующий тестовой выборке, сильно отличается от более ранних временных отрезков с точки зрения волатильности ИПЦ год к году (г/г) в регионах ПФО – тем интереснее оценить прогностическую точность моделей в наиболее волатильный период.

В качестве основной метрики эффективности прогноза моделей мы используем среднеквадратическую ошибку RMSE, показывающую разницу между реальным и прогнозным значениями. Метрика будет рассчитана для следующих периодов – 1, 3, 6, 12, 18 и 24 месяца. В данном случае методика оценивания моделей будет проходить следующим образом (рис. 2).

В качестве модели ориентира используем модель наивного прогноза. Путем создания простого базового прогноза можно определить, насколько точными будут более сложные модели. Если более сложная модель дает результаты, которые существенно лучше наивного прогноза, это можно считать показателем того, что более сложная модель является более эффективной и точной.

Эмпирически модель наивного прогноза (Naïve Forecast) временного ряда является самым распространенным ориентиром. В своей работе (Morlidge, 2013) отмечает, что одно из преимуществ наивного прогноза в качестве ориентира состоит в том, что он неявно учитывает концепцию волатильности. Это связано с тем, что

**Рис. 2.**

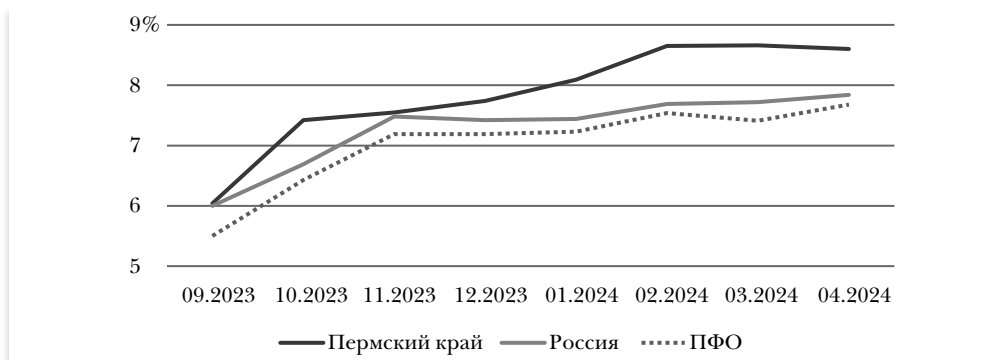
Процесс методики оценивания моделей

Источник: составлено авторами.

наивный прогноз проявляет такой же уровень изменчивости, как и сама переменная. Кроме того, ошибки, связанные с наивным прогнозом, вероятно, лучше отражают предсказуемость для анализа временных рядов. Например, в серии, где последовательные наблюдения сильно положительно коррелированы (что является признаком высокой прогнозируемости), ряд может отклоняться от своего среднего значения в течение нескольких периодов, что приводит к высокому коэффициенту вариации. В отличие от этого ошибки в наивном прогнозе оказываются относительно небольшими, поскольку последовательные наблюдения будут сходными.

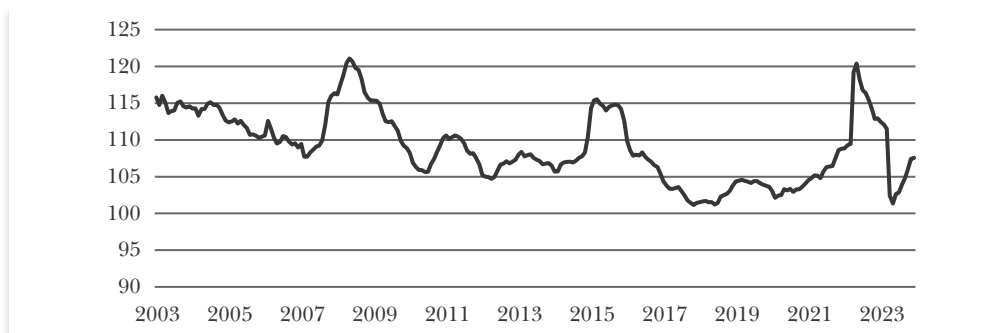
4. Результаты исследования

Далее подробно описаны результаты исследования на примере Пермского края. Регион был выбран в качестве иллюстративного примера, так как с октября 2023 г. по апрель 2024 г. инфляция в Пермском крае превышала показатели как в ПФО, так и по стране в целом (рис. 3), что подчеркивает особенность региона по сравнению с остальными.

**Рис. 3.**

Инфляция в годовом выражении, %

Источник: составлено авторами на основе данных ЕМИСС.

**Рис. 4.**

Исходные данные по ИПЦ г/г в Пермском крае

Отметим, что для каждого региона применена одинаковая процедура – приведение к стационарной форме, построение моделей и сравнение результатов.

В работе также рассматриваются спецификации эконометрических моделей временных рядов, для которых не выполняется требование к остаткам модели. Это связано с тем, что такие модели оказались точнее в сравнении с теми, для которых требование к остаткам выполнялось.

Динамика ИПЦ по Пермскому краю отличается заметным нисходящим линейным трендом и сезонностью (рис. 4). В данном случае процесс не обладает свойствами слабо стационарного ряда. Прежде всего речь идет о свойстве постоянного математического ожидания: $E(Y_t) = E(Y_{t+j}) = \mu$.

Трансформация данных позволила избавиться от тренда – процесс стал более похожим на слабо стационарный. Однако наблюдались существенные выбросы, которые необходимо заменить. Отметим, что процесс замены выбросов осуществлялся с помощью функции `tsoutlier` в R Studio. С ее помощью автоматически происходит как поиск, так и замена выбросов. Процесс идентификации выбросов состоит из следующих этапов:

- 1) функция сначала подгоняет ARIMA-модель к данным временного ряда;
- 2) после подгонки модели функция вычисляет остатки, т.е. разницу между наблюдаемыми значениями и значениями, предсказанными моделью. Остатки показывают, насколько хорошо модель описывает данные;
- 3) далее выбросы определяются на основе остатков. Остатки анализируются на наличие значений, которые значительно превышают стандартное отклонение (например, более трех стандартных отклонений от среднего);
- 4) в конце функция возвращает информацию о найденных выбросах, включая их индексы и значения;
- 5) после этого используется медиана соседних значений для замены найденных выбросов.

В данной работе для каждого региона ПФО в данных по годовой инфляции в первой разности была использована данная функция для поиска и замены выбросов без преобразований Бокса–Кокса. В результате нам удалось получить более устойчивые к выбросам данные, которые представлены на рис. 5.

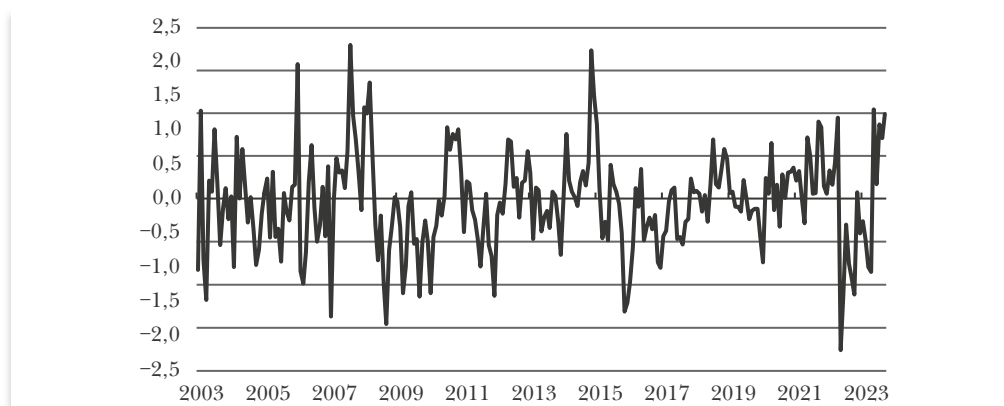


Рис. 5.

ИПЦ г/г в Пермском крае (первая разность и замена выбросов)

Обменный курс доллара США в рублях также был проанализирован на предмет соблюдения свойств слабой стационарности. Далее были применены инструменты трансформации данных – первая разность и замена выбросов, которые позволили привести рассматриваемый ряд к слабо стационарной форме с точки зрения результатов статистических тестов.

Следующий шаг эмпирической части исследования является построение моделей – ориентира, эконометрических моделей временных рядов и моделей машинного обучения, а также их комбинаций.

В табл. 1 представлены рассчитанные на основе прогноза по тестовой выборке метрики RMSE⁶ на рассматриваемые периоды прогноза для каждой из построенных моделей.

Таблица 1.

RMSE на рассматриваемые периоды прогноза для каждой модели

Модель \ Период	1	3	6	12	18	24
AR(1)	0,010	0,268	1,061	1,028	0,971	0,976
AR(12)	0,292	0,436	0,964	0,875	0,858	0,874
ARIMA(1)	0,024	0,252	1,059	1,027	0,971	0,976
ARIMA(12)	0,287	0,374	0,949	0,884	0,847	0,877
SARIMA [12]	0,169	0,202	0,963	0,883	0,876	0,917
ARIMAX(1)	0,111	0,247	1,059	1,036	0,997	0,980
ARIMAX(12)	0,209	0,349	0,939	0,874	0,855	0,865
SARIMAX [12]	0,130	0,309	0,963	0,813	0,837	0,885
Naïve	0,110	0,148	1,074	1,133	1,076	1,014
Seasonal Naïve	0,380	0,263	1,444	1,559	1,354	1,206
RF1	0,034	0,148	1,079	0,910	0,875	0,831
RF2	0,107	0,146	1,004	0,850	0,839	0,773
XgB1	0,044	0,043	1,421	1,080	1,069	0,980
XgB2	0,065	0,108	0,943	0,766	0,796	0,742

⁶ Здесь и далее в работе рассматриваются RMSE, полученные на основании прогноза на тестовой выборке.

Эконометрические модели и методы машинного обучения превосходят ориентир (Naïve и Seasonal Naïve) на всех периодах прогноза. Модель AR(1) оказалась наилучшей для прогноза на 1 месяц. Модель XgB1 оказалась наилучшей для прогноза на три месяца. Модель ARMAX(12) оказалась наилучшей для прогноза на шесть месяцев. Модель XgB2 оказалась наилучшей для прогноза на 12, 18 и 24 месяца.

Гипотеза о том, что модели машинного обучения точнее прогнозируют инфляцию на 1 период, чем эконометрические модели временных рядов, не подтверждается – величина RMSE модели AR(1) в несколько раз меньше, чем среди моделей машинного обучения. Тем не менее гипотеза о превосходстве моделей машинного обучения в прогнозе на три периода подтверждается, поскольку наилучшей оказалась модель XgB1. Более того, результаты согласуются с выводами из обзора литературы – модели машинного обучения обеспечивают более точный прогноз инфляции на долгосрочный период. Так, с помощью модели XgB2 удалось получить наименьшие значения RMSE для периода прогноза на 12, 18 и 24 месяца. На рис. 6 представлены результаты прогноза на тестовой выборке наилучших из рассмотренных моделей.

Отметим, что модели достаточно точно предсказывают паттерны динамики ИПЦ в период до 12 месяцев. Однако прогнозы на более длительный период кажутся менее эффективными. Также стоит отметить, что модели машинного обучения действительно лучше способны описывать связи между переменными, если линейная зависимость слабо выражена между ними. Это подтверждается прогнозом с большей волатильностью, что на некоторых периодах позволяет достаточно близко соотнести результаты с исходными данными.

Результаты сравнения моделей для всех регионов ПФО представлены на рис. 7.

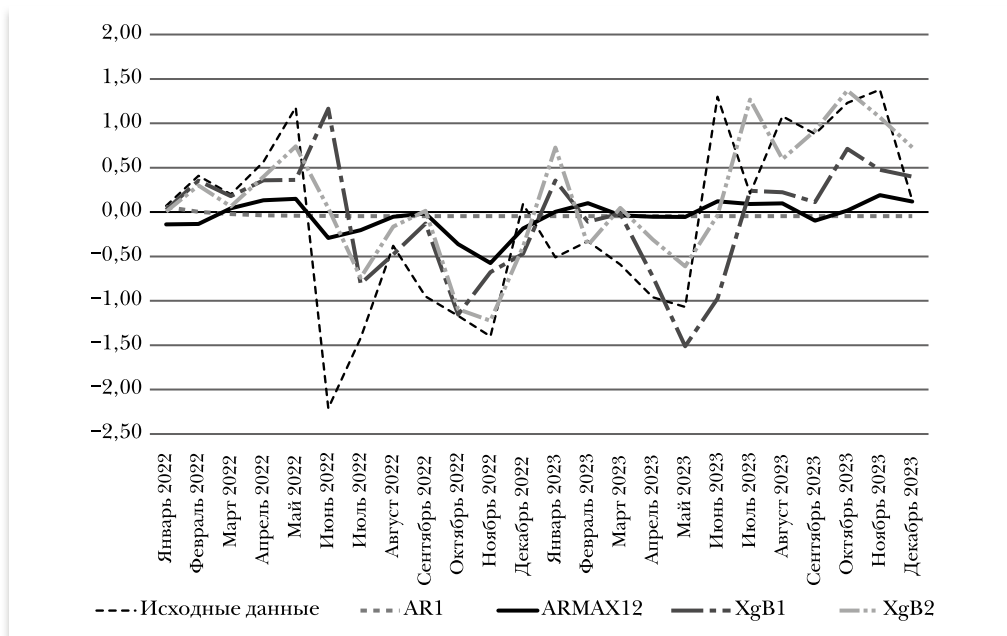
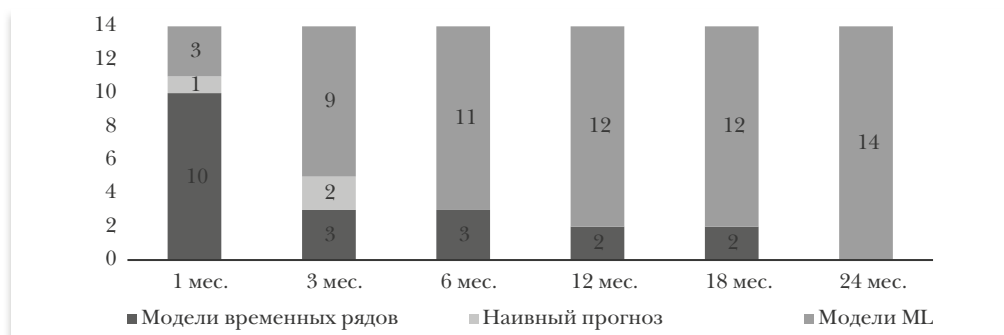


Рис. 6.

Прогнозирование ИПЦ г/г в Пермском крае (первая разность и замена выбросов)

**Рис. 7.**

Сравнение моделей прогнозирования ИПЦ г/г в регионах ПФО, разы

Рис. 7 демонстрирует, сколько раз подходы оказались наилучшими в прогнозе на конкретный период. По мере увеличения периода модели машинного обучения оказывались наиболее точными, чем эконометрические модели временных рядов и наивные модели, в прогнозе на 24 месяца ни одна спецификация эконометрических моделей не превзошла моделей машинного обучения. В то же время в прогнозе на один месяц эконометрические модели временных рядов оказались наилучшими в десяти случаях, на три месяца – в трех, на 6 месяцев – в трех, на 12 и 18 месяцев – только в двух случаях. Также стоит отметить, что в трех случаях наивные модели превзошли как эконометрический подход, так и подход, основанный на моделях машинного обучения.

Дополнительная оценка сравнения эконометрического подхода и подхода, основанного на моделях машинного обучения, в прогнозировании ИПЦ г/г в регионах ПФО представлена в табл. 2 (см. Приложение 2, п. 1, в части моделей по каждому из субъектов ПФО).

Таблица 2.

Средневзвешенная величина RMSE по результатам всех регионов ПФО на рассматриваемые периоды прогноза моделей

Модель	Период					
	1	3	6	12	18	24
AR(1)	0,318	0,437	0,719	0,882	0,853	0,881
AR(12)	0,324	0,570	0,714	0,783	0,784	0,816
ARIMA(1)	0,332	0,432	0,716	0,882	0,853	0,880
ARIMA(12)	0,316	0,519	0,679	0,768	0,766	0,805
SARIMA [12]	0,338	0,430	0,716	0,752	0,764	0,820
ARIMAX(1)	0,338	0,430	0,711	0,894	0,879	0,889
ARIMAX(12)	0,303	0,513	0,671	0,774	0,784	0,808
SARIMAX [12]	0,357	0,427	0,716	0,758	0,782	0,827
Naïve	0,436	0,400	0,726	1,078	1,070	1,000
Seasonal Naïve	0,570	0,470	0,930	1,263	1,153	1,063

Окончание таблицы 2.

Модель	Период					
	1	3	6	12	18	24
RF1	0,355	0,413	0,647	0,727	0,677	0,666
RF2	0,306	0,365	0,623	0,709	0,704	0,662
XgB1	0,408	0,460	0,683	0,755	0,719	0,701
XgB2	0,511	0,449	0,700	0,769	0,758	0,714

Модель ARIMAX(12) в среднем обеспечила наименьшую ошибку прогноза на 1 период по сравнению с остальными. Однако важно отметить, что модель RF2 в целом сопоставима с моделью ARIMAX(12) – средневзвешенная величина RMSE отличается всего на 0,003 ед. Более того, модель RF2 в среднем оказалась наилучшей из рассмотренных в прогнозе на 3, 6, 12 и 24 периода, а модель RF1 – на 18 периодов.

В среднем модели машинного обучения действительно обеспечивают более точные прогнозы региональной инфляции, чем эконометрические модели временных рядов. В 72,7% случаев наилучшими оказались модели машинного обучения, в 23,8% – модели временных рядов, в 3,5% – наивные модели. Однако в краткосрочном периоде эконометрический подход в большинстве случаев превосходит современные модели прогнозирования – особенно в прогнозе на 1 период.

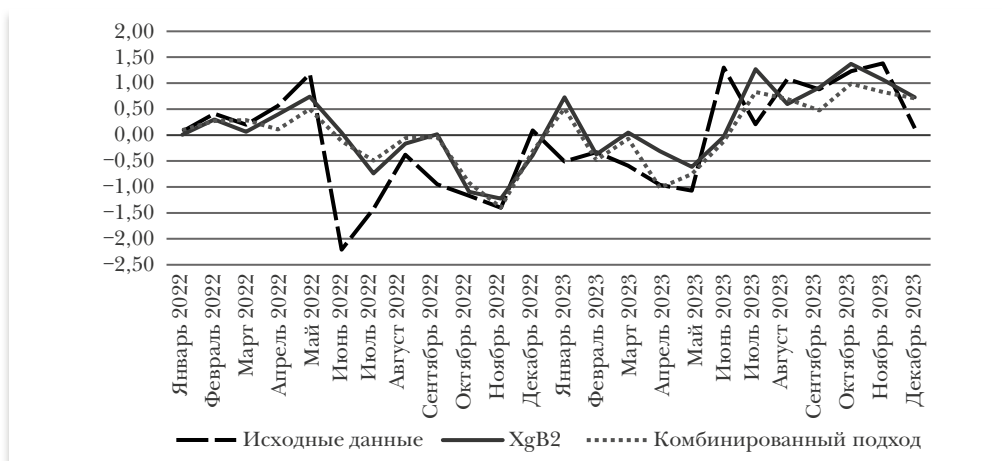
Далее рассмотрим комбинированный подход, основанный на наилучших моделях в прогнозе на рассматриваемый период, и сравним полученные результаты. Используя методику, описываемую ранее, получим метрики RMSE на все рассматриваемые периоды и сравним их с метриками лучших моделей в периоде t (табл. 5).

Результаты, представленные в табл. 3, свидетельствуют о том, что отдельно взятые наилучшие модели более точно прогнозируют ИПЦ г/г в Пермском крае, чем их комбинации с равными весами. На рис. 8 сопоставляются прогнозы комбинированного подхода и модели XgB2.

Таблица 3.

RMSE комбинированного подхода № 1 и наилучших из рассмотренных моделей

Период прогноза в t периоде (месяц)	Наилучшая модель в периоде t	RMSE наилучшей модели	RMSE комбинированного подхода (с равными весами)
3	XgB1	0,043	0,162
6	ARIMAX(12)	0,939	1,085
12	XgB2	0,766	0,876
18	XgB2	0,796	0,862
24	XgB2	0,742	0,802

**Рис. 8.**

Прогнозирование ИПЦ г/г в Пермском крае (первая разность и замена выбросов)

Из рис. 8 можно отметить, что комбинированный прогноз в целом сопоставим с прогнозом модели XgB2 – подход заметно повторяет паттерны модели машинного обучения, однако не обеспечивает более точного прогнозирования.

Тем не менее имеет смысл рассмотреть комбинированный подход, для которого веса моделей назначаются в соответствии с тем, сколько раз та или иная модель оказалась наилучшей в прогнозе на 3, 6, 12, 18 и 24 месяца. В случае с прогнозированием ИПЦ г/г в Пермском крае можно определить следующие веса, соответствующие наилучшим моделям при их сочетании:

XgB1 – наилучший прогноз только в одном периоде (3 месяца), поэтому вес при комбинировании составит $1/5$. ARMAX(12) – наилучший прогноз в одном периоде (6 месяцев), поэтому вес при комбинировании составит $1/5$. XgB2 – наилучший прогноз только в трех периодах (12, 18 и 24 месяца), поэтому вес при комбинировании составит $3/5$.

При таком соотношении весов метрики ошибок прогноза в комбинированном подходе изменились (табл. 4).

Таблица 4.

RMSE комбинированного подхода № 2 и наилучших из рассмотренных моделей

Период прогноза в периоде t (месяц)	Наилучшая модель в периоде t	RMSE наилучшей модели	RMSE комбинированного подхода (с неравными весами)
3	XgB1	0,043	0,137
6	ARIMAX(12)	0,939	1,027
12	XgB2	0,766	0,829
18	XgB2	0,796	0,831
24	XgB2	0,742	0,762

**Рис. 9.**

Регионы ПФО, для которых в ряде случаев комбинированный подход оказался наилучшим

Результаты табл. 4 чуть лучше, чем в табл. 3, однако ни в одном из периодов комбинированный подход не оказался лучше, чем одна отдельно взятая наилучшая модель в прогнозе на период t . Важно отметить, что для некоторых регионов ПФО с помощью комбинированного подхода удалось обеспечить более точные результаты (рис. 9).

Таким образом, полученные результаты не подтверждают гипотезы о том, что комбинированный подход, основанный на наиболее точных моделях прогнозирования с точки зрения метрик ошибок прогноза, обеспечивает более точные прогнозы региональной инфляции в Пермском крае на рассматриваемые периоды.

На основании данных табл. 5 можно сделать вывод о том, что комбинированный подход может выступать в качестве наиболее точного инструмента прогнозирования ИПЦ г/г лишь для Оренбургской области — для данного региона

Таблица 5.

Регионы ПФО, для которых комбинированный подход к прогнозированию ИПЦ г/г с равными/неравными весами оказался наилучшим

Регион	Комбинированный подход				
	Период прогноза				
	3 месяца	6 месяцев	12 месяцев	18 месяцев	24 месяца
Республика Марий Эл				№ 2	№ 2
Пензенская область					№ 1
Оренбургская область	№ 1		№ 1	№ 2	№ 2
Республика Мордовия		№ 2			
Удмуртская Республика				№ 1	№ 1
Нижегородская область			№ 1		№ 1
Самарская область			№ 2	№ 2	№ 2
Саратовская область		№ 2	№ 2		№ 1
Ульяновская область				№ 1	
Республика Татарстан			№ 2	№ 2	№ 2

ПФО такой подход оказался наилучшим в четырех из пяти рассматриваемых случаев (комбинация из моделей XgB1 и XgB2). Важно отметить, что для Республики Башкортостан и Кировской области комбинированный подход не применялся, по причине того что в прогнозе на рассматриваемые периоды только одна модель оказалась наилучшей – XgB2и RF2 соответственно.

Тем не менее, ни для одного региона ПФО не удалось подтвердить гипотезу о том, что комбинированный подход, основанный на моделях, обеспечивающих наибольшую точность прогноза на период t , способен более точно спрогнозировать ИПЦ г/г в регионах ПФО на 3, 6, 12, 18 и 24 периода (рис. 10).

По результатам, представленным на рис. 10, мы делаем вывод о том, что в более чем половине случаев модели машинного обучения (53,6%) оказались

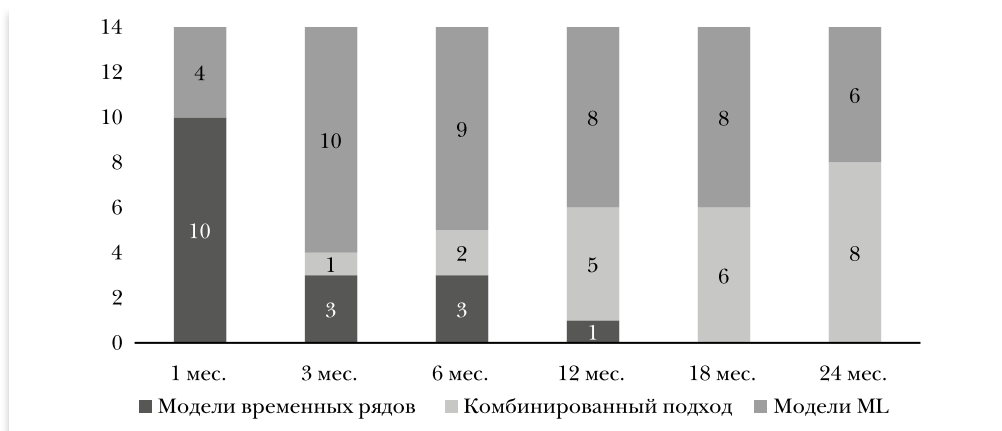


Рис. 10.

Сравнение моделей прогнозирования ИПЦ г/г в регионах ПФО, разы

наилучшими в прогнозе ИПЦ г/г в регионах ПФО, эконометрические модели временных рядов — в 20,6%. Комбинированный подход как с равными, так и с неравными весами, оказался наиболее точным в 25,8% случаев.

Вместе с тем результаты исследования были проверены на предмет устойчивости. В Приложении 2 представлены значения *p-value* парных тестов Диболда–Мариано для каждого региона. Результаты свидетельствуют о том, что зачастую остатки прогнозов по эконометрическим моделям временных рядов значимо отличаются от подхода с применением моделей машинного обучения. Это подчеркивает надежность и достоверность полученных выводов об эффективности того или иного подхода в прогнозировании инфляции в регионах ПФО.

Выводы, полученные в ходе исследования, согласуются с более ранней работой (Букина, Кашин, 2024), что дополнительно подчеркивает устойчивость результатов моделирования. В статье (Букина, Кашин, 2024) авторы также установили преимущество классических эконометрических моделей над методами машинного обучения в прогнозе инфляции в регионах ПФО на краткосрочном горизонте, а на средне- и долгосрочном — более точные прогнозы обеспечивают именно модели машинного обучения. Отметим, что в нашем исследовании с применением комбинированного подхода удалось достичь более точных результатов прогноза, в том числе используя разработанный метод подбора моделей и их весов, что ранее не встречалось в литературе по прогнозированию региональной инфляции.

На означенных ранее результатах исследования также сказалось значительное внимание к классическим эконометрическим моделям. В литературе (см. например, (Семитуркин, Шевелев, 2022; Baybuza, 2018; Pratar, Sengupta, 2019)) зачастую рассматриваются базовые спецификации моделей ARIMA и ARIMAX — эти модели используются в качестве ориентиров для сравнения различных подходов к прогнозированию. В нашей работе применялись альтернативные методы сравнения моделей. В качестве моделей-ориентиров мы использовали наивные прогнозы — модели *Naïve* и *Seasonal Naïve*. Эти модели, несмотря на простоту построения, позволили достаточно точно оценить эффективность более сложных методов прогнозирования.

5. Заключение

Прогнозирование региональной инфляции является важной задачей для всех макроэкономических агентов. Учет будущей динамики уровня цен в регионе позволяет Банку России и правительству страны принимать решения по направлениям экономической политики. Вместе с тем фирмы планируют свое производство, совершают инвестиционные сделки и привлекают заемные средства, принимая во внимание будущие темпы роста цен. Домохозяйства же ориентируются на прогнозные значения инфляции при планировании бюджета и определении заработной платы на рынке труда в регионе.

Для прогнозирования инфляции зачастую используются эконометрические модели временных рядов и модели машинного обучения, а также их комбинирование. Причем для прогноза инфляции в регионах существенных ограничений в использовании того или иного метода нет. Так, в данной работе сравнивались оба подхода к прогнозированию инфляции на основе показателей

индекса потребительских цен в годовом выражении регионов ПФО и среднемесячного рыночного обменного курса доллара США в рублях.

Результаты исследования позволяют сделать вывод о том, что модели машинного обучения действительно обеспечивают более точные прогнозы инфляции в регионах ПФО по сравнению с эконометрическим подходом. Однако необходимо отметить, что эконометрические модели временных рядов отличаются более высокой прогностической силой на один период. Вдобавок комбинированный подход на основе наилучших моделей не обеспечил более точных прогнозов на все рассматриваемые периоды, однако его применение в ряде случаев может быть более эффективным — особенно в прогнозе на 24 периода.

Кроме того, для некоторых регионов наивные модели прогнозирования оказались лучше, чем эконометрические и машинного обучения⁷. Данный подход хотя и является универсальным, однако нельзя сказать, что он может стабильно обеспечивать точные прогнозы инфляции в регионах. Превосходство метода оказалось скорее случайным, чем закономерным.

Отметим, что полученные результаты обусловлены спецификой рассматриваемых методов. Как отмечалось в обзоре литературы, в сравнении с эконометрическим подходом модели машинного обучения точнее прогнозируют числовую переменную ввиду относительной гибкости к выбросам. Так, с увеличением периода прогноза и общей волатильности прогнозной выборки⁸ модели машинного обучения оказывались более эффективны, чем другие методы.

Вместе с тем результаты показали, что они не являются одинаковыми для всех регионов ПФО. Ранее предполагалось, что отличия могут быть обусловлены гетерогенностью в динамике инфляции или структуре экономики. Тем не менее, в работе не удалось обнаружить устойчивого эффекта региональных особенностей в результатах прогнозирования, вызванного спецификой региона, а именно волатильностью инфляции, структурой экономики региона.

В работе присутствуют некоторые ограничения. Во-первых, при спецификации прогнозных моделей использовалась только одна макроэкономическая переменная — рыночный обменный курс доллара США. Другие показатели, которые могли бы выступать в качестве факторов, определяющих динамику инфляции⁹, не были рассмотрены ввиду ограниченности данных по месяцам в разрезе регионов в период с 2003 по 2023 г. Во-вторых, применение моделей машинного обучения предполагает процесс подбора гиперпараметров, который необходим для достижения оптимальных прогнозов. Учитывая, что для каждого региона ПФО рассмотрено по четыре спецификации моделей машинного обучения (по две модели Random Forest и XgBoost), вычисление оптимального набора параметров для 64 моделей (в совокупности по всем регионам ПФО) значительно увеличило бы загруженность вычислительных мощностей. В связи с этим подбор гиперпараметров был ограничен небольшим набором возможных вариантов, что могло повлиять на точность прогнозирования моделей. Более того, возможно провести настройку коэффициентов моделей на отложенной выборке

⁷ С помощью наивных моделей удалось получить наиболее точный прогноз инфляции на один период в Республике Мордовия, на три периода — в Республике Мордовия и Чувашской Республике. Отметим, что в данных регионах ИПЦ г/г (первая разность и замена выбросов) отличается сравнительно низкой динамикой в первые три месяца периода тестовой выборки. Так, инертность в динамике цен была успешно спрогнозирована наивной моделью, прогноз по которой сводится к константе.

⁸ Тестовая выборка была сформирована на основе данных за 2022–2023 гг. — период, в котором инфляция отличалась высокой волатильностью.

⁹ Например данные по инфляционным ожиданиям потребителей и производителей в разрезе регионов с начала 2000-х годов.

для достижения большей объективности и предотвращения искажения результатов. В-третьих, прогнозы инфляции с помощью моделей RF2 и Xgb2 складываются из прогнозных значений обменного курса¹⁰, что вызывает определенные сложности в практическом применении данных спецификаций.

В качестве будущих направлений исследования имеет место применение большего числа моделей машинного обучения – например метод KNN (k-nearest neighbors) или метод опорных векторов (support vector machine), которые зачастую используются в научной литературе. Продолжая развитие темы сравнительного анализа моделей прогнозирования, стоит также рассмотреть новые подходы – структурные модели DSGE или квартально-прогнозные модели (QPM), которые на практике используются в центральных банках при прогнозировании инфляции как в регионе, так и по стране в целом. Важным направлением развития работы также может стать выявление региональных особенностей, ввиду которых возникают отличия в моделях, обеспечивающих точный прогноз инфляции в регионах ПФО на конкретный период.

ПРИЛОЖЕНИЕ 1¹¹

П1. ИПЦ год к году

Республика Башкортостан характеризуется наименьшим уровнем ИПЦ г/г среди остальных регионов ПФО (99,38 п.п.) за 2023 г., в то время как Ульяновская область – наибольшим значением ИПЦ г/г (113,50 п.п.). В среднем уровень ИПЦ г/г в регионах ПФО составил от 104,76 до 107,33 п.п. в 2023 г. Оренбургская область отличается повышенной волатильностью ИПЦ г/г в сравнении с остальными регионами ПФО – величина стандартного отклонения ИПЦ г/г за 2023 г. составила 4,21 п.п. (рис. П1).

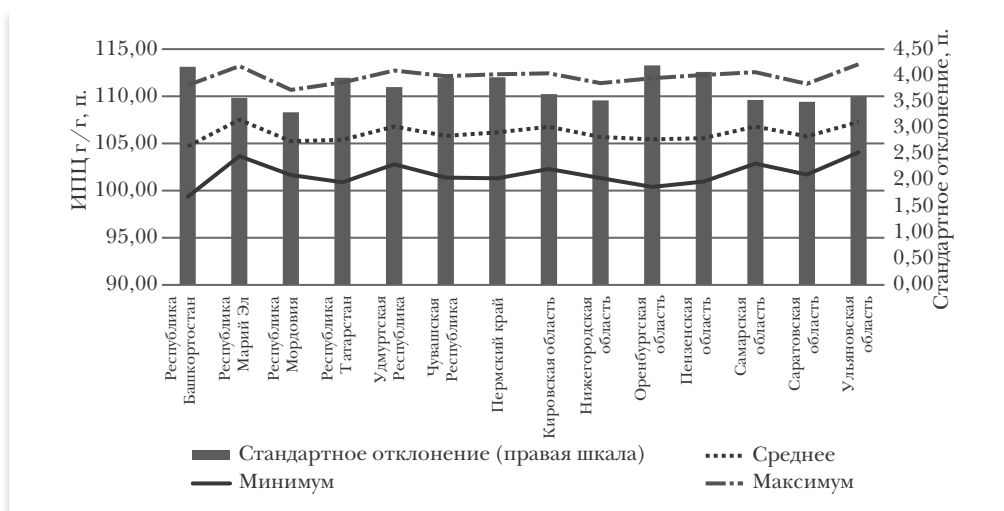


Рис. П1.

ИПЦ г/г регионов ПФО за 2023 г.

Источники: ЕМИСС, расчеты авторов.

¹⁰ Рыночный обменный курс доллара США (первая разность и замена выбросов) в рублях в периоде t выступает в качестве фактора в моделях RF2 и Xgb2, в которых зависимой переменной является ИПЦ г/г в периоде t (первая разность и замена выбросов).

¹¹ Приложение 2 доступно по адресу (http://journal.econorus.org/japp.phtml#Gabov_Bukina_Kashin_JNEA_4_2025).

П2. Рыночный обменный курс доллара

Наименьшее значение обменного курса доллара США в рублях в основном приходится на 2007–2008 гг. – в этот период среднегодовой уровень находится в пределах от 24,86 до 25,58 руб. за 1 долл. США. Первый резкий скачок курса пришелся на 2014 г. и составил 38,47 руб. в среднем за год при волатильности в 6,47 руб. В 2022 г. курс достиг исторического максимума – 103,47 руб. за 1 долл. США в среднем за месяц. Более того, амплитуда волатильности в 2022 г. составила 13,40 руб. – наибольшее значение за весь рассматриваемый период. В 2023 г. курс находился в диапазоне от 68,88 до 97,12 руб. за 1 долл. США при волатильности в 9,64 руб. (рис. П2).

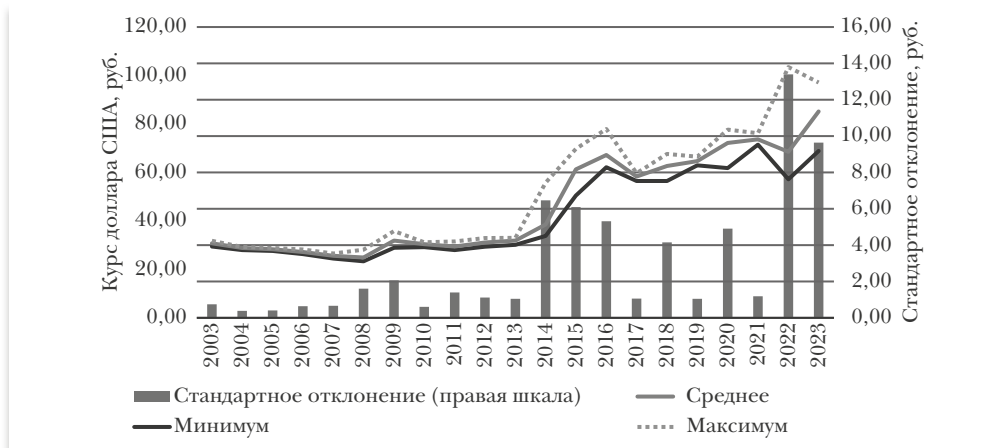


Рис. П2.

Рыночный обменный курс доллара США, руб.

Источники: Банк России, расчеты авторов.

П3. Прогнозирование ИПЦ (первая разность и замена выбросов)

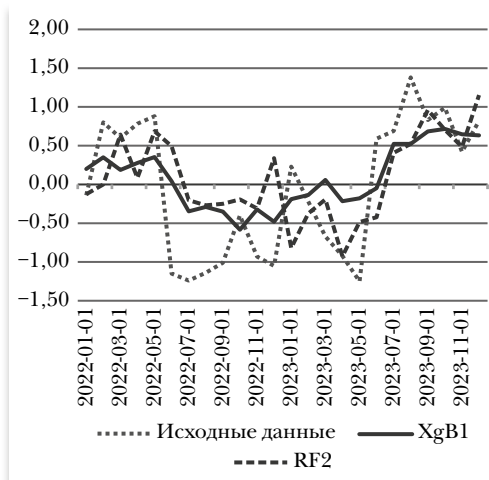


Рис. П3.

Прогнозирование ИПЦ, Кировская область



Рис. П4.

Прогнозирование ИПЦ, Нижегородская область

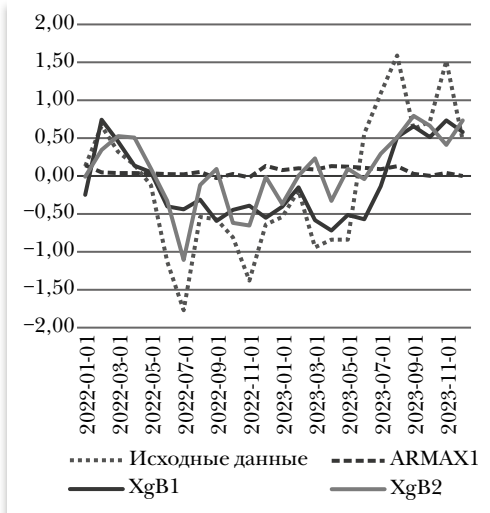


Рис. П5.

Прогнозирование ИПЦ, Оренбургская область

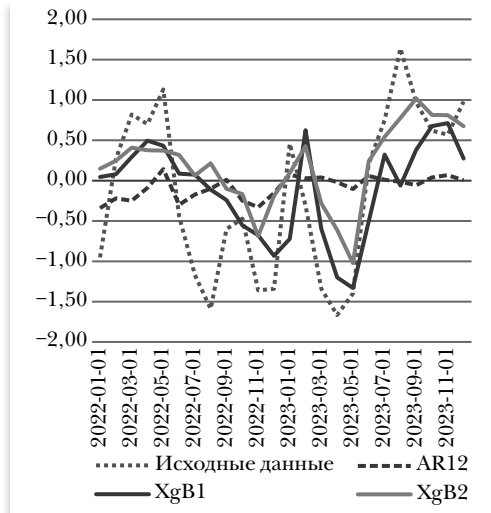


Рис. П6.

Прогнозирование ИПЦ, Пензенская область

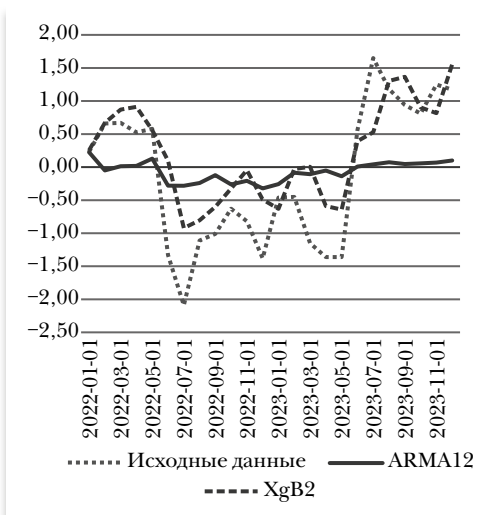


Рис. П7.

Прогнозирование ИПЦ,
Республика Башкортостан

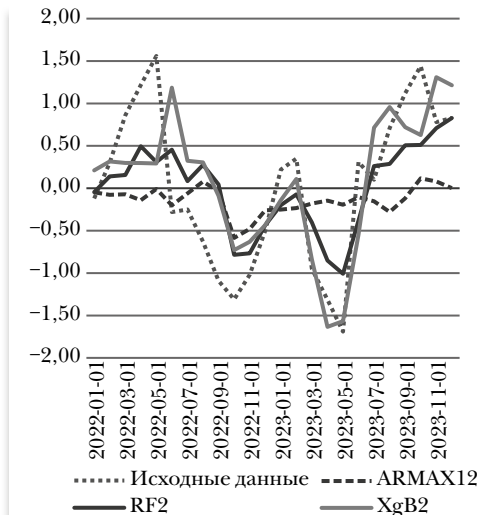


Рис. П8.

Прогнозирование ИПЦ, Республика Марий Эл

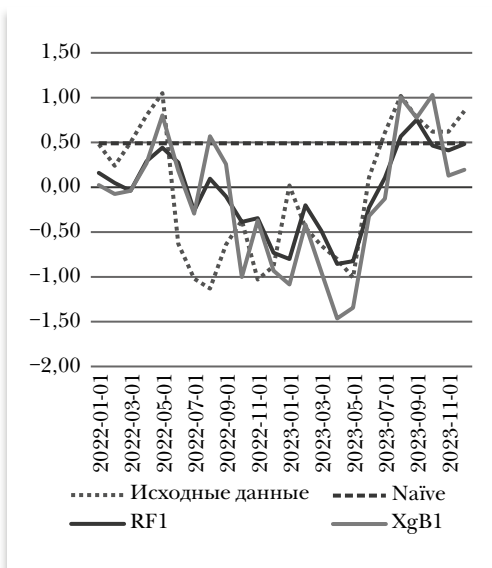


Рис. П9.

Прогнозирование ИПЦ, Республика Мордовия

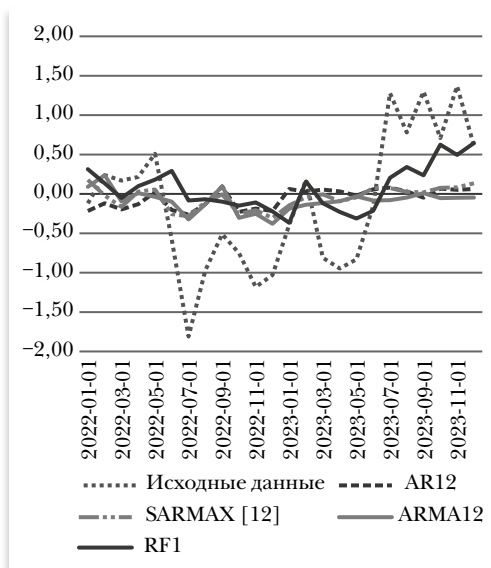


Рис. П10.

Прогнозирование ИПЦ, Республика Татарстан

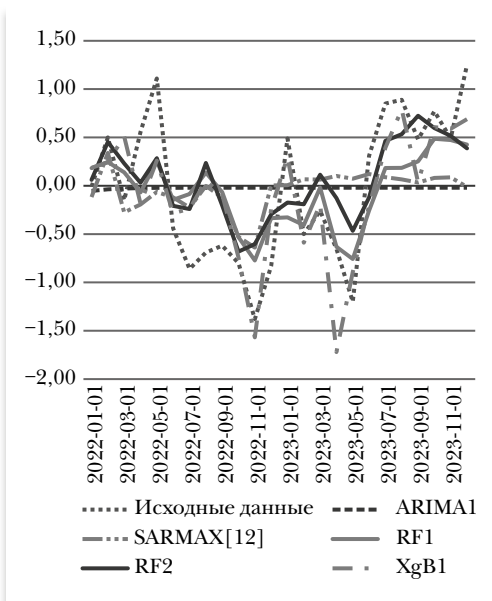


Рис. П11.

Прогнозирование ИПЦ, Самарская область

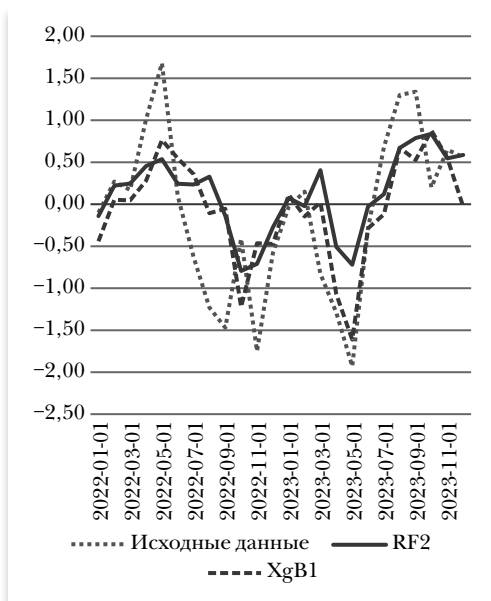


Рис. П12.

Прогнозирование ИПЦ, Саратовская область

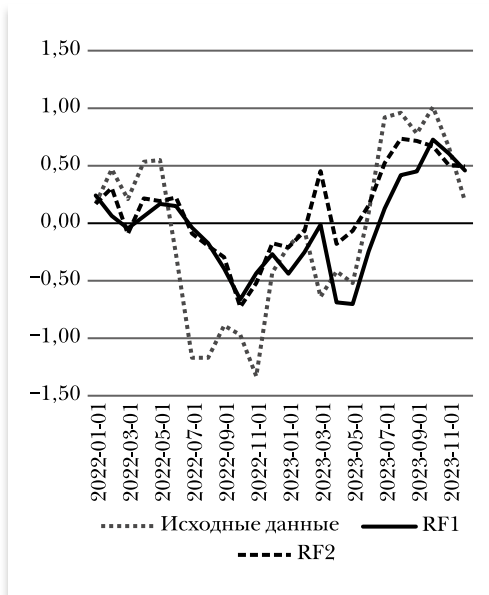


Рис. П13.
Прогнозирование ИПЦ, Республика Удмуртия

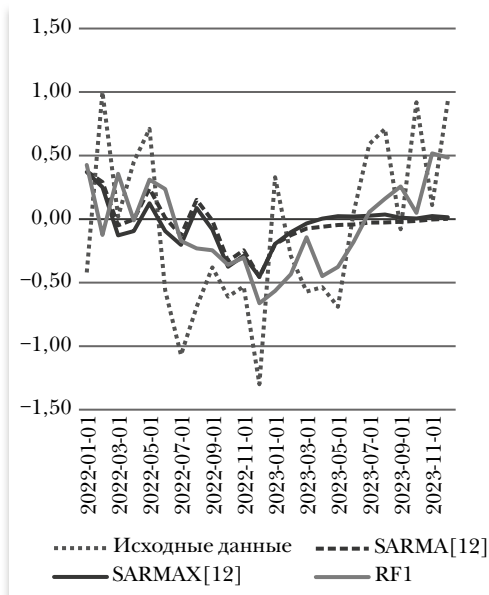


Рис. П14.
Прогнозирование ИПЦ, Ульяновская область

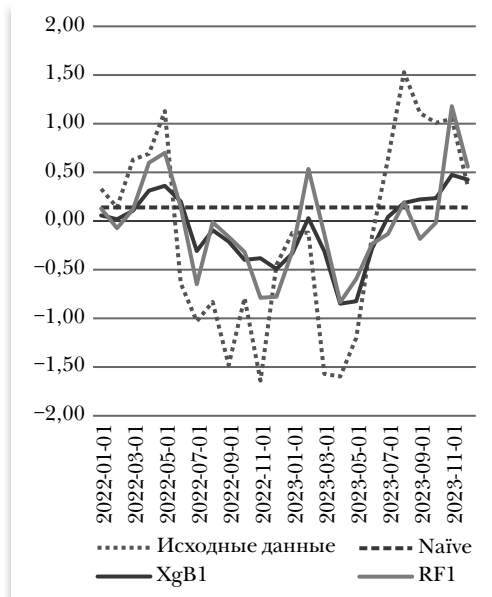


Рис. П15.
Прогнозирование ИПЦ, Республика Чувашия

П4. Средневзвешенные величины RMSE

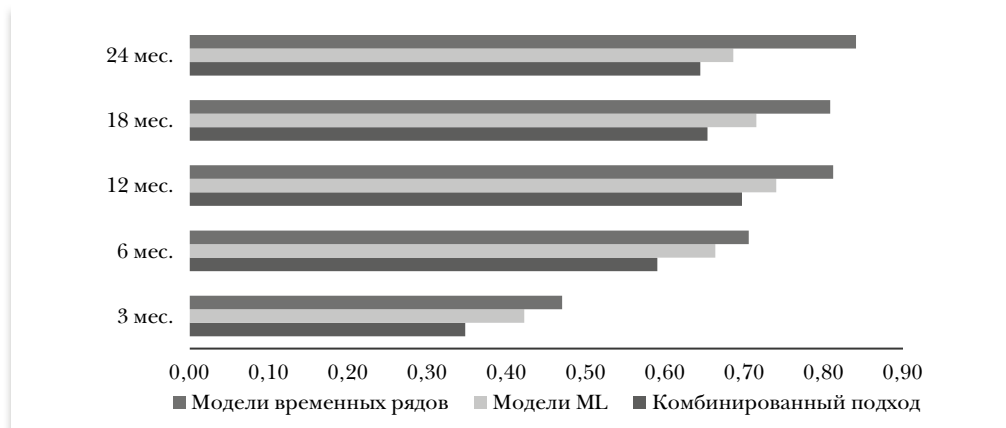


Рис. П16.

Средневзвешенные величины RMSE трех подходов к прогнозированию ИПЦ г/г в регионах ПФО на рассматриваемые периоды

ЛИТЕРАТУРА / REFERENCES

- Букина Т.В., Кашин Д.В.** (2024). Прогнозирование региональной инфляции: эконометрические модели или методы машинного обучения? // *Экономический журнал ВШЭ*. № 28 (1). С. 81–107. [Bukina T.V., Kashin D.V. (2024). Regional inflation forecasting: Econometric models versus machine learning methods? *The HSE Economic Journal*, 28 (1), 81–107 (in Russian).]
- Семитуркин О., Шевелев А.** (2022). Прогнозирование региональной инфляции с помощью методов машинного обучения на примере макрорегиона Сибирь // *Банк России*. № 91/март. [Semiturkin O., Shevelev A. (2022). Forecasting regional inflation rates using machine learning methods: The case of Siberia macroregion. *Bank of Russia*, 91/March (in Russian).]
- Almosova A., Andresen N.** (2023). Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, 42 (2), 240–259.
- Baciu I.C.** (2015). Stochastic models for forecasting inflation rate. Empirical evidence from Romania. *Procedia Economics and Finance*, 20, 44–52.
- Baybuza I.** (2018). Inflation forecasting using machine learning methods. *Russian Journal of Money and Finance*, 77 (4), 42–59.
- Chakraborty C., Joseph A.** (2017). Machine learning at central banks. *Bank of England Working Paper no. 674*. Available at SSRN: <https://ssrn.com/abstract=3031796> or <http://dx.doi.org/10.2139/ssrn.3031796>
- Doguwa S.I., Alade S.O.** (2013). Short-term inflation forecasting models for Nigeria. *CBN Journal of Applied Statistics*, 4 (3), 1–29.
- Elliott G., Timmermann A.** (2005). Optimal forecast combination under regime switching. *International Economic Review*, 46 (4), 1081–1102.
- Faust J., Wright J.H.** (2013). Forecasting inflation. *Handbook of Economic Forecasting*, 2, 2–56. Amsterdam: Elsevier.

- Kapetanios G., Labhard V., Price S.** (2008). Forecast combination and the Bank of England's suite of statistical forecasting models. *Economic Modelling*, 25 (4), 772–792.
- Lidiema C.** (2017). Modelling and forecasting inflation rate in Kenya using SARIMA and Holt-Winters triple exponential smoothing. *American Journal of Theoretical and Applied Statistics*, 6 (3), 161–169.
- Mihretu M.E.** (2023). Inflation forecasting models and forecasting combination analysis: The case of Ethiopia. *National Bank of Ethiopia*, April, 136. Available at: <https://nbe.gov.et/wp-content/uploads/2023/07/Birritu-Magazine-136.pdf>
- Morlidge S.** (2013). How good is a “good” forecast? Forecast errors and their avoidability. *Foresight: The International Journal of Applied Forecasting*, 30, 5–11.
- Mustapha A.M., Kubalu A.I.** (2016). Application of box-inflation dynamics. *Ilimi Journal of Arts and Social Sciences*, 2 (1), 127–142.
- Nyoni T., Mutongi C.** (2019). Modeling and forecasting inflation in the Gambia: An ARMA approach. *MPRA Paper 93980*. University Library of Munich, Germany.
- Nyoni T., Nathaniel S.P.** (2018). Modeling rates of inflation in Nigeria: An application of ARMA, ARIMA and GARCH models. *Munich Personal RePEc Archive, Paper no. 91351*.
- Öğünç F., Akdoğan K., Başer S., Chadwick M.G., Ertuğ D., Hülagü T., Kösem S., Özmen M.U., Tekatlı N.** (2013). Short-term inflation forecasting models for Turkey and a forecast combination analysis. *Economic Modelling*, 33, 312–325.
- Otu O.A., Osuji G.A., Opara J., Mbachu H.I., Iheagwara A.I.** (2014). Application of Sarima models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, 2 (1), 16–28.
- Popoola O.P., Ayanrinde A.W., Rafiu A.A., Odusina M.T.** (2017). Time series analysis to model and forecast inflation rate in Nigeria. *Annals. Computer Science Series*, 15 (1), 174–178.
- Pratap B., Sengupta S.** (2019). Macroeconomic forecasting in India: Does machine learning hold the key to better forecasts? *RBI Working Paper Series, no. 04/2019*, Reserve Bank of India, Mumbai, India.
- Saz G.** (2011). The efficacy of SARIMA models for forecasting inflation rates in developing countries: The case for Turkey. *International Research Journal of Finance and Economics*, 62, 111–142.
- Tamuke E., Jackson E.A., Sillah A.** (2018). Forecasting inflation in Sierra Leone using ARIMA and ARIMAX: A comparative evaluation. Model building and analysis team. *Theoretical and Practical Research in Economic Fields*, 9 (1), 63–74.
- Tena J.D., Espasa A., Pino G.** (2010). Forecasting Spanish inflation using the maximum disaggregation level by sectors and geographical areas. *International Regional Science Review*, 33 (2), 181–204.
- Udom P., Phumchusri N.** (2014). A comparison study between time series model and ARIMA model for sales forecasting of distributor in plastic industry. *IOSR Journal of Engineering*, 4 (2), 32–38.
- Uwilingiyimana C., Munga'tu J., Harerimana J.D.** (2015). Forecasting inflation in Kenya using ARIMA-GARCH models. *International Journal of Management and Commerce Innovations*, 3 (2), 15–27.

Поступила в редакцию 20.11.2024

Received 20.11.2024

M.A. Gabov

Perm Territorial Division of the Ural Main Branch of the Central Bank of the Russian Federation; HSE University, Perm, Russia

T.V. Bukina

HSE University, Perm, Russia

D.V. Kashin

HSE University, Perm, Russia

Comparative analysis of regional inflation forecasting models¹²

Abstract. The study aims to compare approaches to forecasting the monthly level of consumer price index (CPI y/y) in the regions of the Volga Federal District using time series models and machine learning methods. This study attempts to select the most appropriate and efficient models for predicting the regional general price level index. The paper also shows the use of a combined approach, which is based on the combination of both methods. The results show that machine learning models provide more stable and accurate forecasts than econometric models – especially over long forecasting periods (6 months or more). However, for several regions, we found evidence of the effectiveness of time series models for the short term – for several regions, different specifications of extended autoregressive models perform better than the machine learning model approach when forecasting for 1 and 3 months. The results of the combined approach are comparable to the forecasts of machine learning models and more often provide more accurate forecasts for 12 and 24 months. The study showed that it was not possible to detect a sustainable effect of regional characteristics in the forecasting results caused by the specifics of the region, namely the volatility of inflation and the structure of the regional economy.

Keywords: *forecasting, regional inflation, machine learning, econometric modeling, combined approach.*

JEL Classification: E31, E37.

For reference: **Gabov M.A., Bukina T.V., Kashin D.V.** (2025). Comparative analysis of regional inflation forecasting models. *Journal of the New Economic Association*, 4 (69), 87–117.

DOI: 10.31737/22212264_2025_4_87-117

EDN: WBWFAU

¹² The views expressed herein are solely those of the authors. The content and results of this research should not be considered or referred to in any publications as the Bank of Russia's official position, official policy, or decisions. Any errors in this document are the responsibility of the authors.

We thank the anonymous reviewers for comments and suggestions which helped us improve the paper.