

УДК 332.1

Прогнозирование региональной инфляции: эконометрические модели или методы машинного обучения?¹

Букина Т.В., Кашин Д.В.

В статье строится прогноз региональной инфляции на примере субъектов, входящих в Приволжский федеральный округ (ПФО)². Цель исследования – определить модель, которая точнее остальных прогнозирует региональную инфляцию. В работе производится сравнение инструментов машинного обучения – метода опорных векторов, градиентного бустинга и случайного леса с эконометрическими моделями временных рядов, авторегрессией и интегрированной авторегрессией-скользящего среднего, – моделями, которые чаще используются для прогнозирования общероссийской инфляции. По результатам исследования выявлены ключевые макроэкономические показатели, наиболее существенно влияющие на региональную инфляцию. В тройку таких для каждого региона ПФО входит значение инфляции в прошлом месяце, среднее значение инфляции и номер месяца. Показано, что модели машинного обучения не хуже справляются с задачей прогнозирования региональной инфляции на длительных временных периодах, в то время как эконометрические модели достаточно точно прогнозируют на краткосрочных временных горизонтах.

Ключевые слова: прогнозирование; региональная инфляция; машинное обучение; градиентный бустинг; случайный лес; эконометрическое моделирование.

DOI: 10.17323/1813-8691-2024-28-1-81-107

¹ Авторы выражают благодарность анонимному рецензенту за полезные замечания, рекомендации и ценные советы, позволившие доработать рукопись.

² Данные для исследования подготовлены выпускником НИУ ВШЭ-Пермь, Шаймухаметовым Д.М., под руководством к.г.н., доцента департамента экономики и финансов Букиной Т.В.

Татьяна Витальевна Букина – к.г.н., доцент, доцент департамента экономики и финансов. Национальный исследовательский университет «Высшая школа экономики». E-mail: tbukina@hse.ru
Дмитрий Викторович Кашин – к.э.н., доцент департамента экономики и финансов. Национальный исследовательский университет «Высшая школа экономики». E-mail: dvkashin@hse.ru

Статья поступила: 24.10.2023/Статья принята: 08.02.2024.

Для цитирования: Букина Т.В., Кашин Д.В. Прогнозирование региональной инфляции: эконометрические модели или методы машинного обучения? *Экономический журнал ВШЭ*. 2024; 28(1): 81–107.

For citation: Bukina T.V., Kashin D.V. Regional Inflation Forecasting: Econometric Models Versus Machine Learning Methods? *HSE Economic Journal*. 2024; 28(1): 81–107. (In Russ.)

Введение

Один из важнейших индикаторов состояния экономики – инфляция – рассматривается на основе индекса потребительских цен (ИПЦ)³. Прогнозирование инфляции сегодня является актуальной задачей, которая приобрела еще большую значимость после перехода к режиму таргетирования инфляции в рамках денежно-кредитной политики Банка России, принятой в 2014 г.⁴ Прогноз значений инфляции на будущие периоды важен не только для государственных структур в части принятия решений о планируемых сценариях развития экономики, но и для бизнеса и домохозяйств – будь то решение об утверждении корпоративной стратегии, разработке бизнес-модели или плана крупных покупок.

Прогнозирование инфляции в регионах России – не менее важная задача. С одной стороны, сочетание региональных прогнозов может повысить качество национального прогноза инфляции [Tena et al., 2010], с другой – местные стейкхолдеры больше заинтересованы в изучении инфляции в конкретном регионе, непосредственно учитывая сведения об уровнях региональной инфляции в своей деятельности.

В работе на примере 14 субъектов Приволжского федерального округа анализируются региональные и внешние макроэкономические показатели – ИПЦ, уровень безработицы, курс доллара США, процентная ставка за период с января 2010 г. по декабрь 2022 г. Данные выгружены из Единой межведомственной информационно-статистической системы (ЕМИСС), использован R Studio для прогнозирования временных рядов, и Python для моделей машинного обучения (в частности, для метода опорных векторов, градиентного бустинга и случайного леса). Для оценки качества каждой модели мы использовали ошибку прогноза RMSE. Базовой моделью является модель авторегрессии первого порядка AR(1), которую мы использовали в качестве бенчмарка. Для сравнения моделей мы рассчитали отношение RMSE каждой модели к RMSE базовой модели. Если значение показателя получалось меньше единицы, то прогноз сравниваемой модели считался лучше бенчмарка, а если больше единицы, то хуже.

В результате исследования выявлено, что модель градиентного бустинга является наиболее точной среди всех рассмотренных моделей машинного обучения при прогнозировании региональной инфляции. Модель представляет более точные прогнозы, чем авторегрессионные модели на большем количестве периодов прогнозирования. В частности, на горизонтах прогнозирования в 3, 6, 21 и 24 месяца модель градиентного бустинга превосходит базовую модель AR(1) и лишь на горизонте в 1 месяц показывает качест-

³ ИПЦ отражает общий уровень цен по достаточно разнородному набору товаров и услуг. Рост ИПЦ показывает, как растет стоимость жизни в стране или в отдельно взятом регионе.

⁴ Заявление Председателя Банка России Э.С. Набиуллиной по итогам заседания Совета директоров, 12 сентября 2014 г. (<https://www.cbr.ru/press/event/?id=5208> (дата обращения 17.10.2023)).

во, сравнимое с базовой моделью. Прогноз на горизонты в 9 и 12 месяцев незначительно хуже (на 6,9% и 3,5% соответственно) в сравнении с базовой моделью. Модели случайного леса и метод опорных векторов также представляют точные прогнозы в сравнении с эконометрическими моделями, но только на длительных горизонтах в 21 и 24 месяца. Полученные результаты подтверждают гипотезу о том, что модели машинного обучения могут быть эффективным инструментом для прогнозирования региональной инфляции на разных временных горизонтах. Однако использование эконометрических моделей также оправдано, особенно при прогнозировании инфляции на более короткие горизонты (до 12 месяцев). Для достижения наилучшего качества прогнозов рекомендуется комбинировать прогнозы на основе моделей машинного обучения и эконометрических моделей.

Обзор подходов к прогнозированию инфляции

В 1958 г. У. Филлипс впервые предложил взаимосвязь между уровнем безработицы и уровнем цен, показав, что при росте безработицы цены снижаются. Исследуя данные в Великобритании за период с 1861 г. по 1957 г., Филлипс подтвердил эту зависимость эмпирически [Phillips, 1958]. Наличие взаимосвязи между уровнем цен и уровнем безработицы обнаружили и в других странах мира, например, в США [Samuelson, Solow, 1960]. В дальнейшем модель усложнялась, появлялись модификации, которые включали ожидания экономических агентов, инфляцию предыдущих периодов, естественный уровень безработицы (NAIRU). К примеру, Сток и Ватсон в 1999 г. использовали модифицированные кривые Филлипса для прогнозирования инфляции на помесечных данных в США в период с 1959 г. по 1997 г. [Stock, Watson, 1999]. Сравнивая базовые кривые Филлипса с многофакторными моделями, авторы показали, что базовая модель кривых Филлипса прогнозирует точнее на горизонте в один год, в сравнении с прогнозами по моделям с использованием макроэкономических показателей. Другие авторы, Аткинсон и Оганян, в своем исследовании использовали в качестве базовой модели модель наивного прогноза и сравнивали с ошибкой прогноза этой модели, ошибки прогнозов, рассчитанные по кривым Филлипса [Atkeson, Ohanian, 2001]. На помесечных и поквартальных данных за период с 1959 г. по 2000 г. авторы выявили, что кривые Филлипса в любой спецификации не являются более точными методами прогнозирования, чем наивный прогноз.

В 2008 г. Сток и Ватсон впервые предложили классификацию моделей прогнозирования инфляции, выделив 4 категории моделей [Stock, Watson, 2008]. В первую группу авторы отнесли модели, основанные исключительно на исторических данных об инфляции, а именно модели авторегрессии-скользящего среднего (ARMA), модели случайного блуждания (RW) и модель стохастической волатильности с ненаблюдаемыми компонентами (USCV). Вторая группа – модели, в которые включаются показатели экономической активности, в основном это модели по кривым Филлипса. В третью группу авторы отнесли модели, в основе которых лежат прогнозы профессиональных участников рынка или инфляционные ожидания экономических агентов. В четвертую группу вошли многофакторные модели, основанные на индикаторах экономической активности, не сочетаемые с моделями из второй группы. Основной вывод исследователей следующий – модели, которые используют только прошлые значения инфляции, в среднем строят более точный прогноз, чем все остальные группы моделей. В продолжение исследования Стока и Ватсона, более полное сравнение методов прогнозирования инфляции представлено в работе [Faust,

Wright, 2013]. Авторы сравнили 17 моделей, среди них: кривые Филлипса и их модификации, динамические стохастические модели общего равновесия (DSGE), модели равновзвешенного усреднения (EWA) и байесовские модели усреднения (BMA), различные модели временных рядов, как однофакторные – AR и ее модификации, RW и ее модификации, UCSV, так и многофакторные – структурные векторные авторегрессии (SVAR), VAR с изменяющимися во времени параметрами (TVP-VAR) и факторно-дополненные векторные авторегрессии (FAV). Вывод авторов соотносится с предыдущими исследованиями – однофакторные модели прогнозирования инфляции не хуже других, более сложных моделей.

Ряд исследователей предположили, что возможно повысить качество общего прогноза инфляции, объединив дезагрегированные прогнозы, построенные по отдельным компонентам инфляции. Например, Хубрич, изучая гармонизированный индекс потребительских цен Еврозоны (HICP), получил результат, что агрегация, а также комбинирование прогнозов как по отдельным компонентам, так и по различным моделям, не приводит к снижению ошибки общего прогноза инфляции [Hubrich, 2005]. Другие авторы получают противоположный результат: агрегированные прогнозы достаточно точные, однако необходимо подобрать модель под определенный тип данных [Bermingham, D'Agostino, 2011]. Ряд авторов [Tena et al., 2010; Chodorow-Reich, 2020] показывают, что комбинирование прогнозов по отраслям экономики и по отдельным регионам повышают качество прогноза национальной инфляции, поскольку учитывают взаимосвязь между региональными и национальными ценами страны, а также учитывают реакцию на «шоки» экономики локального уровня.

В России использование агрегированных и комбинированных прогнозов инфляции рассмотрено в работе [Андреев, 2016]. Автор сравнил модели: RW, LTAR (AR с линейным трендом), модель ненаблюдаемой компоненты (UCM), VAR, BVAR и линейную регрессию (OLS). На помесечных данных в период с 2002 г. по 2015 г. автор показал, что комбинированные прогнозы имеют высокое качество, в сравнении с отдельно взятыми моделями почти при любом горизонте планирования, а в случаях, когда комбинирование прогнозов уступает по качеству, значения ошибок прогноза отличаются незначительно (менее 1,5% значения U-статистики). Также в работе [Аверина и др., 2018] на российских данных, используя модифицированную методологию [Mehrotra et al., 2006], показано, что инфляционные процессы в группе состоятельных регионов связаны прежде всего с региональной экономической спецификой и/или расположением регионов, и для таких регионов не получается описать инфляцию известными моделями. Однако для остальных 72 регионов авторы построили кривые Филлипса на данных за период с 2000 г. по 2015 г., показав достаточно высокое качество прогнозов.

Таким образом, авторы сходятся во мнении, что сложные модели прогнозирования инфляции, включающие различные индикаторы экономической активности, имеют ошибку прогноза порой больше, чем модели, основанные только на временном ряде инфляции. Комбинирование и агрегирование прогнозов, в целом, способствует повышению качества прогноза общей инфляции, однако необходимо валидировать модель на конкретных данных и правильно подобрать веса как при агрегировании, так и при комбинировании моделей.

Модели машинного обучения для прогнозирования инфляции, в отличие от эконометрических моделей, стали применяться с недавнего времени [Jung, 2018]. Перед исследователями стоит задача сравнить качество прогнозов, получаемых методами машинно-

го обучения, с прогнозами, построенными эконометрическими моделями. Одной из первых работ, демонстрирующих превосходство моделей машинного обучения над классическими эконометрическими моделями, является исследование [Chakraborty, Joseph, 2017]. На поквартальных данных в Великобритании с 1988 г. по 2015 г., используя такие методы машинного обучения как *k*-ближайших соседей (*k*NN), метод опорных векторов (SVM), метод дерева решений (DT), метод случайного леса (RF), гребневую регрессию (Ridge regression), модель нейронных сетей (NN), авторы показали, что все рассматриваемые методы превосходят по качеству прогнозирования базовые эконометрические модели AR и VAR [Chakraborty, Joseph, 2017]. На российских данных методы машинного обучения впервые рассматриваются в статье [Байбуза, 2018]. Автор рассматривает модель случайного леса (RF), модель градиентного бустинга (GB), модели с регуляризацией (LASSO и Elastic Net), модели RW и AR. В качестве входных данных автор использует ИПЦ, а также 92 временных ряда, отражающих состояние деловой активности, промышленного производства, денежного рынка. Результат работы – методы машинного обучения показывают достаточно точные прогнозы инфляции в России: на горизонте прогнозирования в 1 месяц модель случайного леса (RF) и модель градиентного бустинга (GB) показывают сопоставимые результаты с эконометрическими моделями, однако на более длительных горизонтах планирования ошибка прогноза в методах машинного обучения существенно ниже, чем при прогнозировании эконометрическими моделями [Байбуза, 2018]. В статье [Павлов, 2020] автор также сравнивает методы машинного обучения с эконометрическими моделями, строя прогнозы моделями AR, методом опорных векторов (SVM), нейронной сетью с одним слоем и гребневой регрессией. Помимо ИПЦ в выборку включались следующие макроэкономические факторы: ВВП, производительность труда, денежная масса, объем выданных кредитов и другие. Ключевой результат работы – метод опорных векторов (SVM) выдает самый качественный прогноз на 1 месяц вперед, однако на более длинном временном горизонте точность SVM схожа с качеством прогноза нейронной сетью [Павлов, 2020].

Таким образом, многие авторы подтверждают состоятельность методов машинного обучения в целях прогнозирования инфляции (см., например: [Ozgun, 2021]). Исследователи отмечают, что такие методы как модель случайного леса (RF), метод опорных векторов (SVM) и модель нейронных сетей (NN) позволяют получить достаточно точные прогнозы [Barkan et. al., 2023]. Однако использование методов машинного обучения для прогнозирования не национальной, а региональной инфляции на данный момент изучено лишь частично. Одна из немногих работ на эту тему – статья [Семитуркин, Шевелев, 2022], в которой авторы сравнивали прогнозы региональной инфляции, полученные моделями машинного обучения, для сибирских регионов России. Используя ежемесячные данные об ИПЦ и курсе доллара США, а также об их лагированных значениях, средних значениях и стандартных отклонениях за период с 2002 по 2020 гг., авторы построили следующие модели: ARIMAX, модель градиентного бустинга (GB), модель случайного леса (RF), Ridge и Lasso регрессии, Elastic Net. В качестве модели-бенчмарка авторы использовали модель ARIMA и проверили качество прогнозов двумя критериями, посчитав корень из среднеквадратичной ошибки прогноза (RMSE), и провели тест Диболда – Мариано. Основной вывод статьи – методы машинного обучения позволяют, как минимум, строить прогнозы региональной инфляции, сопоставимые с прогнозами по эконометрическим моделям [Meireiros, 2021], и комбинирование прогнозов помогает снизить ошибку прогноза.

Рассмотрев основные подходы к прогнозированию инфляции, на основе литературы мы выдвигаем две ключевые гипотезы: (1) модели машинного обучения показывают превосходство над моделями временных рядов при построении прогнозов региональной инфляции; (2) модели временных рядов позволяют точнее прогнозировать инфляцию на горизонте в 1 месяц, а на более длительных горизонтах планирования ошибка прогноза этих моделей больше, чем у прогнозов, построенных методами машинного обучения.

Данные и методология

В статье используются помесечные данные об ИПЦ и макроэкономических показателях за период с января 2010 г. по декабрь 2022 г. Как и в работах [Байбуза, 2018; Семитуркин, Шевелев, 2022], в исследовании анализируются такие макроэкономические показатели, как уровень безработицы, курс доллара США, процентная ставка. Основными процедурами, традиционно используемыми для корректировки сезонности в макроэкономических показателях, являются методы семейства X-11 (X-12, X-12ARIMA и др.), а также TRAMO/SEATS [Пильник и др., 2015]. Несмотря на то, что данные процедуры имеют определенные нюансы: ложные сигналы в данных, смещение в тестах на единичные корни, неустойчивость к выбросам и к структурным сдвигам, они остаются основными при решении задачи удаления сезонности из данных [Банк России, 2020]. В нашей статье для учета возможной сезонности моделируемых показателей эконометрическими моделями использовались фиктивные (дамми) переменные на номер месяца года (см., например: [Пильник и др., 2015], в части использования фиктивных переменных для решения проблемы сезонности в моделях общего экономического равновесия). В случае с моделями машинного обучения использовался порядковый номер месяца года, что в дальнейшем нашло отражение при выявлении наиболее значимых факторов в моделях машинного обучения (номер месяца как значимый фактор, к примеру, в табл. 3, встречается достаточно часто). Данные получены из Единой межведомственной информационно-статистической системы (ЕМИСС). Для учета различных временных промежутков, в которые выходят макроэкономические индикаторы, применяется метод прогнозирования в режиме псевдорекурсивного времени, что позволяет анализировать все показатели, как будто они доступны в единый момент времени несмотря на то, что данные по ИПЦ выходят в начале месяца, а другие факторы – в конце. Для анализа уровня безработицы, который доступен только в квартальном виде, была выполнена интерполяция с сезонной корректировкой при помощи библиотеки `imputena` для Python.

Для корректной работы моделей временных рядов необходимо обеспечить стационарность данных. Для этого в работе применяется натуральный логарифм к данным по инфляции, а также проведен расширенный тест Дики – Фуллера, чтобы определить необходимый порядок разности для каждого временного ряда. Прогнозирование моделями временных рядов осуществлялось в R Studio. Модели машинного обучения реализованы в Python. Для оценки качества прогнозов каждой модели проведен расчет корня из среднеквадратической ошибки прогноза (RMSE), который показывает разницу между реальными и спрогнозированными значениями инфляции. Прогнозы были выполнены на тестовых выборках, т.е. на периодах, которые находятся за пределами выборки, на которой обучались модели (out-of-sample). Это необходимо для того, чтобы, во-первых, избежать переобучения моделей машинного обучения и, во-вторых, получить более точную оценку

качества прогнозов. Были выбраны горизонты прогнозирования в 1, 3, 6, 9, 12, 21 и 24 месяца, которые следуют за последним доступным месяцем в обучающей выборке.

Бенчмарк (базовая модель) в работе – модель авторегрессии первого порядка AR(1). В целях сравнения моделей машинного обучения и эконометрических моделей по каждой модели машинного обучения мы рассматриваем отношение $RMSE_{model}$ к $RMSE_{AR}$. Если значение показателя меньше единицы, то прогноз по сравниваемой модели лучше бенчмарка, если больше единицы – хуже. В исследовании используется прямая модель авторегрессии, в которой прогноз строится сразу на момент времени $(t + k)$. Количество лагов в модели (p) подбирается с помощью Байесовского информационного критерия (BIC). Следуя методологии в исследовании [Семитуркин, Шевелев, 2022], помимо AR(1) в работе мы рассматриваем модель ARIMAX – модификацию модели авторегрессии-скользящей средней с включением дополнительных факторов. Как и в работе Семитуркина и Шевелева, в качестве внешнего фактора мы используем курс доллара США. Количество лагов в модели (p, q) выбрано на основании значения критерия BIC. Поскольку прогнозирование осуществляется в режиме псевдореального времени, для прогноза подставляется фактическое значение курса доллара. Такой подход имеет определенные ограничения, наиболее важное из которых заключается в том, что при переходе к прогнозированию в режиме реального времени значения валютного курса на несколько периодов вперед не представляется спрогнозировать с достаточной точностью. Тем не менее модельная оценка валютного курса может быть еще более смещенной, что побудило авторов статьи остановиться на использовании реальных значений курса и прогнозировании в режиме псевдореального времени.

В работе используются методы машинного обучения, такие как метод опорных векторов, градиентный бустинг и случайный лес, которые относятся к категории обучения с учителем (supervised learning). Это означает, что в наборе данных присутствуют как входные данные (макроэкономические показатели), так и целевое выходное значение (ИПЦ). Модели обучаются на тренировочных данных (training set), а затем настраиваются для достижения наилучшей точности на тестовой выборке. Для обеспечения точной работы методов машинного обучения необходимо выбрать оптимальные гиперпараметры в моделях. Гиперпараметры отличаются от других параметров моделей тем, что они устанавливаются до начала обучения и определяют спецификацию модели (например, число деревьев и глубину обучения). Для выбора оптимальных гиперпараметров в работе используется кросс-валидация. При анализе временных рядов отсутствует возможность использовать стандартное разделение на несколько фолдов (k-fold), при котором исходные данные разбиваются на k -подвыборки (где k обычно от 5 до 10), а затем в течение k -итераций модель обучается на одной из подвыборок, а остальные ($k - 1$) используются для тестирования. При кросс-валидации временных рядов тренировочные данные должны идти строго перед тестовыми, они не могут пересекаться как при работе со стандартными данными (рис. 1). Авторы выбирали между двумя подходами: использовать тестовые выборки с одинаковым количеством наблюдений или использовать кросс-валидацию на скользящем окне.

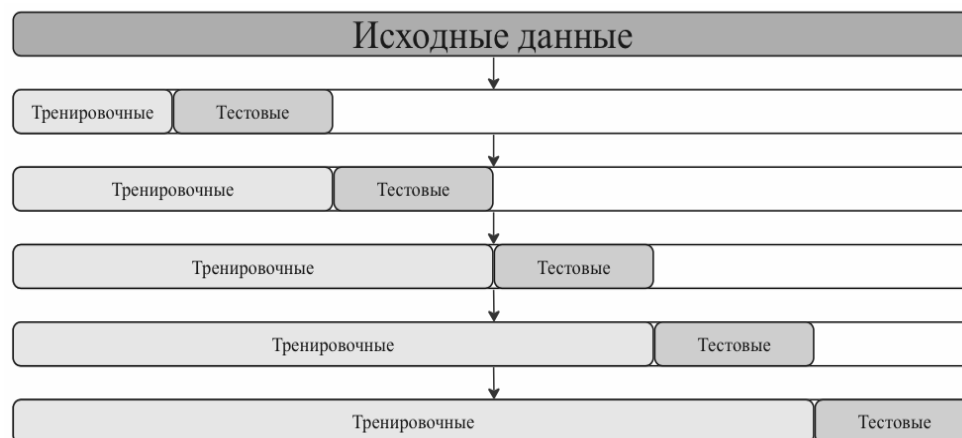


Рис. 1. Принцип кросс-валидации временных рядов

Источник: составлено авторами.

Второй подход заключается в том, что вначале модели обучаются на небольшом отрезке временного ряда, от его начала до периода t , затем делается прогноз на $(t + n)$ шагов вперед и считаются ошибки. Далее обучающая выборка расширяется до $(t + n)$ значения и прогноз строится с $(t + n)$ до $(t + 2 \times n)$, и так продолжается до последнего доступного наблюдения. Количество фолдов в этом случае считается как количество n , уместяющихся в промежуток между началом и концом ряда. Учитывая динамику ИПЦ (и его специфику в регионах ПФО), авторы использовали первый способ кросс-валидации, а именно тестовые выборки с одинаковым числом наблюдений (сбалансированные тестовые выборки). Авторами дополнительно проверено число прогнозов на каждый из временных горизонтов и выявлено, что количество прогнозов совпадает для всех исследуемых временных периодов.

Далее результаты исследования представлены в виде визуализированных прогнозов региональной инфляции для каждого из 14 регионов ПФО с помощью моделей машинного обучения и эконометрических моделей. В моделях случайного леса и градиентного бустинга выделены наиболее значимые факторы, определяющие уровни региональной инфляции. Принимая во внимание значительное число исследуемых регионов, в целях подробной интерпретации результатов, выбран один регион – Пермский край. Остальные регионы ПФО описываются в той части результатов, которая отличается от наблюдений, полученных по Пермскому краю.

Результаты исследования

На рис. 2 представлена помесечная инфляция в Пермском крае с января 2010 г. по декабрь 2022 г. Заметим, что несколько месяцев характеризуются резким ростом инфляции. Данный паттерн характерен и для всех остальных 13 исследуемых регионов ПФО. Инфляция в Пермском крае – нестационарный процесс, что подтвердил расширенный тест Дики – Фуллера.

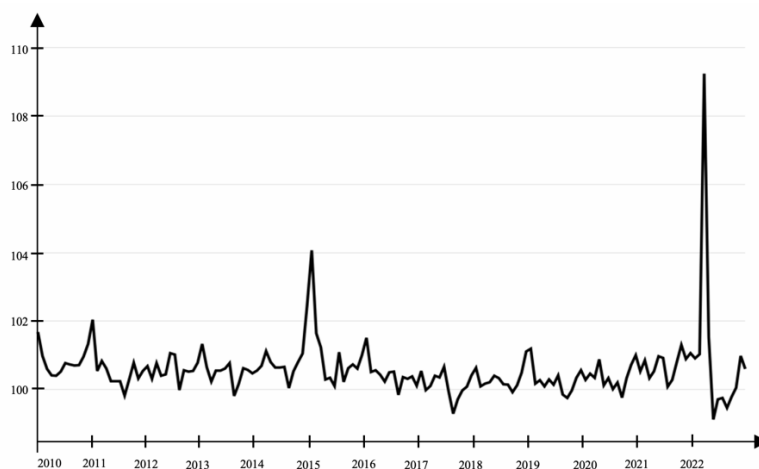


Рис. 2. Помесячная инфляция в Пермском крае за период 2010–2022 гг.

В расчетах инфляция приведена к стационарному процессу через натуральный логарифм и первую разность в логарифмах для всех регионов в выборке. Анализ рис. 2 показал, что в марте 2022 г. наблюдается всплеск инфляции, который остается заметным даже после приведения ряда к стационарности, хотя в остальные периоды месячная инфляция изменяется близко к нулю. Модель AR(1) строит прогноз, схожий с наивным прогнозом, при котором значения инфляции в прогнозных периодах соответствуют значению последнего известного периода, за исключением первых трех месяцев. Модель ARIMAX с включением курса доллара, как внешнего фактора, стремится предсказать значения инфляции менее прямолинейно, повторяя форму реальной инфляции. Аналогичные результаты получены для всех регионов. Инфляция в каждом из регионов является интегрированным процессом первого порядка, и спецификация модели ARIMAX отличается от региона к региону (табл. 1). Тем не менее можно выделить наиболее часто встречающуюся модификацию модели – ARIMAX (1, 1, 1).

Таблица 1.

Модификации модели ARIMAX по регионам ПФО

Регион	Спецификация ARIMAX(p,d, q)
Республика Башкортостан	(0,1,3)
Республика Марий Эл	(0,1,3)
Республика Мордовия	(5,1,0)
Республика Татарстан	(1,1,1)
Удмуртская Республика	(0,1,2)
Чувашская Республика	(1,1,1)
Пермский край	(5,1,0)

Окончание табл. 1.

Регион	Спецификация ARIMAX(p,d,q)
Кировская область	(1,1,1)
Нижегородская область	(1,1,1)
Оренбургская область	(0,1,2)
Пензенская область	(1,1,1)
Самарская область	(0,1,5)
Саратовская область	(1,1,1)
Ульяновская область	(3,1,2)

После предварительного выбора эконометрических моделей для каждого региона был построен прогноз по моделям AR(1) и ARIMAX соответствующей спецификации. В табл. 2 рассчитаны относительные показатели $RMSE_{ARIMAX}/RMSE_{AR}$ по каждому из регионов, а также среднее значение по всем регионам.

Таблица 2.

Отношение $RMSE_{ARIMAX}/RMSE_{AR}$ по регионам ПФО

Регион	Горизонт прогноза (месяц)						
	1	3	6	9	12	21	24
Республика Башкортостан	0,036	0,891	0,868	0,994	0,991	0,903	0,903
Республика Марий Эл	0,334	0,839	0,874	0,985	1,008	0,869	0,869
Республика Мордовия	1,276	4,906	3,042	1,203	1,120	0,824	0,824
Республика Татарстан	0,222	1,003	1,038	1,010	1,010	0,948	0,948
Удмуртская Республика	0,038	0,775	0,778	0,946	1,042	0,851	0,852
Чувашская Республика	1,623	1,140	1,107	1,069	1,066	0,882	0,882
Пермский край	0,366	0,632	0,810	0,981	1,014	0,848	0,850
Кировская область	0,783	0,965	0,967	0,988	1,014	1,014	0,873
Нижегородская область	1,487	1,261	1,120	1,102	1,065	0,838	0,838
Оренбургская область	0,171	0,884	0,964	1,006	1,010	0,962	0,962
Пензенская область	0,434	0,811	0,991	1,006	1,030	0,882	0,882
Самарская область	0,016	0,778	0,788	0,982	1,013	0,925	0,925
Саратовская область	0,574	0,848	0,894	0,984	1,001	0,923	0,923
Ульяновская область	0,471	0,934	1,111	1,069	1,073	0,910	0,909
Среднее значение	0,559	1,190	1,097	1,023	1,033	0,899	0,889

Анализ табл. 2 позволяет сделать следующие выводы: модель ARIMAX в среднем превосходит модель-бенчмарк AR(1) как на достаточно коротких горизонтах планирования (1 месяц), так и на длинных временных промежутках в 21 и 24 месяца. На длинных временных горизонтах такой результат объясняется тем, что значения предыдущей инфляции в периодах $t - 24$ и $t - 21$ перестает оказывать влияние на текущую инфляцию и возрастает роль курса доллара, непосредственно влияющего на импорт и экспорт в конкретном регионе. Объясняя более точный прогноз моделью ARIMAX на 1 месяц, заметим, что курс доллара достаточно быстро реагирует на изменения в экономике, а значит, основываясь на его динамике, можно строить более точный прогноз краткосрочной региональной инфляции.

На рис. 3 представлен прогноз инфляции для Пермского края, рассчитанный на 24 месяца из декабря 2020 г. несколькими моделями машинного обучения (далее в статье такие визуализированные прогнозы представлены для всех регионов ПФО). Модели случайного леса и градиентного бустинга, аналогично модели ARIMAX, способны прогнозировать значительные изменения инфляции, однако не могут в точности определить реальный масштаб изменений показателя⁵.

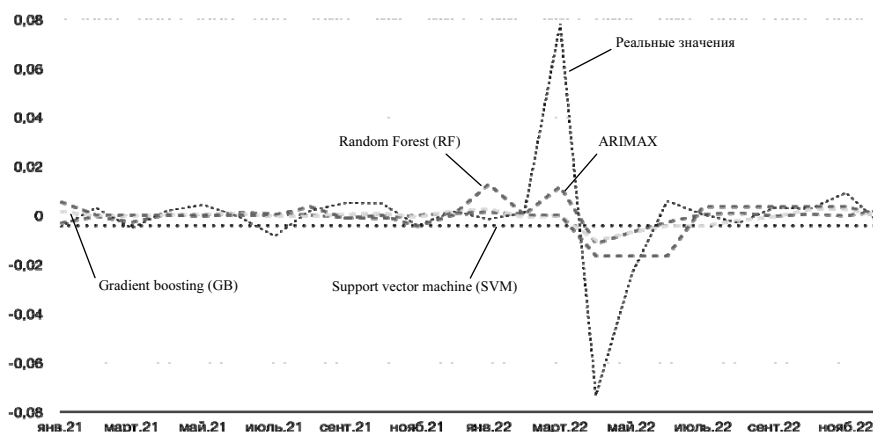


Рис. 3. Инфляция и ее прогноз для Пермского края

Прогнозы, полученные моделью случайного леса для Пермского края, не имеют длительных периодов с линейным прогнозом. Вместо этого прогнозные значения инфляции в районе реального всплеска инфляции достаточно сильно изменяются от месяца к месяцу. Подобная динамика прогнозов характерна и для других регионов, за исключением Удмуртской Республики, Республики Татарстан и Самарской области (см. рис. 4–6).

Для Удмуртской Республики прогнозы моделями случайного леса и градиентного бустинга линейны на периодах прогноза с 8 по 16 месяц, а также с 18 месяца по 24 месяц (рис. 4).

⁵ Отметим, что ни для одного из 14 регионов ПФО не удалось ни одной моделью спрогнозировать пиковые значения инфляции за рассматриваемый период.

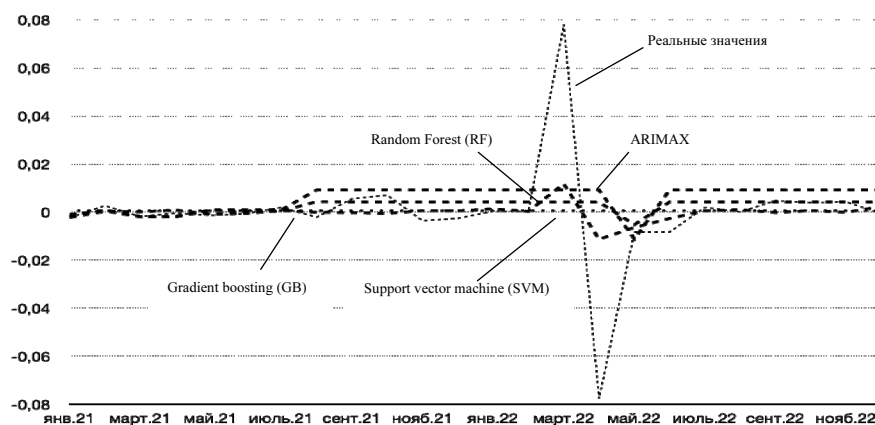


Рис. 4. Инфляция и ее прогноз для Удмуртской Республики

Для Республики Татарстан (рис. 5) линейные прогнозы моделью случайного леса получились с 4 по 6 месяц, с 8 по 10 месяц, с 11 по 15 месяц, а также с 18 по 21 месяц. Прогнозы моделью градиентного бустинга оказались линейны только на промежутке с 4 по 6 месяц. Такая динамика в этих двух регионах могла получиться в виду того, что используемый набор макроэкономических показателей не способен описать природу инфляции в данных регионах.

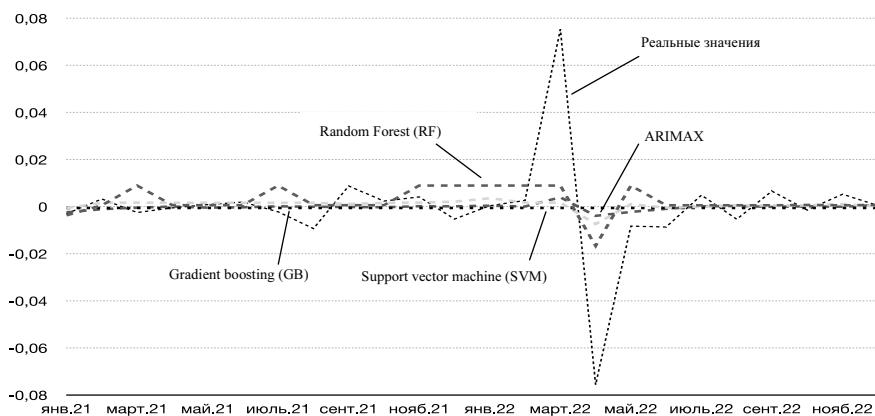


Рис. 5. Инфляция и ее прогноз для Республики Татарстан

Для Самарской области (рис. 6) линейные части прогнозов получились моделью случайного леса на периодах прогноза с 1 месяца по 3 месяц, в 4 и 5 месяцы, с 12 месяца по 15 месяц, а также с 19 месяца по 21 месяц и с 22 месяца по 24 месяц⁶.

⁶ Среднеквадратичная ошибка прогноза значительно выше у модели случайного леса (более в 1,5 раза), в сравнении с моделью AR(1) в следующих регионах: Республика Мордовия, Чувашская

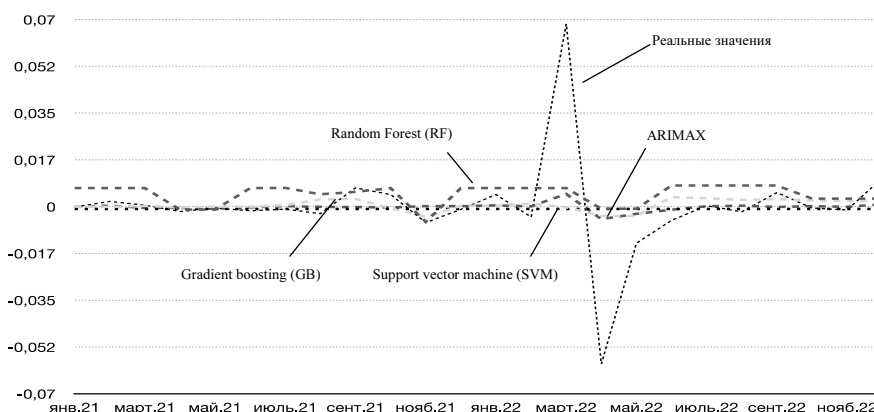


Рис. 6. Инфляция и ее прогноз для Самарской области

Далее по каждому региону, как и в случае со сравнением эконометрических моделей, рассчитаны относительные показатели $RMSE_{RF}/RMSE_{AR}$, а также среднее значение по всем регионам для модели случайного леса (см. табл. П1 в Приложении). Модель случайного леса позволяет получить гораздо более точные прогнозы для будущих периодов (21 и 24 месяца) для всех регионов без исключения. Для половины регионов (Республика Башкортостан, Республика Марий Эл, Республика Татарстан, Удмуртская Республика, Нижегородская область, Пензенская область и Ульяновская область) прогноз на 1 месяц по модели случайного леса также превосходит бенчмарк, однако для другой половины регионов прогноз хуже, чем модель AR(1). Случайный лес предоставляет достаточно точные региональные прогнозы инфляции относительно AR(1): от трех периодов (Республика Мордовия, Оренбургская область, Пензенская область и Самарская область) до шести периодов (Республика Башкортостан и Ульяновская область).

В отличие от модели случайного леса, модель градиентного бустинга демонстрирует более точные прогнозы как на коротких и средних периодах в 1, 3 и 6 месяцев, так и на длинных периодах в 21 и 24 месяца (см. табл. П2 в Приложении). Для следующих регионов: Республика Марий Эл, Республика Татарстан, Удмуртская Республика, Нижегородская область, Оренбургская область, Самарская область и Ульяновская область показатель $RMSE_{GB}/RMSE_{AR}$ больше единицы, однако для оставшихся регионов прогноз получился хуже. Более того, ошибка прогноза не такая высокая (ошибка менее 1,5 раза), как в ранее рассмотренной модели случайного леса. Тем не менее модель градиентного бустинга для прогнозов на ближайший квартал для десяти регионов ПФО демонстрирует более точные прогнозы относительно модели AR(1), и только для четырех регионов прогноз оказался хуже модели бенчмарка (Республика Татарстан, Пермский край, Кировская область и Нижегородская область).

Третья используемая в работе модель – модель, основанная на методе опорных векторов, выдает линейный прогноз с постоянно повторяющимся значением, однако прогнозы на 21 и 24 месяца оказались лучше, чем прогнозы по AR(1). Заметим, что для прогнозов

Республика, Пермский край, Кировская область, Оренбургская область, Самарская область, Саратовская область.

на 24 месяца для всех регионов ПФО прогноз методом опорных векторов оказался лучше эконометрической модели, на горизонте в 21 месяц для 13 регионов ПФО прогноз оказался лучше и только для одного региона (Саратовская область) прогноз оказался чуть хуже бенчмарка (см. табл. ПЗ в Приложении). Данный результат можно объяснить тем, что во второй половине 2022 г. помесечная инфляция замедлилась в российских регионах.

Два из трех используемых метода, а именно модель градиентного бустинга и модель случайного леса, позволили помимо прогноза получить важность (частоту) факторов при моделировании. В тройку наиболее важных факторов для каждого региона ПФО входит значение инфляции в прошлом месяце (табл. 3).

Таблица 3.

Частота факторов в моделях машинного обучения

Случайный лес		Градиентный бустинг	
фактор	частота	фактор	частота
Инфляция $t - 1$	14	Инфляция $t - 1$	14
Среднее значение инфляции $t - 3$	4	Месяц	4
Месяц	4	Среднее значение инфляции $t - 3$	3
Стандартное отклонение инфляции $t - 3$	2	Инфляция $t - 2$	2
Стандартное отклонение ключевой ставки $t - 9$	2	Ключевая ставка $t - 12$	2

Из данных о важности (частоте) признаков для регионов ПФО можно отметить следующее: в модели случайного леса сезонность инфляции характерна только для четырех регионов (Пермский край, Нижегородская область, Пензенская область и Саратовская область), а в модели градиентного бустинга номер месяца важен в Республике Татарстан; для остальных трех регионов номер месяца также является важным фактором.

Среднее значение инфляции за три прошедших месяца – достаточно значимый фактор в модели градиентного бустинга, что подтверждается на данных по Республике Мордовия, Нижегородской области и Ульяновской области. В модели случайного леса помимо представленных регионов среднее значение инфляции за три предыдущих месяца также является значимым фактором при прогнозировании инфляции в Чувашской Республике.

Значимость признаков для прогнозирования инфляции разная почти для каждого региона в выборке. Если не учитывать общий фактор для всех регионов – инфляцию в прошлый месяц, остальные факторы встречаются не так часто среди всех регионов. К примеру, в модели случайного леса четыре раза встречаются средние значения инфляции за прошедшие три месяца и номер месяца, остальные факторы встречаются гораздо реже и зачастую характерны только для одного региона. Аналогично и в модели градиентного бустинга, наиболее часто встречающийся фактор, следующий по частоте после инфляции в прошлый месяц – это номер месяца (встречается 4 раза) и среднее значение инфляции за прошедшие три месяца (встречается 3 раза), остальные факторы также специфичны для каждого региона. Высокая дифференциация наиболее важных факторов для каждого из регионов связана со спецификой региональной экономики, ее размером и общим уровнем развития региона.

Рассмотрим усредненные результаты оценки качества моделей для регионов Приволжского федерального округа, приведенные в табл. 4.

Таблица 4.

**Усредненные результаты оценки качества моделей
по регионам ПФО**

	Средние значения $RMSE_{Model}/RMSE_{AR}$						
	период прогноза (месяц)						
	1	3	6	9	12	21	24
ARIMAX	0,559	1,190	1,097	1,023	1,033	0,899	0,889
RF	2,207	1,556	1,587	1,023	1,464	0,226	0,220
GB	0,999	0,797	0,838	1,069	1,035	0,275	0,223
SVM	2,940	1,680	1,656	2,060	1,046	0,389	0,271

Среди всех рассмотренных моделей модель градиентного бустинга является наиболее качественной, так как предоставляет более точные прогнозы, чем остальные модели на большем количестве периодов прогнозирования. Так, для прогнозов на 3, 6, 21 и 24 месяца модель превосходит бенчмарк, на горизонте в 1 месяц показывает качество сравнимое с бенчмарком. На горизонтах прогнозирования в 9 и 12 месяцев качество прогноза хуже на 6,9% и 3,5% соответственно. Модель случайного леса и метод опорных векторов предоставляют точные прогнозы только на длинных горизонтах в 21 и 24 месяца. Таким образом, все три модели машинного обучения способны предоставлять качественные прогнозы на более долгих горизонтах прогнозирования относительно моделей временных рядов, как AR, так и ARIMAX, что частично подтверждает основную гипотезу о том, что модели машинного обучения превосходят эконометрические модели при прогнозировании региональной инфляции.

Модель случайного леса показывает наиболее точные прогнозы (наименьшее относительное $RMSE_{Model}/RMSE_{AR}$) инфляции на горизонтах в 21 и 24 месяца относительно всех остальных моделей и наихудшее качество в прогнозах на 12 месяцев. Метод опорных векторов имеет наименьшую точность в прогнозировании инфляции на периодах в 21 и 24 месяца среди рассмотренных моделей машинного обучения, также для прогнозов на 1, 3, 6 и 9 месяцев обладает худшим качеством прогноза среди всех моделей. Прогноз моделью ARIMAX на 1 месяц вперед оказался значительно лучше прогноза бенчмарка AR и лучше, чем прогнозы всех моделей машинного обучения. К модели AR(1) по качеству прогноза близка только модель градиентного бустинга, а другие модели машинного обучения прогнозируют хуже, и можно утверждать, что для прогнозов на один месяц модели временных рядов показывают наилучшее качество. Таким образом, гипотеза о том, что модели временных рядов позволяют точнее прогнозировать инфляцию на горизонте в один месяц, подтверждена.

В итоге можно сделать вывод о целесообразности использования методов машинного обучения для прогнозирования региональной инфляции в субъектах ПФО на более долгих временных горизонтах. Модель градиентного бустинга по качеству подходит для любого из рассмотренных горизонтов прогнозирования. Вместе с тем использование эконометрических моделей также оправдано из-за их качества прогнозов на горизонтах про-

гноза до 12 месяцев и в особенности при прогнозе инфляции в следующем месяце. Особенности структуры экономики в регионах, их конституционно-правовой статус, обеспеченность природными ресурсами и географическое расположение объясняют дифференциацию регионов даже внутри одного федерального округа.

Заключение

Для принятия рациональных экономических решений субъектам экономических отношений необходимо учитывать будущую инфляцию в различных регионах. Для прогнозирования значений будущей инфляции используются различные модели, в том числе модели временных рядов и модели машинного обучения. При этом прогнозирование региональной инфляции с помощью моделей машинного обучения является актуальной задачей, научные работы на данную тему только начинают появляться. В статье проведена оценка возможности использования моделей машинного обучения для прогнозирования инфляции в регионах ПФО и проведено сравнение моделей машинного обучения с классическими эконометрическими моделями.

Результаты исследования показали, что модели машинного обучения могут эффективно прогнозировать инфляцию на более длительных временных периодах (21 и 24 месяца), в то время как классические эконометрические модели AR и ARIMAX показывают низкое качество прогнозов на таких дальних горизонтах планирования. По совокупности результатов мы делаем вывод, что при краткосрочном прогнозировании лучшее качество прогнозов демонстрируют эконометрические модели. На горизонтах до 1 года исключение составляют только модели градиентного бустинга (GB), однако при их тестировании в моделях использовались истинные будущие значения экзогенных переменных. При среднесрочном прогнозировании (горизонт 2 года) эконометрические модели явно уступают моделям машинного обучения. Данные результаты соответствуют практике прогнозирования макроэкономических показателей, которая свидетельствует о том, что для среднесрочных прогнозов обычно рекомендуется использовать либо структурные модели (например, DSGE), либо их полуструктурные аналоги (например, QPM). Таким образом, выдвигаемая ранее гипотеза (2) подтвердилась, как и подтвердилась гипотеза (1) о том, что модели машинного обучения могут быть весьма эффективным инструментом для прогнозирования региональной инфляции на разных временных горизонтах. Среди всех использованных моделей наиболее точной оказалась модель градиентного бустинга, которая превосходит модель-бенчмарк AR(1) в 4-х из 7 рассматриваемых периодов, а прогноз на один месяц аналогичен модели AR(1) и немного уступает в качестве только при прогнозах на 9 и 12 месяцев. Авторы приходят к выводу, что для достижения наилучшего качества прогнозов региональной инфляции возможно комбинировать прогнозы на основе моделей машинного обучения и эконометрических моделей.

Следует отметить, что рассматриваемый прогнозный период, 2021–2022 гг., является относительно нетипичным для российской и мировой экономики, инфляция в 2021 г. характеризуется и эффектом низкой денежной базы 2020 г., и повышением мировых цен на продовольственные товары, и дефицитом предложения из-за логистических проблем и проблем с производством на фоне остающихся коронавирусных ограничений, так и в целом ростом инфляции в мире. Экономические санкции, введенные в адрес России в 2022 г., усложнили экономическую ситуацию, и чтобы поддержать экономику и не допустить от-

ток капитала из страны, были приняты новые, ранее не принимаемые меры: остановка торгов на ММВБ, обязательная продажа валютной выручки экспортерами, ограничение переводов валюты за границу физическим лицам и ограничение продажи и снятия валюты. Подобные экономические условия и политические шоки трудно прогнозировать, они не нашли прямого отражения и в исследуемом наборе макроэкономических показателей из-за невозможности прогнозирования таких шоков и их влияния на экономику. Модели машинного обучения, которые использованы в работе, не так хорошо справились со всплеском инфляции в марте 2022 г. и ростом инфляции в 2021 г., тем не менее в более стабильных экономических условиях модели способны предоставлять достаточно качественные прогнозы.

В исследовании присутствуют некоторые ограничения. Во-первых, доступный набор макроэкономических показателей может быть расширен в дальнейших исследованиях. Помесячные данные на региональном уровне представлены более скромно в открытом доступе, чем национальные, поэтому в работе использовались преимущественно доступные данные, и некоторые показатели дополнялись интерполяцией. Во-вторых, качество данных может вызывать вопросы как из-за пересмотра расчета показателей в разные временные промежутки, обновления ранее представленных данных в конце года, так и специфики отдельно взятого региона. В-третьих, ограничение в вычислительной мощности, подбор гиперпараметров в моделях машинного обучения может был выполнен с различными значениями самих гиперпараметров. Однако добавление каждого нового значения существенно увеличивает время для вычисления оптимального набора параметров, поэтому подбор оптимальных значений происходил из достаточно ограниченного набора значений, что могло повлиять на спецификацию модели и качество прогнозов. Наконец, специфика каждого отдельного региона влияет как на данные, так и на необходимые факторы для прогноза. Как показал анализ самых значимых факторов в моделях машинного обучения, только один фактор (ИПЦ в прошлом месяце) входит в тройку наиболее важных для всех регионов в моделях случайного леса и градиентного бустинга. Остальные факторы либо являются общими только для четырех регионов (например, номер месяца), либо являются индивидуальными и важны в моделировании инфляции только у конкретных регионов.

Возможным развитием исследования может послужить использование дополнительных моделей машинного обучения, как, например, нейронных сетей, так и комбинирование прогнозов по нескольким моделям в один прогноз. Важным представляется тестирование устойчивости результатов прогнозов к разным тестовым выборкам и методам кросс-валидации временных рядов (например, методам скользящего или расширяющегося окна). Кроме того, перспективным представляется использование экспертной корректировки прогнозов, что может дополнительно повысить качество прогнозов, скорректировать работу моделей. Важным направлением развития работы может стать исследование рассмотренных моделей на других регионах России при решении задачи прогнозирования региональной инфляции.

Приложение

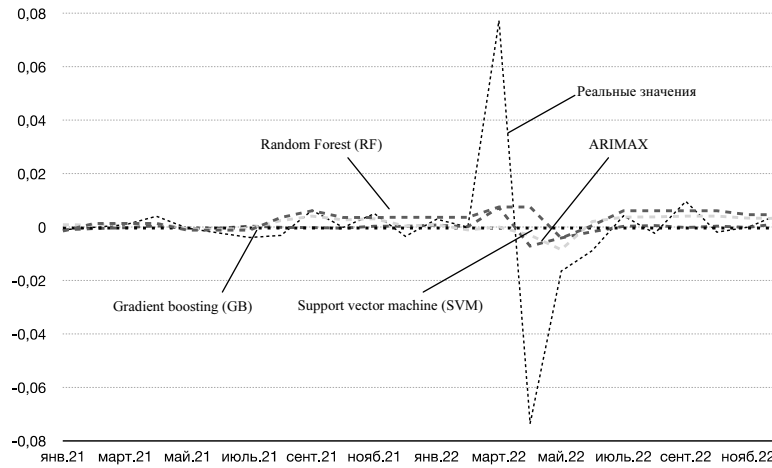


Рис. П1. Реальная инфляция и прогнозы инфляции для Республики Башкортостан

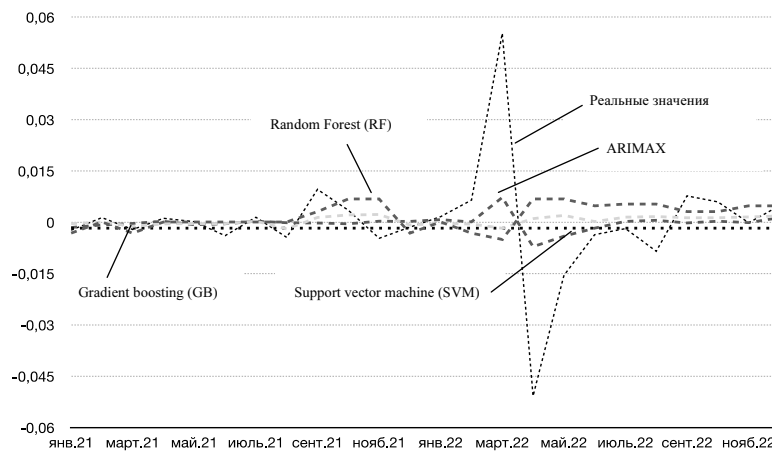


Рис. П2. Реальная инфляция и прогнозы инфляции для Республики Марий Эл

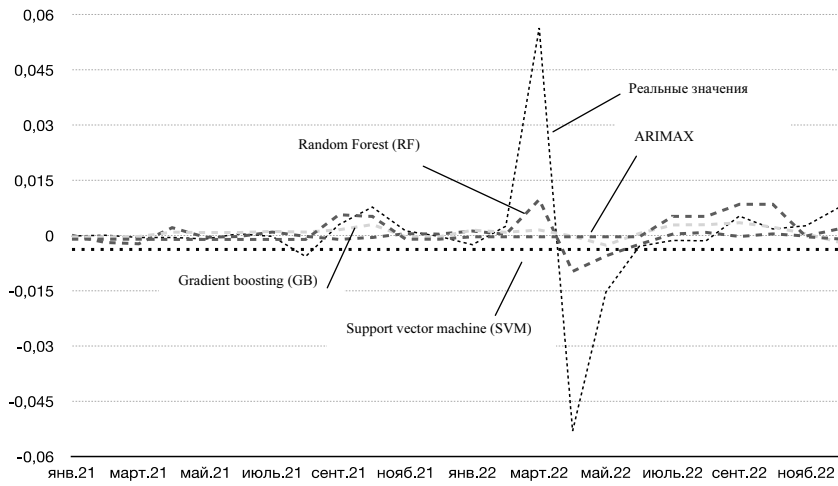


Рис. П3. Реальная инфляция и прогнозы инфляции для Республики Мордовия

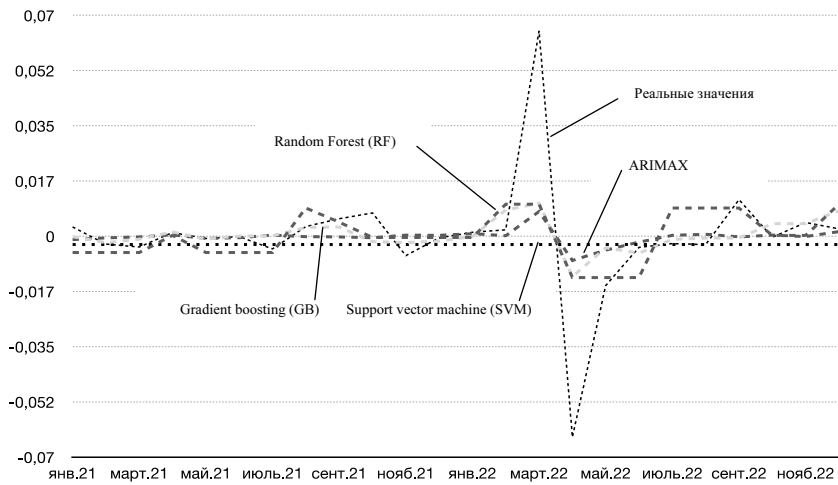


Рис. П4. Реальная инфляция и прогнозы инфляции для Республики Чувашия

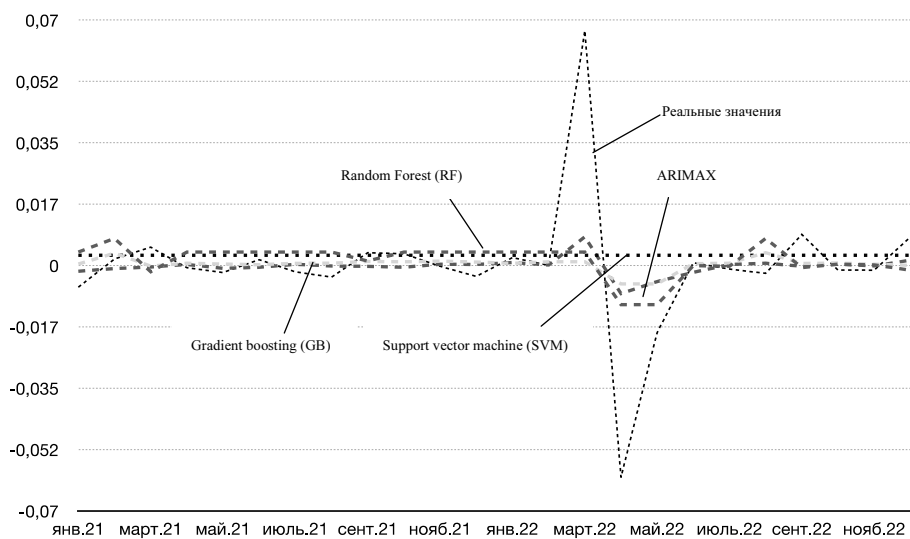


Рис. П5. Реальная инфляция и прогнозы инфляции для Кировской области

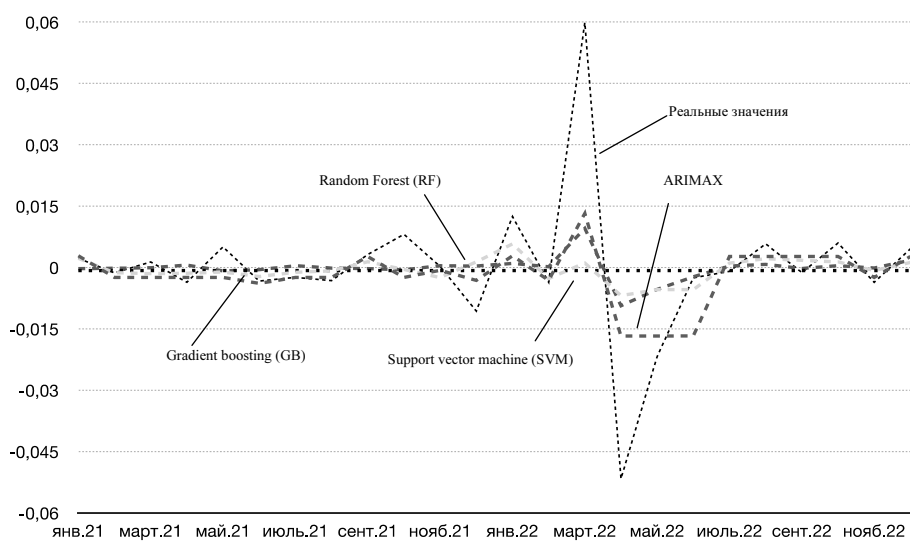


Рис. П6. Реальная инфляция и прогнозы инфляции для Нижегородской области

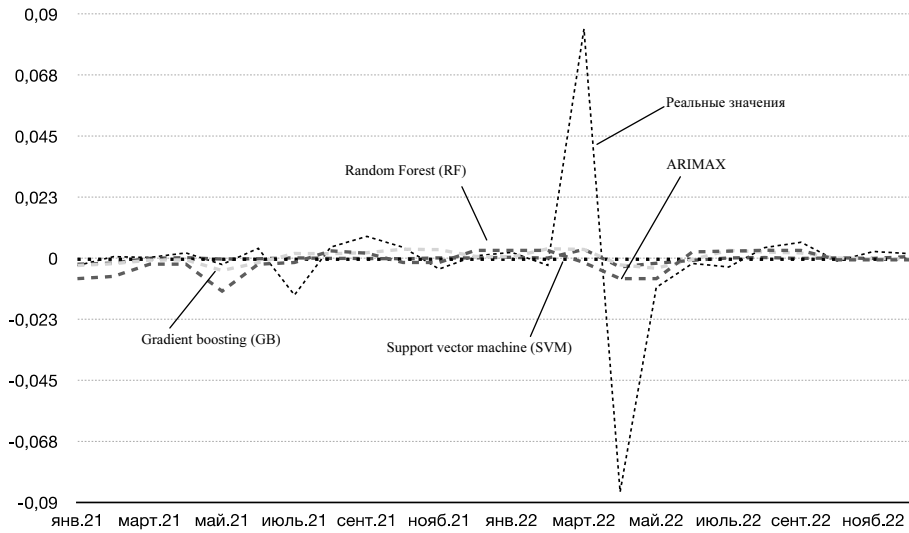


Рис. П7. Реальная инфляция и прогнозы инфляции для Оренбургской области

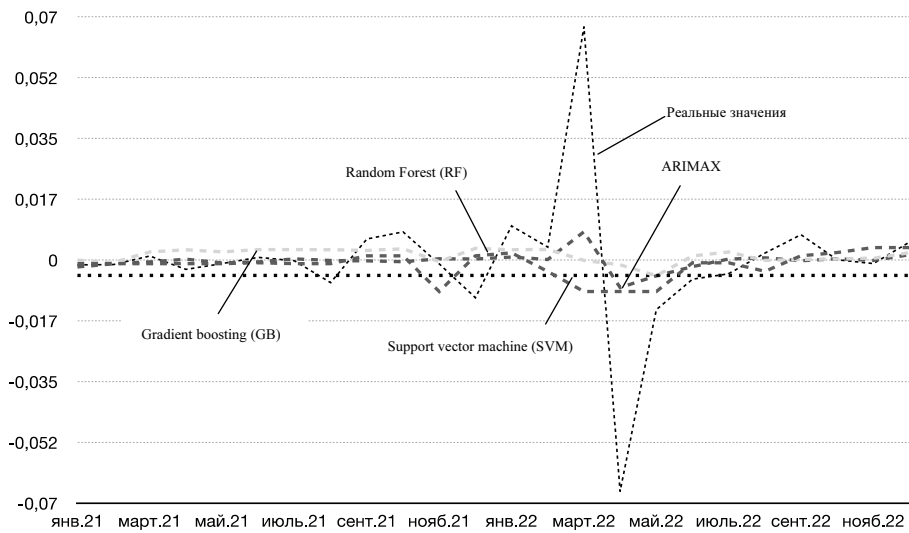


Рис. П8. Реальная инфляция и прогнозы инфляции для Пензенской области

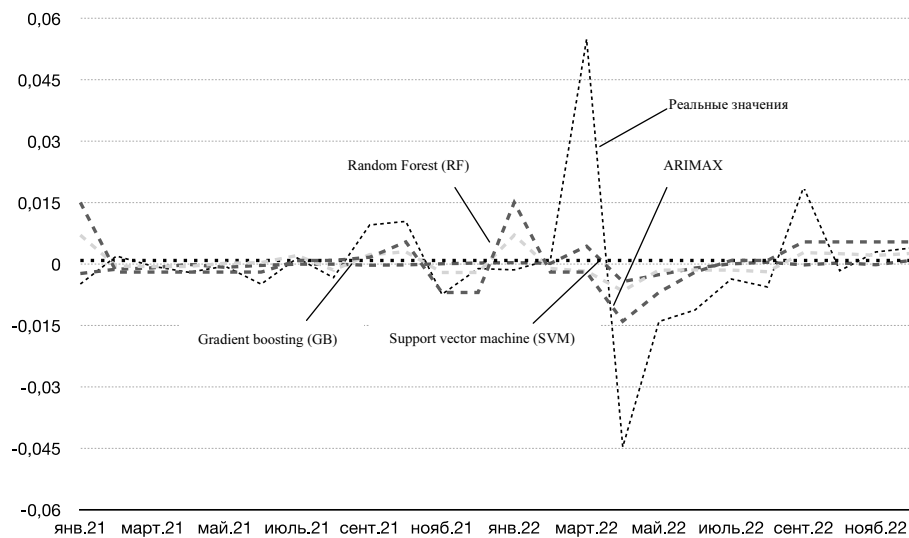


Рис. П9. Реальная инфляция и прогнозы инфляции для Саратовской области

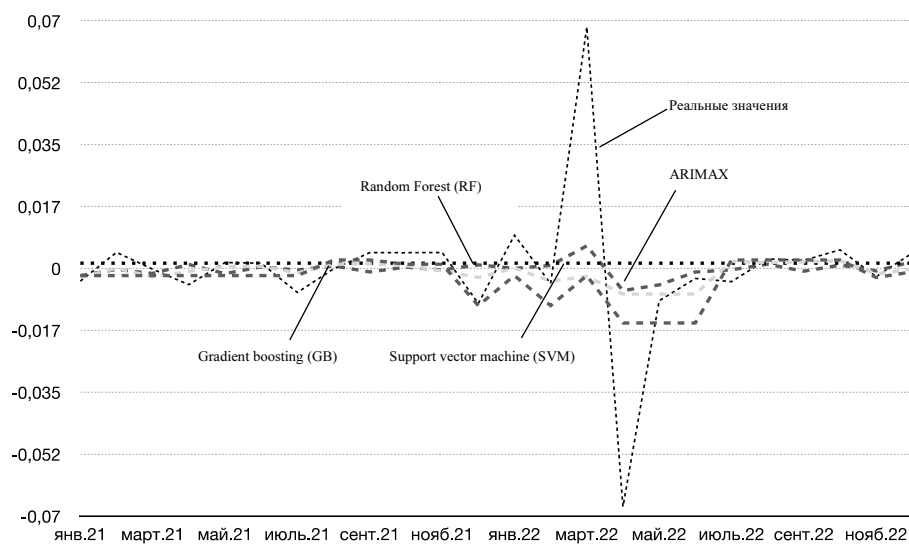


Рис. П10. Реальная инфляция и прогнозы инфляции для Ульяновской области

Таблица П1.

Отношение $RMSE_{RF}/RMSE_{AR}$ по регионам ПФО

Регион	Горизонт прогноза (месяц)						
	1	3	6	9	12	21	24
Республика Башкортостан	0,474	0,616	0,534	0,048	2,211	0,147	0,006
Республика Марий Эл	0,368	0,533	1,839	1,634	0,454	0,267	0,001
Республика Мордовия	6,061	1,527	2,999	1,228	0,260	0,182	0,527
Республика Татарстан	0,265	4,130	0,603	1,736	3,136	0,253	0,012
Удмуртская Республика	0,202	0,016	0,896	1,468	3,647	0,182	0,432
Чувашская Республика	3,284	0,638	2,464	0,353	0,037	0,130	0,404
Пермский край	2,496	1,270	0,186	1,418	0,354	0,013	0,173
Кировская область	1,715	1,506	0,624	0,710	2,184	0,465	0,495
Нижегородская область	0,359	2,660	0,192	0,160	1,612	0,211	0,108
Оренбургская область	1,812	1,566	2,618	1,080	0,384	0,112	0,089
Пензенская область	0,263	1,725	1,016	1,504	2,509	0,283	0,078
Самарская область	8,679	4,515	6,114	0,538	2,455	0,131	0,406
Саратовская область	4,458	0,518	0,994	1,890	1,123	0,785	0,093
Ульяновская область	0,468	0,566	1,137	0,559	0,124	0,005	0,264
Среднее значение	2,207	1,556	1,587	1,023	1,464	0,226	0,220

Таблица П2.

Отношение $RMSE_{GB}/RMSE_{AR}$ по регионам ПФО

Регион	Горизонт прогноза (месяц)						
	1	3	6	9	12	21	24
Республика Башкортостан	1,889	0,526	0,682	0,693	1,196	0,231	0,067
Республика Марий Эл	0,814	0,471	1,616	2,061	0,168	0,371	0,185
Республика Мордовия	1,692	0,865	0,647	0,649	0,041	0,104	0,583
Республика Татарстан	0,914	1,533	0,151	1,633	1,627	0,253	0,016
Удмуртская Республика	0,604	0,073	0,475	0,524	2,092	0,023	0,210
Чувашская Республика	1,239	0,889	0,090	0,866	0,276	0,593	0,299
Пермский край	1,544	1,214	0,339	1,049	0,055	0,153	0,157

Окончание табл. П2.

Регион	Горизонт прогноза (месяц)						
	1	3	6	9	12	21	24
Кировская область	1,134	1,161	0,411	0,749	1,292	0,407	0,421
Нижегородская область	0,007	2,002	0,362	0,583	2,580	0,160	0,192
Оренбургская область	0,004	0,619	2,361	1,048	0,041	0,147	0,075
Пензенская область	1,009	0,914	1,458	1,027	2,952	0,333	0,152
Самарская область	0,147	0,190	1,094	1,496	0,408	0,120	0,456
Саратовская область	2,670	0,034	1,782	1,734	0,182	0,939	0,083
Ульяновская область	0,320	0,672	0,272	0,855	1,578	0,020	0,225
Среднее значение	0,999	0,797	0,838	1,069	1,035	0,275	0,223

Таблица П3.

Отношение $RMSE_{SVM}/RMSE_{AR}$ по регионам ПФО

Регион	Горизонт прогноза (месяц)						
	1	3	6	9	12	21	24
Республика Башкортостан	0,648	1,587	0,932	2,126	0,967	0,418	0,230
Республика Марий Эл	0,256	0,201	1,006	2,848	0,077	0,547	0,405
Республика Мордовия	25,522	10,584	8,259	3,087	1,202	0,520	0,700
Республика Татарстан	0,922	0,698	1,171	1,999	1,037	0,301	0,047
Удмуртская Республика	1,277	1,068	0,706	1,973	0,991	0,172	0,054
Чувашская Республика	2,243	0,286	1,208	2,703	0,548	0,692	0,258
Пермский край	0,133	0,251	1,017	2,097	1,354	0,303	0,015
Кировская область	1,559	0,500	0,361	0,223	1,896	0,293	0,273
Нижегородская область	1,772	1,488	0,838	1,306	2,124	0,025	0,306
Оренбургская область	0,858	0,435	1,817	1,475	0,245	0,235	0,083
Пензенская область	2,234	4,371	3,427	3,270	1,329	0,548	0,479
Самарская область	0,914	1,020	0,469	2,857	0,046	0,305	0,610
Саратовская область	1,286	0,495	1,959	2,080	0,384	1,053	0,193
Ульяновская область	1,537	0,529	0,007	0,803	2,443	0,037	0,142
Среднее значение	2,940	1,680	1,656	2,060	1,046	0,389	0,271

* *
*

СПИСОК ЛИТЕРАТУРЫ

- Аверина Д.С., Горшкова Т.Г., Синельникова-Мурылева Е.В.* Построение кривой Филлипса на региональных данных // Экономический журнал ВШЭ. 2018. Т. 22. № 4. С. 609–630.
- Андреев А.* Прогнозирование инфляции методом комбинирования прогнозов в Банке России // Банк России: серия докладов об экономических исследованиях. 2016. № 14. С. 2–11.
- Байбуза И.* Прогнозирование инфляции с помощью методов машинного обучения // Деньги и кредит. 2018. № 4. С. 42–59.
- Банк России.* Методика сезонной корректировки индекса потребительских цен Банка России, 2020.
- Павлов Е.* Прогнозирование инфляции в России с помощью нейронных сетей // Деньги и кредит. 2020. № 1. С. 57–73.
- Пильник Н.П., Поспелов И.Г., Станкевич И.П.* Об использовании фиктивных переменных для решения проблемы сезонности в моделях общего экономического равновесия // Экономический журнал ВШЭ. 2015. Т. 19. № 2. С. 249–270.
- Семитуркин О., Шевелев А.* Прогнозирование региональной инфляции с помощью методов машинного обучения на примере макрорегиона Сибирь // Банк России: серия докладов об экономических исследованиях. 2020. № 91. С. 2–18.
- Atkeson A., Ohanian L.* Are Phillips Curves Useful for Forecasting Inflation? // Federal Reserve Bank of Minnesota Quarterly Review. 2001. Vol. 25. № 1. P. 2–11.
- Barkan O., Benchimol J., Caspi I., Cohen E., Hammer A., Koenigstein N.* Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks // International Journal of Forecasting. 2023. Vol. 39. № 3. P. 1145–1162.
- Bermingham C., D'Agostino A.* Understanding and Forecasting Aggregate and Disaggregate Price Dynamics // European Central Bank Working Paper. 2011. № 1365.
- Chakraborty C., Joseph A.* Machine Learning at Central Banks // Bank of England Working Papers. 2017. № 674.
- Chodorow-Reich G.* Regional Data in Macroeconomics: Some Advice for Practitioners // Journal of Economic Dynamics and Control. 2020. Vol. 115. № 643386.
- Dickey D.A., Fuller W.A.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root // Journal of the American Statistical Association. 1979. № 74. P. 427–431.
- Faust J., Wright J.H.* Forecasting Inflation // Handbook of Economic Forecasting. 2013. № 2. P. 2–56.
- Hubrich K.* Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy? // International Journal of Forecasting. 2005. Vol. 21. № 1. P. 119–136.
- Jung J., Patnam M., Ter-Martirosyan A.* An Algorithmic Crystal Ball: Forecast-Based on Machine Learning // IMF Working Paper. 2018. № 18/230.
- Medeiros M.C., Vasconcelos G.F., Veiga A., Zilberman E.* Forecasting Inflation in a Data-Rich Environment: the Benefits of Machine Learning Methods // Journal of Business & Economic Statistics. 2021. Vol. 39. № 1. P. 98–119.
- Mehrotra A.N., Peltonen T.A., Rivera A.S.* Modelling Inflation in China – A Regional Perspective // ECB Working Paper. 2007. № 829.
- Ozgun O., Akkoç U.* Inflation Forecasting in an Emerging Economy: Selecting Variables with Machine Learning Algorithms // International Journal of Emerging Markets. 2021. Vol. 17. № 8. P. 1889–1908.
- Phillips A.* The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957 // *Economica*. 1958. № 100. P. 283–299.
- Samuelson P.A., Solow R.M.* Analytical Aspects of Anti-Inflation Policy // American Economic Review Papers and Proceedings. 1960. Vol. 50. № 2. P. 177–194.

Stock J.H., Watson M.W. Forecasting Inflation // Journal of Monetary Economics. 1999. № 44. P. 293–335.

Stock J.H., Watson M.W. Phillips Curve Inflation Forecasts // NBER Working Paper. 2008. № 14322.

Tena J.D., Espasa A., Pino G. Forecasting Spanish Inflation Using the Maximum Disaggregation Level by Sectors and Geographical Areas // International Regional Science Review. 2010. № 33. P. 181–204.

Regional Inflation Forecasting: Econometric Models Versus Machine Learning Methods?

Tatyana Bukina¹, Dmitriy Kashin²

¹ National Research University Higher School of Economics,
37, Boulevard Gagarina str., Perm, 614070, Russian Federation.
E-mail: tbukina@hse.ru

² National Research University Higher School of Economics,
37, Boulevard Gagarina str., Perm, 614070, Russian Federation.
E-mail: dvkashin@hse.ru

The paper reveals the forecasts for regional inflation based on the regions of the Privolzhskiy Federal District (PFD). The purpose of the study is to determine the model that most accurately predicts regional inflation. The paper compares the tools of machine learning – support vector machines, gradient boosting, and random forest – with econometric models of time series, autoregression and integrated autoregression-moving average – models that are often used to predict national inflation in Russia. The results of the study help to identify the key macroeconomic indicators that most significantly affect regional inflation. The top three of them for each region include the value of inflation within last month, the average inflation, and the count number of analyzed month. We show that machine learning models are not as bad at predicting regional inflation over long time periods (21 months and 24 months), while econometric models predict quite accurately on short time horizons.

Key words: forecasting; regional inflation; machine learning; gradient boosting; random forest; econometric modeling.

JEL Classification: E31, E37.

* *
*

References

Andreev A. (2016) Forecasting Inflation Using the Method of Combining Forecasts in the Bank of Russia. *Bank of Russia: Series of Reports on Economic Research*, 14, pp. 2–11. (In Russ.)

- Averina D.S., Gorshkova T.G., Sinelnikova-Muryleva E.V. (2018) Construction of the Phillips Curve Using Regional Data. *HSE Economic Journal*, 22, 4, pp. 609–630. (In Russ.)
- Atkeson A., Ohanian L. (2001) Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minnesota Quarterly Review*, 25, 1, pp. 2–11.
- Bank of Russia (2020) *Methodology for Seasonal Adjustment of the Consumer Price Index of the Bank of Russia*. (In Russ.)
- Barkan O., Benchimol J., Caspi I., Cohen E., Hammer A., Koenigstein N. (2023) Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*, 39, 3, pp. 1145–1162.
- Baybuza I. (2018) Forecasting Inflation Using Machine Learning Methods. *Russian Journal of Money & Finance*, 4, pp. 42–59. (In Russ.)
- Bermingham C., D'Agostino A. (2011) *Understanding and Forecasting Aggregate and Disaggregate Price Dynamics*. European Central Bank Working Paper, 1365.
- Chakraborty C., Joseph A. (2017) *Machine Learning at Central Banks*. Bank of England Working Papers, 674.
- Chodorow-Reich G. (2020) Regional Data in Macroeconomics: Some Advice for Practitioners. *Journal of Economic Dynamics and Control*, 115, 643386.
- Dickey D.A., Fuller W.A. (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, pp. 427–431.
- Faust J., Wright J.H. (2013) Forecasting Inflation. *Handbook of Economic Forecasting*, 2, pp. 2–56.
- Hubrich K. (2005) Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy? *International Journal of Forecasting*, 21, 1, pp. 119–136.
- Jung J., Patnam M., Ter-Martirosyan A. (2018) *An Algorithmic Crystal Ball: Forecast-Based on Machine Learning*. IMF Working Paper, 18/230.
- Medeiros M.C., Vasconcelos G.F., Veiga A., Zilberman E. (2021) Forecasting Inflation in a Data-Rich Environment: the Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics*, 39, 1, pp. 98–119.
- Mehrotra A.N., Peltonen T.A., Rivera A.S. (2007) *Modelling Inflation in China – A Regional Perspective*. ECB Working Paper, 829.
- Ozgun O., Akkoç U. (2021) Inflation Forecasting in an Emerging Economy: Selecting Variables with Machine Learning Algorithms. *International Journal of Emerging Markets*, 17, 8, pp. 1889–1908.
- Pavlov E. (2020) Forecasting Inflation in Russia Using Neural Networks. *Russian Journal of Money & Finance*, 1, pp. 57–73. (In Russ.)
- Phillips A. (1958) The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957. *Economica*, 100, pp. 283–299.
- Pilnik N.P., Pospelov I.G., Stankevich I.P. (2015) On the Use of Dummy Variables to Address Seasonality in General Economic Equilibrium Models. *HSE Economic Journal*, 19, 2, pp. 249–270. (In Russ.)
- Samuelson P.A., Solow R.M. (1960) Analytical Aspects of Anti-Inflation Policy. *American Economic Review Papers and Proceedings*, 50, 2, pp. 177–194.
- Semiturkin O., Shevelev A. (2022) Forecasting Regional Inflation Using Machine Learning Methods Using the Example of the Siberia Macroregion. *Bank of Russia: series of economic research reports*, 91, pp. 2–18. (In Russ.)
- Stock J.H., Watson M.W. (1999) Forecasting Inflation. *Journal of Monetary Economics*, 44, pp. 293–335.
- Stock J.H., Watson M.W. (2008) *Phillips Curve Inflation Forecasts*. NBER Working Paper, 14322.
- 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, pp. 181–204.