

DOI: <https://doi.org/10.15688/re.volsu.2025.4.11>

UDC 336.748.12

LBC 65.24



Submitted: 29.07.2025

Accepted: 10.09.2025

DISAGGREGATED APPROACH TO CONSUMER PRICE INDEX MODELING: REGIONAL ASPECT¹

Yuri A. Metel

Stavropol Territory Branch of the Southern Main Directorate of the Central Bank of the Russian Federation,
Stavropol, Russian Federation

Natalia N. Kunitsyna

North Caucasus Federal University, Stavropol, Russian Federation

Abstract. The paper solves the problem of accuracy improvement of forecasting the consumer price index (CPI), which is a key indicator of inflationary processes. It is used for monetary policy instruments. In contrast to the traditional approach, in which the forecast is made for the aggregated indicator “All goods and services,” which does not allow taking into account the heterogeneity of the dynamics of goods and services categories, the authors implemented disaggregated CPI modeling at the regional level. Forecasting was carried out at three levels: CPI as a whole, by basic components (“Food products,” “Non-food products,” and “Services”) and by 78 product groups. Such an approach makes it possible to identify hidden interrelations and specific factors affecting price dynamics in various segments of the consumer basket, taking into account regional characteristics. Calculations with monthly breakdowns were carried out on the basis of a wide range of methods: traditional econometric models (SARIMA, Prophet, and Ridge regression) and modern machine learning algorithms (Random Forest and CatBoost). The authors applied the principal component method (PCA) and recursive feature elimination (RFE) to improve the accuracy of the predicted data. Accuracy was assessed on the basis of cross-validation. The statistical significance of differences between models was checked using the Diebold-Mariano test. The results revealed that a disaggregated prediction approach provides higher accuracy compared to aggregated models. Particularly noticeable improvements are observed for commodity categories with high price volatility. The findings confirm that detailing the CPI structure in forecasting allows not only to increase the accuracy of estimates but also to obtain a more reliable analytical basis for making economic decisions in monetary policy. *Authors' contribution.* Yu.A. Metel – development of forecasting methodology, model implementation, calculation, and sampling methods; N.N. Kunitsyna – literature review, description of results, drawing conclusions, and editing article text.

Key words: consumer price index, inflation, SARIMA, decisive trees, gradient boosting, regularization, Feature Selection, principal component method, disaggregated forecasting, inflation expectations.

Citation. Metel Yu.A., Kunitsyna N.N., 2025. Disaggregated Approach to Consumer Price Index Modeling: Regional Aspect. *Regionalnaya ekonomika. Yug Rossii* [Regional Economy. South of Russia], vol. 13, no. 4, pp. 118-129. (in Russian). DOI: <https://doi.org/10.15688/re.volsu.2025.4.11>

УДК 336.748.12

ББК 65.24

Дата поступления статьи: 29.07.2025

Дата принятия статьи: 10.09.2025

ДЕЗАГРЕГИРОВАННЫЙ ПОДХОД К МОДЕЛИРОВАНИЮ ИНДЕКСА ПОТРЕБИТЕЛЬСКИХ ЦЕН: РЕГИОНАЛЬНЫЙ АСПЕКТ¹

Юрий Андреевич Метель

Отделение по Ставропольскому краю Южного главного управления Центрального банка Российской Федерации,
г. Ставрополь, Российская Федерация

© Метель Ю.А., Куницына Н.Н., 2025

Наталья Николаевна Куницына

Северо-Кавказский федеральный университет, г. Ставрополь, Российская Федерация

Аннотация. В работе решается задача повышения точности прогнозирования индекса потребительских цен (ИПЦ), который является ключевым показателем инфляционных процессов и используется при формировании денежно-кредитной политики. В отличие от традиционного подхода, в котором прогноз строится для агрегированного показателя «Все товары и услуги», не позволяющего учесть разнородность динамики отдельных категорий товаров и услуг, авторы предложили дезагрегированное моделирование ИПЦ на региональном уровне. Прогнозирование осуществлялось по трем уровням: ИПЦ в целом, по базовым компонентам («Продовольственные товары», «Непродовольственные товары», «Услуги») и по 78 товарным группам. Такой подход позволяет выявлять скрытые взаимосвязи и специфические факторы, влияющие на динамику цен в различных сегментах потребительской корзины с учетом региональных особенностей. Расчеты с помесячной разбивкой осуществлены на основе широкого спектра методов: традиционные эконометрические модели (SARIMA, Prophet, Ridge regression), современные алгоритмы машинного обучения (Random Forest, CatBoost). Для повышения точности предсказываемых данных применен метод главных компонент (PCA) и отбор признаков Recursive Feature Elimination (RFE). Оценка точности осуществлялась на основе кросс-валидации, статистическая значимость различий между моделями проверялась с помощью теста Диболда – Мариано. Результаты показали, что дезагрегированный подход к прогнозированию обеспечивает более высокую точность по сравнению с агрегированными моделями. Особенно заметные улучшения наблюдаются для товарных категорий с высокой волатильностью цен. Полученные выводы подтверждают, что детализация структуры ИПЦ при прогнозировании позволяет не только повысить точность оценок, но и получить более надежную аналитическую базу для принятия экономических решений в области денежно-кредитной политики. *Вклад авторов.* Ю.А. Метель – разработка методологии прогнозирования, построение моделей, проведение расчетов и описание используемых методов. Н.Н. Куницына – обзор литературы, описание результатов, формулировка выводов, редактирование текста статьи.

Ключевые слова: индекс потребительских цен, инфляция, SARIMA, решающие деревья, градиентный бустинг, регуляризация, Feature Selection, метод главных компонент, дезагрегированное прогнозирование, инфляционные ожидания.

Цитирование. Метель Ю. А., Куницына Н. Н., 2025. Дезагрегированный подход к моделированию индекса потребительских цен: региональный аспект // Региональная экономика. Юг России. Т. 13, № 4. С. 118–129. DOI: <https://doi.org/10.15688/re.volsu.2025.4.11>

Введение

Моделирование инфляционных процессов представляет собой важнейший инструмент экономического анализа и прогнозирования, позволяющий оценивать динамику цен, предвидеть макроэкономические риски и корректировать направления денежно-кредитной политики. Инфляция оказывает непосредственное влияние на уровень жизни населения, инвестиционную активность бизнеса и стабильность финансовой системы, что делает ее прогноз критически значимым как для государственных экономических институтов, так и для частного сектора.

Особую сложность представляет региональное прогнозирование важного критерия инфляции – индекса потребительских цен, поскольку ценовая динамика в субъектах Федерации существенно отличается от общероссийской из-за различий в структуре потребления, локальных экономических условий и внешних шоков. Традиционные подходы, основанные на агрегированных показателях, зачастую не учитывают эту неодно-

родность, что снижает точность прогнозов. В связи с этим применение современных методов машинного обучения и многоуровневого моделирования ИПЦ позволяет не только повысить качество предсказаний, но и выявить ключевые факторы инфляции на уровне отдельных товарных групп.

Данное исследование направлено на разработку методики дезагрегированного прогнозирования ИПЦ с использованием алгоритмов SARIMA, Prophet, Ridge regression, Random Forest и CatBoost, а также оценку влияния различных уровней детализации данных на точность моделей.

Обзор литературы

Прогнозирование инфляции остается одной из ключевых задач экономической науки и практики, особенно в условиях нестабильности мировых рынков и сохранения региональных дисбалансов. Точные прогнозы позволяют центральным банкам корректировать денежно-кредитную

политику, бизнесу – планировать стратегии, правительству – разрабатывать антиинфляционные меры.

Вместе с тем большинство работ, посвященных моделированию инфляции, имеют общую черту: в качестве искомой переменной выступает только индекс потребительских цен по группе «Все товары и услуги», так называемая инфляция в целом.

В экономической литературе существуют разные подходы к прогнозированию ИПЦ. Так, для предсказания годовых темпов инфляции инструментами машинного обучения У. Джункеев [Джункеев, 2024] применил модели градиентного бустинга, решающих деревьев, сверточную и рекурентную нейронные сети (CNN, RNN). В качестве эндогенной переменной автор использовал ИПЦ в процентах к соответствующему месяцу предыдущего года (г/г). Аналогичного подхода придерживались Е. Павлов [Павлов, 2020] и И. Байбуза [Байбуза, 2018].

Е. Волгина в своих исследованиях также применила модели машинного обучения и нейронные сети [Волгина, 2025]. В качестве эндогенной переменной автор использовала тот же ИПЦ в процентах, но помесячно (м/м). Отличительной особенностью работы является введение дополнительного фактора – тональности СМИ.

О. Семитуркин и А. Шевелев также моделировали инфляцию в месячном выражении, но уже на региональном уровне – по макрорегиону Сибирь [Семитуркин, Шевелев, 2023]. В качестве инструментария авторы использовали ML-модели и нейронные сети.

Традиционные методы обычно фокусируются на агрегированных показателях инфляции, однако покомпонентный подход, предполагающий анализ отдельных составляющих индекса потребительских цен, приобретает все большую популярность благодаря своей потенциальной точности и информативности, что подтверждают результаты исследований В. Крамкова [Крамков, 2023], который в качестве эндогенных переменных использует 3 компонента (продовольственные, непродовольственные товары и услуги), 6 компонент (отдельно выделены плодоовощная продукция, топливо моторное и услуги ЖКХ), 18 компонент (выделены еще несколько наиболее весомых и/или волатильных товарных групп) и 45 компонент. Аналогичные результаты представлены в работе Р. Латыпова, Е. Ахмедовой, Е. Постолита и М. Микитчук [Латыпов и др., 2024], где авторы используют разложение ИПЦ на 80 компонентов.

Анализ публикаций свидетельствует, что некоторые ученые для моделирования инфляции опираются на различные архитектуры нейронных сетей. Для целей настоящего исследования их применение мы посчитали нецелесообразным ввиду недостаточного количества точек для качественного обучения модели, даже на нескольких эпохах: согласно Е. Choi с соавторами, для достижения хорошей производительности RNN требуется минимум 1 000–2 000 наблюдений (83 года с помесячными данными); их число зависит от сложности задачи и специфики данных; на практике зачастую выстраиваются более длинные ряды – от 5 000 наблюдений [Choi et al., 2016].

Кроме того, как отмечают М. Mamedli и D. Shhibitov [Mamedli, Shhibitov, 2021], при построении прогноза важно учитывать задержки в публикации макроэкономических данных: например, показатели промышленного производства за определенный месяц становятся доступными лишь в начале, а данные по инфляции – в середине следующего месяца. Игнорирование этих лагов может привести к переоценке точности прогноза. В нашей работе данный эффект учитывается, что позволяет получить оценки вневыборочной ошибки, в большей мере соответствующие условиям реального прогнозирования инфляции.

Методология исследования

В целях построения прогноза месячного индекса потребительских цен (ИПЦ) на региональном уровне (в Ставропольском крае) были использованы различные модели временных рядов и алгоритмы машинного обучения:

- 1) SARIMA (Seasonal ARIMA);
- 2) Prophet;
- 3) Ridge regression (регрессия с L2-регуляризацией);
- 4) Random Forest Regression;
- 5) CatBoost Regressor.

Модель SARIMA представляет собой расширение классической модели ARIMA с учетом сезонных колебаний. Ее обобщенная форма записывается следующим образом [Box et al., 2015]:

$$\Phi_p(B) * \Phi_p(B^S) * (1 - B)^d * (1 - B^S)^D y_t = \\ = \theta_q(B) * \theta_Q(B^S) * \varepsilon_t, \quad (1),$$

где $\Phi_p(B)$ и $\theta_q(B)$ – полиномы операторов авторегрессии и скользящего среднего; $\Phi_p(B^S)$ и $\theta_Q(B^S)$ – полиномы сезонной авторегрессии и сезонного скользящего среднего; B – оператор сдвига назад: $B_y = y_{t-1}$; d и D –

порядок обычного (несезонного) дифференцирования временного ряда; S – длина сезонного периода; y_t – наблюдаемое значение временного ряда в момент времени t ; ε_t – белый шум (случайная ошибка).

Prophet – метод прогнозирования временных рядов, разработанный Facebook, ориентированный на данные с выраженной сезонностью и праздничными эффектами. Он основан на аддитивной декомпозиции, представлен следующей формой:

– тренд (*Trend*): используется два типа трендов: линейный и логарифмический (с ограничением сверху), который подстраивается под данные и может меняться во времени;

– сезонность (*Seasonality*): учитывается как годовая сезонность, так и дополнительные сезонные колебания (еженедельные или ежемесячные циклы);

– праздники (*Holidays*): позволяет учитывать влияние праздников и других специальных событий на временной ряд, например, продажи в праздничные дни;

– ошибки (*Residuals*): остатки модели предполагаются случайными, и их можно моделировать с использованием нормального распределения или других подходов для оценки неопределенности.

Математическая форма модели *Prophet* выглядит следующим образом [Taylor, Letham, 2017]:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (2)$$

где $y(t)$ – наблюдаемое значение во времени t ; $g(t)$ – тренд; $s(t)$ – сезонность; $h(t)$ – влияние праздников и специальных событий; ε_t – случайная ошибка.

Метод *Ridge regression*, также известный как L2-регрессия, представляет собой разновидность линейной регрессии, включающую в себя регуляризацию для улучшения качества модели и предотвращения переобучения.

В классической линейной регрессии модель определяется уравнением:

$$y = X\beta + \varepsilon, \quad (3)$$

где y – вектор наблюдаемых признаков; X – матрица признаков; β – вектор коэффициентов модели; ε – ошибка модели.

Метод находит такие значения коэффициентов β , которые минимизируют сумму квадратов ошибок. Однако при большом количестве признаков или наличии мультиколлинеарности коэффициенты модели могут становиться нестабиль-

ными и чрезмерно чувствительными к незначительным изменениям в данных, что приводит к переобучению. Метод *Ridge regression* вводит дополнительный штраф за величину коэффициентов, что позволяет уменьшить их влияние на модель и улучшить обобщающую способность. Модифицированная цель оптимизации с L2-регуляризацией выглядит следующим образом [Bishop, 2006; James et al., 2013]:

$$\hat{\beta}_{ridge} = argmin_{\beta} (\|y - X\beta\|^2 + \lambda \|\beta\|^2), \quad (4)$$

где λ – параметр регуляризации (гиперпараметр), который контролирует величину штрафа; $\|\beta\|^2$ – сумма квадратов коэффициентов.

Метод *Random Forest* (случайный лес) создает несколько деревьев на основе случайно выбранных подмножеств данных и случайных признаков, используемых для построения каждого дерева. Для прогнозирования модель агрегирует результаты всех деревьев. Для задачи регрессии используется усреднение прогнозов [Breiman, 2001; Liaw, Wiener, 2002].

CatBoost (градиентный бустинг) позволяет строить аддитивную функцию в виде суммы деревьев решений итерационно, по аналогии с методом градиентного спуска [Dorogush, Ershov, Gulin, 2017]. Его основной идеей является устранение ошибок предыдущего шага. Обобщенный алгоритм его работы предполагает определение набора данных $\{(x_i, y_i)\}_{i=1, \dots, n}$; числа итераций M ; выбор функции потерь $L(y, f)$ с выписанным градиентом; выбор семейства функций базовых алгоритмов $h(x, \theta)$, с процедурой их обучения; определение дополнительных гиперпараметров $h(x, \theta)$, например, глубины дерева у дерева решений [Prokhorenkova et al., 2019].

Затем проводится инициализация модели градиентного бустинга константным значением $\hat{f}(x) = \hat{f}_0, \hat{f}_0 = \gamma, \gamma \in R$:

$$\widehat{f}_0 = argmin \sum_{i=1}^n L(y_i, \gamma), \quad (5)$$

где n – количество обучающих примеров (объектов).

После чего для каждой итерации $t = 1, \dots, M$ повторяется ряд действий:

1) определение псевдо-остатков r_t :

$$r_{it} = - \left[\frac{\delta L(y_i, f(x_i))}{\delta f(x_i)} \right]_{f(x)=\hat{f}(x)}, \quad i = 1, \dots, n, \quad (6)$$

где δ – частная производная функции потерь по предсказанию модели;

2) построение нового базового алгоритма $h_t(x)$ на псевдо-остатках $\{(x_i, r_{it})\}_{i=1,\dots,n}$;
 3) поиск оптимального коэффициента ρ_t при $h_t(x)$ относительно исходной функции потерь:

$$\rho_t = \operatorname{argmin} \sum_{i=1}^n L(y_i, \hat{f}(x) + \rho * h(x_i, \theta)), \quad (7)$$

$$4) \text{сохранение } \hat{f}_t(x) = \rho_t * h_t(x).$$

После этого формируется итоговая GBM-модель $\hat{f}(x)$:

$$\hat{f}(x) = \sum_{i=0}^M \hat{f}_i(x) \quad (8)$$

В качестве исходных данных для проведения исследования выступила информация об индексе потребительских цен в Ставропольском крае и России в целом в месячном выражении (далее – ИПЦ, МоМ²) по 82 компонентам инфляции: ИПЦ, МоМ в целом («Все товары и услуги») (1-й уровень), компоненты «Продовольственные товары», «Непродовольственные товары» («Услуги») (2-й уровень) и 78 основных товарных групп (3-й уровень).

В качестве регрессоров выбраны следующие группы факторов: 1) статистические данные Росстата, характеризующие социально-экономическое положение субъектов Российской Федерации; 2) курс доллара в среднем за месяц³; 3) индекс ММВБ в среднем за месяц; 4) данные мониторинга предприятий Банка России; 5) динамика банковского кредитования и привлечения средств клиентов. Таким образом, в итоговый набор экзогенных переменных без учета лаговой структуры вошли более 60 показателей.

Отметим, что в настоящем исследовании мы не используем сезонно-корректированный таргет (индекс потребительских цен и его компоненты), поскольку его применение может исказить как интерпретацию результатов, так и сам процесс прогнозирования. По нашему мнению, сглаживание сезонной компоненты до этапа моделирования фактически лишает модель возможности учитывать повторяющиеся закономерности, присущие временным рядам. Кроме того, сезонная коррекция предполагает наличие априорного знания о структуре сезонности, тогда как в реальных условиях сезонные колебания могут быть нестабильными и изменяться со временем. Важно также отметить, что прогноз в модели, обученной на скорректированном таргете, требует последующего восстановления сезонной компоненты, что влечет дополнительные источники ошибки. Следовательно, в таком случае присутствует искусственное завышение качества модели, которое нивелируется

при восстановлении сезонно-корректированного ряда. Поскольку в прикладных задачах, как правило, интерес представляют именно фактические (некорректированные) значения, использование оригинального ИПЦ позволяет получать более реалистичные оценки и отражает поведение показателя в будущем.

Очевидно, что региональные экономики в едином национальном пространстве функционируют в условиях высокой степени открытости, иными словами, продукция, произведенная в одном субъекте России, свободно перемещается и реализуется на рынках других регионов. Это свидетельствует о том, что объем производства товаров в регионе не всегда коррелирует с уровнем цен на данной территории и в ряде случаев может не оказывать на него существенного влияния. В этой связи в исследовании использованы экзогенные переменные как по Ставропольскому краю, так и в целом по Российской Федерации.

Непосредственно для моделирования исходная выборка разделена на обучающую (train) и тестовую (test) части.

Важно отметить, что тренировочная выборка по ИПЦ и его компонентам включает период с января 2010 г. по декабрь 2023 года. При этом для экзогенных переменных – с января 2009 г. по сентябрь 2023 г. включительно. Такое отличие в периодичности от таргета необходимо для учета лаговой структуры в процессе моделирования.

Тестовая часть не задействована в обучении моделей и их валидации, а используется исключительно для оценки прогностических свойств модели. В качестве тестового периода рассматривались последовательно расширяющиеся временные интервалы в пределах 2024 г., а именно: январь – март (1–3 месяца), январь – июнь (1–6 месяцев), январь – сентябрь (1–9 месяцев), январь – декабрь (1–12 месяцев).

Обучение моделей осуществлялось с использованием процедуры кросс-валидации с тремя фолдами, что обеспечивало устойчивую оценку их прогностических характеристик и позволяло выбрать наилучшую модель для прогнозирования отдельных компонент индекса потребительских цен (ИПЦ) в помесячной динамике (МоМ) на различных горизонтах прогнозирования.

Важным элементом методики явилось требование предварительной известности всех экзогенных переменных, используемых в моделях. В частности, при построении прогноза на h месяцев вперед все экзогенные переменные включались в модель с минимальным лагом не менее h

месяцев, что исключало использование информации «из будущего» (табл. 1).

Прогнозирование индекса потребительских цен (ИПЦ) осуществлялось поэтапно, с использованием трех уровней агрегирования, отражающих различные степени детализации:

1) Моделирование агрегированного ИПЦ – построение прогноза непосредственно по совокупному индексу, без предварительного разложения на составляющие компоненты.

2) Моделирование укрупненных компонент – прогнозирование по трем основным категориям: «Продовольственные товары», «Непродовольственные товары» и «Услуги». Последующее агрегирование общего значения ИПЦ осуществлялось посредством взвешенного суммирования прогнозных значений указанных категорий с использованием официальных весов, опубликованных Федеральной службой государственной статистики (Росстат) на 2024 год.

3) Моделирование на уровне 78 товарных групп – прогнозирование по каждому из субкомпонент ИПЦ, включающих товары и услуги. Итоговое значение индекса формировалось путем агрегирования прогнозов с учетом соответствующих весов, отражающих структуру потребительской корзины.

Для каждого прогнозного горизонта (1, 3, 6, 9 и 12 месяцев) производился отбор наилучшей модели на основе процедуры трехфолдовой кросс-валидации. Выбор модели осуществлялся отдельно для каждого уровня детализации (агрегированный, укрупненный и детализированный уровни), что обеспечивало адаптивность подхода к различным аспектам динамики потребительских цен.

Методические подходы к моделированию

Перед подачей данных на вход прогнозным моделям все числовые переменные были стандартизированы с использованием метода *StandardScaler*, обеспечивающего приведение

значений признаков к единой шкале. Для линейных моделей дополнительно проведена проверка временных рядов на стационарность с применением теста Дики – Фуллера. Все нестационарные ряды преобразованы методом первых разниц.

Для повышения вычислительной эффективности и сокращения времени обучения моделей размерность пространства экзогенных переменных снижена методом *главных компонент (PCA)*. Данный метод позволяет трансформировать исходный набор коррелированных признаков в ограниченное число новых ортогональных переменных – главных компонент, которые упорядочиваются по степени объясняемой дисперсии [Ringnér, 2008].

В настоящем исследовании сохраненный набор компонент объяснял не менее 95 % дисперсии исходных данных. При моделировании ИПЦ на горизонте 12 месяцев с учетом лагов было получено свыше 600 предикторов, их удалось сократить до 75 компонент без существенной потери качества (см. рис. 1).

Поскольку не все предикторы вносят значимый вклад в формирование ИПЦ и его компонент, для отбора наиболее информативных признаков применялся алгоритм Recursive Feature Elimination (RFE). Этот метод итеративно исключает наименее значимые признаки на основе внутренних оценок важности модели (например, коэффициентов регрессии или важности признаков в деревьях решений), до достижения заданного числа признаков [Kuhn, Johnson, 2013; Pedregosa et al., 2011].

Для повышения качества прогнозов, приближенных к условиям реального времени, отбор оптимальных моделей для каждого прогнозического горизонта и компоненты ИПЦ осуществлялся на основе *трехфолдовой кросс-валидации*.

Дополнительно проводилась настройка гиперпараметров моделей при помощи процедуры GridSearchCV из библиотеки scikit-learn, обеспечивающей перебор заданных комбинаций пара-

Таблица 1

Прогнозные горизонты и используемые лаги экзогенных переменных

Период прогноза	Лаги
1–3 месяца	3–12
4–6 месяцев	6–12
7–9 месяцев	9–12
9–12 месяцев	12

Примечание. Составлено авторами.

метров и оценку результатов на основе кросс-валидации. В качестве целевой метрики использовалась средняя абсолютная ошибка (MAE).

Подбор гиперпараметров с помощью GridSearchCV (кроме SARIMA и Prophet) осуществлялся отдельно для каждой из следующих моделей⁴:

1. Модель SARIMA характеризуется набором гиперпараметров, определяющих ее структуру: p, d, q – порядки авторегрессии, интегрирования (дифференцирования) и скользящего среднего соответственно; P, D, Q – их сезонные аналоги с периодом сезонности s . Эти параметры не обучаются напрямую, а подбираются до построения модели, поскольку определяют конфигурацию временного ряда, включая тренды, сезонные колебания и шум.

2. Модель Prophet имеет ряд гиперпараметров, регулирующих ее поведение и степень адаптации к данным временного ряда. Ключевыми из них являются: `changepoint_prior_scale` – управляет чувствительностью модели к изменениям тренда; `seasonality_prior_scale` – регулирует силу сезонной компоненты; `seasonality_mode` – определяет тип сезонности (аддитивная или мультиплексиативная); `n_changepoints` – задает число возможных точек смены тренда.

3. Ridge Regression: основным настраиваемым параметром выступал коэффициент L2-регуляризации, варьируемый в логарифмически равномерной сетке значений. Это позволяло эффективно проводить переобучение при наличии мультиколлинеарности среди регрессоров.

4. Random Forest Regressor: подбор включал следующие параметры: количество деревьев (`n_estimators`), максимальная глубина деревьев (`max_depth`), минимальное количество наблюдений в листе (`min_samples_leaf`), а также доля признаков, используемых при каждом разбиении (`max_features`). Данный подбор позволял адаптировать модель к различным структурам данных и уровню шумов.

5. CatBoost Regressor: для данной модели использовалась обертка `CatBoostRegressor` совместно с `GridSearchCV` через `scikit-learn API`. Оптимизировались параметры: количество итераций (`iterations`), шаг обучения (`learning_rate`), глубина деревьев (`depth`), коэффициент L2-регуляризации (`l2_leaf_reg`) и минимальное число наблюдений в листе (`min_data_in_leaf`). Это обеспечивало устойчивость модели и высокую точность при наличии категориальных или разреженных признаков.

Каждая модель проходила отдельный цикл подбора гиперпараметров для каждого горизонта прогнозирования и уровня детализации структуры ИПЦ. Такой подход позволил учесть различия в статистических свойствах компонент индекса и обеспечить адаптивность моделей к конкретным условиям прогнозирования.

Результаты моделирования

В рамках эмпирического исследования протестированы три различных подхода к прогнозированию индекса потребительских цен, различа-

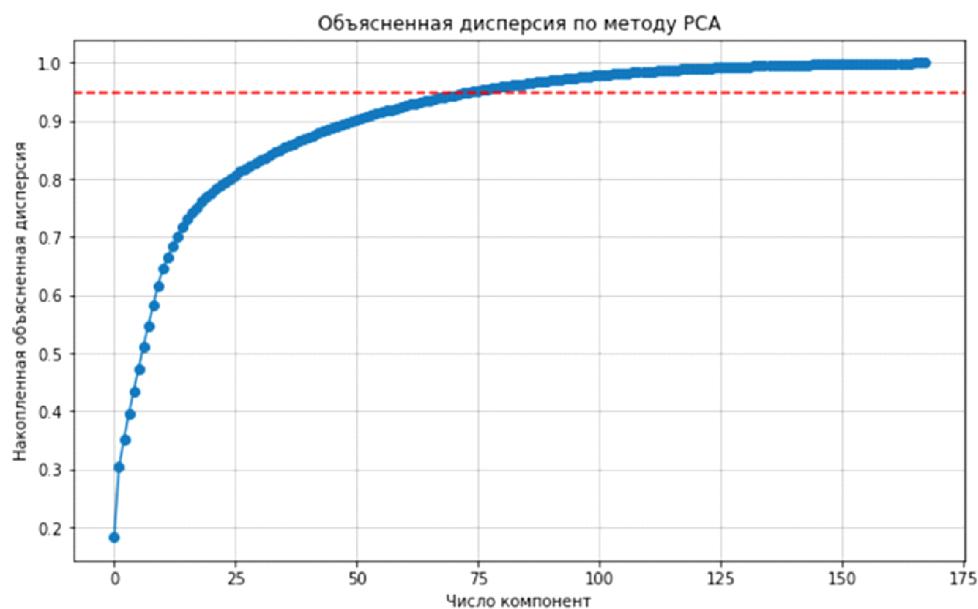


Рис. 1. Количество главных компонент РСА для сохранения 95 % дисперсии

Примечание. Составлено авторами.

ющиеся уровнем агрегирования моделируемых компонент:

1) агрегированный подход – прямое прогнозирование ИПЦ в целом, без предварительной декомпозиции на подгруппы;

2) средний уровень агрегации – прогнозирование по трем укрупненным категориям: продовольственные товары, непродовольственные товары и услуги. Финальное значение ИПЦ восстанавливается путем агрегирования соответствующих прогнозов с использованием официальных весов потребительской корзины по формуле (9):

$$\text{ИПЦ} = \sum_{i=1}^n I_i * w_i \quad (9)$$

где I_i – индекс потребительских цен на конкретный товар / услугу; w_i – вес конкретного товара / услуги в потребительской корзине; n – количество компонент ИПЦ;

3) дезагрегированный подход – это индивидуальное моделирование 78 товарных и сервисных групп с последующим агрегированием аналогичным образом. Результаты расчетов представлены на рисунке 2.

В таблице 2 приведены данные об ошибке моделей на тестовой выборке.

Полученные результаты демонстрируют устойчивое преимущество покомпонентного модели-

рования, при котором каждая товарная группа прогнозируется отдельно. Данный подход обеспечил наименьшее значение MAE на всех прогнозных горизонтах, особенно в кратко- и среднесрочной перспективе (до 9 месяцев). Отметим, что даже при более длительном горизонте (1–12 месяцев) точность данного метода остается сопоставимой с краткосрочными прогнозами по другим подходам.

Для формальной оценки значимости различий в точности моделей был применен тест Диболда – Мариано, основанный на сравнении среднеквадратичных ошибок (MSE) различных стратегий прогнозирования (см. табл. 3).

Таким образом, статистическая проверка подтверждает, что различия в точности прогнозов между подходами являются *не случайными*, а статистически значимыми. Особенно заметным является преимущество детализированного подхода над агрегированными.

Вклад в ошибку прогноза мы рассчитали по формуле (10):

$$\text{Вклад} = \sum_{i=1}^n I_i * w_i \quad (10)$$

где I_i – отклонение фактического значения от прогнозного для отдельного товара / услуги; w_i – вес конкретного товара / услуги в потребительской корзине; n – количество компонент ИПЦ.

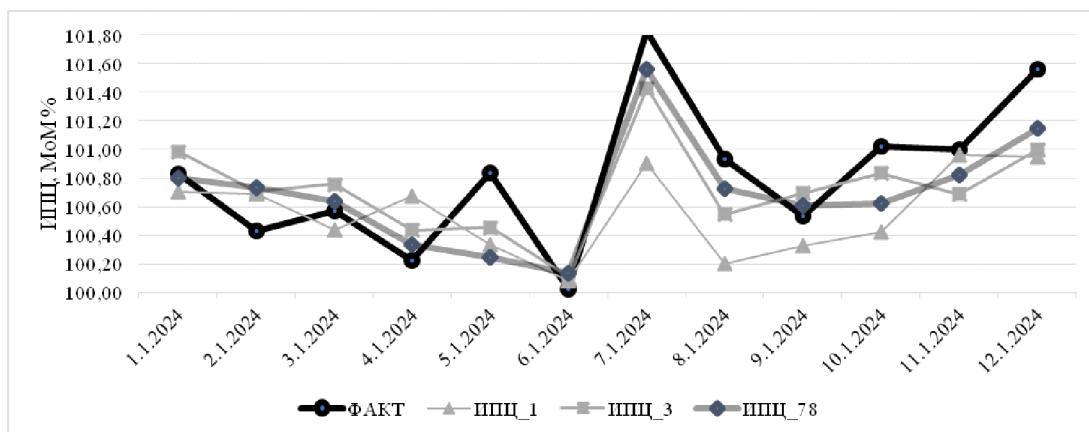


Рис. 2. Сравнительная оценка качества прогнозов ИПЦ в Ставропольском крае на основе агрегированного и дезагрегированного подходов

Примечание. Составлено авторами.

Таблица 2

Сравнение MAE различных подходов к моделированию ИПЦ

Подход	MAE			
	1–3 мес.	1–6 мес.	1–9 мес.	1–12 мес.
ИПЦ в целом	0,17	0,26	0,38	0,39
ИПЦ по 3 компонентам	0,21	0,22	0,25	0,28
ИПЦ по 78 компонентам	0,13	0,20	0,20	0,23

Примечание. Составлено авторами.

Согласно проведенной декомпозиции прогнозистических ошибок (табл. 3), наибольший вклад в среднюю абсолютную ошибку вносили такие компоненты, как услуги пассажирского транспорта и мясопродукты. Это обусловлено, с одной стороны, их высокой волатильностью, с другой – существенной долей в структуре потребительской корзины, что усиливает их влияние на итоговое значение ИПЦ (рис. 3).

Дополнительно данные категории подвержены воздействию факторов, имеющих слабо предсказуемую природу. В частности, на динамику цен мясной продукции и плодовоощной группы значительное влияние оказывают эпизоотологические риски (например, вспышки заболеваний среди животных), а также погодные условия, критически важные для сельскохозяйственного производства.

Так, например, в мае – июне 2024 г. в Ставропольском крае наблюдалась аномально холодная погода, что привело к росту издержек производителей тепличных овощей на отопление. Это, в свою очередь, обусловило ускоренный рост цен на соответствующую продукцию по сравнению с предыдущими годами.

Заключение

В настоящем исследовании реализован многоуровневый подход к прогнозированию регионального индекса потребительских цен (ИПЦ) на примере Ставропольского края с использованием современных методов машинного обучения. Прогнозирование осуществлялось по трем уровням детализации: по агрегированному ИПЦ, по основным категориям (продовольственные и непродовольственные товары, услуги), а также по 78 товарным подгруппам. Такой подход позволил оценить влияние структуры и уровня детализации модели на точность прогнозов.

В качестве прогнозных алгоритмов использовались как классические статистические модели (SARIMA, Prophet), так и модели машинного обучения (Ridge Regression, Random Forest, CatBoost). Гиперпараметры подбирались с использованием GridSearchCV в сочетании с трехфолдовой кросс-валидацией, основной метрикой выступила средняя абсолютная ошибка (MAE). Дополнительно применялись методы снижения размерности (PCA, где это применимо) и отбора

Таблица 3

Результаты теста Диболда – Мариано для оценки различий в точности прогнозов

Сравнение подходов	DM-статистика	P-value	Значимость ($\alpha = 0,05$)
ИПЦ 1 vs ИПЦ 3	-2,347	0,0189	Различаются
ИПЦ 1 vs ИПЦ 78	-3,128	0,0018	Различаются
ИПЦ 3 vs ИПЦ 78	-1,892	0,0485	Различаются

Примечание. Составлено авторами.

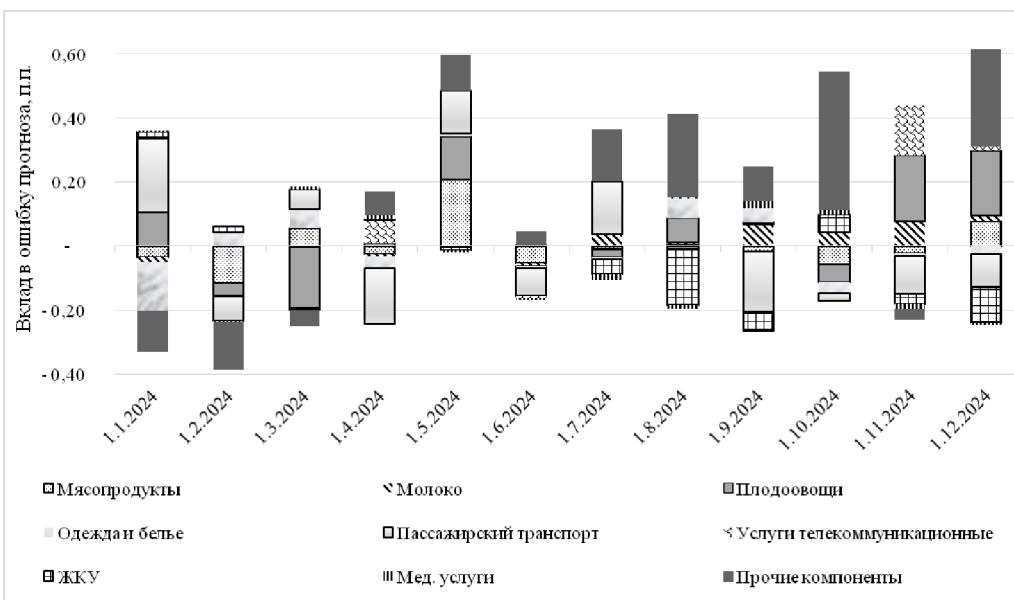


Рис. 3. Декомпозиция ошибки прогноза по вкладу отдельных компонент ИПЦ

Примечание. Составлено авторами.

признаков (RFE, где это применимо), что позволило повысить интерпретируемость моделей.

Результаты эмпирического анализа показали, что на всех прогнозных горизонтах наиболее точные результаты демонстрирует подход, основанный на моделировании по отдельным товарным группам с последующим агрегированием. Подход, использующий три укрупненные категории, оказался менее точным, однако превзошел по качеству прогнозы, полученные при моделировании общего ИПЦ без дезагрегации.

Дополнительный анализ ошибок показал, что наибольший вклад в отклонение прогнозов вносили компоненты с высокой волатильностью и значительным весом в структуре потребительской корзины, что подчеркивает важность учета специфики отдельных товарных групп и использования гибких моделей, способных адаптироваться к различным источникам неопределенности, включая внешние шоки и погодные аномалии.

Полученные выводы подтверждают репрезентативность покомпонентного подхода к прогнозированию инфляции и обосновывают важность его применения в региональной практике, особенно в условиях ограниченности данных и повышенной волатильности отдельных рынков.

ПРИМЕЧАНИЯ

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 решение регулятора.

² Month-over-Month, то есть «из месяца в месяц».

³ Investing.com.

⁴ В классическом понимании модели SARIMA и Prophet не имеют гиперпараметров, подобных ML-моделям. Вместо этого обе модели могут настраиваться с помощью структурных компонент, определяющих форму и гибкость аппроксимации временного ряда.

СПИСОК ИСТОЧНИКОВ

- Байбуза И., 2018. Прогнозирование инфляции с помощью методов машинного обучения // Деньги и кредит. № 4. С. 42–59. DOI: 10.31477/rjmf.201804.42
- Волгина Е., 2025. Прогнозирование инфляции с использованием новостных индексов // Деньги и кредит. № 84. С. 26–59.
- Джункеев У., 2024. Прогнозирование инфляции в России на основе градиентного бустинга и нейронных сетей // Деньги и кредит. № 1. С. 53–76.

- Крамков В., 2023. Улучшает ли учет компонентов ИПЦ качество прогнозов инфляции? // Серия докладов об экономических исследованиях. № 112. URL: https://cbr.ru/StaticHtml/File/145659/wp_112.pdf
- Латышов Р., Ахмедова Е., Постолит Е., Микитчук М., 2024. Прогнозирование компонент инфляции методами машинного обучения // Деньги и кредит. № 83. С. 23–44.
- Павлов Е., 2020. Прогнозирование инфляции в России с помощью нейронных сетей // Деньги и кредит. № 1. С. 57–73. DOI: 10.31477/rjmf.202001.57
- Семитуркин О., Шевелев А., 2023. Корректное сравнение предиктивных свойств моделей машинного обучения на примере прогнозирования инфляции в Сибири // Деньги и кредит. № 82. С. 87–103.
- Bishop C. M., 2006. Pattern Recognition and Machine Learning. Springer. 736 p.
- Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M., 2015. Time Series Analysis: Forecasting and Control. 5th ed. Wiley. 709 p.
- Breiman L., 2001. Random Forests // Machine Learning. Vol. 45, № 1. P. 5–32. DOI: 10.1023/A:1010933404324
- Choi E., Bahadori M. T., Schuetz A., Stewart W. F., Sun J., 2016. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks // Proceedings of the 1st Machine Learning for Healthcare Conference, PMLR. Vol. 56. P. 301–318. URL: <https://proceedings.mlr.press/v56/Choi16.pdf>
- Dorogush A. V., Ershov V., Gulin A., 2017. CatBoost: Gradient Boosting with Categorical Features Support // ArXiv. Vol. abs/1810.11363. DOI: 10.48550/arXiv.1810.11363
- James G., Witten D., Hastie T., Tibshirani R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer. 426 p.
- Kuhn M., Johnson K., 2013. Applied Predictive Modeling. N. Y.: Springer. 600 p.
- Liaw A., Wiener M., 2002. Classification and Regression by Random Forest // R News. Vol. 2/3. P. 18–22. URL: <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>
- Mamedli M., Shibitov D., 2021. Forecasting Russian CPI with Data Vintages and Machine Learning Techniques // Working Paper Series. URL: https://cbr.ru/StaticHtml/File/120016/wp-apr21_e.pdf
- Pedregosa F. et al., 2011. Scikit-Learn: Machine Learning in Python // Journal of Machine Learning Research. Vol. 12. P. 2825–2830. DOI: 10.48550/arXiv.1201.0490
- Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A., 2019. CatBoost: Unbiased Boosting with Categorical Features // ArXiv: 1706.09516v5. DOI: 10.48550/arXiv.1706.09516
- Ringnér M., 2008. What is Principal Component Analysis? // Nature Biotechnology. Vol. 26, № 3. P. 303–304. DOI: 10.1038/nbt0308-303
- Taylor S. J., Letham B., 2017. Forecasting at Scale // PeerJ Preprints. Vol. 5: e3190v2. DOI: 10.7287/peerj.preprints.3190v2

REFERENCES

- Baybuza I., 2018. Prognozirovaniye infljaci s pomoshchju metodov mashinnogo obuchenija [Inflation Forecasting Using Machine Learning Methods]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 4, pp. 42-59. DOI: 10.31477/rjmf.201804.42
- Volgina E., 2025. Prognozirovaniye infljaci s ispolzovaniem novostnyh indeksov [Forecasting Inflation Using News Indices]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 84, pp. 26-59.
- Dzhunkeev U., 2024. Prognozirovaniye infljaci v Rossii na osnove gradientnogo bustinga i nejronnyh setej [Forecasting Inflation in Russia Using Gradient Boosting and Neural Networks]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 1, pp. 53-76.
- Kramkov V., 2023. Uluchshayet li uchet komponentov IPC kachestvo prognozov infljaci? [Does Taking into Account CPI Components Improve the Quality of Inflation Forecasts?]. *Serija dokladov ob ekonomicheskikh issledovanijah* [Series of Reports on Economic Research], no. 112. URL: https://cbr.ru/StaticHtml/File/145659/wp_112.pdf
- Latypov R., Akhmedova E., Postolit E., Mikitchuk M., 2024. Prognozirovaniye komponent infljaci metodami mashinnogo obuchenija [Bottom-up Inflation Forecasting Using Machine Learning Methods]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 83, pp. 23-44.
- Pavlov E., 2020. Prognozirovaniye infljaci v Rossii s pomoshchju nejronnyh setej [Forecasting Inflation in Russia Using Neural Networks]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 1, pp. 57-73. DOI: 10.31477/rjmf.202001.57
- Semiturkin O., Shevelev A., 2023. Korrektnoe sravnenie prediktivnyh svojstv modelej mashinnogo obuchenija na primere prognozirovaniya infljaci v Sibiri [Correct Comparison of Predictive Features of Machine Learning Models: The Case of Forecasting Inflation Rates in Siberia]. *Dengi i kredit* [Russian Journal of Money and Finance], no. 82, pp. 87-103.
- Bishop C.M., 2006. *Pattern Recognition and Machine Learning*. Springer. 736 p.
- Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M., 2015. *Time Series Analysis: Forecasting and Control*. Wiley. 709 p.
- Breiman L., 2001. Random Forests. *Machine Learning*, vol. 45, no. 1, pp. 5-32. DOI: 10.1023/A:1010933404324
- Choi E., Bahadori M.T., Schuetz A., Stewart W.F., Sun J., 2016. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *Proceedings of the 1st Machine Learning for Healthcare Conference, PMLR*, vol. 56, pp. 301-318. URL: <https://proceedings.mlr.press/v56/Choi16.pdf>
- Dorogush A.V., Ershov V., Gulin A., 2017. CatBoost: Gradient Boosting with Categorical Features Support. *ArXiv*, vol. abs/1810.11363. DOI: 10.48550/arXiv.1810.11363
- James G., Witten D., Hastie T., Tibshirani R., 2013. *An Introduction to Statistical Learning: with Applications in R*. New York, Springer. 426 p.
- Kuhn M., Johnson K., 2013. *Applied Predictive Modeling*. Springer. 600 p.
- Liaw A., Wiener M., 2002. Classification and Regression by Random Forest. *R News*, vol. 2/3, pp. 18-22. URL: <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>
- Mamedli M., Shabitov D., 2021. Forecasting Russian CPI with Data Vintages and Machine Learning Techniques. *Working Paper Series*. URL: https://cbr.ru/StaticHtml/File/120016/wp-apr21_e.pdf
- Pedregosa F. et al., 2011. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830. DOI: 10.48550/arXiv.1201.0490
- Prokhorenkova L., Gusev G., Vorobyev A., Dorogush A.V., Gulin A., 2019. CatBoost: Unbiased Boosting with Categorical Features. *ArXiv*, vol. 1706.09516v5. DOI: 10.48550/arXiv.1706.09516
- Ringnér M., 2008. What Is Principal Component Analysis? *Nature Biotechnology*, vol. 26, no. 3, pp. 303-304. DOI: 10.1038/nbt0308-303
- Taylor S.J., Letham B., 2017. Forecasting at Scale. *PeerJ Preprints*, vol. 5: e3190v2. DOI: 10.7287/peerj.preprints.3190v2

Information About the Authors

Yuri A. Metel, Candidate of Sciences (Economics), Chief Economist, Stavropol Territory Branch of the Southern Main Directorate of the Central Bank of the Russian Federation, Lenina St, 286, 355003 Stavropol, Russian Federation, akademik.st.2018@mail.ru, <https://orcid.org/0009-0001-5334-0020>

Natalia N. Kunitsyna, Doctor of Sciences (Economics), Professor, Head of the Department of Finance and Credit, North Caucasus Federal University, Pushkina St, 1, 355017 Stavropol, Russian Federation, nkunitcyna@ncfu.ru, <https://orcid.org/0000-0001-9336-8100>

Информация об авторах

Юрий Андреевич Метель, кандидат экономических наук, главный экономист, Отделение по Ставропольскому краю Южного главного управления Центрального банка Российской Федерации, ул. Ленина, 286, 355003 г. Ставрополь, Российская Федерация, akademik.st.2018@mail.ru, <https://orcid.org/0009-0001-5334-0020>

Наталья Николаевна Куницына, доктор экономических наук, профессор, заведующий кафедрой финансов и кредита, Северо-Кавказский федеральный университет, ул. Пушкина, 1, 355017 г. Ставрополь, Российская Федерация, nkunitcyna@ncfu.ru, <https://orcid.org/0000-0001-9336-8100>