

УДК 330.45

DOI 10.34755/IROK.2026.25.76.058

Алёшечкина Таисия Александровна

Федеральное Государственное Бюджетное Образовательное Учреждение Высшего
Образования «Национальный исследовательский университет «МЭИ»

Молокова София Сергеевна

Федеральное Государственное Бюджетное Образовательное Учреждение Высшего
Образования «Национальный исследовательский университет «МЭИ»

Денисенко Вера Константиновна

Федеральное Государственное Бюджетное Образовательное Учреждение Высшего
Образования «Национальный исследовательский университет «МЭИ»

Сравнительный анализ моделей ARIMA и XGBOOST для прогнозирования валютного курса

Аннотация. В статье представлены результаты сравнительного анализа двух моделей прогнозирования временных рядов — классического эконометрического подхода ARIMA и алгоритма градиентного бустинга XGBoost — применительно к курсу китайского юаня к российскому рублю. Охарактеризована роль прогнозных моделей в алгоритмическом трейдинге. Рассмотрены преимущества и ограничения использования методов машинного обучения на валютном рынке. Выполнена интерполяция пропусков в данных, осуществлено разбиение на обучающую и тестовую выборки в пропорции 80% к 20%. Проведено обучение и тестирование моделей, рассчитаны метрики точности. Все расчёты реализованы в среде Python с использованием библиотек Pandas, NumPy, Matplotlib, Statsmodels и Scikit-learn. В результате установлено, что XGBoost обеспечивает снижение средней абсолютной ошибки на 24,4% по сравнению с ARIMA, что подтверждает его преимущество в условиях ограниченной выборки и нестационарности ряда.

Ключевые слова: валютный курс, ARIMA, XGBoost, алгоритмический трейдинг, прогнозирование временных рядов, машинное обучение.

Alyoshechkina Taisiya Alexandrovna

Federal State Budgetary Educational Institution of Higher Education "National Research
University "MPEI"

Molokova Sofia Sergeevna

Federal State Budgetary Educational Institution of Higher Education "National Research
University "MPEI"

Denisenko Vera Konstantinovna

Federal State Budgetary Educational Institution of Higher Education "National Research
University "MPEI"

Comparative Analysis of ARIMA and XGBoost models for exchange rate forecasting

Abstract. The article presents the results of a comparative analysis of two time series forecasting models — the classical ARIMA econometric approach and the XGBoost gradient boosting algorithm — applied to the exchange rate of the Chinese yuan against the Russian ruble. The role of forecasting models in algorithmic trading is described. The advantages and limitations of using machine learning methods in the foreign exchange market are considered. The data was interpolated, and the training and test sets were divided in a 80% to 20% ratio. The models were trained and tested, and accuracy metrics were calculated. All calculations were performed in the Python environment using the Pandas, NumPy, Matplotlib, Statsmodels, and Scikit-learn libraries.

The results showed that XGBoost achieved a 24.4% reduction in mean absolute error compared to ARIMA, confirming its superior performance in limited sample sizes and non-stationary series.

Keywords: exchange rate, ARIMA, XGBoost, algorithmic trading, time series forecasting, machine learning.

Прогнозирование временных рядов представляет собой процесс определения будущих значений исследуемого показателя на основе анализа его ретроспективных данных. Данный подход заключается в выявлении устойчивых закономерностей, трендов и циклических компонент, после чего на их основе строятся математические модели, позволяющие экстраполировать ряд на заданный горизонт упреждения.

Временной ряд понимается как последовательность числовых значений, упорядоченных во времени и отражающих динамику развития некоторого явления или процесса. В экономике и финансах прогнозирование временных рядов используется для предсказания цен активов, объёмов спроса, макроэкономических индикаторов и других показателей, необходимых для принятия обоснованных управленческих и инвестиционных решений.

Цель настоящей научной статьи заключается в проведении сравнительного анализа двух моделей прогнозирования — ARIMA и XGBoost — для исследования динамики валютного курса и определении наиболее подходящего инструмента для краткосрочного прогнозирования.

Задачи научной статьи:

1. провести обзор моделей прогнозирования временных рядов, применяемых в эконометрике и машинном обучении;
2. охарактеризовать валютный рынок и роль алгоритмической торговли;
3. выбрать датасет с котировками CNY/RUB, выполнить его предварительную обработку и интерполяцию пропусков;
4. реализовать модели ARIMA и XGBoost, выполнить прогнозирование;
5. сравнить полученные результаты на основе метрик точности;
6. оценить перспективы использования машинного обучения в задачах валютного прогнозирования.

Актуальность работы обусловлена высокой волатильностью валютных рынков и возрастающей долей китайского юаня в расчётах России со странами Азии, что создаёт практический запрос на точные и интерпретируемые прогнозные модели.

Алгоритмическая торговля представляет собой процесс совершения операций на финансовых рынках по заранее определённым правилам с использованием специализированных компьютерных систем, часто называемых торговыми роботами [3]. В последние годы фондовый и валютный рынки демонстрируют устойчивую тенденцию к цифровизации, что выражается в росте доли автоматизированных сделок. По данным Московской биржи, на срочном рынке доля роботизированных операций достигает 60%, на валютной секции — около 65%. Российские банки, включая СберБанк, ВТБ и Альфа-Банк, предоставляют клиентам доступ к роботам-советникам, что свидетельствует о широком внедрении алгоритмических подходов.

Прогнозные модели являются ключевым элементом алгоритмических стратегий, поскольку качество предсказания будущей цены напрямую влияет на финансовый результат. Применение методов машинного обучения на валютном рынке обладает рядом преимуществ: способность обрабатывать большие объёмы данных, автоматическое выявление нелинейных зависимостей, адаптация к изменяющимся условиям. Однако существуют и ограничения: риск переобучения на исторических данных, чувствительность к выбросам и структурным сдвигам, а также сложность интерпретации некоторых моделей, которые воспринимаются как «чёрный ящик».

В рамках настоящего исследования сопоставляются два подхода, принадлежащих к принципиально различным методологическим традициям. С одной стороны, модель ARIMA

(AutoRegressive Integrated Moving Average) — инструмент классической эконометрики, накопивший обширную практику применения при анализе временных рядов на протяжении нескольких десятилетий [1]. С другой — XGBoost (eXtreme Gradient Boosting), алгоритм из арсенала машинного обучения, завоевавший признание благодаря стабильно высоким результатам в задачах предиктивного моделирования [6].

В основе ARIMA лежит представление ряда как линейной комбинации собственных лагированных значений и накопленных ошибок прогноза. Модель параметризуется тремя величинами: p — глубина авторегрессионной составляющей, d — число операций разностного преобразования, необходимых для достижения стационарности, q — порядок компоненты скользящего среднего.

XGBoost строит прогноз посредством последовательного наращивания ансамбля деревьев решений, где каждое новое дерево нацелено на компенсацию невязок, оставшихся от предыдущих. В качестве входных признаков для данной задачи использованы: пять лаговых значений курса, скользящие средние по трём и пяти торговым дням, а также календарные переменные — день недели и порядковый номер дня в месяце. Подобная конфигурация позволяет одновременно учитывать инерционность курса на коротком горизонте и потенциальные внутринедельные закономерности.

Качество прогнозов оценивалось с помощью четырёх стандартных метрик регрессионного анализа [2]:

1. RMSE (среднеквадратическая ошибка) — измеряет среднее расстояние между прогнозными и фактическими значениями, придавая больший вес крупным ошибкам:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

2. MAE (средняя абсолютная ошибка) — представляет собой среднее абсолютных отклонений и является линейной оценкой, одинаково взвешивающей все ошибки:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

3. MAPE (средняя абсолютная процентная ошибка) — выражает отклонение прогноза от факта в процентах, что удобно для интерпретации:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|}$$

4. R^2 (коэффициент детерминации) — показывает долю дисперсии зависимой переменной, объяснённую моделью:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2}$$

Исходный датасет содержал 61 официальную котировку курса китайского юаня к российскому рублю за период с 3 февраля по 30 апреля 2026 года. Данные получены из открытых источников Центрального банка РФ [4]. Поскольку котировки публикуются только в рабочие дни, ряд содержал пропуски, соответствующие выходным и праздничным дням. Для устранения этого недостатка выполнена линейная интерполяция, в результате которой сформирован непрерывный ряд из 87 ежедневных значений.

Набор исходных данных разделён на обучающую и тестовую выборки в соотношении 80% к 20%: на первых 69 днях модели обучались, последующие 18 дней использовались для проверки качества прогнозов. Такое разбиение имитирует реальный сценарий, когда по историческим данным строится прогноз на ближайший период.

Реализация всех этапов — загрузки данных, интерполяции, визуализации, обучения моделей и расчёта метрик — выполнена на языке программирования Python с использованием библиотек Pandas для обработки данных, NumPy для математических операций, Matplotlib для

построения графиков, Statsmodels для реализации ARIMA и Scikit-learn для реализации XGBoost [5].

На рисунке 1 представлен график исходного ряда после интерполяции, на котором отчётливо видны три фазы: стабильный февраль, резкий рост в марте с пиком 20 марта и последующая коррекция в апреле.



Рисунок 1. График исходного ряда – Составлено авторами

Далее приведены результаты теста Дики–Фуллера, подтверждающие нестационарность исходного ряда ($p\text{-value} = 0,6756 > 0,05$), что обосновывает необходимость взятия первой разности.

ТЕСТ ДИКИ-ФУЛЛЕРА (стационарность ряда)

ADF Statistic: -1.195336

p-value: 0.675599

Критические значения:

1%: -3.509736

5%: -2.896195

10%: -2.585258

Рисунок 2. Тест Дики-Фуллера (стационарность ряда) – Составлено авторами

На рисунке 3 показаны автокорреляционная и частная автокорреляционная функции первой разности, на основе которых определена структура модели ARIMA(1,1,1).

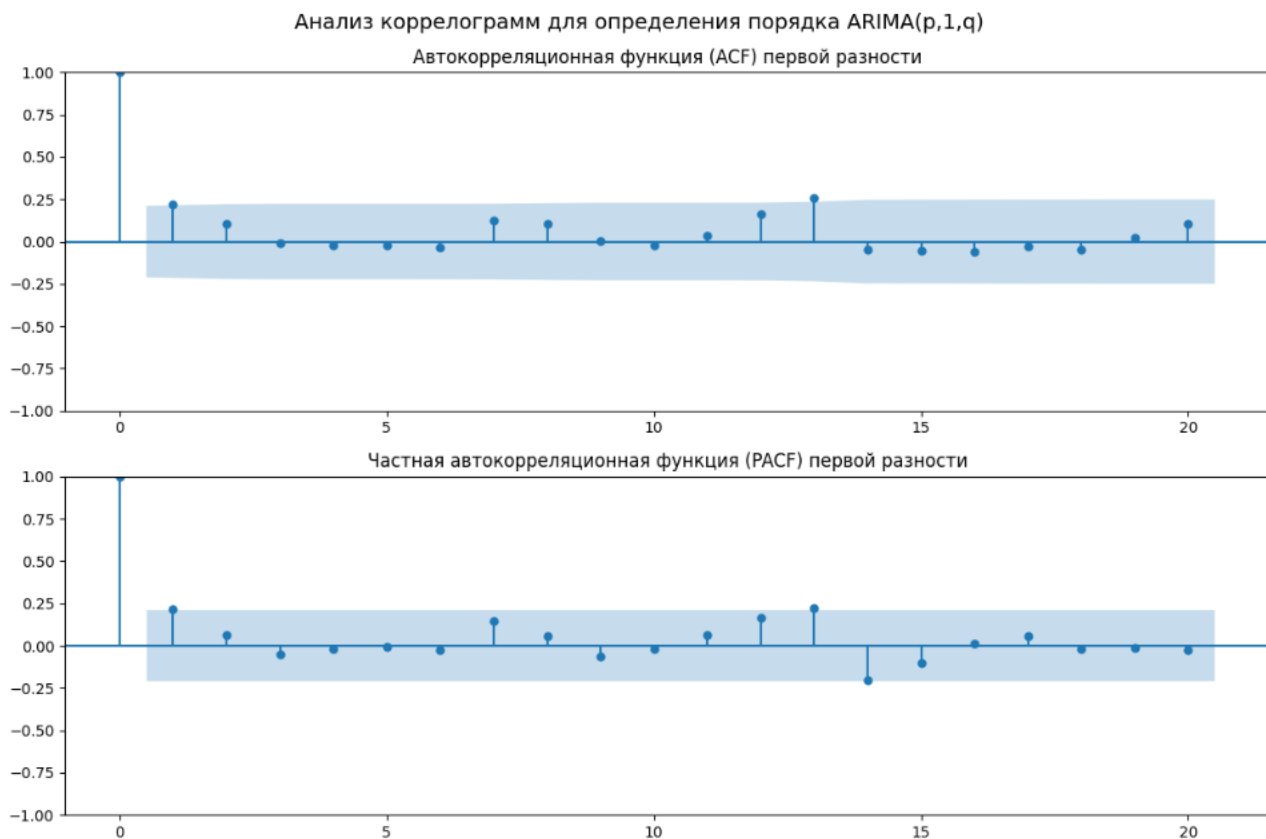


Рисунок 3. ACF И PACF – Составлено авторами

После обучения моделей выполнено прогнозирование на тестовой выборке из 18 дней. На рисунке 4 представлено сравнение прогнозных значений ARIMA(1,1,1) с фактическими данными.



Рисунок 4. Прогноз и метрики точности – Составлено авторами

Модель корректно воспроизвела общее направление движения курса вниз, но не уловила локальных ускорений падения. На рисунке 5 приведены диагностические графики остатков ARIMA, подтверждающие их соответствие белому шуму.

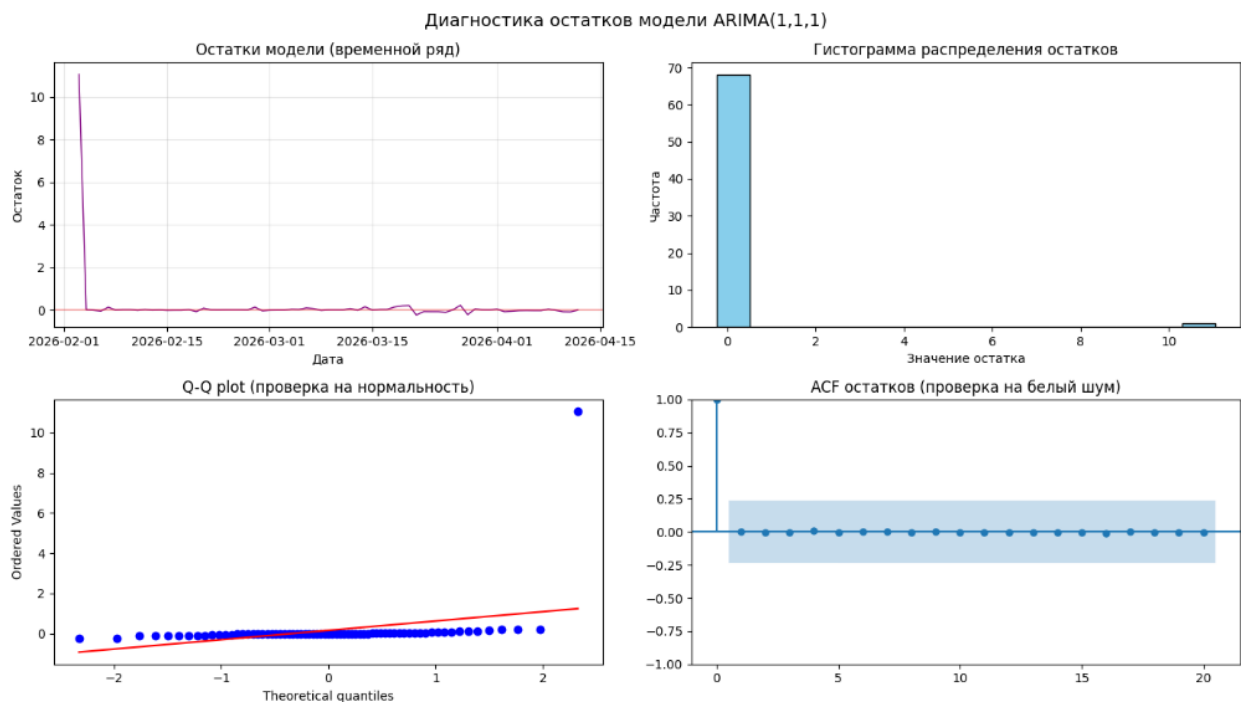


Рисунок 5. Диаграмма остатков модели – Составлено авторами

На рисунке 6 показан прогноз, полученный с помощью XGBoost. Данная модель значительно точнее отразила динамику курса, включая оба ускорения падения во второй половине апреля.

MAE: 0.0881 P
RMSE: 0.1067 P
MAPE: 0.80%



Рисунок 6. Прогноз XGBoost на тестовой выборке – Составлено авторами

Для количественной оценки точности обеих моделей рассчитаны метрики, представленные в таблице 1.

Таблица 1

Метрики точности прогнозных моделей

Метрика	ARIMA	XGBoost
MAE, руб.	0,1551	0,1172
RMSE, руб.	0,1741	0,1290
MAPE, %	1,41	1,06
R ²	0,823	0,891

Исходя из полученных метрик, можно сделать вывод, что модель XGBoost превосходит ARIMA по всем показателям точности. Средняя абсолютная ошибка снижена на 24,4%, что

является экономически значимым результатом. При объёме сделки в 1 миллион юаней разница в ошибке между моделями составляет около 40 тысяч рублей на одну сделку.

Анализ важности признаков XGBoost показал, что наибольший вклад в прогноз вносят лаг 1 (важность 0,32), скользящее среднее с окном три дня (0,27) и лаг 2 (0,18). Календарные признаки имеют важность менее 0,08, что говорит об их незначимости на рассматриваемом временном интервале. Таким образом, модель случайного леса может считаться более надёжной в условиях ограниченной выборки и нестационарности ряда. Преимущество XGBoost заключается в способности ансамбля деревьев охватывать множество возможных исходов, что при усреднении результатов позволяет нивелировать эффект переобучения и повысить устойчивость к выбросам.

Практическая значимость работы состоит в обосновании целесообразности использования XGBoost в задачах краткосрочного прогнозирования валютного курса, особенно в периоды коррекции после резких движений. Полученные результаты могут быть применены при разработке алгоритмических торговых стратегий.

Дальнейшие исследования предполагают расширение набора признаков за счёт включения макроэкономических показателей (разница ключевых ставок, цены на энергоносители), а также тестирование моделей на других валютных парах и временных интервалах.

Список источников

1. Бокс Дж., Дженкинс Г. Анализ временных рядов: прогноз и управление. — М.: Мир, 1974. — 406 с.
2. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — 3-е изд. — М.: Вильямс, 2016. — 912 с.
3. Малыхин Е.М. Алгоритмический трейдинг для профессионалов. — СПб.: БХВ-Петербург, 2021. — 176 с.
4. Официальные курсы валют Центрального банка Российской Федерации [Электронный ресурс]. — URL: https://cbr.ru/currency_base/dynamics/?UniDbQuery.Posted=True&UniDbQuery.so=1&UniDbQuery.mode=1&UniDbQuery.date_req1=&UniDbQuery.date_req2=&UniDbQuery.VAL_NM_RQ=R01375&UniDbQuery.From=01.02.2026&UniDbQuery.To=30.04.2026 (дата обращения: 01.05.2026).
5. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python. — СПб.: Диалектика, 2020. — 320 с.
6. Хэсти Т., Тибширани Р., Фридман Дж. Элементы статистического обучения. — М.: ДМК Пресс, 2020. — 764 с.

Сведения об авторах

Алёшечкина Таисия Александровна, магистр, кафедра Менеджмента в энергетике и промышленности, Федеральное Государственное Бюджетное Образовательное Учреждение Высшего Образования «Национальный исследовательский университет «МЭИ», Москва, Россия.

Молокова София Сергеевна, магистр, кафедра Менеджмента в энергетике и промышленности, Федеральное Государственное Бюджетное Образовательное Учреждение Высшего Образования «Национальный исследовательский университет «МЭИ», Москва, Россия.

Денисенко Вера Константиновна, научный руководитель, научный сотрудник, ассистент кафедры безопасности и информационных технологий, Федеральное Государственное Бюджетное Образовательное Учреждение Высшего Образования «Национальный исследовательский университет «МЭИ», Москва, Россия.

Information about the author

Alyoshechkina Taisiya Alexandrovna, master's degree, Department of Management in Energy and Industry, Federal State Budgetary Educational Institution of Higher Education "National Research University "MPEI", Moscow, Russia.

Molokova Sofia Sergeevna, master's degree, Department of Management in Energy and Industry, Federal State Budgetary Educational Institution of Higher Education "National Research University "MPEI", Moscow, Russia.

Denisenko Vera Konstantinovna, Scientific Advisor, Research Fellow, Assistant of the Department of Security and Information Technologies, Federal State Budgetary Educational Institution of Higher Education "National Research University "MPEI", Moscow, Russia.