

*Осипенко Александр Анатольевич,
Студент кафедры прикладной математики,
РТУ МИРЭА*

ОЦЕНКА ВЛИЯНИЯ МАКРОЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ДИНАМИКУ ЦЕН РОССИЙСКОЙ ЖИЛОЙ НЕДВИЖИМОСТИ: СОПОСТАВЛЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

Аннотация. Статья посвящена количественной оценке связи макроэкономических показателей с динамикой цен на жилую недвижимость в Российской Федерации. На основе открытых квартальных рядов Росстата и Банка России (2014Q1–2022Q4) построена воспроизводимая модель, в которой целевой переменной выступает индекс цен на вторичном рынке жилья (% к аналогичному кварталу предыдущего года). Сравниваются линейная регрессия, Ridge-регрессия, случайный лес, градиентный бустинг, а также эконометрические ориентиры $AR(1)$ и $ARDL/Ridge$ при единой схеме временной валидации. Показано, что ансамблевые методы машинного обучения снижают ошибку относительно простой линейной спецификации, однако на коротком шоковом тестовом интервале 2021–2022 гг. уступают наивному прогнозу и ряду эконометрических моделей. Основной практический результат — согласованное ранжирование макрофакторов (ключевая ставка, доходы населения, безработица, курс доллара США, ипотечный канал) с использованием SHAP-анализа и сценарного моделирования. Выводы интерпретируются как статистические ассоциации, а не как строгие причинно-следственные эффекты.

Ключевые слова: рынок недвижимости, макроэкономические факторы, машинное обучение, временные ряды, случайный лес, ARDL, SHAP, официальная статистика.

Abstract. *The paper quantifies the association between macroeconomic indicators and residential real estate price dynamics in the Russian Federation. Using open quarterly data from Rosstat and the Bank of Russia (2014Q1–2022Q4), a reproducible pipeline is implemented with the secondary housing price index (year-on-year) as the target variable. Linear regression, Ridge, Random Forest, Gradient Boosting, AR(1), and ARDL/Ridge are compared under a unified time-based validation scheme. Ensemble machine learning models reduce forecast error relative to a simple linear specification, but on the short shock-sensitive test window of 2021–2022 they underperform naive forecasts and several econometric benchmarks. The main practical contribution is a consistent ranking of macro factors (policy rate, household income, unemployment, USD exchange rate, mortgage channel) supported by SHAP analysis and scenario experiments. Results are interpreted as statistical associations rather than causal effects.*

Keywords: *housing market, macroeconomic factors,; machine learning, time series, Random Forest, ARDL, SHAP, official statistics.*

Введение

Рынок жилой недвижимости занимает особое место в экономике Российской Федерации: динамика цен на жильё отражает баланс спроса и предложения, доступность ипотечного кредитования, инвестиционные ожидания и общую макроэкономическую конъюнктуру. Для аналитических и прогнозных задач необходимы модели, связывающие движение цен с агрегированными макроэкономическими индикаторами, публикуемыми в системе официальной статистики.

Традиционный подход опирается на линейные регрессии, модели с распределёнными лагами (ARDL), векторные авторегрессии (VAR) и модели коррекции ошибок (VECM). Их преимущество — прозрачная интерпретация коэффициентов и развитая процедура диагностики временных рядов. Вместе с тем совместное воздействие нескольких макрофакторов, нелинейности и лаговые эффекты могут быть описаны недостаточно полно в рамках линейных спецификаций. Методы машинного обучения (ML) для задач регрессии позволяют аппроксимировать более сложные зависимости, но их применение к макроэкономическим данным требует строгой временной валидации и сопоставления с базовыми моделями.

Цель настоящей статьи — разработать и апробировать воспроизводимую модель оценки влияния макроэкономических факторов на динамику цен на жилую недвижимость в РФ с применением ML и сравнением с эконометрическими подходами на единой эмпирической базе открытых данных Росстата и Банка России. Гипотеза исследования формулируется как частичное превосходство ансамблевых ML-моделей над линейной регрессией по точности на отложенной выборке при сохранении информативности макропризнаков для ранжирования факторов.

Обзор литературы

В зарубежных исследованиях рынка жилья подчёркивается роль доходов населения, доступности кредита и ожиданий участников рынка (Case, Shiller, 2003; Glaeser et al., 2008). Для российских данных характерны выраженные режимные сдвиги (кризисы 2008, 2014, 2020 гг., санкционный шок 2022 г.), высокая роль ипотечных программ и решений Банка России по ключевой ставке (Лapidус, Малинецкая, 2018; Кириллов, 2023). В последние годы усилился интерес к сравнению эконометрических и ML-подходов на одних и тех же макроэкономических рядах (Aye et al., 2020; Bork, Møller, 2020).

Общим методологическим требованием для временных рядов является исключение «утечки» информации из будущего при обучении и настройке гиперпараметров (Hyndman, Athanasopoulos, 2021; Plosser, 2021). Для интерпретации «чёрных ящиков» деревьев и бустинга применяются методы относительной важности признаков и SHAP-значения (Lundberg, Lee, 2017; Molnar, 2022). Настоящее исследование дополняет литературу по российскому рынку жилья единым воспроизводимым контуром: от описания показателей официальной статистики до сопоставления линейных, регуляризованных и ансамблевых моделей на одной матрице признаков.

Материалы и методы

Эмпирическая база сформирована из открытых источников: показатели Росстата (индекс цен на вторичном рынке жилья по РФ, «все типы квартир», % к АППГ; индекс потребительских цен; реальные денежные доходы; уровень безработицы по методологии МОТ; индекс физического объёма ВВП; показатели ввода жилья) и Банка России (ключевая ставка, курс USD, объём ипотечного жилищного кредитования). Для расширенного эксперимента использована квартальная панель за 2014Q1–2022Q4; дополнительно включены регистрируемая безработица, индекс предпринимательской уверенности в строительстве, физический объём работ в строительстве и фиктивные переменные для 2020 и 2022 гг.

Подготовка данных включает унификацию периодичности (агрегирование месячных рядов до квартальных), построение лагов макропризнаков (1, 2 и 4 квартала), стандартизацию признаков на обучающей выборке и формирование матрицы X и вектора отклика y . После формирования лагов модельная выборка содержит $N = 24$ наблюдения; для базового эксперимента с лагом 1 квартала — $N = 19$. Программная реализация выполнена на Python (pandas, scikit-learn, LightGBM/XGBoost) с фиксацией `random_state` и версий библиотек.

Сравниваются следующие модели: (1) линейная регрессия; (2) Ridge-регрессия с подбором коэффициента регуляризации по временным фолдам; (3) случайный лес (Random Forest); (4) градиентный бустинг (Gradient Boosting); (5) AR(1) и (6) ARDL/Ridge как эконометрические ориентиры. Все модели обучаются на одной матрице признаков. Выборка разбивается по времени: обучение — 2018Q2–2021Q3 (расширенная спецификация: до 2021Q2), тест — 2021Q4–2022Q4 (расширенная: 2021Q3–2022Q4). Подбор гиперпараметров выполняется только на обучающей и валидационной частях.

Качество оценивается по MAE, RMSE и R^2 на тестовом интервале. Дополнительно проводятся скользящая временная кросс-валидация (TimeSeriesSplit), анализ чувствительности (исключение отдельных признаков), сценарное моделирование (изменение ключевой ставки, курса USD, доходов и объёма ИЖК) и SHAP-анализ для градиентного бустинга. Макропризнаки сгруппированы по экономическим каналам: денежно-кредитные (ставка, курс), спросовые (доходы, безработица, ипотека), предложенческие (ввод жилья, строительные индикаторы), шоковые (dummy-переменные).

Результаты

На базовом тестовом интервале 2021Q4–2022Q4 ($N = 19$ после лагов) случайный лес показал наименьшую ошибку среди ML-моделей: MAE = 10,485; RMSE = 10,798 против MAE = 12,53 у линейной регрессии и MAE = 14,386 у градиентного бустинга (табл. 1). Вместе с тем все спецификации уступили наивному прогнозу последнего значения индекса (MAE = 3,544), что отражает структурный сдвиг 2021–2022 гг. и высокую инерцию целевого ряда в шоковый период.

В расширенном эксперименте (лаг 4 квартала, $N = 24$) наилучшими ориентирами точечного прогноза выступили линейная регрессия (MAE = 2,243) и ARDL/Ridge (MAE = 2,564); случайный лес дал MAE = 8,331, AR(1) — MAE = 4,882. Таким образом, гипотеза о превосходстве ML над линейными моделями по

ошибке подтверждается лишь частично и зависит от спецификации лагов и выбора тестового окна.

Таблица 1. Метрики качества моделей на тестовом интервале (базовая спецификация, лаг 1 кв.)

Модель	MAE	RMSE	R ² (тест)
Линейная регрессия	12,53	13,189	-10,705
Ridge-регрессия	13,028	13,703	-11,636
Случайный лес	10,485	10,798	-6,846
Градиентный бустинг	14,386	14,702	-13,546
Наивный прогноз (y_{t-1})	3,544	4,061	-0,110

Анализ важности признаков (табл. 2) демонстрирует согласованность линейной и нелинейной интерпретаций: наибольший вклад вносят ключевая ставка Банка России (SHAP = 4,115; нормированный коэффициент Ridge = 1,052), индекс потребительских цен, реальные доходы, индекс физического объёма ВВП и объём ипотечного кредитования. В расширенной модели случайного леса доминирует ключевая ставка (важность 0,621), за ней следуют безработица (0,138), доходы (0,095) и курс USD (0,069). Результаты согласуются с экономической логикой каналов спроса и стоимости заимствования.

Таблица 2. Топ макропризнаков по вкладу в объяснение динамики индекса цен

Признак (лаг 1 кв.)	Ridge	SHAP (знач.)	Важность RF (расш.)
Ключевая ставка ЦБ	1,052	4,115	0,621
ИПЦ	1,223	0,451	0,002
Реальные доходы	0,185	0,437	0,095
Курс USD	0,208	0,553	0,069
Индекс физ. объёма ВВП	0,705	0,327	0,028
Объём ИЖК (ln)	0,805	0,236	0,014
Безработица	3,291	0,082	0,138

Ablation-анализ показал, что исключение ключевой ставки снижает MAE случайного леса с 10,118 до 8,267, что подтверждает её центральную роль в спецификации. Сценарный анализ (повышение ставки на 2 п.п., ослабление рубля

на 10 %, снижение доходов на 2 п.п.) демонстрирует локальную реакцию модели в ожидаемых направлениях, однако не интерпретируется как причинный прогноз. Наибольшая ошибка RF приходится на 2022Q1 (факт 123,33; прогноз 109,94), когда индекс достиг локального максимума на фоне резкого изменения монетарных условий.

Обсуждение

Полученные результаты следует трактовать с учётом ограничений эмпирической базы. Малый объём квартальных наблюдений после формирования лагов ($N = 19-24$), агрегирование по РФ без региональной детализации и наличие структурных сдвигов (регулирование ипотеки, пересмотр методик индексов, шок 2022 г.) снижают надёжность точечного прогноза уровня индекса. Отрицательные значения R^2 на коротком тестовом окне у всех моделей указывают на доминирование инерции целевого ряда над макроэкономическими предикторами в выбранном периоде.

С методологической точки зрения ML-модели в данной постановке сохраняют исследовательскую ценность даже при уступании ARDL/Ridge по MAE: они проверяют нелинейную гипотезу, дают независимое ранжирование факторов и поддерживают сценарный анализ. Эконометрические модели остаются важным ориентиром для коротких рядов благодаря более устойчивой точности и явной параметрической интерпретации. Практическая значимость работы — в воспроизводимом конвейере подготовки данных и сравнительной оценке факторов рынка жилья на основе исключительно открытых источников.

Перспективы дальнейших исследований связаны с расширением горизонта данных, построением региональной панели, разделением первичного и вторичного рынков и включением моделей с явным учётом структурных сдвигов (Markov-switching, dummy-переменные расширенного набора, VECM при достаточной длине ряда).

Заключение

В статье разработана и апробирована модель оценки связи макроэкономических факторов с динамикой цен на жилую недвижимость в Российской Федерации на квартальных данных Росстата и Банка России. Сопоставление линейных, регуляризованных, ансамблевых ML-моделей и эконометрических ориентиров показало, что макропризнаки и их лаги информативны для описания динамики индекса цен и ранжирования факторов; наиболее значимы ключевая ставка, доходы населения, безработица, курс USD и ипотечный канал.

Гипотеза о безусловном превосходстве ML над линейными моделями по точности прогноза подтверждена лишь частично: случайный лес снижает MAE относительно линейной регрессии, но на коротком шоковом ряду уступает наивному прогнозу и ряду эконометрических спецификаций. Основной вклад исследования — методический: единая временная валидация, SHAP-интерпретация и сценарный анализ на официальной статистической базе. Выводы носят ассоциативный характер и не заменяют специализированные прогнозные комплексы или инвестиционные рекомендации.

Список литературы:

1. Aye G.C. et al. Forecasting housing prices using machine learning // *Research in International Business and Finance*. — 2020. — Vol. 52.
2. Bork L., Møller S. Housing prices and unemployment: A forecasting approach // *International Journal of Forecasting*. — 2020. — Vol. 36. — № 2.
3. Case K.E., Shiller R.J. Is there a bubble in the housing market? // *Brookings Papers on Economic Activity*. — 2003. — № 2.
4. Glaeser E.L. et al. Housing bubbles // *NBER Working Paper*. — 2008. — № 14193.

5. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. — 3rd ed. — Melbourne : OTexts, 2021.
6. James G. et al. An Introduction to Statistical Learning. — 2nd ed. — New York : Springer, 2021.
7. Кириллов П.Н. Рынок жилья и ипотечное кредитование в России // Финансы и кредит. — 2023. — № 29 (829).
8. Кислицына О.В. Моделирование динамики цен на жильё с учётом макроэкономических факторов // Экономика и математические методы. — 2022. — Т. 58. — № 2.
9. Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // NeurIPS. — 2017. — Vol. 30.
10. Molnar C. Interpretable Machine Learning. — 2nd ed. — 2022.
11. Plosser C. Machine Learning for Macroeconomics // NBER Working Paper. — 2021.
12. Цены на рынке жилья : стат. сборник / Росстат. — М., 2023.
13. Единая межведомственная информационно-статистическая система (ЕМИСС). — URL: <https://www.fedstat.ru/> (дата обращения: 01.06.2026).