

ПРОГНОЗИРОВАНИЕ ИНДЕКСА ЦЕН НА НЕДВИЖИМОСТЬ В РОССИИ

Н. С. НИКИТИНА

Статья посвящена выбору наилучшей модели для краткосрочного прогнозирования индекса цен на недвижимость в России. Были рассмотрены популярные методы машинного обучения: Ridge и Lasso regressions, Elastic Net regression и методы работы с временными рядами: Naive, Exponential smoothing, ARIMA, OLS. Набор переменных включает в себя значения ВВП, инфляции, эффективного обменного курса, ставки межбанковского кредитования и цен на нефть. Методы машинного обучения — Ridge regression и Elastic Net regression — показывают высокое качество прогнозирования индекса цен на недвижимость по сравнению со стандартными методами работы с временными рядами — Naive, Exponential smoothing, ARIMA.

Статья подготовлена в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС при Президенте Российской Федерации.

Ключевые слова: прогнозирование, индекс цен на недвижимость, машинное обучение.

JEL: C32, C53, R30.

Введение и обзор литературы

Существенный рост индекса цен на недвижимость в последние годы привлекает все большее внимание исследователей. В то время как большинство из них пытаются описать поведение рынка недвижимости в зависимости от различных факторов и выявить причины, вызвавшие такой рост, мы ставим основной целью своей работы поиск и оценку модели, которая давала бы наименьшую ошибку прогноза индекса цен на недвижимость, построенного наиболее достоверным способом, не ставя перед собой задачу факторизации динамики цен по детерминантам.

Данной теме посвящено множество исследований, представленных обширным списком литературы, стимулом к появлению которой послужил мировой экономический кризис 2008–2009 гг. Сегодня невозможно представить себе современную макроэкономическую модель без включения в нее жилищного сектора, являющегося каналом влияния традиционных структурных шоков на динамику агрегированных переменных.

Ключевые работы по макроэкономическому моделированию состояния рынка недвижимости базируются на построении как моделей с коинтегрированными временными рядами — VECM [8], так и векторных авторег-

рессий различного вида — SVAR [3], PVAR [4], BVAR [13], FAVAR [7]. Отдельно можно выделить литературу, в которой особое внимание уделяется анализу влияния на рынок недвижимости состояния рынка нефтепродуктов. Так, было выявлено гетерогенное влияние шоков нефтяного сектора в зависимости от положения страны в мировой торговле нефтью, в соответствии с которым страна—экспортер нефти более подвержена шокам на рынке нефтепродуктов, что переносится и на рынок недвижимости [6; 9].

По рынку недвижимости в России имеется несколько работ. Например, в [1] используется МНК-модель для описания процесса ценообразования на рынке недвижимости в Москве. В работе показано значимое положительное влияние цен на нефть марки Urals на динамику цен на недвижимость, а также отмечено влияние на них валютного курса и склонности к сбережению.

Что касается литературы, посвященной собственно прогнозированию индекса цен на недвижимость, то обзор современных техник в этой области представлен в [12]. Авторы выделяют следующие техники: «...Artificial neural network (ANN), Hedonic price model (HPM), Fuzzy logic system (FLS), Support vector machine (SVM), Linear regression (LR), Decision tree (DT),

Никитина Наталья Сергеевна, младший научный сотрудник РАНХиГС при Президенте Российской Федерации (Москва), e-mail: nikitina-ns@ranepa.ru

Random forest (RF), K-nearest neighbour (KNN), Partial least square (PLS), Naive bayes (NB), Multiple regression analysis (MRA), Spatial analysis (SA), Gradient boosting (GB), Ridge regression, Lasso regression и Ensemble learning model (ELM)».

Кроме того, в этой работе определены преимущества и недостатки каждой из моделей: ANN позволяет использовать нелинейные функции, но имеет низкую точность прогнозирования; HPM адаптивна к добавлению различных рыночных товаров и учету внешних факторов; FLS – альтернативная модель традиционной оценки имущества, снижающая неопределенность прогноза, но не учитывающая факторы риска; SVM может помочь найти оптимальное решение, но в ограниченном объеме выборки; LR требует наличия предпосылки о структуре данных и использует линейную форму функции, которая неизменна во времени; DT подходит для выбора переменных и работы с категориальными данными, но может приводить к созданию сложного «дерева решений», результаты которого невозможно интерпретировать; RF обладает высокой точностью, но набор данных формируется случайно; KNN учитывает меру сходства на основе функции расстояния от похожего по исходным характеристикам объекта, чтобы получить наиболее точную оценку стоимости, поэтому возникает проблема поиска таких соответствий; PLS подходит для работы с коррелированными переменными, но может быть ненадежна, когда данные не распределены нормально; NB – простая в использовании модель, не требующая большого количества данных; MRA позволяет оценивать коэффициенты с большим количеством данных, но при наличии коррелированных переменных оценка может быть невозможна; SA сложно проводить из-за требуемого качества пространственных данных; GB работает лучше, чем RF, но требует качественных данных; Ridge и Lasso regressions позволяют работать с мультиколлинеарностью, но со смещенной

оценкой дисперсии; ELM обладает высокой точностью и надежностью прогнозирования, но чувствительна к наличию наблюдений, существенно отличающихся по характеристикам от общей выборки. Поскольку большинство представленных выше моделей хорошо прогнозируют поведение индекса цен на недвижимость, выбор конкретной техники прогнозирования должен производиться исходя из конечных целей исследования.

В эмпирической работе [2] прогнозирование осуществляется с помощью обычного регрессионного анализа. Существенным ограничением при этом является качество данных в моделях: временные ряды доступны лишь за 2014–2017 гг., а исследуемые дома расположены в одном из городов Индонезии; также стоит отметить, что работа сфокусирована не на макроэкономических аспектах формирования стоимости жилья, а на микроданных конкретных районов по таким показателям, как расстояние от центра города, количество ресторанов, игровых площадок, медицинских учреждений и др. Авторы построили отдельные модели для каждого из районов исследуемого города, но получили сравнительно большие ошибки для прогнозных цен.

Существенным улучшением прогнозных моделей может служить учет нелинейных взаимосвязей между переменными, поскольку традиционный регрессионный анализ не позволяет учесть эти особенности. В [16] используется ARIMA-модель для прогнозирования индивидуальной стоимости дома, представляющая собой систему для автоматизированной оценки стоимости, чтобы не допустить предвзятых оценок из-за личной заинтересованности. В этой работе использовались также микроданные описания местоположения квартиры. В результате прогнозные значения здесь в основном соответствуют истинным значениям, а прогноз тенденции стоимости жилья совпадает с общей тенденцией развития данного рынка.

Логичным расширением применяемых техник прогнозирования считается машинное

обучение. Например, в [15] используются подходы DT и MRA. Однако ограничением этого исследования можно считать географическую локацию привлекаемых данных — модели построены только для одного города. Расчеты показали, что для задач такого типа с точки зрения ошибки прогноза применение MRA предпочтительнее DT.

Большее количество способов прогнозирования в рамках машинного обучения представлено в [11]: MRA, GB, Ridge и Lasso regression, Elastic Net и Ada Boost regression. Наилучшей моделью с точки зрения MSE и RMSE в этой работе оказалась GB — она показала наибольшую точность прогнозирования данных. В [17] задействованы практически все возможные модели машинного обучения для прогнозирования цен на недвижимость; наилучшую прогностическую способность здесь показали модели GB и RF — при их использовании среднеквадратическая ошибка прогноза наименьшая.

В литературе существует большой пласт исследований, посвященных анализу взаимосвязи между ценами на недвижимость и фазами кредитного цикла. В частности, данному вопросу посвящены [5; 10; 14]. Действительно, разные фазы кредитного цикла по-разному влияют на состояние рынка недвижимости. Однако главной проблемой в подобных исследованиях является то, что сами показатели кредитного цикла — ненаблюдаемые и их выделение требует отдельного эконометрического анализа. Поэтому данную проблему мы оставляем для будущих исследований, а в статье ограничиваемся набором вышеуказанных фундаментальных переменных.

Данные и используемые модели

Для расчетов были использованы квартальные данные — с первого квартала 2001 г. по четвертый квартал 2020 г., а в качестве переменных — следующие показатели (при необходимости переменные были очищены от сезонности с помощью процедуры X-13 ARIMA):

- *house* — индекс номинальной стоимости вторичной¹ недвижимости в целом по Российской Федерации для всех типов квартир;
- *gdp* — ВВП; данный показатель определяет агрегированный уровень деловой активности и доходов домохозяйств и, соответственно, выступает одним из факторов со стороны спроса на недвижимость;
- *inf* — инфляция; для этого показателя были использованы данные «квартал к соответствующему кварталу предыдущего года» — это позволяет получить сезонно дифференцированные ряды, что означает отсутствие необходимости очищать полученный ряд от сезонности дополнительно. Инфляция с экономической точки зрения важна для прогнозирования цен на недвижимость из-за возникающих инфляционных ожиданий относительно будущих периодов, поскольку домохозяйства, стремясь застраховаться от инфляционных рисков, покупают недвижимость;
- *oil* — цена на нефть марки Brent; данный показатель является важнейшей детерминантой уровня экономического развития, что подтверждается и проведенным выше обзором литературы. Это во многом очевидно, так как при повышении цен на нефть доходы населения возрастают даже при неизменном уровне ВВП, что, в свою очередь, увеличивает спрос на недвижимость;
- *miacr* — ставка MIACR, квартальные значения которой получены усреднением ежедневных значений средневзвешенной фактической ставки по кредитам, предоставленным московскими банками; срок кредита — 1 день. Конечно, хорошим показателем, отражающим ситуацию на рынке недвижимости и ипотечного кредитования в России, будет значение средневзвешенной ставки по жилищным или ипотечным кредитам, выданным в течение месяца

¹ Результаты для индекса цен первичного жилья практически идентичны и могут быть предоставлены отдельно по запросу.

физическим лицам-резидентам (такие сведения публикуются на официальном сайте Банка России только с 2009 г.²); кроме того, это важный показатель как для домохозяйств, которые берут ипотечный кредит, так и для строительных компаний. Данная переменная позволила бы использовать еще, например, и спред между ставками. Но мы решили использовать именно MIACR, для того чтобы не сокращать исследуемый временной ряд, поскольку эта ставка во многом определяет условия на рынке кредитования и является его индикатором;

- *reer* – реальный эффективный валютный курс, отражающий покупательную способность домохозяйств и тем самым влияющий на спрос на недвижимость.

Для наших моделей требуется разделить выборку на обучаемую (*train*) и тестовую (*test*); в качестве границы обучаемого набора данных мы берем четвертый квартал 2014 г. Для оценки качества прогнозов применяется RMSE: чем ниже его значение, тем точнее прогноз:

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

где \hat{y}_i – прогнозное значение индекса цен на недвижимость в период t ; y_i – истинное значение

Качество прогнозов

Модель	Значение RMSE
Naive	0,0190
Exponential smoothing	0,0162
ARIMA	0,0145
OLS train	0,0143
OLS test	0,0143
Ridge regression train	0,0163
Ridge regression test	0,0138
Lasso regression train	0,0165
Lasso regression test	0,0140
Elastic Net regression train	0,0159
Elastic Net regression test	0,0111

Источник: расчеты автора.

индекса цен на недвижимость в период t ; T – число периодов прогнозирования.

Для прогнозирования была использована следующая эконометрическая спецификация:

$$house_t - house_{t-1} = \alpha + \beta_1[x_{t-1} - x_{t-2}] + \beta_2[x_{t-2} - x_{t-3}] + \varepsilon_t,$$

где $house_t$ – логарифм индекса цен на недвижимость в период t ; x_{t-i} – матрица независимых логарифмированных (кроме *miacr*) переменных (*gdp, inf, oil, miacr, reer*) в период $t - i$.

Мы оценили следующие модели: Naive, Exponential smoothing (ES), ARIMA, OLS, Ridge regression (RR), Lasso regression (LR), Elastic Net regression (EN).

Результаты прогнозирования

В таблице представлены значения RMSE для всех указанных выше моделей. Из нее видно, что наиболее точными оказываются модели EN и RR, имеющие наименьшую ошибку прогноза среди всех использованных. Наихудшей моделью оказалась Naive – наивный прогноз, при котором прогнозным значением является значение этого же показателя в предыдущем периоде; соответственно, значение RMSE у этой модели оказалось наибольшим среди всех рассмотренных – 0,0190. Наилучшей моделью на тестовой подвыборке стала Elastic Net regression: значение RMSE – 0,0111 – по сравнению с наивным прогнозом уменьшилось у нее на 42%. Не плохой результат на тестовой подвыборке показала также Ridge regression: значение RMSE по сравнению с наивным прогнозом уменьшилось у нее на 27%.

Заключение

В статье были построены различные модели прогнозирования индекса цен на недвижимость в России: Naive, Exponential smoothing, ARIMA, OLS, Ridge regression, Lasso regression

² Сайт Центрального банка РФ. URL: <https://cbr.ru/statistics/table/?tableId=4-1>

и Elastic Net regression. Для этого были использованы квартальные данные основных макроэкономических переменных: ВВП, цен на нефть, инфляции, ставки MIACR и эффективного обменного курса. Было выявлено, что наилучшей прогностической способностью для индекса цен на недвижимость обладают модели Ridge и Elastic Net regressions — они

показывают наименьшую ошибку прогноза среди всех моделей.

Представленные методы могут быть использованы в дальнейшем для принятия решений о проведении последовательной экономической политики и построения прогнозов для индекса цен на недвижимость в краткосрочном периоде. ■

Литература / References

1. Сальников В.А., Михеева О.М. Модели прогнозирования цен на московском рынке жилой недвижимости // Проблемы прогнозирования. 2018. № 1 (166). С. 129–139 / Salnikov V.A., Mikheeva O.M. Price forecasting models in the Moscow residential real estate market // Forecasting Problems. 2018. No. 1 (166). Pp. 129–139.
2. Alfiyatina A.N. et al. Modeling house price prediction using regression analysis and particle swarm optimization // International Journal of Advanced Computer Science and Applications. 2017. Vol. 8. No. 10. Pp. 323–326.
3. Bian T.Y., Gete P. What drives housing dynamics in China? A sign restrictions VAR approach // Journal of Macroeconomics. 2015. No. 46. Pp. 96–112.
4. Carstensen K., Hulsegger O., Wollmershauser T. Monetary policy transmission and house prices: European cross-country evidence // CESifo Working Paper. 2009.
5. De Stefani A. House price history, biased expectations, and credit cycles: The role of housing investors // Real Estate Economics. 2021. Vol. 49. No. 4. Pp. 1238–1266.
6. Grossman V. et al. Drilling down: the impact of oil price shocks on housing prices // The Energy Journal. 2019. Vol. 40. Special Issue. Pp. 59–84.
7. Hirata H. et al. Global house price fluctuations: Synchronization and determinants // NBER International Seminar on Macroeconomics. 2013. Vol. 9. No. 1. Pp. 119–166.
8. Iacoviello M. House prices and business cycles in Europe: A VAR analysis // Working Papers in Economics. 2002.
9. Killins R.N., Egly P.V., Escobari D. The impact of oil shocks on the housing market: Evidence from Canada and US // Journal of Economics and Business. 2017. Vol. 93. Pp. 15–28.
10. Kuang P. A model of housing and credit cycles with imperfect market knowledge // European Economic Review. 2014. Vol. 70. Pp. 419–437.
11. Madhuri C.H.R., Anuradha G., Pujitha M.V. House price prediction using regression techniques: a comparative study // 2019 International Conference on Smart Structures and Systems (ICSSS) IEEE. 2019. Pp. 1–5.
12. Mohd T. et al. An overview of real estate modelling techniques for house price prediction // Charting a Sustainable Future of ASEAN in Business and Social Sciences. 2020. Pp. 321–338.
13. Nocera A., Roma M. House prices and monetary policy in the euro area: evidence from structural VARs // ECB Discussion Paper. 2017. No. 2073.
14. Runstler G., Vlekke M. Business, housing, and credit cycles // Journal of Applied Econometrics. 2018. Vol. 33. No. 2. Pp. 212–226.
15. Thamarai M., Malarvizhi S.P. House Price Prediction Modeling Using Machine Learning // International Journal of Information Engineering & Electronic Business. 2020. Vol. 12. Pp. 15–20.
16. Wang F. et al. House price prediction approach based on deep learning and ARIMA model // 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT) IEEE. 2019. Pp. 303–307.
17. Wang X., Zhai S., Chen J.L. Research on House Price Forecast Based on Hyper Parameter Optimization Gradient Boosting Regression Model // 2020 8th International Conference on Orange Technology (ICOT) IEEE. 2020. Pp. 1–6.

Forecasting the Real Estate Price Index in Russia

Natalia S. Nikitina – Junior Researcher of the Russian Presidential Academy of National Economy and Public Administration (Moscow, Russia). E-mail: nikitina-ns@ranepa.ru

This article is devoted to choosing the best model for short-term forecasting of Russia's real estate price index. Popular machine learning methods: Ridge and Lasso regressions, Elastic Net regression and methods of working with time series were considered: Naive, Exponential smoothing, ARIMA, OLS. The set of variables includes the values of GDP, inflation, effective exchange rate, interbank lending rates, and oil prices. Machine learning methods – Ridge Regression and Elastic Net regression – show the high quality of forecasting the real estate price index compared to standard ways of working with time series – Naive, Exponential smoothing, ARIMA.

Key words: forecasting, real estate price index, machine learning.

JEL: C32, C53, R30.