International Journal of Ecosystems and Ecology Science (IJEES) **Vol. 14 (4):** 31-40 (2024)
https://doi.org/10.31407/ijees https://doi.org/10.31407/ijees14.4

PREDICTIVE MODELING OF HOUSING CONSTRUCTION IN RUSSIA: INSIGHTS FROM ARIMA AND LOCAL REGRESSION ANALYSIS

Natalia A. Sadovnikova[1*], Olga G. Lebedinskaya[1], Alexander V. Bezrukov[1], Galina L. Popova[1], Elvira A. Yarnykh[1]

[1]Plekhanov Russian University of Economics, 117997, Moscow, Stremyannyi Alleyway, 36, Russia;*

[*]Corresponding Author Natalia A. Sadovnikova, e-mail: Sadovnikova.NA@rea.ru;

Received September 2024; Accepted October 2024; Published November 2024;

DOI: https://doi.org/10.31407/ijees14.405

ABSTRACT

The purpose of this work is to identify the fundamental factors influencing the dynamics of residential real estate construction in order to further predict the population's housing supply. The problem that analysts face is the relatively short time series with annual data characterizing this market. Earlier methodological approaches to forecasting short-term series suggested the use of exponential moving averages, Holt-Winters modeling, ARIMA, LSTM neural networks, LOESS and others, mainly in Europe and the USA. In this study, the authors present the results of forecasting the dynamics of residential real estate construction in Russia using the ARIMA model, as well as a variety of series decomposition (STL) methods and taking into account all the rules of the Holt-Winters model.

Keywords: Housing market, regression analysis, short series, indicator system.

INTRODUCTION

The problem of modeling the dynamics of housing construction volume is the subject of many studies. The purpose of all-time series forecasting methods is to build a process model, and then use this model on the latest values of the time series to extrapolate past behavior into the future. This approach works well for long time series, where the amount of available data allows you to build forecasting methods that use hundreds or even thousands of data points to build a mathematical (usually approximating) model of the process. But the situation is completely different when the available amount of data is small, which is typical for evaluating the results of only implemented government programs, including in housing construction. To solve the problem of such forecasting, the authors proposed at the first stage to investigate the structure of the series in terms of autocorrelation and partial autocorrelation, at the second stage three models were proposed for comparison: ARIMA, adaptive Holt-Winters, decomposition with local weighted regression based on ARIMA

The content side of forecasting the dynamics and structure of construction is related to the demographic, socio-economic, industrial and technological components of the region's development, namely: the number (an integral indicator of natural population growth (loss) and migration intensity) and the family structure of the region's population, income, demanded living comfort and investment potential of the region for social construction (public housing programs), as well as the specifics of the design solutions of mass residential buildings and industrial facilities. In addition to the problem of selecting correlating factors, there are certain features in the formation of forecasting methods. Let's list the main problems of forecasting short time series, including the volume of housing construction:

- The inability to identify long-term trends in the development of the process based only on past observations.
- Unreliability of estimates of the parameters of the series model.
- The impossibility of using traditional statistical methods to assess the accuracy of the forecast (determination of confidence intervals, hypothesis testing).
- Loss of the meaning of a detailed statistical analysis of the remnants of the model and, consequently, the inapplicability of the adequacy criteria.
- The inefficiency of complex models for describing trends, since their evaluation requires large samples, can be limited to linear or quadratic trends.

Short-term time series forecasting methods are based on various smoothing techniques, such as a moving average filter and exponential smoothing [Christiaanse, 1971], methods based on the use of an ARPSS-type model [Kim, Sin, 2007], artificial intelligence methods such as HC (artificial neural network) [Casolari, Colajanni, 2007]. The unsteadiness of many short BP, compared with socio-economic indicators, leads to the impossibility of obtaining adequate predictive solutions using classical models and algorithms

Usually, an essential part of the task of forecasting the volume of housing construction in the Russian Federation is the problem of modeling "short" time series based on annual data. As a result, forecasting such data excludes the possibility of detecting and analyzing short-term patterns. Approaches to predicting short series of dynamics usually include exponential moving average models, Holt-Winters models, autoregressive moving average models of the ARSS/ARISS families, neural networks of direct communication, local weighted regression, singular spectrum decomposition, the method of local empirical modes, classical 4-component decomposition and seasonal locally weighted decomposition. A comparative assessment of the simulation results showed that when predicting short time series for housing construction indicators in the Russian Federation, it is most advisable to use the STL model, preceded by an analysis of the autoregressive structure of the series.
Also, the formation of the forecasting strategy was influenced by the fact that when predicting short series of dynamics, criteria for the quality of model fitting, such as the Akaike criterion and the Schwartz criterion, acquire special importance.

According to the authors, the Akaike criterion has the informative value of using predictive models to assess the quality in short time series, since cross-validation and creation of test data are difficult in short time series. The Akaike criterion, at the same time, is an analogue of the estimate of the average square error "out of sample" for one step of the forecast. As a result, choosing a model with a minimum akaike criterion allows you to take into account both the number of model parameters and the amount of "random noise" in short time series.
It should be borne in mind, however, that the lowest values of the Akaike criterion correspond to overly simplified models, since, for example, due to an error in fitting an autoregressive integrated moving average model with a number of parameters greater than 1-2, they will give significantly incorrect forecasts.

MATERIAL AND METHOD

Methodology

Part of the essence of the problem of forecasting living space construction in Russian Federation is the relatively short time series with annual data. Therefore, forecasting such data nigh-eliminates detection and, consequently, accounting for short-term regularities.
In this study, the methodological approaches that have been tested are the ARIMA modelling, the Holt-Winters modelling and the STL modelling.
The forecasting of Russian Federation living space construction was performed upon four models: ARIMA decomposition, Holt-Winters with and without damping of the trend component, and Seasonal-Trend LOESS decomposition with ARIMA forecasting. The authors propose that the Seasonal-Trend LOESS decomposition model

with ARIMA forecasting is the most optimal suggestion for forecasting annual data on very short time series. Sources conclude that short-term time series can be optimally fitted with an ARIMA process. Further observation of the apparent time series structure suggests that annual living space construction data correspond to autoregressive integrated moving average process, as their differences exhibit structure in time-scale dependency for both autocorrelation and partial autocorrelation. However, because the living space construction is closely related with economic movements and conditions, the authors applied the Seasonal-Trend LOESS decomposition to extract and forecast the cyclical component upon locally weighted regression.

The article proposes the following recursive strategy for forecasting short time series in the field of housing construction in the Russian Federation:

- Analyzing the time series structure for unit root and autocorrelation;
- Analyzing the differenced time series and testing for unit root;
- Proposing an ARIMA (p, d, q) model based upon the analysis of the time series structure;
- Fitting the model and obtaining the forecast.

The parameters of the ARIMA model are determined by studying autocorrelation functions. To find the parameter q, we analyze the values of the function. It can be noted that when changing the order of autocorrelation, the value of the function goes beyond the confidence limits with a lag equal to 1.
The forecasting of housing construction in the Russian Federation has been carried out according to the following indicators:

X1 - Commissioning of residential buildings in the Russian Federation (millions of square meters of total residential area)
X2 - Commissioning of residential buildings in urban areas in the Russian Federation, (millions of square meters of total residential area)
X3 - Commissioning of residential buildings in rural areas in the Russian Federation,(millions of square meters of the total area of residential premises)
X4 - The number of apartments built in the Russian Federation, thousand
X5 - The average actual cost of construction per square meter of the total area of residential premises in commissioned residential buildings in the Russian Federation, rubles (apartment-type residential buildings without extensions, superstructures and built-in premises and without residential buildings built by the population).

Table 1. Dynamics of housing market indicators.

Times	X1	X2	X3	X4	X5
2000	30,3	23,1	7,2	373	4779
2001	31,7	24,3	7,4	382	7244
2002	33,8	26,2	7,6	396	9025
2003	36,4	28,3	8,1	427	10037
2004	41	32,3	8,7	477	11720
2005	43,6	34,1	9,5	515	13812
2006	50,6	40,6	10	609	16840
2007	61,2	47,5	13,7	722	20720
2008	64,1	49	15,1	768	26622
2009	59,9	43,8	16,1	702	30312
2010	58,4	43,7	14,7	717	31877
2011	62,3	46,8	15,5	786	33320
2012	65,7	50	15,7	838	34354
2013	70,5	53	17,5	929	36439
2014	84,2	62,2	22	1124	39447
2015	85,3	62	23,3	1195	39258
2016	80,2	58,8	21,4	1167	40890
2017	79,2	57,2	22	1139	41459
2018	75,7	54,7	21	1076	41358
2019	82	56,7	24,5	1120	42551
2020	82,2	57,5	24,7	1122	44518
2021	92,6	65,6	27	1205	49200
2022	102,7	69,7	33	1290	53200
2023	110,4	73,8	36,6	1449	54050

The forecast of housing construction dynamics was made using the adaptive Holt-Winters model and its modification. This method is also known as linear exponential smoothing, the use of which is advisable if there is a trend in the levels of the series. The model "belatedly" reflects changes in empirical data, while the forecast characteristics are extremely overestimated (according to it, by 2033 the volume of housing construction will more than double compared to 2023). Therefore, it cannot be argued that the method qualitatively approximates the data.

RESULT

The methodology proposed by the authors allowed us to construct forecast values for five indicators that form the conditions of housing construction in Russia: commissioning of residential buildings in the Russian Federation; commissioning of residential buildings in urban areas in the Russian Federation; commissioning of residential buildings in rural areas in the Russian Federation; the number of apartments built in the Russian Federation; the average actual cost of construction per square meter of the total area of residential premises in commissioned residential buildings in the Russian Federation (apartment-type residential buildings without extensions, superstructures and built-in premises and without residential buildings built by the population). The article presents step-by-step results of modeling the dynamics of commissioning residential buildings in the Russian Federation (millions of square meters of total living space) (X1). The simulation results for indicators X2-X4 are presented in Table 3 and Figures 7-11. The series is non-stationary, has a trend and other components (Dickey-Fuller statistics is 0.662, p-value is 0.989).

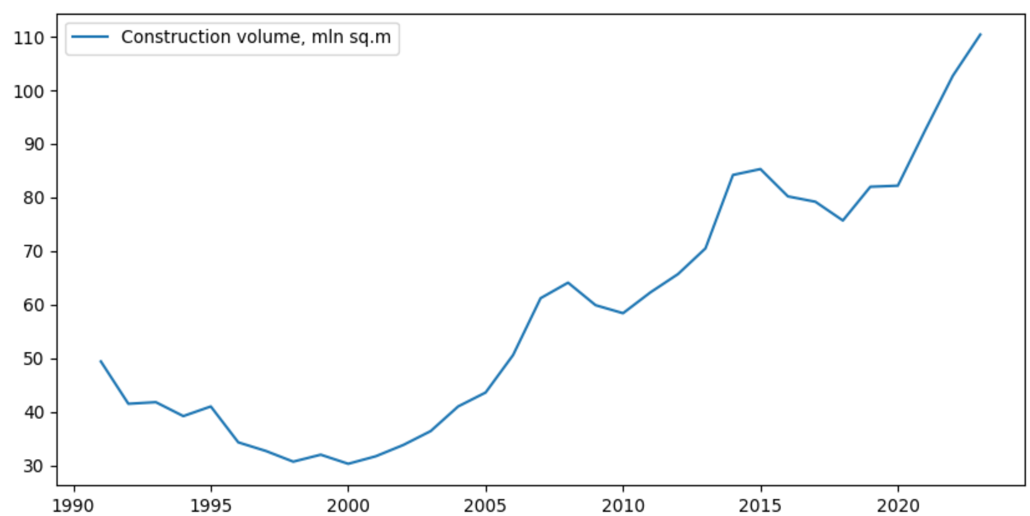
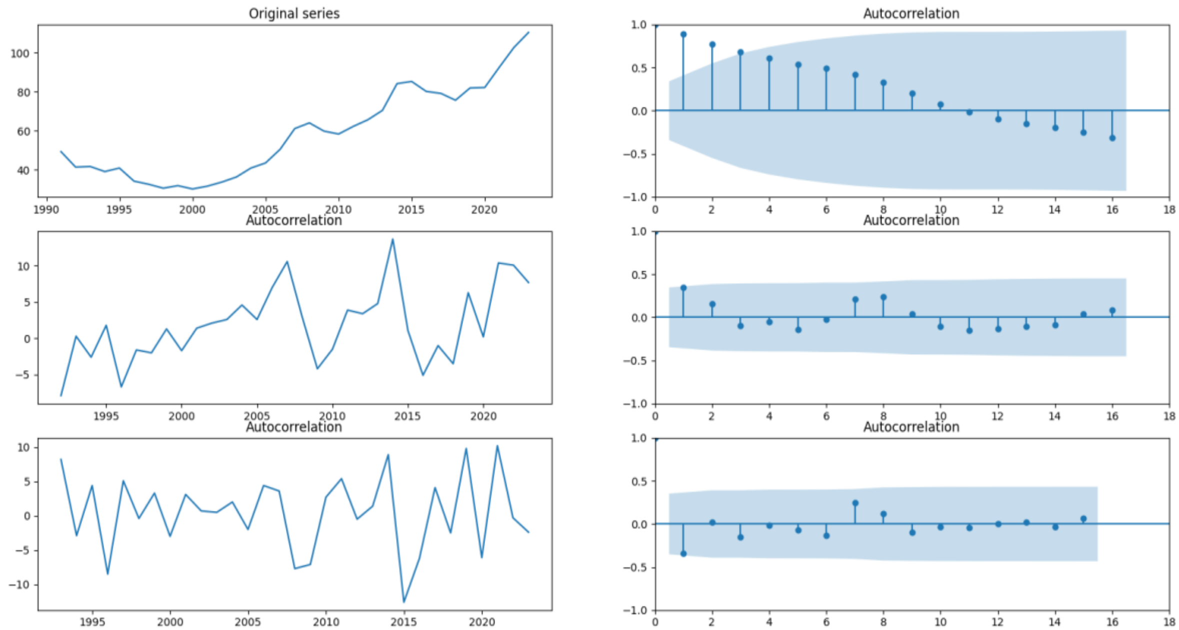


Figure 1. Dynamics of commissioning of residential buildings in the Russian Federation.

The graph of autocorrelation coefficients with 33 lags (since the length of the time series is 33 levels) indicates that the degree of interconnection of the values of the series is gradually decreasing. The highest dependence is observed until about the fifth lag, then it begins to decrease. Given that the chart is moving towards zero relatively slowly, it can be assumed that there is a trend in the series.

High autocorrelation of the initial time series suggests the presence of a single root, which is confirmed by the supplemented Dickey-Fuller test (ADF statistics: 0.662298, p-value: 0.989055). Autocorrelation remains positive up to 11 years and exceeds 0.5 with a 6-year lag and remains statistically significant ($p < 0.05$) with a 7-year lag.

When changing the order of autocorrelation, the value of the function goes beyond the confidence limits with a lag equal to 1. Since the relationship of neighboring values of the time series level is the most significant, this lag was used in the construction of models. High autocorrelation of the initial time series suggests the presence of a single root, which is confirmed by the supplemented Dickey-Fuller test (ADF statistics: 0.662298, p-value: 0.989055). Autocorrelation remains positive up to 11 years and exceeds 0.5 with a 6-year lag and remains statistically significant ($p < 0.05$) with a 7-year lag.



Graphic 1. Autocorrelation.

The matrix of multidimensional trajectory vectors is decomposed by the singular value decomposition (SVD) method. It is believed that the first 11-12 components of the time series make the greatest contribution, but in our case, only the first 4-5 components have the greatest information content.

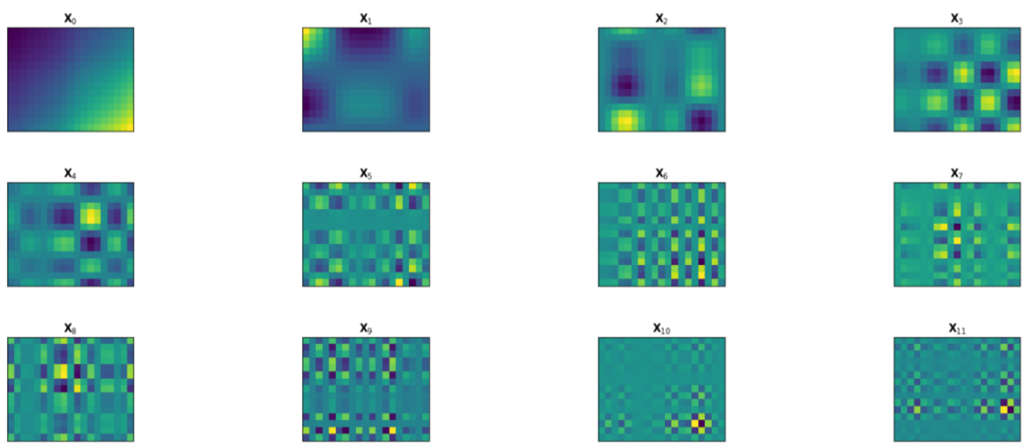


Figure 2. Trajectory matrix (components).

Information content ceases to increase significantly after the inclusion of 1-2 components, therefore, such a number is usually sufficient.
However, we can note that, although the results of SVD decomposition indicate the informativeness of the main components of the initial series, the W-correlation matrix contains only strongly correlated components, and therefore we conclude that for short time series of residential construction, SVD analysis for the purpose of modeling and forecasting may be ineffective.

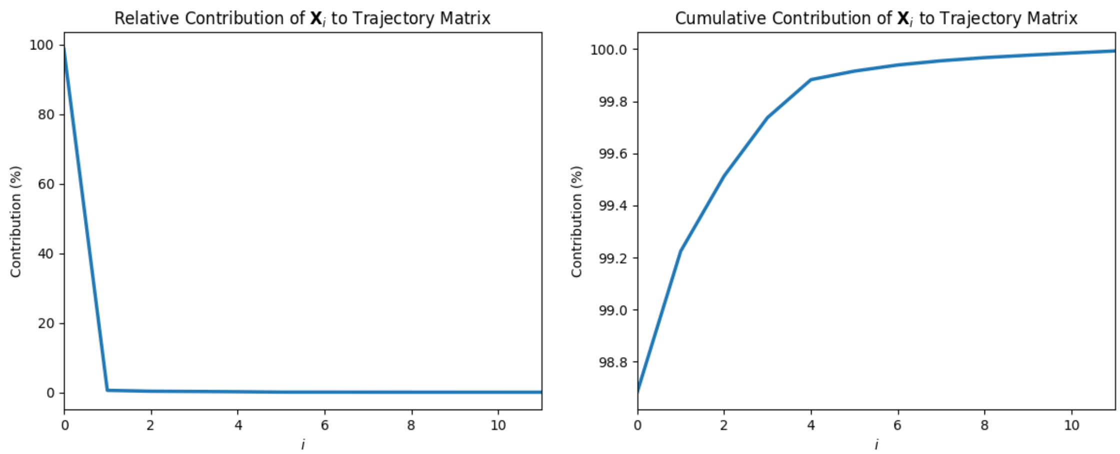


Figure 3. Relative and contribution of X1 to Trajectory matrix.

When differentiating the series, a lag of 6 was used to give it stationarity by two orders of magnitude, since the value of the partial autocorrelation function with it goes beyond the confidence limits. To bring the series to a stationary form, second-order differentiation was used.

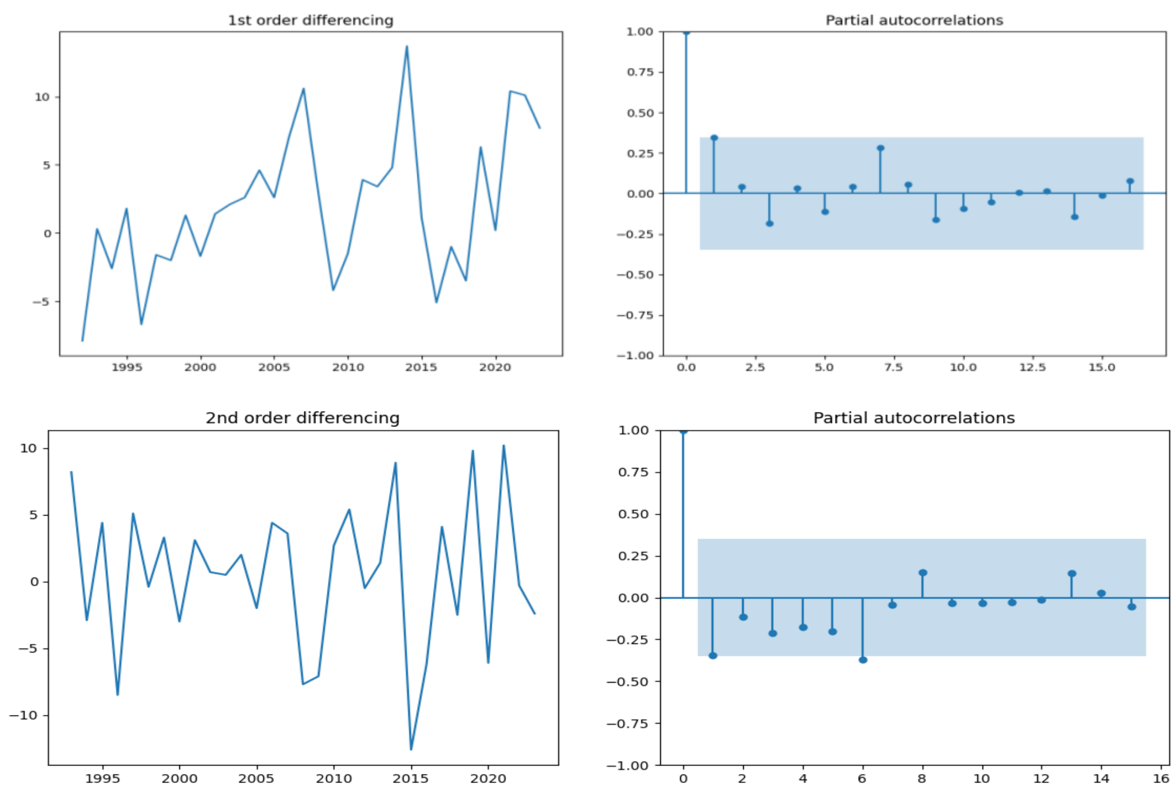


Figure 4. 1 and 2nd order differencing.

The analysis of partial autocorrelations suggests structure for 2nd order differencing, with lags 1 and 6 being statistically significant and lags 2-5 displaying some structural time-scale dependency. ARIMA (6,2,1) model was proposed to perform the autoregressive decomposition and then, subsequently, to obtain the forecast upon the STL decomposition. The forecasted values are presented on the graph.

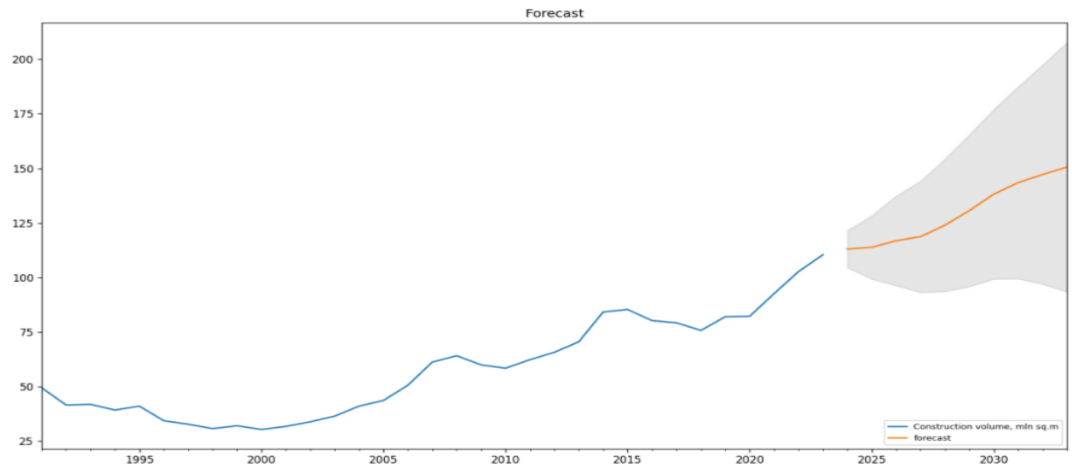


Figure 5. Forecast.

The "auto-tuning" of the model according to the Akaika criterion gave the best results: For the ARIMA model (6, 2, 1), the Akaika criterion is 187.49, and for the ARIMA model (0, 1, 1) 197.09. Therefore, the model with manually configured parameters shows itself better.

The analysis of the residuals for the constructed model showed that the error distribution is quite close to normal, and the autocorrelation in the residuals is extremely low, therefore, there is no dependence in the residuals, they are random.

The model diagnostics suggest that there is no autocorrelation in residuals and their distribution corresponds to Gaussian structure. According to formal criteria, the model describes empirical data on the volume of housing construction well.

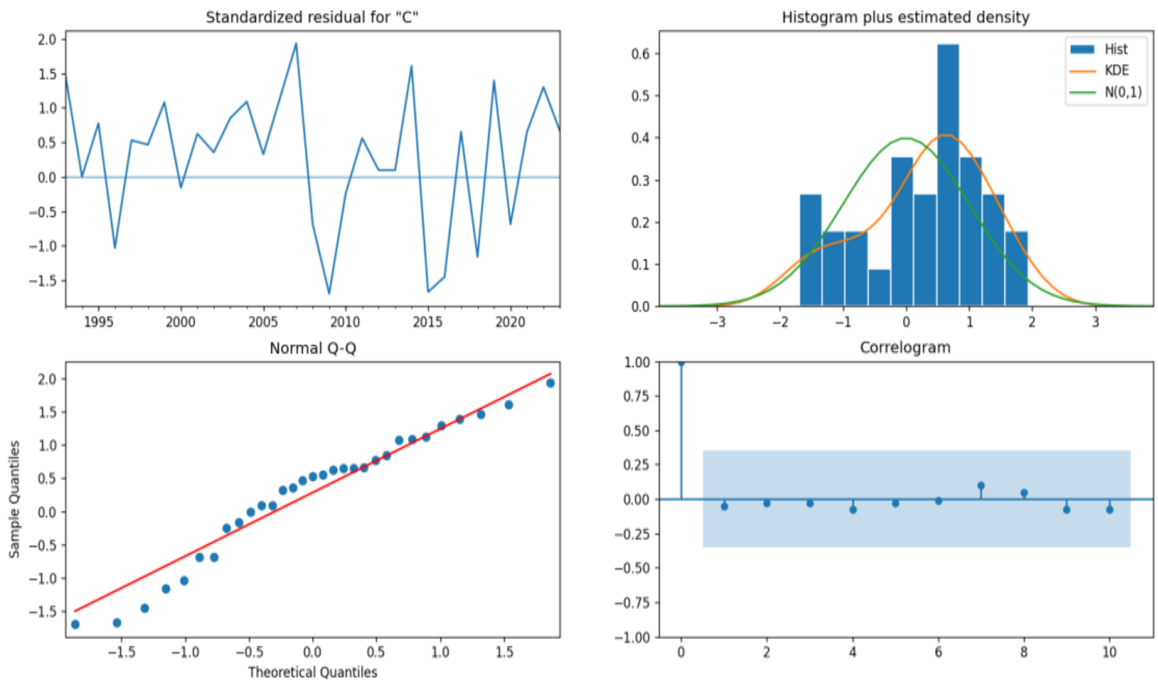


Figure 6. Error distribution.

Similarly, the results of the autoregressive structure for other factors were obtained

Table 2. Results of the analysis of the autoregressive structure of the series.

Indicator	Evaluation
X1 - Commissioning of residential buildings in the Russian Federation (millions of square meters of total living space)	ADF-statistic: 0.662298 p-value: 0.989055
X2 - Commissioning of residential buildings in urban areas in the Russian Federation, (millions of square meters of total living space)	ADF-statistic: 0.057288 p-value: 0.963021
X3 - Commissioning of residential buildings in rural areas in the Russian Federation, (millions of square meters of total living space).	ADF-statistic: 2.023878 p-value: 0.998703
X4 - The number of apartments built in the Russian Federation, thousand	ADF-statistic: -0.886241 p-value: 0.792470
X5 is the average actual cost of construction of one square meter of the total area of residential premises in commissioned residential buildings in the Russian Federation, rubles (apartment-type residential buildings without extensions, superstructures and built-in premises and without residential buildings built by the population).	ADF-statistic: -0.982822 p-value: 0.759478

Taking into account the above, the following forecast values were obtained.

Table 3 – the results of forecasting the studied indicators using the STL method.

Indicator	2024	2025	2026	2027	2028	2029	2030
X1	113,01	113,81	116,83	118,70	123,98	130,67	138,20
X2	75,04	79,16	80,40	84,52	85,76	89,88	91,12
X3	41,31	45,16	49,83	53,69	58,36	62,21	66,89
X4	1376,30	1409,76	1433,65	1504,00	1422,92	1460,36	1486,04
X5	55277,65	58204,92	58301,22	60705,02	60368,47	62414,37	61781,88

The results obtained indicate a generally positive dynamics of all the considered indicators characterizing the housing construction market.

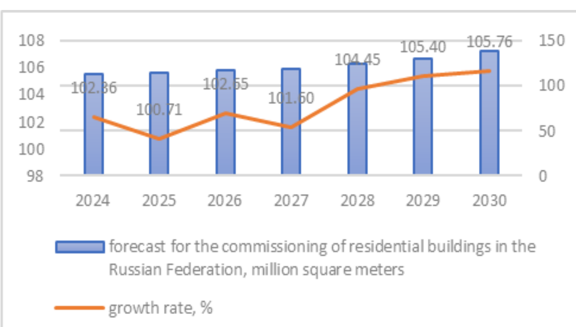


Figure 7 - forecast for the commissioning of residential buildings in the Russian Federation

Figure 8 - forecast for the commissioning of residential buildings in urban areas in the Russian Federation

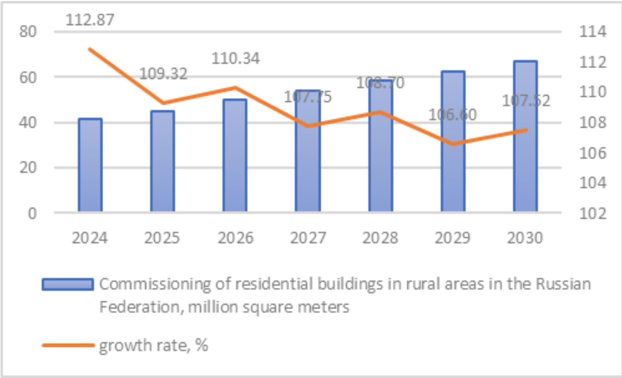


Figure 9 - Commissioning of residential buildings in rural areas in the Russian Federation

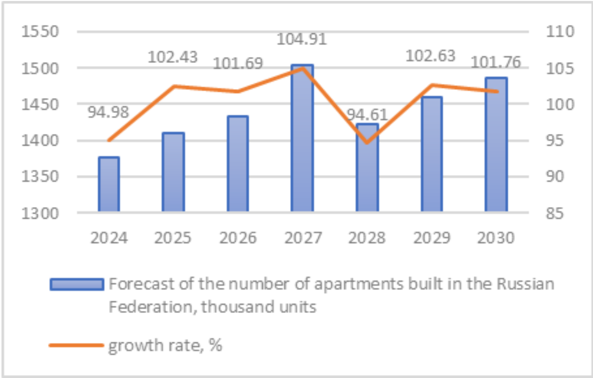


Figure 10 - Forecast of the number of apartments built in the Russian Federation

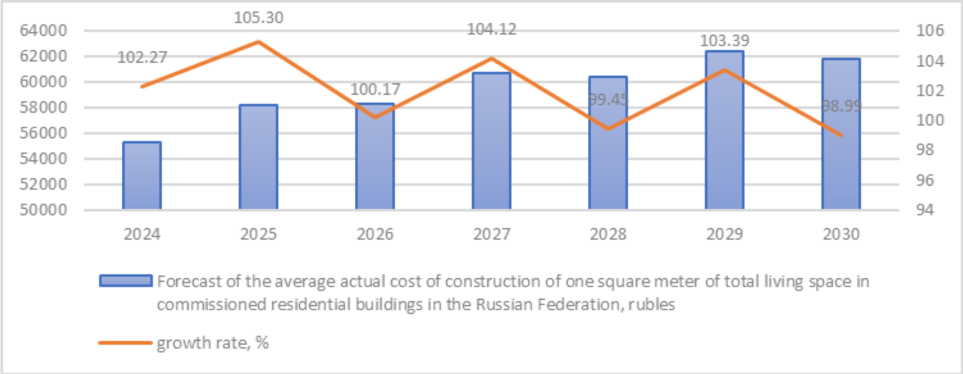


Figure 11. Forecast of the average actual cost of construction of one square meter of total living space in commissioned residential buildings in the Russia.

CONCLUSION

The article presents the results of forecasting the volume of housing construction in Russia, taking into account the characteristics of short rows. The most accurate results in predicting the volume of housing construction are demonstrated by the ARIMA model (6, 2, 1) and local regression based on the decomposition of a time series. According to the models obtained, the value of all indicators is growing, but the growth rates vary. Thus, according to the developed model, the commissioning of residential buildings in the Russian Federation will grow to 138.2 million square meters of the total area of residential premises, the commissioning of residential buildings in urban areas in the Russian Federation – up to 91.12 million square meters. The commissioning of residential buildings in rural areas in the Russian Federation from 2024 to 2030 will grow to 66.89 million square meters of the total area of residential premises). However, the growth rate of the indicator for the forecast period will slow down year by year. This is due to mass urbanization.
The number of apartments built will grow no less significantly (from 1,378 to 1,486 thousand units and, unfortunately, the average actual cost of building one square meter of the total area of residential premises – from 55.3 to 61.8 thousand rubles, which is associated with both demand and supply inflation.

Acknowledgements. This research was carried out within the framework of an internal grant from the Plekhanov Russian University of Economics.

REFERENCES

1. Christiaanse, W. R. (1971). Short term load forecasting using general exponential smoothing. IEEE Transactions on Power Apparatus and Systems, 90(2), 900–911, https://doi.org/10.1109/TPAS.1971.293572;
2. Casolari, S., & Colajanni, M. (2009). Short-term prediction models for server management in Internet-based contexts. Decision Support Systems, 48(1), 212–223. https://doi.org/10.1016/j.dss.2009.05.001;
3. Kim, H., & Sin, H. K. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. Applied Soft Computing, 7(2), 569–576. https://doi.org/10.1016/j.asoc.2006.02.005;
4. Aieb, A., Liotta, A., Jacob, A., & Yaqub, M. A. (2024). Short-term forecasting of non-stationary time series. Engineering Proceedings, 68, 34. https://doi.org/10.3390/engproc2024068034;
5. Fung, K. F., Huang, Y. F., Koo, C. H., & Soh, Y. W. (2020). Drought forecasting: A review of modelling approaches (2007–2017). Journal of Water and Climate Change, 11(3), 771–799, https://doi.org/10.2166/wcc.2019.236;