TIME SERIES PREDICTION FOR THE HOUSING CONSTRUCTION MARKET WITH THE USE OF NARNET

Daria Wotzka
Opole University of Technology
Faculty of Electrical Engineering
ul. Prószkowska 76
45-758 Opole, Poland
E-mail: d.wotzka@po.edu.pl

Grażyna Suchacka
University of Opole
Institute of Informatics
ul. Oleska 48
45-052 Opole, Poland
E-mail: g.suchacka@gmail.com

Łukasz Mach
Opole University of Technology
Faculty of Economics and Management
ul. Prószkowska 76
45-758 Opole, Poland
E-mail: l.mach@po.edu.pl

Paweł Frącz
University of Opole
Faculty of Economics
ul. Oleska 48
45-040 Opole, Poland
E-mail: p.fracz@gmail.com

Joachim Foltys
Humanitas University
in Sosnwiec
ul. Kilińskiego 43
41-200 Sosnowiec, Poland
E-mail: joachimfol@onet.pl

Ionela Maniu
Lucian Blaga University of Sibiu
Dept. of Mathematics and Informatics
str. Dr. Ioan Ratiu, 550012
Sibiu, Romania
E-mail: ionela.maniu@ulbsibiu.ro

KEYWORDS

Nonlinear Autoregressive Neural Network, NARNet, NARNN, time series prediction, machine learning, real estate forecast

ABSTRACT

The paper discusses the study to develop and tune parameters of a nonlinear autoregressive neural network (NARNet or NARNN) model for predicting the number of housing units. Predictions were made for the housing construction market in Poland, which is a dynamically growing European market. Three stages of the housing construction process have been taken into consideration: permits issued for house construction, houses under construction, and completed new houses. Experimental results have shown that a NARNet model can be a very effective tool in the considered scenario. A network model using the Levenberg-Marquardt backpropagation training function achieved the best model fit, as well as the most accurate one-month predictions.

INTRODUCTION

According to a report by real estate consultancy JLL, the real estate market in Poland has been in a good shape and the number of dwelling offers on the primary and secondary markets have been increasing (JLL 2021). About 88,000 new apartments were available on the primary market in Poland. Moreover, the number of new apartments offered by developers has continued to grow and apartment prices also has increased by about 5-10%. As for the secondary market, the number of listings was about 172,000 at the same time, and most of them were from individual investors. Prices of apartments on the secondary market varied depending on the location and condition of the apartments. It should be noted, however, that the real estate market situation in Poland is changing dynamically and can vary from region to region.

The housing construction market is characterized by high seasonality, determined, among other things, by the seasons of the year and related weather conditions (Frącz et al. 2023; Mach et al. 2021). At the same time, there is a great need to predict the actual situation in this market and to make predictions over different time horizons for its key components. Thus, it is crucial to utilize effective predictive tools and constantly enhance existing ones, considering the unique features of the real estate market and the recurring emergence of abnormal situations, such as pandemics or armed conflicts (Hozer et al. 2019; Kokot 2022; Mach 2019; Mach et al. 2020).

The real estate market has seen a successful implementation of predictive techniques utilizing artificial intelligence, including neural networks and linear regression (Huang et al. 2011; Lin et al. 2021).

The housing construction market is known for its heterogeneity and low flexibility, which pose challenges to its size prediction. Conventional methods of time series analysis often fail to provide acceptable results in this area, leading researchers to explore advanced techniques, such as those based on artificial intelligence. This study supports the application of NARNet method for the housing construction market forecasts, demonstrating its utility in this context. By leveraging the capabilities of NARNet, researchers can better address the complex nonlinear relationships and patterns

Communications of the ECMS, Volume 37, Issue 1,
Proceedings, ©ECMS Enrico Vicario, Romeo Bandinelli,
Virginia Fani, Michele Mastroianni (Editors) 2023
ISBN: 978-3-937436-80-7/978-3-937436-79-1 (CD) ISSN 2522-2414

inherent in the housing construction market, improving the accuracy of forecasting models. These findings may have important implications for stakeholders in the housing construction industry, enabling them to make more informed decisions and mitigate risks associated with market fluctuations.

In this paper, we consider three main stages of the housing construction process, which translate into three components of the housing construction market:

- number of permits issued for house construction,
- number of houses under construction,
- number of new houses that have been put into use.

These components, along with sales of construction and assembly products, are key determinants of the housing market with a direct impact on it.

Motivated by recent successful results of time series forecasting with a nonlinear autoregressive neural network, we develop and test a NARNet model for the housing construction market. A research hypothesis investigated in the paper assumes that a nonlinear autoregressive neural network can be a reliable tool to build forecasts for the three main components of the housing construction market.

PRELIMINARIES

A *time series* is a sequence of discrete-time data, i.e., a sequence of values measured in time.

A *univariate time series* is a series consisting of a single-feature (scalar) observation with equally spaced values (taken at successive equally spaced points in time, e.g., every month).

The *nonlinear autoregressive neural network* (*NARNet* or *NARNN*) model can be expressed as

$$y_t = f(y_{t-1}, \ldots, y_{t-d}),$$

where y is a variable to be forecasted, t is the time index, d is the number of delays, and f is the function. A big advantage of the NARNet model, besides its high forecasting efficiency, is that it uses its own universal predicting algorithm without the need for a mathematical model of a process being modeled. Thus, function f is unknown in advance and is estimated with the use of a neural network (Xu and Zhang 2022). The model is autoregressive so it is able to capture hidden patterns in data-driven predictions by using historical time series data to predict future values in a time series (Adedeji et al. 2019; Padilla et al. 2021).

RESEARCH METHODOLOGY

The methodology applied in the study included the following steps: data preparation, developing a NARNet model, running experiments with parameter tuning, and evaluating prediction performance.

Data Preparation

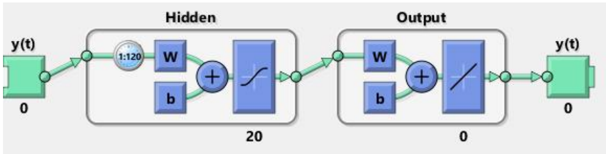
The input for a model has a form of a time series, containing natural numbers. We consider univariate time series, where each element of the series represents the number of houses (permitted, initiated, or completed ones, depending on a construction stage under consideration) in the one-month interval.

The original series have been normalized with the use of z-score standardization.

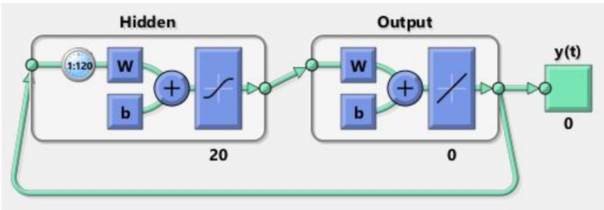
Developing a NARNet Model

A NARNet model was developed using Deep Learning Toolbox available in MATLAB (Narnet 2023). Two types of networks were used in the study to train the models: an open loop network (Fig. 1a) and a closed loop network, in which the feedback input is replaced with a direct connection from the output layer (Fig 1b). Forecasts were made using open-loop networks with one-step-ahead prediction (Fig. 1c).

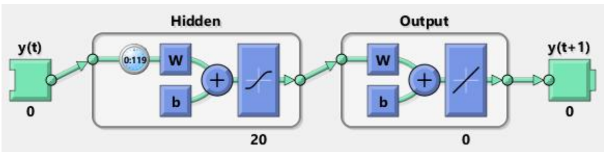
During the model tuning, a whole time series is randomly divided into training, validation, and testing data, with multiple repetitions. The following data ratios were applied: a training data ratio of 0.7, a validation data ratio of 0.15, and a testing data ratio of 0.15.



(a) Open Loop Network



(b) Closed Loop Network



(c) Open Loop Network with One Step Ahead Prediction

Figure 1: Block Diagram of the NARNet Models Used

Four various functions have been used for neural network training. These functions differ in the way in which weight and bias values are updated during the training:

- *trainscg* – scaled conjugate gradient backpropagation,
- *trainoss* – one step secant backpropagation,
- *traincgb* – conjugate gradient backpropagation with Powell-Beale restarts,
- *trainlm* – Levenberg-Marquardt backpropagation.

For all four functions, the training procedure stopped when any of the following conditions occurred: the maximum number of epochs was reached, the maximum training time was exceeded, the minimum performance or the minimum performance gradient was reached, or the maximum number of validation performance increases after its last decrease was exceeded. An additional stop condition for *trainlm* function was when an adaptive μ value surpassed the assumed maximum.

The following parameter values were applied during model training:

- number of hidden layers: 1,
- the hidden layer size: 20,
- number of epochs: 1000,
- performance goal: 1e-6,
- maximum validation checks: 50,
- feedback delays: 1-120,
- feedback type: open/closed loop,
- prediction: closed loop (Fig. 1b); one step ahead open loop (Fig. 1c).

Prediction Performance Measures

To assess the quality of the prediction model, two measures have been used: *MSE* (*Mean Squared Error*) and the regression coefficient.

MSE is calculated based on differences between actual and predicted values of a time series:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2,$$

where n is the number of time series values (elements), Y and $\widehat{Y}$ are vectors of observed and predicted values of the time series, respectively.

The *regression coefficient* describes the average functional relationship between actual and predicted time series values Y and $\widehat{Y}$. It is calculated by minimizing the sum of squared residuals of a linear model:

$$R = \frac{covariance(Y,\widehat{Y})}{variance(Y)}.$$

RESULTS AND DISCUSSION

Data Description

The data used in the study refer to the Polish housing construction market (GUS 2022). They had been collected over the course of 206 months (more than 17 years) in the period from May 2005 to June 2022.

Two types of dwellings were taken into account: houses constructed by individual investors for their own needs and apartments raised by developers for sale or rent. Given the marginal quantitative importance of municipal, cooperative, social rental, and company objects, these premises were not taken into account in this study.

The data used correspond to three stages of the housing construction process:

(a) "***permits***" – number of permits issued for house construction,

(b) "***initiated***" – number of houses under construction,

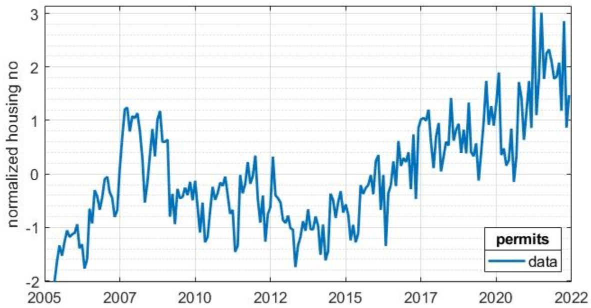
(c) "***completed***" – number of new houses that have been put into use.

The data were elaborated as monthly time series, with successive elements of a series denoting the corresponding numbers of houses (permitted, initiated, or completed ones, depending on a construction stage) in successive time slots (1-month intervals).

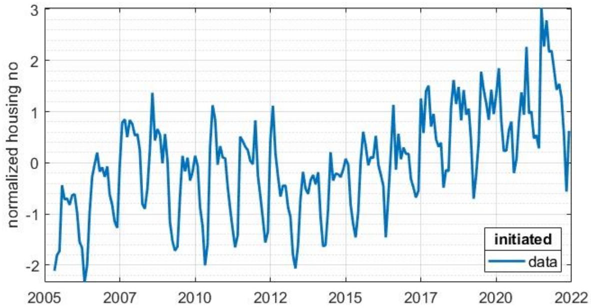
The three original time series are visualized in Fig 2. It can be seen that the number of permits for the beginning of house construction is characterized by cyclicality and has a trend. The analysis of these characteristics of the waveforms was not considered in the present work due to the space limit. Instead, the purpose of the research was to develop a model for data prediction without the need for differentiation and detrending the processes (as in ARIMA-type models).

Analysis Results

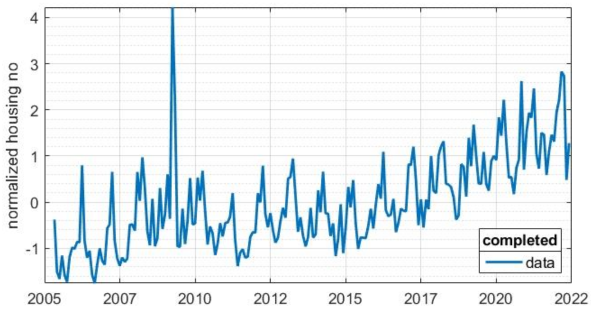
After building NARNet models for the three-time series (Fig. 3 – Fig. 6), errors in the models have been assessed.



(a) House Construction Permits



(b) Houses Under Construction

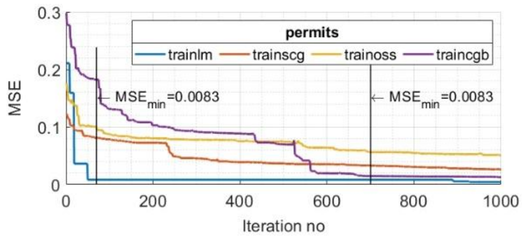


(c) Completed Houses

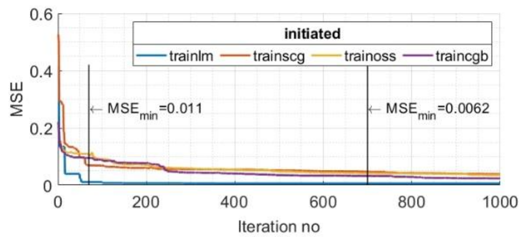
Figure 2: Visualization of the Original Time Series

First, the model fit was investigated for various numbers of iterations, ranging from 1 to 1000. Fig. 3 presents values of *MSE* obtained for various numbers of iterations during the model training for the four training functions.

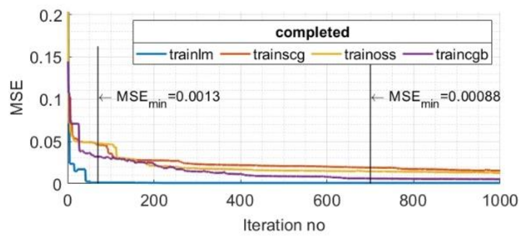
.



(a) House Construction Permits



(b) Houses Under Construction



(c) Completed Houses

Figure 3: *MSE* Values for Various Training Functions Depending on the Number of Iterations

As it can be seen, the model performance depends on the training function applied: these differences are greater for *permits* model (Fig. 3a) and smaller for *initiated* (Fig. 3b) and *completed* (Fig. 3c) models. In all three cases, a clear outperformance of *trainml* training function may be observed (a blue line)

Furthermore, the quality of the models depends on the number of iterations applied during the model training: the higher the number of iterations is, the lower *MSE* values are obtained. In Fig. 3 minimal *MSE* values have been highlighted for the 70th iteration, when *MSE* levels start to stabilize, as well as for the 700th iteration, when *MSE* levels are just fairly stabilized. In all the cases, the minimum *MSE* values were achieved for *trainml* function. It can be also observed that for the house construction permits (Fig. 3a), the *MSE* stabilizes only relatively late, after about the 600th iteration. For both remaining series, related to houses under construction and completed houses, the *MSE* stabilizes much earlier, after only about 100 iterations.

To sum up this part of the analysis, the best results were obtained for the *trainml* model. The results in Fig. 3 confirm that a sufficiently large number of iterations is required to properly train the NARNet model.

Regardless of the aforementioned *MSE* stabilization results for individual numbers of iterations of the learning process, in further parts of the paper neural network models trained after 1000 iterations have been analyzed and discussed.

An example model response for the initiated number of houses, with the use of *trainlm* function, after 1000 iterations in the course of 85 consecutive months, is shown in Fig. 4.

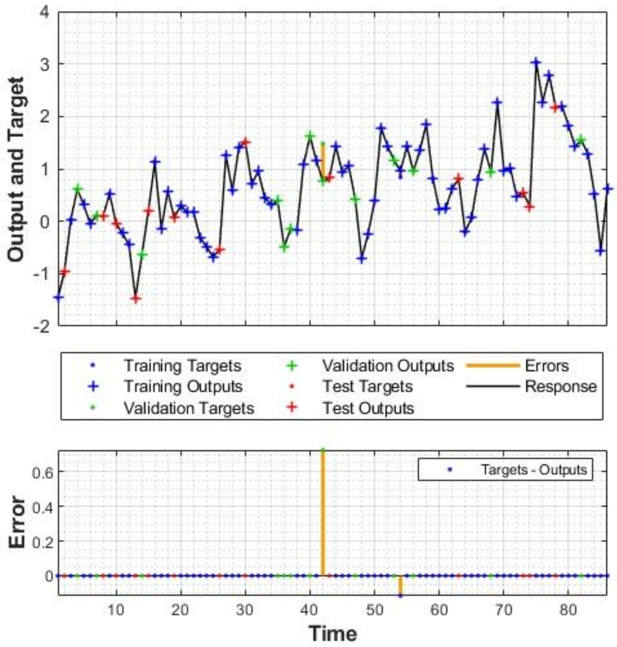


Figure 4: Example Model Response for the Number of Initiated Houses (*trainlm* Function, 1000 Iterations of the Model Training)

The presented charts show the target and the output housing number values of the considered data sets for the following experiment phases: learning (blue dots and pluses), validation (green dots and pluses), and testing (red dots and pluses), as well as the values of the residual error, marked with yellow bars. In the given example it can be observed that in most cases the points overlap, representing a well-chosen and well-trained model of the neural network.

Fig. 5 shows an example scatter plot of the model output and target for the houses under construction, with the use of *trainlm* function, after 1000 iterations of training. The chart shows four graphs for the three datasets: the learning, validation, and test datasets, as well as for the total dataset. The circles represent the samples in each case and the solid line indicates the regression model approximated by a linear function, whose parameters are given in each case on the y-axis. In most cases, a very good fit of the NARNet model to the empirical data was obtained. It has been observed that an outlier has been included in the validation data set, which has resulted in an underestimation of the regression curve. However, the impact of this outlier may be deemed insignificant and has not considerably affected the overall correlation coefficient, which still remains high, exceeding 0.96. The research team acknowledges the presence of this outlier and has considered various statistical methods to account for its effect on the analysis.

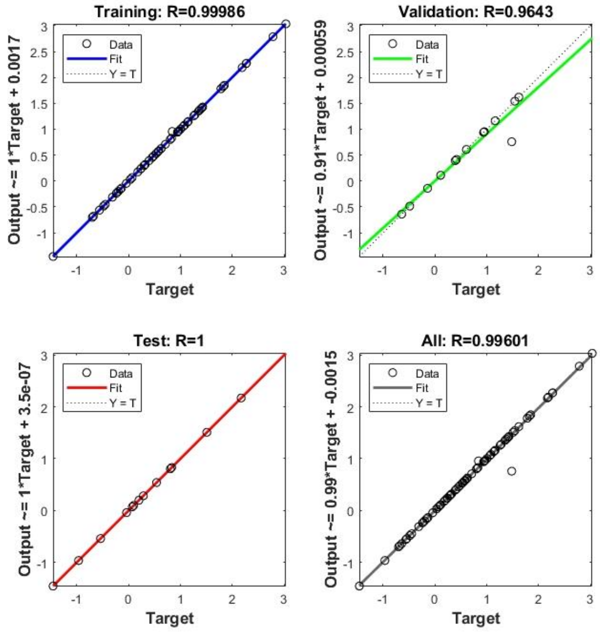


Figure 5: Scatter Plot for Example Model Output and Target (for the Number of Initiated Houses, *trainlm* Function, 1000 Iterations of the Model Training)

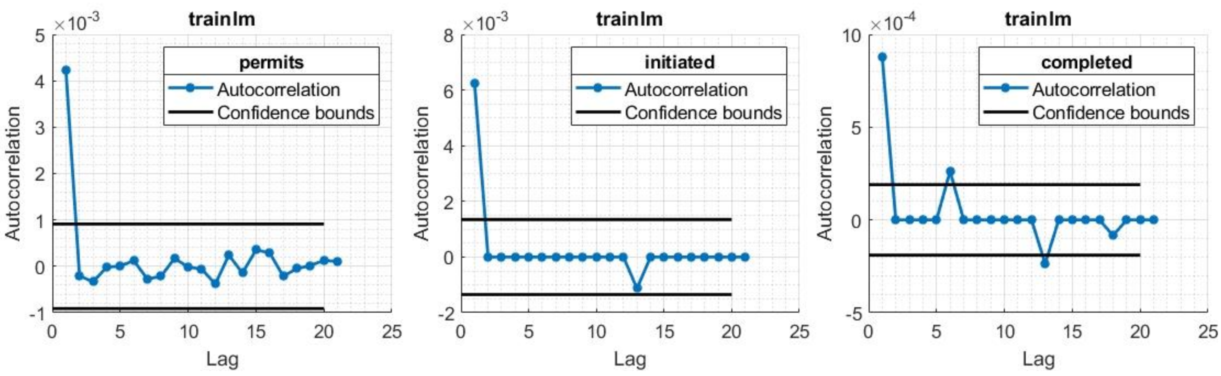
Despite this, the results obtained from the analysis remain valid and reliable for future research in this field. Further investigations will be conducted to understand the cause of the outlier and its potential impact on the results.

In the next step, the autocorrelation of residuals has been investigated. A good prediction model requires that the errors be random, normally distributed, and not correlated with each other. Furthermore, the autocorrelation values for various lags should fall within confidence intervals. Fig. 6 visualizes sample error autocorrelation for various house construction stages for the best training function, i.e., *trainlm* (due to the space limits results obtained for other training functions are not presented in the figure; nevertheless, they are included in the analysis).

In most cases the autocorrelations of the errors have the correct course, i.e., they decrease with the lag increase and fall within the confidence intervals. Exceptions are the data related to houses under construction (the *initiated* model), where for the *trainoss* and *traincgb* functions autocorrelation increases appear. In contrast, the *trainlm* function for the number of building permits issued (the *permits* model) and housing units finished (the *completed* model) do not exceed the confidence intervals. In the case of the number of completed houses, small autocorrelations appear for the 6- and 12-month periods.

To provide a summary of prediction results obtained, their visualization is provided in Fig. 7. Individual bars on the graph represent the Pearson correlation coefficient values that were obtained in the regression process of the learning and the actual data. The colors of the bars distinguish results obtained for individual data sets: the train, validation, test, and all-data ones whereas the x-axis spreads out groups of values obtained for individual training functions: *trainscg*, *trainoss*, *traincgb*, and *trainlm*. The range of the y-axis was limited from 0.9 to 1 since very good fits were obtained for all the cases, with differences of about 0.04 at most. The best fit, characterized by the highest correlation coefficient, was obtained for the number of housing units put into service. Regardless of the dataset, the best fit was achieved for the network with the integrated *trainlm* algorithm.

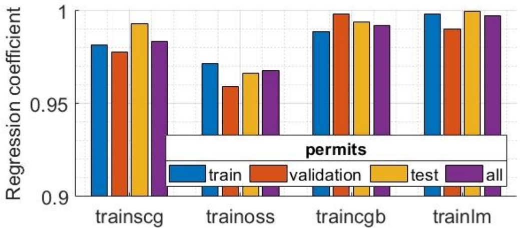
Fig. 8 presents *MSE* results obtained in an open loop network and for delayed one-step ahead prediction, which provides an additional measure for evaluating the quality of the neural network's fit to the empirical data. Let us note that the y-axes were not standardized for better observation of differences between the training functions. The best results were obtained for the



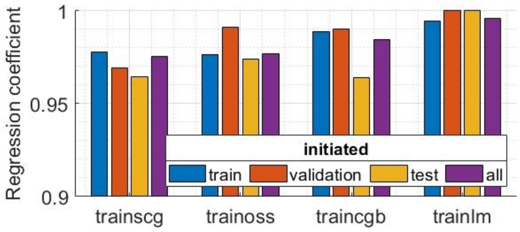
(a) Construction Permits (b) Houses Under Construction (c) Completed Houses

Figure 6: Error Autocorrelation for the Considered House Construction Stages, *trainlm* Function

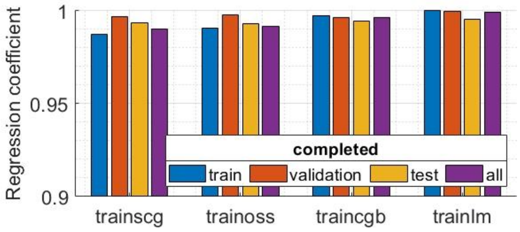
prediction of the number of completed housing units, for which the smallest *MSE* values were obtained (below 0.025). Regardless of the type of data, the best fit was obtained for *trainlm* function, for which *MSE* scores were close to zero. The highest error rates, exceeding 0.08, were obtained for the number of house construction permits predicted with the use of *trainoss* function.



(a) Construction Permits

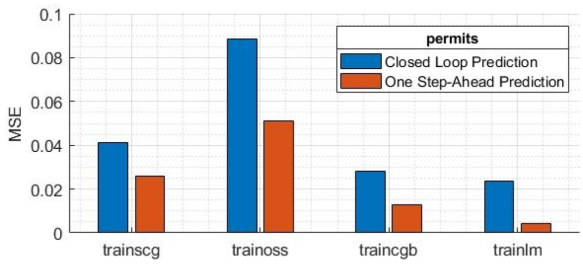


(b) Houses Under Construction

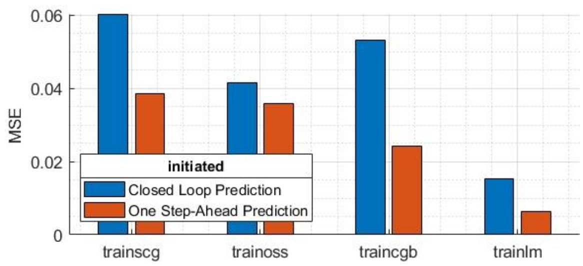


(c) Completed Houses

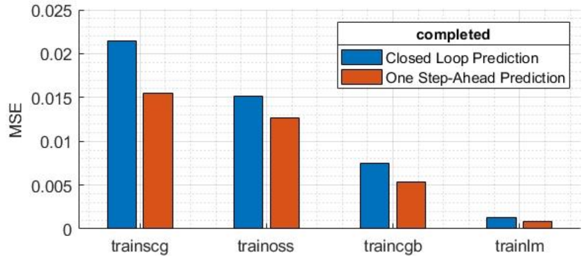
Figure 7: Prediction Results – Regression Coefficient



(a) Construction Permits



(b) Houses Under Construction



(c) Completed Houses

Figure 8: Prediction Results – *MSE*

CONCLUSION

This paper discusses a study aimed at the application of the nonlinear autoregressive neural network (NARNet) for the housing construction market forecasts. The study focuses on predicting the number of houses related to the main stages of housing construction, such as permits issued for house construction, houses under construction, and completed new houses. Real data from the Polish housing construction market was used and three groups of NARNet models were developed for the three construction stages, with the use of four various neural network training functions.

The model efficiency was verified through a number of experiments and analyzed in terms of the model fit and prediction error rates (mean squared error, regression coefficient). The results show that NARNet can be a very effective tool in this scenario, with a network model using the Levenberg-Marquardt backpropagation training function achieving the best model fit and the most accurate one-month predictions.

One limitation of a NARNet model is that it relies heavily on the quality and quantity of input data. Incomplete or error-containing data may introduce bias into the model and result in inaccurate predictions. Additionally, NARNet method assumes that the data follows a specific pattern, which may not always hold true in real-world situations, making it challenging to capture sudden changes or irregularities in the housing construction market, such as unexpected economic shocks or policy changes. Another limitation of developing a NARNet model is that it requires a significant computational cost to train and optimize the model, especially for large datasets with many input variables. To overcome these limitations, researchers may explore alternative approaches to the housing construction market forecasts, such as hybrid models that combine multiple methods or machine learning algorithms that can handle high-dimensional data more effectively. However, NARNet remains a popular choice due to its flexibility and ability to capture complex nonlinear relationships between variables. Besides, the application area under consideration is not a case of a real-time environment so the computation time is not critical.

Further research will focus on expanding the forecasting models used, resulting in ensemble or hybrid methods, which should further improve the quality of the forecasting models built. Moreover, as part of future work we are going to test the quality of forecasts using the equivalent NARNet models for data obtained from different regions of the country and from other countries.

REFERENCES

Adedeji, P. A.; Akinlabi, S.; Ajayi, O.; and Madushele, N. 2019. "Non-linear autoregressive neural network (NARNET) with SSA filtering for a university energy consumption forecast." *Procedia Manufacturing*, 33, 176-183.

Frącz, P.; Dąbrowski, I.; Wotzka, D.; Zmarzły, D.; and Mach, Ł. 2023. "Identification of differences in the seasonality of the developer and individual housing market as a basis for its sustainable development." *Buildings* 2023, 13, 316.

GUS. 2022. Główny Urząd Statystyczny – Publications, https://stat.gov.pl/publikacje/publikacje-a-z

Hozer, J.; Gnat, S.; Kokot, S; and Kuźmiński, W. 2019. "The problem of designating elementary terrains for the purpose of Szczecin algorithm of real estate mass appraisal." *Real Estate Management and Valuation* 27(3), 42-58.

Huang, Y.; Wang, Y.; and Shu, G. 2011. "The application and research of a new combinatorial analysis and forecasting method in real estate area based on grey system theory and multivariate linear regression." *Procedia Engineering* 15, 4532-4537.

JLL. 2021. Badania rynku nieruchomości, https://www.jll.pl/pl/trendy-i-analizy/badanie.

Kokot, S. 2022. "Identification of regularities in relation between prices on primary and secondary housing market in selected cities in Poland." *Real Estate Management and Valuation* 30(3), 45-60.

Lin, R. F. Y.; Ou, C.; Tseng, K. K.; Bowen, D.; Yung, K. L.; and Ip, W. H. 2021. "The spatial neural network model with disruptive technology for property appraisal in real estate industry." *Technol. Forecast Soc. Change* 173, 121067.

Mach, Ł. 2019. "Measuring and assessing the impact of the global economic crisis on European real property market." *J. Business Economics and Management* 20(6), 1189-1209.

Mach, Ł.; Zmarzły, D.; Dąbrowski, I.; and Frącz, P. 2020. Comparison of forecasting models for dwellings granted building permissions in Poland. In *Proc. of IBIMA*, 11279-11294.

Mach, Ł.; Zmarzły, D.; Dąbrowski, I.; and Frącz, P. 2021. "Identification and parametrization of polycyclicity in the primary housing market." *J. Hous. and the Built Environ.* 36, 1699-1712.

Narnet Documentation, MathWorks – https://www.mathworks.com/help/deeplearning/ref/narnet.html (access date: 06.02.2023).

Padilla, C.; Hashemi, R.; Mahmood, N. H.; and Latva-aho, M. 2021. "A Nonlinear Autoregressive Neural Network for interference prediction and resource allocation in URLLC scenarios". https://doi.org/10.48550/arXiv.2111.15630.

Xu, X. and Zhang, Y. 2022. "Soybean and Soybean oil price forecasting through the Nonlinear Autoregressive Neural Network (NARNN) and NARNN with exogenous inputs (NARNN–X)." *Intelligent Systems with Applications* 13, 200061.

AUTHOR BIOGRAPHIES

DARIA WOTZKA received an M.Sc. degree in Computer Science from the Technische Universität Berlin, Germany, and a Ph.D. degree in Electrical Engineering from the Opole University of Technology, Poland. She is a lecturer and research fellow at the Opole University of Technology, Poland. Her research interests include data mining, modeling, and simulation applied in engineering and medicine. Her e-mail address is `d.wotzka@po.edu.pl`.

GRAŻYNA SUCHACKA received M.Sc. degrees in Computer Science and Management, as well as a Ph.D. degree in Computer Science with distinction from Wroclaw University of Science and Technology, Poland. Now she is an Assistant Professor at the Institute of Informatics at the University of Opole, Poland. Her research interests include data analysis and modeling, data mining, machine learning, and Quality of Web Service with special regard to bot detection and electronic commerce support. Her e-mail address is `g.suchacka@gmail.com`.

ŁUKASZ MACH received a Ph.D. degree in Economics and Finance from the Warsaw School of Economics, Poland. He is a Professor, researcher, and lecturer at the Faculty of Economics and Management at the Opole University of Technology, Poland. His research interests include quantitative methods in economic research, time series analysis, modeling of the residential real estate market, and the analysis of its cyclicality. His e-mail address is `l.mach@po.edu.pl`.

PAWEŁ FRĄCZ is a full Professor, researcher, and lecturer at the Faculty of Economics at the University of Opole, Poland. His research interests include methods of time series analysis, issues in the area of modeling real estate markets, and applications of signal analysis methods in the time and frequency domain in economic sciences. His e-mail address is `p.fracz@gmail.com`.

JOACHIM FOLTYS is a professor extraordinary at Humanitas University in Sosnowiec. His research interests include outsourcing, digital environment (virtual), and shaping and valuation of contemporary organizations, including intangible assets, and family companies. His e-mail address is `joachimfol@onet.pl`.

IONELA MANIU received a Ph.D. degree in Cybernetics and Economic Statistics from the Bucharest University of Economic Studies, Romania. She is a lecturer and research fellow in the Mathematics and Informatics Department at the Lucian Blaga University of Sibiu, Romania. Her current research interests include data mining, bioinformatics, educational policy, network analysis. Her e-mail address is: `ionela.maniu@ulbsibiu.ro`.