Central European Journal of Economic Modelling and Econometrics

Artificial Neural Networks vs Spatial Regression Approach in Property Valuation

Damian Przekop*

Submitted: 6.08.2021, Accepted: 20.02.2022

Abstract

The purpose of this paper is to compare two approaches applied in property valuation: artificial neural networks and spatial regression. Despite the fact that artificial neural networks are often the first choice for modeling in the big data era, spatial econometrics methods offer incorporation of information on dependences between multiple objects in the studied space. Although this dependency structure can be incorporated into artificial neural network via feature engineering, this study is focused on abilities of reproducing it with machine learning method from crude coordinate data. The research is based on the database of 18,166 property sale transactions in Warsaw, Poland. According to this study, such volume of data does not allow artificial neural networks to compete in reflecting spatial dependence structure with spatial regression models.

Keywords: artificial neural networks, spatial regression, SDEM, GNS, property valuation

JEL Classification: C21, C45

*Warsaw School of Economics; e-mail: dprzek@sgh.waw.pl; ORCID: 0000-0002-3151-4667

1 Introduction

In his 1970 paper, American-Swiss geographer W.R. Tobler remarks that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This idea is referred to as the "first law of geography" (Annamoradnejad et al., 2019) and was a starting point for construction and verification of many research hypotheses in many fields: from education, through the labor market and even to organized crime (Dietz, 2002). This observation had an impact on the achievements in economics, as well as in its auxiliary science: econometrics. As part of the latter, a separate trend in the area of quantitative research has been developed, which is more broadly known as spatial econometrics.
One of the areas within the application of spatial econometrics is the identification of sources of real estate price volatility – both for land and housing (Krause and Bitter, 2012). The importance of this issue is reflected in the amount of capital that is related to the real estate market. According to the Report on the situation of the residential and commercial real estate market in Poland in 2018 (prepared by the Economic Analysis Department of the National Bank of Poland), the estimated value of residential real estate assets in Poland – at the end of 2018 – amounted to over PLN 3.8 trillion; see `https://www.nbp.pl/publikacje/rynek_nieruchomosci/raport_2018.pdf`. The research problem, that economists are trying to explore, is whether the neighborhood has an impact on real estate prices. While intuition provides many arguments to support this hypothesis, it is the task of econometricians to confirm it quantitatively.
However, these past few years seem to open up a new era of modelling, one that is based on the concept of big data. The definition of this new phenomena, as is proposed by De Mauro et al., requires dedicated analytical methods in order to make proper use of information kept in vast volumes of data (De Mauro et al., 2016). Oswald et al. enumerate big data analytical methods that can be applied in the field of social sciences: random forests and gradient boosted trees, artificial neural networks (ANNs), and support vector machines (Oswald and Putka, 2017). Due to the little assumptions that must be made when using these machine learning (ML) methods, their application has become very popular. The use of ML methods in property valuation constantly grows in popularity. Accoriding to Masias and Valle, ML techniques can provide a useful set of tools for acquiring information on housing markets (Masias and Valle, 2016). It can be found in papers of Lin et al. (2021), Yacim and Boshoff (2020) or Selim (2009).
According to the universal approximation theorem, a feedforward network with a linear output layer and at least one hidden layer can approximate any continuous function (Cybenko, 1989), however the theorem does not precise how large this network has to be (Goodfellow et al., 2016). Thus, even though the function approximation is possible, sufficient volume of data and neural network complex enough have to be used. That is why artificial neural networks' both explanatory and prediction capabilities can be greater than those of linear models. However,

spatial regression offers the concept of a weight matrix W, which reflects the pattern of dependences between multiple objects in the studied space. The matrix W is the spatial weight matrix that contains non-zero elements w_{ij} if observations j and i are neighbors and zero otherwise. The neighborhood definition is arbitrary and may range from the one based on mere adjacency between two territorial units, to those based on a maximum distance (that is $j \in N(i)$ if $d_{ij} < d_{\max}$), to those based on the nearest neighbor criterion (Arbia, 2014). It is a powerful means of encoding which in fact is a derivative of the spatial location of individual objects. The purpose of this paper is to verify whether a machine learning approach in its standard design, without additional feature engineering, can reproduce the information kept in W matrix from latitude and longitude data, given limited volume of data at hand.

Nikparvar and Thill in their paper persuade there is extensive literature that applies ML to spatial data but research that explicitly features the spatial properties of data in ML remains rather limited (Nikparvar and Thill, 2021). As the way of incorporating spatial information into ML models is to include the spatial components of data in the observation matrix.

One way of spatial information inclusion is adding coordinates alongside remaining attributes. Martin et al. in their work on solar energy forecasting enrich input data with longitude and latitude of each meteorological station (Martin et al., 2016). In their research are employed: support vector machines and gradient boosting. Zanella et al. use latitude and longitude data as an input to random forest model on deforestation and forest fragmentation in the Brazilian Atlantic forest (Zanella et al., 2017). The other way of spatial information inclusion is feature engineering.

Reproducing the information kept in W matrix from latitude and longitude data seem to be an attractive direction for the research due to the at least two reasons. The first one is computational time and complexity of W matrix. The another reason is sensitivity of the results to the different choices for W matrix specification (LeSage and Pace, 2014).

The structure of the paper is as follows: Section 2 is devoted to the literature review; both in the field of spatial modeling and research that focuses on identifying factors that are influencing house prices. In Section 3, the data set that has been analyzed is presented. Section 4 describes the tools that are used in this paper. Section 5 is devoted to the results of the research. Finally, Section 6 concludes and provides directions for further work.

2 Literature review

The problem of property valuation is broadly discussed in the scientific literature. Much attention is paid to the determinants of housing prices, while other researchers focus on the methods that are used.

Wang et al. attempt to identify factors that influence housing prices in Taitung (Taiwan). Based on 3,533 transactions, they verify the influence of several elements;

these include apartment size, floor, building age, distance to major road, distance to train station, and distance to school or park (Wang et al., 2019). The influence of environmental factors like noise, air quality, or window view is verified in the paper of Hui et al. (Hui et al., 2007). In their article, Ceccato and Wilhelmsson prove that in Stockholm, the close proximity of crime hot spots affects housing prices. According to their research, every additional kilometer of distance from such hot spots raises the price of apartments by 3,000 euro (Ceccato and Wilhelmsson, 2019). Many works focus on the influence that public transport infrastructure has on property prices; among them are either the study of Yang et al. (Yang et al., 2019) or of Henneberry (Henneberry, 1998). Urban infrastructure is another factor that is considered by researchers. Zhang et al., in their paper, present the positive impact that the presence of shopping centers has on flat prices. Close proximity to public utility facilities is not only practical, but it also contributes to the reduction of exhaust fumes being generated and – as a result – influences housing decisions (Zhang et al., 2019). Yang et al. point out that the presence of sports and cultural centers or schools positively impact housing prices in the neighborhood. Hospitals, on the other hand, have a negative impact on the price (Yang et al., 2018). Additionally, school input quality affects nearby property prices in the Taipei metropolis (Peng, 2019). A valuable literature review of factors that influence Malaysian housing prices (including building properties, location, neighborhood, and quality of life) can be found in the studies of Hilmi et al. (Hilmi et al., 2016).

The linear hedonic regression model is often used as a starting point in research on property valuation (Copiello, 2020; Annamoradnejad et al., 2019). The main assumption that underlies hedonic models, is that the value of a particular good is treated as a sum of individual utility-bearing attributes (Rosen, 1974). A classic approach is based on OLS estimation and the final model can be written as follows:

$$Y = X\beta + \varepsilon, \tag{1}$$

where Y is property price and X is a vector of utility-bearing attributes.

The turn of both the 1970's and 1980's of the XX century brought incorporation of spatial effects into econometric modeling; Goodman (Goodman, 1978), as well as Li and Brown (Li and Brown, 1980), all measured the influence of the neighborhood on housing prices. The intense development of spatial econometrics took place a decade later (Anselin, 1988). Pace and Gilley prove that spatial modeling improves both the effectiveness and precision of the estimation (Pace and Gilley, 1997). In the research of Osland, a useful overview of spatial models that is applied to Norwegian property valuation problems can be found (Osland, 2010). The author proves that a spatial lag model (SLM; alternatively known as SAR – spatial autoregressive model) and a spatial error model (SEM) offer a better fit and precision of the estimation than an OLS approach. Similar conclusions are drawn by Bourassa et al. (Bourassa et al., 2010) and Palma et al. (Palma et al., 2018). The aim of spatial modeling is not only to find the determinants of property prices, but to also precisely fix their actual contribution.

Cohen and Coughlin compare spatial autocorrelation and autoregressive models with the OLS solution for measuring the impact of airport noise on Atlanta housing prices. Although the differences in estimations are quite small, it turns out that OLS strongly underestimates the impact of the noise on the prices, due to omitting the spatial multiplier factor. Similar conclusions in the field, regarding the impact of schooling quality on property prices, can be found in the paper of Brasington and Haurin (2006).
The last decade has witnessed the domination of machine learning approaches when it comes to modeling most scientific problems – including property valuation. Mimis et al., in their paper, apply artificial neural networks to the problem of property valuation in Athens (Mimis et al., 2013). The application of the gene expression programming approach for house pricing in Iran can be found in the article of Shekarian and Fallahpour (2013). A real estate price estimation, which considers the environmental quality of property location, can be found in paper of Chiarazzo et al. (2014).
Because the machine learning approach has entered the already well-established field of property valuation, some researchers focus on comparing different modeling techniques. Selim compares linear hedonic regression against artificial neural network; proceeding to then prove the superiority of the latter in terms of price prediction power (Selim, 2009). Geographically weighted regression is squared up against gradient boosting approach for house price appreciation evaluation, as seen in the research of Kang et al. (2020). The application of several different approaches for mass appraisal of properties can be found within the article by Yacim and Boshoff (2020). Predictive accuracy is measured for: OLS, geographically weighted regression, spatial error model, spatial lag model, support vector machine, and artificial neural network. The conclusion is that spatial models proved to be a better option in price estimation than ANNs due to their abilities to tackle the problems of both spatial dependence and spatial heterogeneity. However, the research lacks information on various ANNs architectures, the total number of neurons in the hidden layers, and activation functions that distort the conclusions covered in the presented paper as these hyperparameters affect the performance of estimated ANNs. Other research that has focused on method comparison in property valuation include the works of Abidoye and Chan (2018), as well as Embaye et al. (2021).
In some papers, ANNs applications to spatially autocorrelated data incorporate spatial information via including crude coordinate data or some form of feature engineering. Cui et al. in their research on optimal scheduling of interfering links in a dense wireless network base neural network solution solely on the geographic locations of transmitters and receivers (Cui et al., 2019). On the other hand, Lin et al. in their paper on property appraisal in real estate industry enrich Support Vector Machine or Multi-Layer Perceptron approaches with additional features derived from satellite images (Lin et al., 2021). Both ways of feature enrichment aim to reproduce spatial dependencies between entities that in case of spatial regression is handled by W matrix.

Table 1: Descriptive statistics

	Minimum	1st Quartile	Median	Mean	Deviation	3rd Quartile	Maximum	Literature
price of square meter (in PLN)	1 244	7 198	8 357	8 840	2 789	9 845	47 618	-
building								
- year of construction	1 860	1 984	2 016	1 999	28	2 018	2 019	Mok et al. (1995)
- floors overground	1	5	6	7	4	9	54	Tan and Guan (2021)
- floors underground	0	1	1	1	1	1	5	
primary market (binary)	0	0	0	0.5	0.5	1	1	Reichel and Zimčík (2018)
flat								
- floor	-1	2	3	4	3	5	54	Zakaria and Fatine (2021)
- number of rooms	0	2	3	3	1	3	12	Król (2015)
- size (square meters)	15.2	39.7	51.0	56.9	26.0	67.3	363.9	Law (2017)
- has cellar (binary)	0	0	0	0.2	0.4	0	1	Džupka and Gróf (2021)
- has parking (binary)	0	0	0	0.4	0.5	1	1	
- is on attic (binary)	0	0	0	0.0	0.1	0	1	Helbich et al. (2013)
location								
- distance from metro station (in kms)	0.0	1.1	2.5	3.2	2.7	4.5	15.6	Wen et al. (2020)
- distance from CBD (in kms)	0.0	4.8	6.7	7.1	3.2	8.9	16.9	Waddell et al. (2013)
- latitude	52.1	52.2	52.2	52.2	0.1	52.3	52.4	-
- longitude	20.9	21.0	21.0	21.0	0.1	21.1	21.3	-

Table 1: Descriptive statistics cont.

	Minimum	1st Quartile	Median	Mean	Deviation	3rd Quartile	Maximum	Literature
district								
- Bialoleka	0	0	0	0.12	0.33	0	1	
- Bemowo	0	0	0	0.05	0.22	0	1	
- Bielany	0	0	0	0.05	0.22	0	1	
- Mokotow	0	0	0	0.16	0.37	0	1	
- Ochota	0	0	0	0.05	0.22	0	1	
- Praga Polnoc	0	0	0	0.03	0.18	0	1	
- Praga Poludnie	0	0	0	0.10	0.30	0	1	
- Rembertow	0	0	0	0.00	0.06	0	1	
- Srodmiescie	0	0	0	0.05	0.22	0	1	D'Elia et al. (2020)
- Targowek	0	0	0	0.06	0.25	0	1	(district impact)
- Ursus	0	0	0	0.03	0.18	0	1	
- Ursynow	0	0	0	0.04	0.21	0	1	
- Wola	0	0	0	0.10	0.30	0	1	
- Wawer	0	0	0	0.02	0.14	0	1	
- Wesola	0	0	0	0.01	0.09	0	1	
- Wilanow	0	0	0	0.05	0.22	0	1	
- Wlochy	0	0	0	0.03	0.17	0	1	
- Zoliborz	0	0	0	0.02	0.14	0	1	

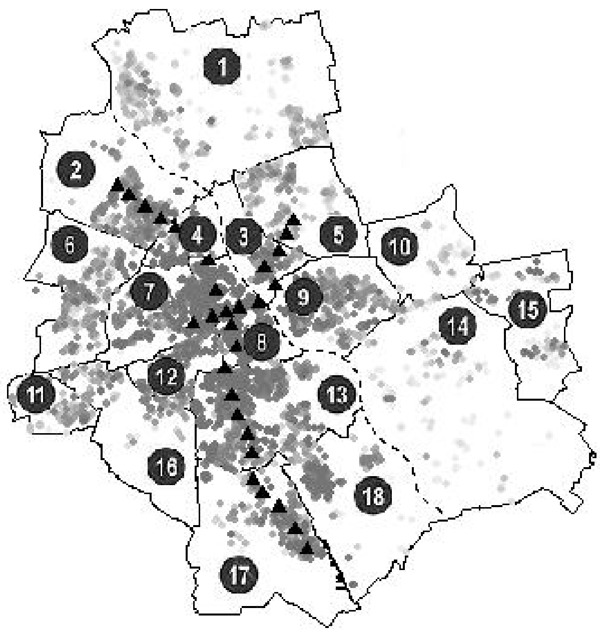
3 Data

The dataset consists of 18,166 flat sale transactions made in Warsaw between July 2018 and April 2019. The time shift in the analyzed dataset is adopted in order to exclude influence of both: COVID-19 pandemic and serious growth of estate prices in Poland beginning from 2^{nd} half of 2019. The data is of limited access as it is granted for a fee by a Warsaw Real Estate Unit. Descriptive statistics of the dataset are presented in Table 1. The last column of Table 1 presents papers in which particular variables were used in the hedonic modelling.
The average cost of a typical flat sold in Warsaw, at that time, was about 500,000 PLN. The cost of the most expensive flat was over 15,000,000 PLN. The number of rooms in flats that are kept in the database vary from 1 to 12. Buildings that are related to transactions are dated from 1860 to 2019. The average flat size is 57 m2 (median: 51 m2). The price of the flat's square meter varies from 1,244 PLN to 47,618 PLN, with the average being 8,840 PLN. The mean distance of property to metro station is 3.2 kilometers and to Central Business District (CBD; calculated as distance to the Palace of Culture and Science) is 7.1 kilometers. Most of the transactions were made in the Mokotow and Bialoleka districts.
Figure 1 presents a map of Warsaw; the lines represent the borders of districts and the bold black triangles represent metro stations (intersection of both lines marks the center of the CBD). Gray dots reflect transactions that are stored in the dataset; darker dots refer to transactions on more expensive properties, whereas lighter dots represent the cheapest ones. Observation of the graphic alone suggests that spatial factor may play an important role in explaining the transaction price of property. Some districts are "darker" than others; furthermore, a closer proximity to a metro station seems to raise the flat's pricing. Additionally, properties that are situated on the left side of the Vistula river (the curved dotted line that is drawn from the bottom right of the map to its top left) seem to be more expensive than those on the right side.

4 Experiment

The purpose of this research is to compare two approaches to property valuation: artificial neural networks and spatial regression. The two spatial models presented here are the spatial Durbin error model (SDEM), and a general nesting spatial model (GNS). For artificial neural networks, 18 different architectures are tested. The aim of the comparison is to verify whether artificial neural networks in its standard design, without additional feature engineering, can reproduce the information kept in W matrix from latitude and longitude data, given the volume of the data at hand. The dataset, that consists of 18,166 transactions, should be perceived as large not only in the context of property valuation but in general – for applications in the field of economics. Thus the actual research question is whether such volume of data

Figure 1: Property sale prices in Warsaw



Districts: 1 – Bialoleka, 2 – Bielany, 3 – Praga Polnoc, 4 – Zoliborz, 5 – Targowek, 6 – Bemowo, 7 – Wola, 8 – Srodmiescie, 9 – Praga Poludnie, 10 – Rembertow, 11 – Ursus, 12 – Ochota, 13 – Mokotow, 14 – Wawer, 15 – Wesola, 16 – Wlochy, 17 – Ursynow, 18 – Wilanow.

lets universal approximation theorem work for neural networks fed with spatially autocorrelated data.

The most crucial concept of spatial econometrics is a weight matrix W, which reflects the pattern of dependences between multiple objects in the studied space. It is a square matrix with number of rows and columns that are equal to a number of observations. Its i-th row can be interpreted as a vector of weights that reflect the influence of particular observations on i-th object. More on information on the W matrix can be found in the works of Arbia (2014).

In this paper, W matrix is based on the concept of k nearest neighbors as database refers to points in space where k is set to 11 due to property valuation practice in Poland fixing that more than 10 properties have to be taken into account when estimating property price.

The spatial autoregressive model (SAR) allows for incorporating the effect of nearby flat prices' influence on the price of the evaluated property. The SAR model assumes an endogenous interaction effect on dependent variable:

$$Y = \rho WY + X\beta + \varepsilon, \tag{2}$$

where Y is the price of square meter and X is a vector of predictors that consists of the remaining variables (except for *longitude* and *latitude*) that are presented in Table 1, W is a spatial weight matrix, WY is a spatially lagged dependent variable, and ρ is a spatial dependence parameter. When $\rho = 0$, SLM reduces to linear hedonic model (1).

SAR model allows for the testing of the hypothesis that property price depends, not only on the vector of flat characteristics, but also on the prices of nearby properties. In the scientific literature, such a phenomenon is known as an adjacency effect (Can, 1990); this is perceived as a way of benchmarking to similar transactions when fixing the price of a particular property (Osland, 2010).

The purpose of the spatial error model (SEM) is to address the issue of spatial autocorrelation of residuals that violate one of the assumptions of the OLS method. By including the spatial autocorrelation factor, the SEM model reduces bias caused by not including unobserveables. The formula of the SEM model is as follows:

$$Y = X\beta + \varepsilon, \tag{3}$$

$$\varepsilon = \lambda W\varepsilon + u, \tag{4}$$

where W is a spatial weight matrix, u is a random term that is independent and identically distributed $u \sim N(0, \sigma^2 I)$.

Spatial autoregressive model with autoregressive disturbances (Kelijan-Prucha or SAC model) cover both above-mentioned sources of variability; it can be written as follows:

$$Y = \rho WY + X\beta + \varepsilon, \tag{5}$$

$$\varepsilon = \lambda W\varepsilon + u, \tag{6}$$

where symbols are consistent with both SEM and SAR model notation. The SAC model is used when models that assume only one source of spatial variability turn out to be insufficient.

Spatial Durbin model (SDM) is formulated as follows:

$$Y = \rho WY + X\beta + WX\theta + \varepsilon \tag{7}$$

and it includes the $WX\theta$ term that represents spatial spillovers. Its incorporation in the model reflects the impact of determinants of prices of nearby properties on the price of other ones.

Spatial Durbin error model (SDEM) is formulated as follows:

$$Y = X\beta + WX\theta + \varepsilon, \tag{8}$$

$$\varepsilon = \lambda W\varepsilon + u. \tag{9}$$

The general nesting spatial model (GNS) is formulated as follows:

$$Y = \rho WY + X\beta + WX\theta + \varepsilon, \tag{10}$$

$$\varepsilon = \lambda W\varepsilon + u, \tag{11}$$

and it nests SAC, SDM and SDEM models.

All the aforementioned approaches to property price modeling require the adoption of explicit functions between both input and output data. Artificial neural networks on the other hand, do not require any assumptions as to the underlying functional form of the model. ANNs are models that are inspired by the biological neural networks that constitute animal brains. According to Hornik et al., what makes ANNs so powerful is the concept of “universal approximation” (Hornik et al., 1989). These are able to imitate different functional forms that reflect a real variability of the data.

In this research a multilayer perceptron (MLP) class of feedforward artificial neural network is employed. This solution has three main components: the input data layer (corresponding to X as is mentioned above – plus *longitude* and *latitude*), the hidden layers, and the output layer (corresponding to Y as is mentioned above). Each of these layers consist of nodes; these nodes are connected to nodes at adjacent layers. The hidden layers contain two processes: the weighted summation functions and the transformation function. Both functions relate the values from the input data to the output measures. Multilayer perceptron is perceived as most popular and widely used neural network type (Zhang, 2000) that is why such a structure is adopted in the presented research.

According to Amirabadi et al. (2020), one of the main difficulties in working with ANN is hyperparameter tuning. Hyperparameters are the design parameters, and could affect the training qualification. The hyperparameters to tune are: the number of neurons and hidden layers, activation function, optimization algorithm (and for some of them also: learning rate, batch size, and number of epochs). As the choice of hyperparameters’ values often depends on the structure of the problem, Aggarwal (2018) in his book states that he most well-known technique is *grid search*, in which a set of values is selected for set of hyperparameters and combinations of their values are tested in order to determine the optimal ANN choice.

In this study, various architectures of ANNs are examined; these are dependent on a different number of hidden layers, number of neurons, optimization algorithm, and activation function. Their role in the modelling part is following:

i) hidden layers – are the layers between input layer and output layer; smaller number of layers may cause underfitting, larger number – may cause overfitting;

ii) neurons – many hidden units within a layer can increase accuracy;

iii) activation function – defines the output of a particular node given an input or set of inputs;

iv) algorithm – optimization algorithms are used to change the weights of artificial neural networks in order to reduce the losses and provide the most accurate results.

A total of 18 architectures were tested, which are presented in Table 2.

Table 2: Artificial neural network architectures tested in the research

id	hidden layers	hidden neurons	algorithm	activation function
nn1	2	6	RPROP+	softplus
nn2	2	10	RPROP+	softplus
nn3	2	14	RPROP+	softplus
nn4	3	9	RPROP+	softplus
nn5	3	15	RPROP+	softplus
nn6	3	21	RPROP+	softplus
nn7	2	6	RPROP-	logistic
nn8	2	10	RPROP-	logistic
nn9	2	14	RPROP-	logistic
nn10	3	9	RPROP-	logistic
nn11	3	15	RPROP-	logistic
nn12	3	21	RPROP-	logistic
nn13	2	6	RPROP-	softplus
nn14	2	10	RPROP-	softplus
nn15	2	14	RPROP-	softplus
nn16	3	9	RPROP-	softplus
nn17	3	15	RPROP-	softplus
nn18	3	21	RPROP-	softplus

More on RPROP+ and RPROP- algorithms can be found in Riedmiller and Braun (1993), as well as in Riedmiller (1994). An exact models' specification in R code is presented in Appendix A. Full code is available upon request.
The prediction power of spatial models vs neural networks is measured with the mean absolute percentage error (MAPE), as well as the root-mean-square error (RMSE).
MAPE is a measure of the prediction accuracy of a forecasting approach. It expresses the accuracy as a ratio, which is defined by the formula:

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{A_i - P_i}{A_i} \right|}{n} * 100\%, \tag{12}$$

where A is the actual value and P is the predicted value.
The RMSE represents the square root of the second sample moment of the differences between predicted values and observed values. RMSE is a measure of accuracy and is used to compare forecasting errors of different models for a particular dataset; its

formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - A_i)^2}{n}}. \tag{13}$$

The prediction power of estimated models is verified via spatial 8-fold cross-validation. Each iteration of the cross-validation procedure trains models on 7 geo-clusters of transactions and performs tests on the remaining one. Clusters consist of comparable volume of transactions performed in adjacent districts. Clusters and their descriptions are listed in Table 3.

Table 3: Geo-clustering used in cross-validation procedure

Iteration / cluster	districts	number of transactions
1	Praga Poludnie, Rembertow, Wawer, Wesola	2435
2	Bialoleka	2250
3	Bemowo, Bielany, Zoliborz	2236
4	Mokotow	2912
5	Ochota, Ursus, Wlochy	2068
6	Ursynow, Wilanow	1704
7	Srodmiescie, Wola	2766
8	Praga Polnoc, Targowek	1795

In the case of spatial models, each iteration of cross-validation procedure assumes model (and thus W matrix) training on 7 geo-clusters of data and predictions are calculated for the remaining cluster with W matrix computed on the entire (covering all 8 geo-clusters) dataset. Such an approach to W matrix computation in the test phase is supposed to reflect the fact that ANNs utilize latitude and longitude data.
All computations have been performed in R version 4.0.2. The maximum time of convergence of each model was set to 72 hours. Part of the neural networks did not converge during that time: nn1, nn2, nn3, nn4, nn5, nn6, nn13, nn14, nn15, nn17, and nn18. As a result, these models are not presented in the following sections of this paper.

5 Results

The application of artificial neural networks does not require any particular assumptions; this is neither on data nor on the formula linking input and output of the model. Using spatial regression models should be preceded by spatial autocorrelation testing. This can be done with the Moran I test (Arbia, 2014). A null hypothesis states that there is no spatial autocorrelation in the residuals. These test results are presented in Table 4.
According to Moran I test results, in case of the OLS model, the null hypothesis should be rejected. For spatial models, the test provides evidence that these models

Table 4: Moran I test statistics

	Moran I statistic	p-value
OLS	0.422	<0.001
SAC	-0.000	0.986
SDM	0.002	0.199
SDEM	0.000	0.522
GNS	0.000	0.479

address the issue of spatial autocorrelation in the residuals. In conclusion, due to the fact that there is spatial dependence in the data, spatial models should be applied, as the OLS solution loses its efficiency and – thus – estimates of coefficients are biased. A series of likelihood ratio tests applied to spatial models are presented in Table 5.

Table 5: LR test results

	Likelihood ratio	p-value
GNS vs SAC	361.9	<0.001
GNS vs SDM	41.1	<0.001
GNS vs SDEM	0.07	0.793

According to LR test results from Table 5, GNS provides better fit to data than SAC and SDM models. There is also no evidence to reject the null hypothesis that SDEM model provides as good fit as more general model. Estimation results of spatial models are presented in Table 6.

According to estimation results from Table 6, in both models most variables are significant at $\alpha = 0.05$ except for: *building – floors overground, location – distance from metro station, lagged: flat – has cellar, lagged: location – distance from metro station, lagged: location – distance from CBD*. In both models WX component is significant (theta) and WY is not significant (rho) in GNS model. Both models are significant in terms of Wald statistic at $\alpha = 0.05$.

Comparison measures for estimated models are presented in Table 7.

The results of the RMSE and MAPE calculation from Table 7 show that, in terms of RMSE and MAPE criterions calculated within cross-validation procedure spatial models are superior to artificial neural networks. Comparison of the cross-validation results to RMSE and MAPE calculated for models trained on the entire dataset also reveal the robustness of spatial models to overfitting. Artificial neural networks prediction accuracy drops significantly on test sets.

Figure 2 presents scatterplots of actual and predicted prices of the competing models. The straight line represents $y = x$ identity function.

Based on the information seen in Figure 2, a few conclusions can be drawn. First, data on neural network predictions are more dispersed than in the case of spatial

Table 6: Estimation results of spatial models

	SDEM		GNS	
	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
Intercept	26 373	<0.000	26 677	<0.000
building				
- year of construction	17	<0.000	17	<0.000
- floors overground	-14	0.067	-13	0.069
- floors underground	623	<0.000	623	<0.000
transaction - primary market (binary)	-1 120	<0.000	-1 119	<0.000
flat				
- floor	99	<0.000	99	<0.000
- number of rooms	-527	<0.000	-527	<0.000
- size (square meters)	14	<0.000	14	<0.000
- has cellar (binary)	133	<0.000	133	<0.000
- has parking (binary)	725	<0.000	725	<0.000
- is on attic (binary)	-639	0.005	-640	0.005
location				
- distance from metro station (in kms)	54	0.666	54	0.666
- distance from CBD (in kms)	-36 370	0.033	-36 301	0.034
district				
- Bialoleka	-10 062	<0.000	-10 063	<0.000
- Bemowo	-5 303	<0.000	-5 304	<0.000
- Bielany	-3 449	<0.000	-3 445	<0.000
- Mokotow	-1 253	<0.000	-1 253	<0.000
- Ochota	-2 351	<0.000	-2 347	<0.000
- Praga Polnoc	-3 953	<0.000	-3 952	<0.000
- Praga Poludnie	-3 788	<0.000	-3 787	<0.000
- Rembertow	-8 990	<0.000	-8 977	<0.000
- Targowek	-4 895	<0.000	-4 896	<0.000
- Ursus	-5 062	<0.000	-5 057	<0.000
- Ursynow	-2 368	0.001	-2 366	0.001
- Wola	-1 541	<0.000	-1 540	<0.000
- Wawer	-4 527	<0.000	-4 517	<0.000
- Wesola	-5 497	<0.000	-5 496	<0.000
- Wilanow	-2 030	<0.000	-2 031	<0.000
- Wlochy	-3 424	0.001	-3 424	0.001
- Zoliborz	34	0.946	32	0.949
lagged: building				
- year of construction	-24	<0.000	-24	<0.000
- floors overground	45	0.034	46	0.033
- floors underground	-550	<0.000	-548	<0.000
lagged: transaction - primary market (binary)	666	<0.000	662	<0.000

Table 6: Estimation results of spatial models cont.

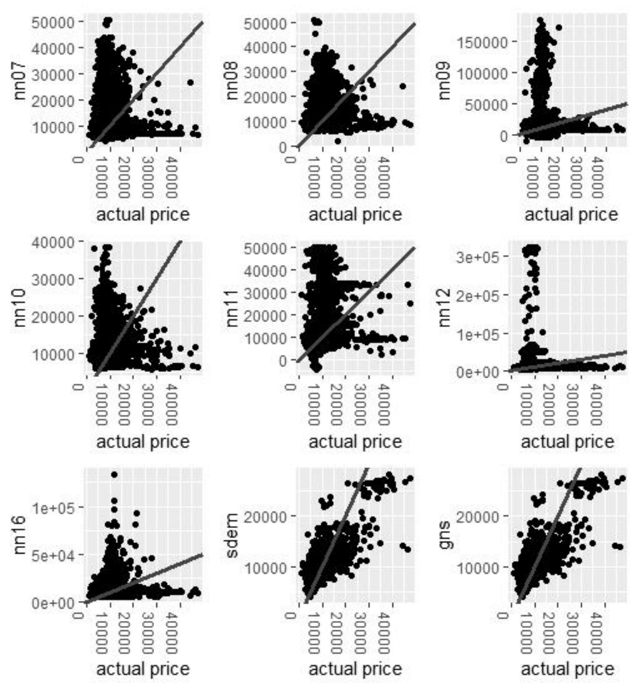
	SDEM		GNS	
	Estimate	Pr(>|z|)	Estimate	Pr(>|z|)
lagged: flat				
- floor	93	0.006	94	0.006
- number of rooms	-578	<0.000	-591	<0.000
- size (square meters)	34	<0.000	34	<0.000
- has cellar (binary)	1	0.996	1	0.997
- has parking (binary)	527	<0.000	546	<0.000
- is on attic (binary)	-2 359	0.023	-2 373	0.023
lagged: location				
- distance from metro station (in kms)	-124	0.365	-124	0.364
- distance from CBD (in kms)	21 519	0.222	21 284	0.228
- Bialoleka	6 308	<0.000	625	<0.000
- Bemowo	2 979	<0.000	2 946	<0.000
- Bielany	1 183	0.043	1 148	0.059
- Mokotow	-915	0.033	-940	0.036
- Ochota	-2	0.997	-40	0.945
- Praga Polnoc	797	0.189	750	0.235
- Rembertow	6 565	<0.000	6 520	<0.000
- Targowek	1 282	0.024	1 234	0.043
- Ursus	2 132	0.044	2 083	0.053
- Ursynow	-532	0.508	-574	0.486
- Wola	-501	0.221	-530	0.218
- Wawer	1 725	0.052	1 671	0.066
- Wesola	3 727	<0.000	3 676	<0.000
- Wilanow	-424	0.509	-453	0.493
- Wlochy	517	0.634	475	0.666
- Zoliborz	-1 989	0.001	-2 009	0.001
theta	0.731	<0.000	0.736	0.017
rho			-0.014	0.794

regressions. Second, the spatial regressions predictions are more aligned to the red line of identity function – neural networks tend to overestimate square meter prices of cheaper properties and underestimate prices of more expensive ones.

Table 7: Comparison measures

	RMSE crossvalid	MAPE crossvalid	RMSE entire set	MAPE entire set	computation time
nn7	6 702	54%	1 621	12%	3h
nn8	5 021	37%	1 477	11%	6h
nn9	12 841	71%	1 393	11%	26h
nn10	6 591	53%	1 583	12%	4h
nn11	9 174	53%	1 460	11%	8h
nn12	17 117	42%	1 429	11%	13h
nn16	6 307	44%	1 606	12%	30h
SDEM	1 762	15%	1 507	11%	13h
GNS	1 757	15%	1 499	11%	25h

Figure 2: Scatterplots of prices



6 Discussion

The purpose of this article was to compare two approaches applied in property valuation: artificial neural networks and spatial regression. This research was based on the database of property sale transactions in Warsaw.

According to the Moran I test, the data that was examined is spatially autocorrelated. That is why spatial regression was taken into consideration and was confronted with artificial neural networks. The aim of this task was to verify whether a machine learning approach in its standard design, without additional feature engineering, can reproduce the information kept in W matrix from latitude and longitude data, given the volume of the data at hand (18,166 transactions).

Prediction abilities of both approaches were examined. In terms of the MAPE and RMSE criterions, spatial models' performance is better than in case of neural networks. Both metrics are similar when calculated for models estimated on the entire dataset. Cross-validation procedure reveals superiority of spatial models. It turns out that for such volume of data (18,166 transactions), artificial neural networks are not able to reproduce the information kept in W matrix from latitude and longitude data and spatial models outperform neural networks.

Another significant advantage of spatial regression approach over neural networks is the interpretability of the results. In case of the former, the exact influence of particular variables on the target can be defined. Artificial neural networks, however, are black boxes that offer little insight into their recommendation criteria.

Based on the Warsaw transaction database, it is recommended to apply the spatial regression approach – if possible. Better performance, as well as the possibility of result interpretation, makes spatial regression superior to artificial neural networks in modeling spatially autocorrelated data.

As for the direction of future works, research can be replicated on another dataset. Furthermore, analysis can be extended on different neural network architectures; stability analysis of both approaches may also be verified. The most attractive direction for future research would be the application of feature engineering on coordinate data for artificial neural networks. Such an activity would reveal what effort should be done to make artificial neural networks comparable with spatial models.

References

[1] Abidoye R. B., Chan A. P., (2018), Improving property valuation accuracy: a comparison of hedonic pricing model and artificial neural network, *Pacific Rim Property Research Journal* 24(1), 71–83.

[2] Aggarwal C., (2018), *Neural Networks and Deep Learning*, Springer Cham.

[3] Amirabadi M., Kahaei M., Nezamalhosseini S., (2020), Novel suboptimal approaches for hyperparameter tuning of deep neural network, *Physical Communication* 41.

[4] Annamoradnejad R., Safarrad T., Annamoradnejad I., Habibi J., (2019), Using Web Mining in the Analysis of Housing Prices: A Case study of Tehran, *5th International Conference on Web Research (ICWR)*, 55–60.

[5] Anselin L., (1988), *Spatial econometrics: Methods and models*, Kluwer Academic, Dordrecht.

[6] Arbia G., (2014), *A Primer for Spatial Econometrics with Applications in R*, Palgrave Macmillan.

[7] Bourassa S. C., Cantoni E., Hoesli M., (2010), Predicting house prices with spatial dependence: A comparison of alternative methods, *Journal of Real Estate Research* 32, 139–160.

[8] Brasington D., Haurin D. R., (2006), Educational outcomes and house values: A test of the value added approach, *Journal of Regional Science* 46, 245–268.

[9] Can A., (1990), The Measurement of Neighborhood Dynamics in Urban House Prices, *Economic Geography* 66, 254–72.

[10] Ceccato V., Wilhelmsson M., (2019), Do crime hot spots affect housing prices?, *Nordic Journal of Criminology*, DOI: 10.1080/2578983X.2019.1662595.

[11] Chiarazzo V., Caggiani L., Marinelli M., Ottomanelli M., (2014), A Neural Network based model for real estate price estimation considering environmental quality of property location, *Transportation Research Procedia* 3, 810–817.

[12] Copiello S., (2020), Spatial dependence of housing values in Northeastern Italy, *Cities* 96.

[13] Cui W., Shen K., Yu W., (2019), Spatial deep learning for wireless scheduling, *IEEE Journal on Selected Areas in Communications* 37, 1248–1261.

[14] Cybenko G., (1989), Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems* 2, 303–314.

[15] D'Elia V., Grand M., León S., (2020), Bus rapid transit and property values in Buenos Aires: Combined spatial hedonic pricing and propensity score techniques, *Research in Transportation Economics* 80.

[16] De Mauro A., Greco M., Grimaldi M., (2016), A formal definition of Big Data based on its essential features, *Library Review* 65(3), 122–135.

[17] Dietz R. D. ,(2002), The estimation of neighborhood effects in the social sciences: An interdisciplinary approach, *Social Science Research* 31, 539–575.

[18] Džupka P., Gróf M., (2021), The influence of the new cultural infrastructure on residential property prices. Evidence from Košice ECoC 2013, *Cities* 110.

[19] Embaye W. T., Zereyesus Y. A., Chen B., (2021), Predicting the rental value of houses in household surveys in Tanzania, Uganda and Malawi: Evaluations of hedonic pricing and machine learning approaches, *PLoS ONE* 16(2), DOI: 10.1371/journal.pone.0244953.

[20] Goodfellow I., Bengio Y., Courville A., (2016), *Deep Learning*, MIT Press.

[21] Goodman A. C., (1978), Hedonic prices, price indices and housing markets, *Journal of Urban Economics* 5, 471–484.

[22] Helbich M., Jochem A., Mücke W., Höfle B. (2013), Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning, *Computers, Environment and Urban Systems* 39, 81–92.

[23] Henneberry J., (1998), Transport investment and house prices, *Journal of Property Valuation and Investment* 16, 144–158.

[24] Hornik K., Stinchcombe M., White H., (1989), Multi-layer feedforward networks are universal approximators, *Neural Networks* 2(5), 359–366.

[25] Hilmi M., Masri M., Nawawi b A. H., Sipan b I., (2016), Review of Building, Locational, Neighbourhood Qualities Affecting House Prices in Malaysia, *Procedia - Social and Behavioral Sciences* 234, 452–460.

[26] Hui E., Chau C. K., Pun L., Law M. Y., (2007), Measuring the neighboring and environmental effects on residential property value: Using spatial weighting matrix, *Building and Environment* 42, 2333–2343.

[27] Kang Y., Zhang F., Peng W., Gao S., Rao J., Duarte F., Ratti C., (2020), Understanding house price appreciation using multi-source big geo-data and machine learning, *Land Use Policy* 111.

[28] Kolbe J., Schulz R., Wersing M., (2021) Real estate listings and their usefulness for hedonic regressions, *Empirical Economics* 61, 3239–3269.

[29] Krause A. L., Bitter C., (2012), Spatial econometrics, land values and sustainability: Trends in real estate valuation research, *Cities* 29, 19–25.

[30] Król A. (2015), *Application of Hedonic Methods in Modelling Real Estate Prices in Poland*, [in:] Data Science, Learning by Latent Structures, and Knowledge Discovery. Studies in Classification, Data Analysis, and Knowledge Organization, [eds.:] B. Lausen, S. Krolak-Schwerdt, M. Böhmer, Springer.

[31] Law S., (2017), Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London, *Cities* 60(A), 166–179.

[32] LeSage J. P., Pace R. K., (2014), The Biggest Myth in Spatial Econometrics, *Econometrics* 2, 217–249.

[33] Li M., Brown H., (1980), Micro-neighborhood externalities and hedonic housing prices, *Land Economics* 56, 125–141.

[34] Lin R., Chiye O., Tseng K., Bowen D., Yung K., Ip W., (2021), The Spatial neural network model with disruptive technology for property appraisal in real estate industry, *Technological Forecasting & Social Change* 173(C).

[35] Martin R., Aler R., Valls J. M., Galván I. M., (2016), Machine learning techniques for daily solar energy prediction and interpolation using numerical weather models, *Concurrency and Computation Practice and Experience* 28, 1261–1274.

[36] Masias V., Valle M., (2016), Property Valuation using Machine Learning Algorithms: A Study in a Metropolitan-Area of Chile, *Conference: AMSE Conference Santiago/Chile*.

[37] Mimis A., Rovolis A., Stamou M., (2013), Property valuation with artificial neural network: the case of Athens, *Journal of Property Research* 30(2), 128–143.

[38] Mok H., Chan P., Cho Y., (1995), A hedonic price model for private properties in Hong Kong, *The Journal of Real Estate Finance and Economics* 10, 37–48.

[39] Nikparvar B., Thill J. C., (2021), Machine Learning of Spatial Data, *International Journal of Geo-Information* 10.

[40] Osland L., (2010), An application of spatial econometrics in relation to hedonic house price modeling, *Journal of Real Estate Research* 32, 289–320.

[41] Oswald F. L., Putka D. J., (2017), Big data methods in the social sciences, *Behavioral Sciences* 18, 103–106.

[42] Pace R. K., Gilley O. W., (1997), Using the spatial configuration of the data to improve estimation, *Journal of Real Estate Finance and Economics* 14, 333–340.

[43] Palma M., Cappello C., De Iaco S., Pellegrino D., (2018), The residential real estate market in Italy: A spatio-temporal analysis, *Quality & Quantity* 53, 1–22.

[44] Peng T. C., (2019), Does the school input quality matter to nearby property prices in Taipei metropolis? An application of spatial analyses, *International Journal of Housing Markets and Analysis* 12, 865–883.

[45] Reichel V., Zimčík P., (2018), Determinants of Real Estate Prices in the Statutory City of Brno, *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 66, 991–999.

[46] Riedmiller M., Braun H., (1993), A direct adaptive method for faster backpropagation learning: The Rprop algorithm, *Proceedings of the IEEE International Conference on Neural Networks*, 586–591.

[47] Riedmiller M., (1994), Advanced supervised learning in multi-layer perceptrons - From backpropagation to adaptive learning algorithms, *Computer Standards and Interfaces* 16(5), 265–278.

[48] Rosen S., (1974), Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy* 82, 34–55.

[49] Selim H., (2009), Determinants of house prices in Turkey: Hedonic regression versus artificial neural network, *Expert Systems with Applications* 36, 2843–2852.

[50] Shekarian E., Fallahpour A., (2013), Predicting house price via gene expression programming, *International Journal of Housing Markets and Analysis* 6(3), 250–268.

[51] Tan M., Guan C., (2021), Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using Twitter data, *Applied Geography* 132.

[52] Tobler A. W. R., (1970), A computer movie simulating urban growth in the Detroit region, *Economic Geography* 46, 234–240.

[53] Waddell P., Berry B., Hoch I., (1993), Residential property values in a multinodal urban area: New evidence on the implicit price of location, *Journal of Real Estate Finance and Economic* 7, 117–141.

[54] Wang W. C., Chang Y. J., Wang H. C., (2019), An Application of the Spatial Autocorrelation Method on the Change of Real Estate Prices in Taitung City, *International Journal of Geo-Information* 8, 249-269.

[55] Wen H., Gui Z., Zhang L., Hui E., (2020), An empirical study of the impact of vehicular traffic and floor level on property price, *Habitat International* 97.

[56] Yacim J. A., Boshoff D. G., (2020), Neural networks support vector machine for mass appraisal of properties, *Property Management* 38(2), 241–272.

[57] Yang L., Wang B., Zhou J., Wang X., (2018), Walking accessibility and property prices, *Transportation Research Part D. Transport and Environment* 62, 551–562.

[58] Yang L., Zhou J., Shyr O. F., Huo D., (2019), Does bus accessibility affect property prices?, *Cities* 84, 56–65.

[59] Zakaria F., Fatine F., (2021), Towards the hedonic modelling and determinants of real estates price in Morocco, *Social Sciences & Humanities Open* 4(1).

[60] Zanella L., Folkard A. M., Blackburn G. A., Carvalho L. M., (2017), How well does random forest analysis model deforestation and forest fragmentation in the Brazilian Atlantic forest?, *Environmental and Ecological Statistics* 24, 529–549.

[61] Zhang L., Zhou J., Hui E., Wen H., (2019), The effects of a shopping mall on housing prices: a case study in Hangzhou, *International Journal of Strategic Property Management* 23, 65–80.

[62] Zhang X. (2000), *Neural Networks in Optimization*, Springer.

Appendix A

nn_01 <- neuralnet(f,data=scaled_train,hidden=c(3,3), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_02 <- neuralnet(f,data=scaled_train,hidden=c(5,5), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_03 <- neuralnet(f,data=scaled_train,hidden=c(7,7), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_04 <- neuralnet(f,data=scaled_train,hidden=c(3,3,3), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_05 <- neuralnet(f,data=scaled_train,hidden=c(5,5,5), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_06 <- neuralnet(f,data=scaled_train,hidden=c(7,7,7), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop+", rep=3, linear.output=T)

nn_07 <- neuralnet(f,data=scaled_train,hidden=c(3,3), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T)

nn_08 <- neuralnet(f,data=scaled_train,hidden=c(5,5), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T)

nn_09 <- neuralnet(f,data=scaled_train,hidden=c(7,7), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T) nn_10 <- neuralnet(f,data=scaled_train,hidden=c(3,3,3), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T)

nn_11 <- neuralnet(f,data=scaled_train,hidden=c(5,5,5), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T)

nn_12 <- neuralnet(f,data=scaled_train,hidden=c(7,7,7), stepmax= 1e+06, act.fct = logistic, algorithm = "rprop-", rep=3, linear.output=T)

nn_13 <- neuralnet(f,data=scaled_train,hidden=c(3,3), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

nn_14 <- neuralnet(f,data=scaled_train,hidden=c(5,5), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

nn_15 <- neuralnet(f,data=scaled_train,hidden=c(7,7), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

nn_16 <- neuralnet(f,data=scaled_train,hidden=c(3,3,3), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

nn_17 <- neuralnet(f,data=scaled_train,hidden=c(5,5,5), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

nn_18 <- neuralnet(f,data=scaled_train,hidden=c(7,7,7), stepmax= 1e+06, act.fct = softplus, algorithm = "rprop-", rep=3, linear.output=T)

GNS <- sacsarlm(prxm~mar_pri+bld_yr+bld_flrd+loc_flr+dis_m+bld_flru+ loc_att+par+cbd+loc_rms+loc_spa+cel+d_.BA+d_.BO+d_.BY+d_.MW+ d_.OA+d_.PC+d_.PE+d_.RW+d_.TK+d_.US+d_.UW+d_.WA+d_.WR+ d_.WS+d_.WW+d_.WY+d_.ZZ+lag.mar_pri+lag.bld_yr+lag.bld_flrd+ lag.loc_flr+lag.dis_m+lag.bld_flru+lag.loc_att+lag.par+lag.cbd+lag.loc_rms+ lag.loc_spa+lag.cel+lag.d_.BA+lag.d_.BO+lag.d_.BY+lag.d_.MW+lag.d_.OA+ lag.d_.PC+lag.d_.PE+lag.d_.RW+lag.d_.TK+lag.d_.US+lag.d_.UW+ lag.d_.WA+lag.d_.WR+lag.d_.WS+lag.d_.WW+lag.d_.WY+lag.d_.ZZ, data=wawa_train, listw=W5_list, Durbin=FALSE, type = "sacmixed")

SDEM <- errorsarlm(prxm ~ mar_pri+bld_yr+bld_flrd+loc_flr+dis_m+ bld_flru+loc_att+par+cbd+loc_rms+loc_spa+cel+d_.BA+d_.BO+ d_.BY+d_.MW+d_.OA+d_.PC+d_.PE+ d_.RW+d_.TK+d_.US+d_.UW+ d_.WA+d_.WR+d_.WS+d_.WW+d_.WY+d_.ZZ+lag.mar_pri+lag.bld_yr+ lag.bld_flrd+lag.loc_flr+lag.dis_m+lag.bld_flru+lag.loc_att+lag.par+lag.cbd+

lag.loc_rms+lag.loc_spa+lag.cel+lag.d__.BA+lag.d__.BO+lag.d__.BY+
lag.d__.MW+lag.d__.OA+lag.d__.PC+lag.d__.PE+lag.d__.RW+lag.d__.TK+
lag.d__.US+lag.d__.UW+lag.d__.WA+lag.d__WR+lag.d__.WS+lag.d__.WW+
lag.d__.WY+lag.d__.ZZ, data=wawa_train, listw = W5_list)