THE APPLICATION OF SPATIAL AUTOREGRESSIVE MODELS FOR ANALYZING THE INFLUENCE OF SPATIAL FACTORS ON REAL ESTATE PRICES AND VALUES

Katarzyna Kobylińska
Department of Geoengineering
University of Warmia and Mazury in Olsztyn
e-mail: katarzyna.kobyliinska@uwm.edu.pl

Abstract

The spatial distribution of real estate in specific geographic locations, real estate transactions, and the prices and values of properties are a highly complex spatial phenomena that should be analyzed with the use of multidimensional methods. Spatial factors are taken into account in the modeling process to increase the reliability of real estate market analyses, and spatial autoregressive models are applied to determine the effect of spatial factors on real estate prices and values.

The present study relies on a review of the literature and the results of an experiment. The concept and principles of market analysis were designed with the use of spatial autoregressive models, and the influence of selected spatial factors on real estate prices was presented on maps. Analyses involving autoregressive models enable reliable modeling and support correct interpretation of the observed processes.

Key words: spatial autoregressive models, spatial phenomena, real estate market.

JEL Classification: C50, R21, R31, R58.

Citation: Kobylińska, K. (2021). The application of spatial autoregressive models for analyzing the influence of spatial factors on real estate prices and values. Real Estate Management and Valuation, 29(4), 23-35.

DOI: https://doi.org/10.2478/remav-2021-0027

1. Introduction

Spatial factors play a very important role in the analyses of the real estate market, but they are not easy to incorporate in models of the evaluated space. In the literature, most market analyses have been conducted with the use of multiple regression models and their derivatives (Isakson, 1998; Czaja, 2001; Benjamin et al., 2004; Sirmans et al., 2005; Adamczewski, 2006; Bitner, 2007; Czaja & Dąbrowski, 2008; Barańska, 2010; Sawilow, 2010; Dąbrowski, 2011). Multiple regression models are a highly useful tool for analyzing transaction prices, but they are not frequently applied in practice, mainly due to problems with meeting formal requirements during model design (Hozer, 2001). The effectiveness and statistical reliability of regression models have to be verified before these models are used to predict transaction prices on the real estate market (Bitner, 2007). The problems associated with the application of multiple regression models in real estate market analyses have been extensively described in the literature (Mark & Goldberg, 1988; Czaja, 2001; Hozer, 2001; Wang & Wolverton 2002; Lis, 2005; Adamczewski, 2006; Bitner, 2007; Barańska, 2010; Sawilow, 2010).

The significance of parameters in classical regression models is not influenced by the spatial structure of the investigated phenomenon, which can lead to an incorrect interpretation of the results (Charlton & Fotheringham, 2009), in particular when real estate markets are assumed to be spatially heterogeneous. Models that do not account for spatial autocorrelation can produce erroneous results (Anselin, 1998; Tu et al., 2007), which is why many researchers have relied on spatial models to
analyze real estate markets and predict transaction prices (Can & Megbolugbe, 1997; Bowen et al., 2001; James et al., 2005; Bourassa et al., 2010). If spatial correlations exist between real estate prices in market datasets, classical regression models produce similar results because they do not meet the assumption of independent observations; the sample therefore carries less information than an independent sample of the same size. These autocorrelations are taken into account in spatial autoregressive models (SAR), where the prices of neighboring real estate constitute additional explanatory variables (Ligas, 2006).

2. Literature review

2.1 Spatial autocorrelation

Spatial autocorrelation denotes the presence of correlations between observations in geographic space (Can & Megbolugbe, 1997). Spatial autocorrelation is expressed by a quantitative relationship between locations and relates to continuous spatial changes which are visualized by clusters of similar values or consistent spatial patterns (Kopczewska, 2006).

Basu and Thibodeau (1998) have identified two main causes of spatial autocorrelation on the real estate market: structural similarity of real estate in a given location, and the influence of the same location factors in a given area. These factors can be associated with municipal infrastructure or environmental factors. The specificity of the real estate market, namely the presence of spatial interactions between property prices in selected locations, is affected by the behavior of market actors. Location is a factor that initiates autocorrelation effects (spatial correlations), which is consistent with Tobler’s first law of geography. The parties to real estate transactions base their decisions on information about similar properties that were traded in the neighborhood, including their individual attributes. These phenomena are responsible for spatial correlations between real estate prices, and these relationships should be considered in the developed model. An autoregressive component that represents the influence of real estate prices in the neighborhood on the price of the traded property should be taken into consideration in the process of modeling real estate prices and values (Kobylińska et al., 2017). For this reason, spatial autoregressive models were applied in the present study.

Spatial autoregressive models can be used to investigate spatial variations based on the mutual interactions between variables describing the studied objects (Goodchild & Haining, 2003). Correlations or interactions can involve explained (endogenous) variables, explanatory variables as well as random effects. Interactions that involve the endogenous variable are indicative of spatial autoregression. The above implies that the values of the endogenous variable in other locations influence the value of the endogenous variable in the studied location. Interactions involving random effects indicate that spatial autocorrelation exists between random effects in the model. The above also applies to explanatory variables when the explanatory variable in a given location is influenced by the values of exogenous variables in other locations. This situation is described by a spatial cross-regressive model (Suchecki, 2010).

A spatial weight matrix is a prerequisite for and the first step in a spatial autocorrelation analysis. Moran’s I is most widely used to test spatial autocorrelation (Longley et al. 2005, Kisiala 2016). This test statistic is calculated with the use of the below formula (1):

\[
I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z)(z_j - z)}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z)^2}
\]  

where:

- \( n \) – number of observations,
- \( z_i \) – value of variable \( z \) in the \( i \)th location,
- \( z \) – mean value of variable \( z \) for all observations (locations),
- \( w_{ij} \) – weight of spatial interactions between observations \( i \) and \( j \).

If the null hypothesis states that variable \( z \) has a random distribution (absence of spatial autocorrelation), the significance of spatial autocorrelation tested with Moran’s I is verified based on parameter \( Z_I \) with normal distribution (expected value – 0, variance – 1) according to formula (2):

\[
Z_I = \frac{I - E(I)}{\sqrt{Var(I)}}
\]  

where:
E(I) = \frac{-1}{\pi-1} is the expected value, 
\text{Var}(I) \quad \text{variance.}

Residual spatial autocorrelation in a linear model, verified by diagnostic tests for linear models, requires the application of spatial estimation methods (Kopczewska, 2006). Depending on the type of spatial interactions, two main types of spatial regression models are generally used: the spatial lag model (SLM) and the spatial error model (SEM) (Anselin, 1998; Wilhelmsson, 2002; Arbia, 2006).

2.2 Spatial lag model

The neighborhood matrix \( W \) exerts different effects on the endogenous variable in the spatial lag model and the spatial error model. The spatial lag model contains a spatially lagged endogenous variable \( Wy \); it therefore is an autoregressive model. The spatial error model is developed on the assumption that spatial autocorrelation exists between the residuals from the model. The general form of the SLM is given by Formula (3) (Anselin, 1998; Wall, 2004; Arbia, 2006; Kopczewska, 2006; Ligas, 2006):

\[
y = \beta X + \rho Wy + u
\]  

which can be reduced to (Formula (4)):

\[
y = (I - \rho W)^{-1}\beta X + e
\]  

where:

\( y \) – vector of the observed explained variable (real estate price, value) (n x 1); 
\( X \) – matrix of independent variables (real estate attributes) (n x k+1); 
\( \beta \) – parameter vector of coefficients (k+1 x 1); 
\( W \) – spatial weight matrix (n x n); 
\( e \) – model error vector, 
\( \rho \) – spatial autoregressive coefficient, 
\( u \) – model error, vector of the random effect 
\( u \sim N (0, \sigma^2 I) \) (n x 1).

When spatial autocorrelation does not occur \( (\rho = 0) \), the result is a classical model of multiple linear regression. The model can be reduced only when matrix \( (I - \rho W)^{-1} \) is an invertible matrix (Cellmer, 2014).

In most cases, the autoregressive coefficient takes on values in the range of (-1, 1). In this case, matrix \( (I - \rho W)^{-1} \) can be expanded (Won Kim et al., 2003) with the use of Formula (5):

\[
(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots
\]  

The attributes of the above decomposition can be used to define the elements of a variance-covariance matrix of the random effect in reduced form (Formula (6)) (Anselin, 1998; Suchecki, 2010):

\[
E(ee^T) = \sigma^2 [(I - \rho W)^{-1} [(I - \rho W)^{-1}]^T] 
\]  

Spatial lag models cannot be estimated by the least squares method due to estimator loading, and have to be described by maximum likelihood estimation (Anselin, 2009). Spatial lag \( Wy \) is interpreted as the value of dependent variable \( y \) in neighboring regions. If the value of \( Wy \) is significant, the value of variable \( y \) in region \( i \) is influenced by the level of the analyzed phenomenon in the neighboring regions \( j \) and by the remaining explanatory variables (Kopczewska, 2006).

2.3 Spatial error model

The general form of the spatial error model is given by Formulas (7) and (8) (Anselin, 1998; Wall, 2004; Arbia, 2006; Kopczewska, 2006; Ligas, 2008):

\[
y = \beta X + u
\]

\[
u = \lambda Wu + e
\]

where:

\( y \) – vector of the observed explained variable (real estate price, value) (n x 1), 
\( X \) – matrix of independent variables (real estate attributes) (n x k+1), 
\( \beta \) – parameter vector of coefficients (k+1 x 1),
The original random effect \( u \) is a function of a “true” random effect that meets standard assumptions \( \zeta \sim \mathcal{N}(0, \sigma^2I) \) (Anselin, 1998), and is described by Formulas (9) and (10):

\[
(I - \lambda W) u = e \quad (9) \\
\quad u = (I - \lambda W)^{-1} e \quad (10)
\]

If matrix \((I-\lambda W)\) is invertible, the SEM can be reduced with the use of Formula (11) (Anselin, 1998; Osland, 2010; Cellmer, 2014):

\[
y = \beta X + ((I - \lambda W)^{-1} e \quad (11)
\]

The value of the explained variable in every location is influenced by error \( \zeta \) and the multiplication factor \((I - \lambda W)^{-1}\). The lower the value of the spatial autocorrelation coefficient, the weaker the effect of the multiplication factor. The above formulas can be used to derive Formula (12):

\[
y = \lambda W y + \beta X - \lambda WX\beta + e \quad (12)
\]

Unlike the spatial lag model, the spatial error model accounts for the effects of spatial lag \( \lambda WX\beta \) of exogenous variables (Cellmer 2014).

In a spatial error model, parameter \( \beta \) is estimated by the generalized least squares (GLS) method, and parameter \( \lambda \) is estimated by optimization. The spatial autocorrelation of the residuals is also determined (Kopczewska, 2006). Both types of spatial correlations can be presented in a single model. A general spatial process model is given by Formulas (13) and (14):

\[
y = \rho W_1 y + X\beta + u \quad (13) \\
u = \lambda W_2 u + e \quad (14)
\]

The above model accounts for the effects of the spatial lag \( \rho W_1 \) of variable \( y \) and the autocorrelation of random effect \( \lambda W_2 u \). Spatial weight matrices \( W_1 \) and \( W_2 \) correspond to the neighborhood matrices of the autoregressive process and the random effects covariance matrix (Arbia, 2006). If matrix \((I - \lambda W_2)\) is invertible, the model is described by Formula (15):

\[
y = \rho W_1 y + \beta X + (I - \lambda W_2)^{-1} e \quad (15)
\]

The model can be reduced with the use of Formula (16):

\[
y = (I - \rho W_1)^{-1} \beta X + (I - \rho W_1)^{-1}(I - \lambda W_2)^{-1} e \quad (16)
\]

The model is transformed similarly to the spatial lag model, and it is expressed by Formula (17) (Anselin, 1998):

\[
y = \beta X + \rho W_1 y + \lambda W_2 y + \lambda W_2 X\beta - \lambda \rho W_1 W_2 y + e \quad (17)
\]

### 3. Data and Methods

The analysis involved spatial vector data and a database of real estate transactions for the city of Olsztyn (north-eastern Poland, Region of Warmia and Mazury, Olsztyn county, Olsztyn municipality) that were acquired from the Department of Geodesy and Real Estate Management of the Olsztyn City Office.

The factors influencing the prices and values of real estate on the Olsztyn market were analyzed, and variables were selected and described. The collected spatial vector data described the attributes of vacant land plots (zoned for residential construction) that are most often included in real estate appraisals and can potentially influence the prices and values of real estate: distribution of utility networks (power network, water supply network, gas network, telecommunications network, sewer network), urban planning attributes – designation in the local zoning plan, public roads, railway lines, development density, forests, water bodies (lakes, main rivers), distribution of public transportation stops, location of public utility services, potential supply – vacant land that can be potentially traded on the local market, distance from the city center, noise levels – acoustic map of Olsztyn. The collected data described a total of 15 real estate attributes. The period of analysis covered 2004 to 2015.
number of transactions in each analyzed year ranged from 40 in 2012 to 221 in 2011, with an average number of 77 transactions per year. Real estate that was not traded on the free market was excluded from the analysis, including land plots for public road projects, land occupied by ditches, ponds, trees, shrubs and forests. Only real estate supporting an objective evaluation of the local market was selected for the study. Data concerning 932 residential property transactions were ultimately included in the analysis. The collected data were integrated and spatial analyses were conducted with the use of ArcGIS 10.0 software. Calculations were performed in Microsoft Excel 2007. Open Street Map and Geoportal.gov.pl services were also used to obtain additional information about the analyzed market.

Fifteen variables describing each traded real estate were adopted for the analysis: noise levels ("noise", in dB), distance from public transportation stops ("stops", Euclidean distance in meters), distance from the city center ("center", junction of Pieniężnego Street and Pilskiego Avenue; Euclidean distance in meters), distance from a forest ("forest", Euclidean distance in meters), as well as variables described by kernel density estimation: gas network ("gas"), water supply network ("water"), telecommunications network ("telecom"), district heating network ("heat"), development density ("density"), potential supply, vacant land that can be potentially traded on the market ("vacant"), power network ("power"), distance from a railway line ("railway"), distance from public roads ("roads") and the availability of public services ("services").

A reference layer containing information about real estate prices (PLN/m²) and attributes (attribute table) was generated in ArcGIS. A fragment of the reference layer is presented in Figure 1.

Fig. 1. A fragment of the attribute table in ArcGIS – reference database of real estate prices and attributes that potentially influence prices. Source: own elaboration.

The influence of location and other spatial attributes on the prices and values of real estate was determined with the use of a classical multiple regression model and spatial autoregressive models. The applicability of spatial autoregressive models, also referred to as spatial regression models, is determined by the presence of spatial autocorrelations.

4. Empirical results

Additive models have been most widely used in real estate valuation. However, models that sum up the analyzed attributes do not always produce optimal results on all real estate markets (Barańska 2005). Non-linear multiplicative models can estimate the results with greater accuracy. A multiplicative model, where the estimated coefficients are raised to the power of the represented attributes, always generates positive market values of real estate for each attribute and describes attribute variability with the use of a monotonic function (constantly increasing or constantly decreasing). One of the main advantages of a multiplicative model is that it presents the percent
variation in real estate value relative to the value estimated based on real estate attributes (Barańska 2005).

The model applied in the study is described by formula (18):

\[ c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot \ldots \cdot B_m^{x_m} \]  

(18)

where:
- \( c \) – unitary price or value of real estate in the database,
- \( x_1, x_2, \ldots, x_m \) – real estate attributes in the database,
- \( B_0 \) – unitary value of real estate for zero value of all attributes,
- \( B_j \) – estimated model coefficients.

To estimate the value of coefficients \( B_j \), the above function has to be transformed to a linear function by taking the natural logarithm on both sides of the equation (5.1) with the use of formulas (19), (20) and (21):

\[ \ln c = \ln B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot \ldots \cdot B_m^{x_m} \]  

(19)

\[ \ln c = \ln B_0 + x_1 \ln B_1 + x_2 \ln B_2 + \ldots + x_m \ln B_m \]  

(20)

\[ \ln c = B_0' + B_1' x_1 + B_2' x_2 + \ldots + B_m' x_m \]  

(21)

These operations generate the parameters of the model (22):

\[ B_1 = e^{B_1'} \]  

(22)

The above transformation produces a classical multiple regression model, where the explained variable is the logarithm of real estate price in PLN/m². The selected spatial attributes are the explanatory variables (Table 1). The results generated by the multiple regression model are presented in Table 1.

| Attribute   | Coefficient | Standard error | \( t \)  | \( p \) value | Influence % (multiplication factor) |
|-------------|-------------|----------------|--------|---------------|----------------------------------|
| Constant    | 5.435       | 0.125          | 43.535 | 0.000         | 228.962                           |
| Noise       | -0.003      | 0.002          | -1.759 | 0.079         | 0.998                            |
| Stops       | 0.000       | 8.09E-05       | -4.412 | 0.000         | 1.000                            |
| Center      | 0.000       | 2.69E-05       | -4.519 | 0.000         | 1.000                            |
| Gas         | -0.003      | 0.006          | -0.330 | 0.743         | 0.998                            |
| Water       | 0.027       | 0.008          | 3.302  | 0.000         | 1.027                            |
| Telecom     | -0.002      | 0.005          | -0.193 | 0.847         | 0.999                            |
| Heat        | 0.122       | 0.015          | 8.343  | 0.000         | 1.128                            |
| Density     | 0.279       | 0.535          | 0.522  | 0.603         | 1.323                            |
| Vacant      | 5.630       | 2.250          | 2.504  | 0.013         | 278.663                           |
| Power       | -0.030      | 0.005          | -6.207 | 0.000         | 0.970                            |
| Railway     | 0.015       | 0.007          | 2.367  | 0.018         | 1.014                            |
| Roads       | -0.035      | 0.023          | -1.513 | 0.132         | 0.967                            |
| Services    | -0.010      | 0.007          | -1.548 | 0.123         | 0.990                            |
| Water bodies| 0.000       | 5.95E-05       | 5.675  | 0.000         | 1.000                            |
| Forests     | 0.000       | 5.50E-05       | 2.874  | 0.004         | 1.000                            |

*Source:* own elaboration.
The percentage influence of selected attributes on real estate prices was estimated in the “Influence % (multiplication factor)” column.

The coefficient of determination measures how well the model fits the analyzed data. In the estimated multiple regression model, the coefficient of determination $R^2$ reached only 0.167, and the standard error of the estimate was determined at 46.06. Based on the values of six variables (noise, gas, telecom, density, roads and services) at the significance level of 0.05, there are not any grounds for rejecting the hypothesis that independent variables do not influence the explained variable. The plus/minus sign before the estimated parameters indicates that these attributes exert a positive or a negative effect on the explained variable, which is not always consistent with expectations (which is the case for variables such as noise, water, telecom or heat). The relatively poor results of the estimated model can be attributed to the spatial structure of the analyzed data which, in turn, can affect the accuracy of parameter estimation. The analyzed model does not support reliable predictions of the explained variable, and it should be used only as a reference point for comparison with spatial models. The presence of spatial correlations expressed by spatial autocorrelation was investigated in the next step of the analysis.

Spatial autocorrelation is tested with global Moran’s $I$ which measures the strength of correlations between the value of the analyzed variable in a given location and the value of the same variable in other locations. The presence of such correlations indicates that the variables are spatially clustered. A positive autocorrelation occurs when the values of the analyzed variable are similar (high or low) and form clusters (groups). In contrast, a negative autocorrelation occurs when high values of the observed variable are located in the proximity of low values, and when low values of the analyzed variable are situated in the proximity of high values (Suchecki 2010).

The calculated values of Moran’s $I$, including the expected value and variance, are presented in Table 2.

| I        | E(I)  | var (I) | Z(I)         | p-value |
|----------|-------|---------|--------------|---------|
| 0.195758 | 0.0011| 0.0001  | 18.8655      | 0.0000  |

Source: own elaboration.

The calculated probability density function of $Z(I)$ is presented in Figure 2.

Fig. 2. Probability density function of $Z(I)$. Source: own calculations in ArcGIS 10.0.

In Figure 2, the critical values of the test statistic at various levels of significance (0.1; 0.05; 0.01) are marked with different colors. There are no grounds for rejecting the null hypothesis when the test statistic is in the critical region. The null hypothesis should be rejected when the test statistic is outside the critical region.
Based on the results, the hypothesis postulating the absence of a global spatial autocorrelation (\( p<\alpha; \alpha=0.05 \)) should be rejected. Statements \( I > E(I) \) and \( Z(I) > 0 \) indicate that a positive autocorrelation exists between real estate prices: high prices occur in the proximity of high prices, and low prices occur in the proximity of low prices. The probability that the observed spatial correlation is accidental is low (\( Z(I)= 18.8655, p <0.001 \)). The examined phenomenon was also analyzed based on the spatial distribution of real estate prices (Fig. 3).

**Fig. 3.** Scatterplot of Moran’s I values for vacant land. *Source: own elaboration.*

The points in the first (I) and third (III) quadrant of the scatterplot represent objects/real estate surrounded by similar neighbors (similar in price and attributes). The objects in quadrant I have high prices and are surrounded by high-priced real estate, whereas the objects in quadrant III have low prices and are surrounded by low-priced real estate. The points in the second (II) and fourth (IV) quadrant denote real estate located in the proximity of non-similar neighbors. The objects in quadrant II have low prices and are surrounded by high-priced real estate, whereas the objects in quadrant IV have high prices and are surrounded by low-priced real estate.

The results of the analysis indicate that transaction prices are spatially autocorrelated; therefore, spatial regression models can be applied. The modeled parameters were estimated in the GeoDa program (developed by the Center for Geospatial Analysis and Computation at Arizona State University; https://geoplan.asu.edu/geodacenter-redirect). The results of spatial autoregressive model estimation are presented in Table 3.

**Table 3**

| Attribute | Coefficient | Standard error | \( t \) | \( p \) value | Influence % (multiplication factor) |
|-----------|-------------|----------------|--------|--------------|-------------------------------------|
| Constant  | 3.794       | 0.420          | 9.034  | 0.000        | 44.38                               |
| Noise     | -0.003      | 0.001          | -2.120 | 0.034        | 0.998                               |
| Stops     | 0.000       | 0.000          | -3.834 | 0.000        | 1.000                               |
| Center    | 0.000       | 0.000          | -3.233 | 0.001        | 1.000                               |
| Gas       | 0.003       | 0.006          | 0.278  | 0.781        | 1.003                               |
The results of spatial error model estimation are presented in Table 4.

**Table 4**

Basic statistics in the spatial error model

| Attribute      | Coefficient | Standard error | $t$  | $p$ value | Influence %  
|----------------|-------------|----------------|------|-----------|---------------
| Constant       | 5.520       | 0.164          | 33.826 | 0.000     | 249.65        |
| Noise          | -0.004      | 0.001          | -2.506 | 0.013     | 0.995         |
| Stops          | 0.000       | 0.000          | -3.848 | 0.000     | 1.000         |
| Center         | 0.000       | 0.000          | -3.532 | 0.000     | 1.000         |
| Gas            | 0.000       | 0.008          | 0.031  | 0.975     | 1.000         |
| Water          | 0.024       | 0.010          | 2.445  | 0.014     | 1.024         |
| Telecom        | -0.003      | 0.006          | -0.241 | 0.810     | 0.998         |
| Heat           | 0.118       | 0.018          | 6.654  | 0.000     | 1.125         |
| Density        | 0.265       | 0.661          | 0.400  | 0.688     | 1.304         |
| Vacant         | 0.006       | 0.004          | 1.840  | 0.066     | 1.006         |
| Power          | -0.030      | 0.006          | -5.084 | 0.000     | 0.970         |
| Railway        | 0.014       | 0.008          | 1.630  | 0.104     | 1.014         |
| Roads          | -0.024      | 0.028          | -0.830 | 0.406     | 0.976         |
| Services       | -0.012      | 0.006          | -1.856 | 0.064     | 0.988         |
| Water bodies   | 0.000       | 0.000          | 4.215  | 0.000     | 1.000         |
| Forests        | 0.000       | 0.000          | 2.456  | 0.014     | 1.000         |

*Source: own elaboration.*

Both the spatial error model and the spatial lag model contain explanatory variables that are not statistically significant. These variables differ from the variables in the multiple regression model. Development density, railways and roads were not significant, and they exerted a weak influence on transaction prices in all models, which could be attributed to the specificity of market data.

The statistical verification of the results of spatial regression model estimation is a complex process. These results were verified with the use of the appropriate statistical tests. Models estimated by the least squares method are compared with spatial error models and spatial lag models with the use of the log-likelihood function that maximizes $\log L$, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC, Schwartz information criterion), rather than the coefficient of determination.
The applied statistical tests are described by formula (23):

$$AIC = -\frac{2\ln L}{N} + \frac{2K}{N}, \quad BIC = -\frac{2\ln L}{N} + \frac{K\log(N)}{N}$$

(23)

The estimated models were analyzed based on the selected criteria, and the results are presented in Table 5.

| Model              | log L  | AIC     | BIC     | R²   |
|--------------------|--------|---------|---------|------|
| Multiple regression| -591.946 | 1215.89 | 1293.29 | 0.167|
| Spatial lag        | -583.246 | 1200.49 | 1282.73 | 0.186|
| Spatial error      | -582.413 | 1196.83 | 1274.22 | 0.190|

Source: own elaboration.

The results of the log-likelihood test indicate that the spatial error model was characterized by the highest goodness of fit. Similar observations were made based on AIC and BIC values. The spatial error model also best fits the data in the analysis based on the values of the coefficient of determination. Despite the above, all models were characterized by relatively poor goodness of fit, and only minor differences were observed between the results of statistical tests.

The analyzed variables exerted relatively varied effects on real estate prices. The spatial distribution of the varied effects exerted by selected real estate attributes on real estate prices is presented graphically in maps in Figure 4.

### Table 5

#### Analysis of the estimated models

| Model              | log L  | AIC     | BIC     | R²   |
|--------------------|--------|---------|---------|------|
| Multiple regression| -591.946 | 1215.89 | 1293.29 | 0.167|
| Spatial lag        | -583.246 | 1200.49 | 1282.73 | 0.186|
| Spatial error      | -582.413 | 1196.83 | 1274.22 | 0.190|

---

The results of the log-likelihood test indicate that the spatial error model was characterized by the highest goodness of fit. Similar observations were made based on AIC and BIC values. The spatial error model also best fits the data in the analysis based on the values of the coefficient of determination. Despite the above, all models were characterized by relatively poor goodness of fit, and only minor differences were observed between the results of statistical tests.

The analyzed variables exerted relatively varied effects on real estate prices. The spatial distribution of the varied effects exerted by selected real estate attributes on real estate prices is presented graphically in maps in Figure 4.
5. Discussion and Conclusions

In the present study, autoregressive models were used to determine spatial variations in the effects exerted by real estate attributes on real estate prices. The effects of the examined attributes were not statistically significant (p=0.05) in all cases.

Real estate market analyses usually involve real estate attributes that are most often considered in the valuation process, but the market of vacant land is characterized by numerous and highly varied attributes, including random factors, behavioral factors as well as uncertainty which is an immanent
The structural complexity of the real estate market impedes the consolidation of market attributes, which does not preclude the applicability of spatial models in real estate market analyses.

6. References

Adamczewski, Z. (2006). *Elementy modelowania matematycznego w wycenie nieruchomości. Podejście porównawcze*. Oficyna Wydawnicza Politechniki Warszawskiej.

Anselin, L. (1998). GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. *Journal of Housing Research*, 9(1), 113–133. [https://doi.org/10.1080/10835547.1998.12091930](https://doi.org/10.1080/10835547.1998.12091930)

Anselin, L. (2009). Spatial Regression, w: Fotheringham A. S., Rogerson P.A. (red.), *The SAGE Handbook of Spatial Analysis*, SAGE Publications. [https://doi.org/10.4135/9780857020130.n14](https://doi.org/10.4135/9780857020130.n14)

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer-Verlag.

Barańska, A. (2005). *Estymacja parametrów nieliniowych modeli funkcyjnych dla potrzeb predykcji rynkowej wartości nieruchomości. UWND AGH*.

Barańska, A. (2010). *Statystyczne metody analizy i weryfikacji proponowanych algorytmów wyceny nieruchomości. Rozprawy i Monografie, Wydawnictwa AGH*.

Basu, S., Thibodeau, T. (1998). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, 17 (1), 61–85.

Benjamin, J. D., Randall, S., Guttery, R. S., & Sirmans, C. F. (2004). Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. *Journal of Real Estate Practice and Education*, 7(1), 65–77. [https://doi.org/10.1080/10835547.2004.12091602](https://doi.org/10.1080/10835547.2004.12091602)

Bitner, A. (2007). Konstrukcja modelu regresji wielorakiej przy wycenie nieruchomości. *Acta Scientiarum Polonorum. Administratio Locorum*, 6(4), 59–66.

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.

Bowen, W. M., Mikelbank, B. A., & Prestegaard, D. M. (2001). Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth and Change*, 32(4), 466–490. [https://doi.org/10.1111/0017-4815.00171](https://doi.org/10.1111/0017-4815.00171)

Can, A., & Megbolugbe, I. (1997). Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, 14, 203–222. [https://doi.org/10.1023/A:1007744706720](https://doi.org/10.1023/A:1007744706720)

Cellmer, R. (2014). Modelowanie przestrzenną w procesie opracowywania map wartości gruntów [Modelowanie przestrzenna w procesie opracowywania map wartości gruntów]. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego.

Charlton, M., & Fotheringham, S. (2009). *Geographically weighted regression*. National Centre for Geocomputation.

Czaja, J. (2001). *Metody szacowania wartości rynkowej i katastralnej [Methods of estimating market and cadastral values]*. Komp-System.

Czaja, J., Dąbrowski, J. (2008). Special algorithms for assessing market value of real estates. *Geomatics and Environmental Engineering*, 2(2).

Dąbrowski, J. (2011). Zastosowanie w modelach regresji wielorakiej wartości czynnikowych do analizy rynku nieruchomości. *Studia i Materiały Towarzystwa Naukowego Nieruchomości*, 19(3), 162–169.

Goodchild, M. F., & Haining, R. P. (2003). GIS and spatial data analysis: Converting perspectives. *Papers in Regional Science*, 83, 363–385. [https://doi.org/10.1007/s10110-003-0190-y](https://doi.org/10.1007/s10110-003-0190-y)

Hozer, J. (2001). Regresja wieloraka a wycena nieruchomości [Multiple regression and property valuation]. *Rzeczoznawca Małżonkowy*, 2, 13–14.

Isakson, H. R. (1998). The Review of Real Estate Appraisals Using Multiple Regression Analysis. *Journal of Real Estate Research*, 15(2), 177–190. [https://doi.org/10.1080/10835547.1998.12090922](https://doi.org/10.1080/10835547.1998.12090922)

James, V., Wu, S., Gelfand, A., & Sirmans, C. F. (2005). Apartment Rent Prediction Using Spatial Modeling. *Journal of Real Estate Research*, 27(1), 105–136. [https://doi.org/10.1080/10835547.2005.12091148](https://doi.org/10.1080/10835547.2005.12091148)

Kisiała, W. (2016). Modele regresji przestrzennej w badaniach czynników korzystania ze świadczeń ratownictwa medycznego [Spatial regression models in the study of emergency medical services utilization factors]. Bogucki Wydawnictwo Naukowe.
Kobylińska, K., Cellmer, R., Żróbek, S., & Lepkova, N. (2017). Using kernel density estimation for modelling and simulating transaction location. *International Journal of Strategic Property Management, 21*(1), 29–40. https://doi.org/10.3846/1648715X.2016.1250837

Kopczewska, K. (2006). *Ekonomia i statystyka przestrzenna z wykorzystaniem programu R CRAN*. CeDeWu, Warszawa.

Ligas, M. (2006). Przestrzenne modele autoregresji w zastosowaniu do wyceny nieruchomości. *Studia i Materiały Towarzystwa Naukowego Nieruchomości, 14*(1), 123–136.

Lis, C. (2005). Ekonometryczne modele cen transakcyjnych lokali mieszkalnych. *Zeszyty Naukowe Uniwersytetu Szczecińskiego Nr 415. Prace Katedry Ekonomii i Statystyki, 16*, 161-174.

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2005). *Geographic Information Systems and Science*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester.

Mark, J., & Goldberg, M. A. (1988). Multiple regression analysis and mass assessment. A review of the issues. *The Appraisal Journal, 56*, 89-109.

Osland, L. (2010). An Application of Spatial Econometrics in Relation to Hedonic House Price Modelling. *Journal of Real Estate Research, 32*(3), 289–320. https://doi.org/10.1080/10835547.2010.12091282

Sawilow, E. (2010). Problematyka określania wartości nieruchomości metodą analizy statystycznej rynku [Problems of determining the value of real estate using the method of statistical market analysis], *Studia i Materiały Towarzystwa Naukowego Nieruchomości, 8*(1), 21-32.

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). *The Composition of Hedonic Pricing Models*. *Journal of Real Estate Literature, 13*(1), 1–46. https://doi.org/10.1080/10835547.2005.12090154

Suchecka, J. (ed.). (2014). *Statystyka przestrzenna. Metody analizy struktur przestrzennych*. Wydawnictwo C. H. Beck, Warszawa.

Suchecki, B. (ed.). (2010). *Ekonomia przestrzenna. Metody i modele analizy danych przestrzennych*. C.H. Beck, Warszawa.

Tu, Y., Sun, H., & Yu, S. (2007). Spatial autocorrelations and Urban housing market segmentation. *The Journal of Real Estate Finance and Economics, 34*, 385–406. https://doi.org/10.1007/s11146-007-9015-0

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference, 121*, 311–324. https://doi.org/10.1016/S0378-3758(03)00111-3

Wang, K. (ed.), & Wolverton, M. L. (ed.). (2002). *Real Estate Valuation Theory*. Springer. https://doi.org/10.1007/978-1-4615-0909-7

Wilhelmsson, M. (2002). Spatial Model in Real Estate Economics. *Housing Theory and Society, 19*, 92–101. https://doi.org/10.1080/140360902760385646

Won Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management, 45*, 24–39. https://doi.org/10.1016/S0095-0696(02)00013-X