ISBN 978-608-4904-04-5(електронско издание)
ISBN 978-608-4904-03-8(печатено издание)
UDC: 519.233.5:332.6]:[725:528.44(497.711)

A GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH IN REGIONAL MODEL FOR REAL ESTATE MASS VALUATION

Natasha Malijanska, Sanja Atanasova, Gjorgji Gjorgjiev, Igor Peshevski, Daniel Velinov

Abstract. Real estate mass valuation models of a market value have a tendency to generate real estate property values as close as to the real market values. Property valuation theory, as one of the primary factors influencing property value, considers location. The main statistical tool used for modelling in this investigation is geographically weighted regression. More precisely, the paper is striving to establish a mass valuation real estate property model considering the implementation of spatial data as a significant factor in determining the market value of condominiums in Skopje.

1. INTRODUCTION

The great importance of real estate, both in economic as well as in social life creates a need for trustworthy data about its own value, which will be helpful in making decisions during its management and usage.

The value, in the publication Uniform standards of professional appraisal practice by the Appraisal Foundation, is defined as "*the monetary relationship between properties and those who buy, sell, or use those properties*". Value expresses an economic concept. As such, it is never a fact, that is, it is always an opinion about the value of the property at a given time in accordance with a certain definition of value. Real estate appraisal or property valuation is *"the act or process of developing an opinion of value of the property"*, [1].

It is important to distinguish the term market value from the term market price, which is the amount for which real estate is sold on a certain date. In addition to the market, investment, liquidation value, value according to the principle of continuity and many other types of real estate value can be also estimated. The market value by the International Valuation Standards Council in their publication International Valuation Standards is defined as *"the estimated amount for which an asset or liability should exchange on the valuation date between a willing buyer and a willing seller in an arm's length transaction, after proper marketing and where the parties had each acted knowledgeably, prudently and without compulsion"*, [8].

Regarding the method of valuation, i.e., the number of real estates that are apprised, there is an individual and mass valuation.

2010 *Mathematics Subject Classification.* Primary: 62H10, Secondary: 62P30, 62J05, 60E99.

Key words and phrases. Geographically weighted regression approach, multivariate regression, sensitive analysis, real estate mass valuation.

The individual valuation is an estimate of the value specifically intended for individual real estate, taking into account its specific characteristics and referring to a specific date. Unlike individual valuation, mass valuation is a process of valuing a group of real estate, on a given date, using common data, applying standardized methods and conducting statistical tests to ensure unity and equality in valuation. When assessing a large number of real estates, it is difficult to emphasize each of their qualities, so special attention is paid to defining what is common to all real estate that is valued, i.e., significant factors for their value. The mass valuation, by the Appraisal Foundation in their publication Uniform Standards of Professional Appraisal Practices, is defined as a *"process of valuing a universe of properties as of a given date using standard methodology, employing common data and allowing for statistical testing"*, [1].

Mass valuation is based on the same basic principles as individual valuation. However, mass valuation includes many real estates for a certain date, which is why mass valuation techniques include equations, tables, and plans, collectively called models.

Mass valuation models attempt to represent the market for a certain type of real estate in a particular area. The structure of such models can be seen as a two-step process:

- Model specification and
- Model calibration.

The model specification provides a framework for simulating supply forces and real estate market demand. This step involves selecting the variables of supply and demand, that need to be considered and defining their correlation towards the value as well as their own correlation. Model calibration is the process of adjusting the mathematical model for mass valuation, the tables, and the estimates for the current market. The structure of the model can be valid for several years, but it is usually calibrated or updated each year. For longer periods, a complete market analysis is required, [3]. The purpose of the mass valuation is to reflect the current conditions in the local market.

When specifying the mass valuation model, firstly the variables are identified (supply and demand) that can impact the value of the real estate and then they are defined as mathematical conversions such as logarithms, which are often used to transform nonlinear data. At the same time, the mathematical form of the model is defined. It can be used in linear (additive) and nonlinear (including multiplier) forms. Next, the model is calibrated, i.e., the data are analysed so we can determine the adjustments or the coefficients that represent the contribution to the value of the real estate of the selected variables.

The construction of the models requires a good theoretical foundation, data analysis, and research methods. The best valuation models are expected to be accurate, rational, and explainable. Regression analysis is one of the most used methods in statistics, it is used for understanding, modelling, predicting, and explaining complex phenomena. In regression analysis, the predicted variable is called a dependent variable, and the variables used for prediction are called independent variables. Regression analysis allows the creation of a model for predicting the values of a dependent variable, based on the values of other independent variables or only one independent variable.

Building a regression model is an iterative process that involves finding effective independent variables to explain the dependent variable we are trying to model or understand. By repeating the regression procedure, we determine which variables are effective predictors, and then we constantly subtract and/or add variables until we find the best possible regression model. The process of building a model is a research process. It is necessary to identify explanatory variables in consultation with theory, experts in the field, and based on common sense. We need to be able to state and justify the expected relationship between each explanatory variable and the dependent variable before the analysis, and we need to question the models where these relationships do not match.

The first law of geography, given by Waldo Tobler, is that "*everything is related to everything else, but near things are more related than distant things*", [12]. Foundations of many spatial statistical methods are based on this law. Geographically weighted regression (GWR) is a method used in spatial statistical analysis, discovering geographical variations in the relationship between a response variable and a set of covariates. GWR has been applied in a variety of disciplines and studies, aided by the increased availability of geo-referenced data at finer scales, and by an appreciation that global regression models can mask substantively important departures from average trends at local levels.

Based on the established infrastructure related to the mass valuation of real estate, the aim of this research paper is defined as the first attempt to establish a model for mass valuation of real estate for parts of the city of Skopje, while explicitly incorporating the spatial factor. The research also focuses on the application of data that the Agency of Cadastre, registers as real estate transactions that take place within the state, and which are the only official and relevant data source. Considering that in the value of the real estate, and consequently in the assessment of the value, the location has a great impact, the intention is to base the research on GeoInformation systems with which the spatial factor will be easily implemented in the model as well as the control of this component will be more extensive.

The rest of the paper is arranged as follows. In Section 2, the basic concerning geographically weighted regression are provided. Central part in this paper is Section 3. In this section are given two different models of real estate mass valuation, both using geographically weighted regression. The impacts of the age of the building and garage area are separately depicted and analysed. At the end, certain comparison of mass valuation models performance is given.

2. GEOGRAPHICALLY WEIGHTED REGRESSION

The main objective of spatial analysis is to identify the nature of the relationships that variables exhibit, [5-7]. Usually, this is made by calculating statistics or estimating parameters with observations taken from different spatial units across a study area, [9-12]. The obtained statistics or estimates of the parameter are assumed to be constant across space although this might be a very questionable assumption to make in many circumstances. In general, it is reasonable to assume that there might be intrinsic differences in relationships over space or that there might be some problem with the specification of the model from which the relationships are being measured and which manifests itself in terms of spatially varying parameter estimates. In either case it would be useful to have a means of describing and mapping such spatial variations as an exploratory tool for developing a better understanding of the relationships being studied, [2].

The most used model in geographical analysis is the model of simple linear regression. Using this technique, a particular variable (the dependent variable), is modeled as a linear function of a set of independent or predictor variables. The model states as follows:

$$y_i = a_0 + \sum_{k=1}^{m} a_k x_{ik} + \varepsilon_i \tag{1}$$

where y_i is the ith observation of the dependent variable, x_{ik} is the ith observation of the kth independent variable, ε_i are independent normally distributed error terms with zero means, and each a_k are determined from the observations. The number of observations is n. Using the least squares method, a_κ, $k = 1,2,...,m$ are estimated. In context of matrices, the upper equation can be written as

$$\hat{a} = (x^t x)^{-1} x^t y \tag{2}$$

where the independent observations are the columns of x and the dependent observations are the single column vector y. The column vector $\hat{a}$ contains the coefficient estimates. Each of these estimates can be looked of as a “rate of change” between one of the independent variables and the dependent variable. For example, if y were agreed condominiums prices, and x contained several variables related to the attributes of the condominiums and its surrounding environment, coefficients could be used to estimate the change in condominiums price for an extra square meter of garage, an extra bedroom, or the condominiums being located one kilometer closer to the nearest school.

Note that these rates of change are assumed to be universal. Wherever an apartment is located, for example, the marginal price increase associated with an additional bedroom is fixed. It might be more reasonable to assume that rates of change are determined by local culture or local knowledge, rather than a global utility assumed for each commodity. Returning to the example, the value added for an additional bedroom might be greater in a neighborhood populated by families with children where extra space is likely to be viewed highly beneficial in a neighborhood populated by singles or elderly couples, in which case extra space might be viewed as a negative feature. These variations in relationships over space, such as those described above, are referred to as spatial nonstationarity, [2].

Geographically weighted regression (GWR) addresses problems like the one in the previous paragraph. It is a relatively simple technique, extending the traditional regression framework of equation (1). Local variations in rates of change are allowed, so that the coefficients in the model are specific to a location i, rather than being global estimates. The regression equation in this case is given by

$$y_i = a_{i0} + \sum_{k=1}^{m} a_{ik} x_{ik} + \varepsilon_i \tag{3}$$

where a_{ik} is the value of the k th parameter at location i. Note that (1) is a special case of (3), by putting all of the functions are constants across space. As will be shown below, the point i at which estimates of the parameters are obtained is completely generalizable and need not only refer to points at which data are collected. Using GWR, it is quite easy to compute parameter estimates. For instance, for locations lying between data points, which makes it possible to produce detailed maps of spatial variations in relationships. Although the model in equation (3) appears to be a simple extension of (1), a problem with calibrating (3) is that the unknown quantities are in fact functions mapping geographical space onto the real line, rather than simple scalars as in (1). In a typical data set, samples of the dependent and independent variables are taken at a set of sample points and it is from these that the parameters must be estimated. In the traditional model, these estimates are constant for all i but in equation (3) this is clearly not the case. For model (3), it seems intuitively appealing to base estimates of a_{ik} on observations taken at sample points close to i. If some degree of smoothness of the a_{ik}, $k = 1, 2, ..., m$ is assumed, then reasonable approximations may be made by considering the relationship between the observed variables in a region geographically close to i.

By the use of a weighted least squares approach to calibrating regression models, different emphases can be placed on different observations in generating the estimated parameters. In ordinary least squares, the sum of the squared differences of predicted and actual y_i, is minimized by the coefficient estimates. In weighted least squares a weighting factor w_i is applied to each squared difference before minimizing, so that the inaccuracy of some predictions carries more of a penalty than others. If w is the diagonal matrix consisting of all w_i, then the estimated coefficients satisfy

$$\tilde{a} = (x^t wx)^{-1} x^t wy \,. \tag{4}$$

In Geographically weighted regression, weighting an observation in accordance with its proximity to i would allow an estimation of a_{ik} to be made that meets the criterion of "closeness of calibration points" set out above. Note that usually in weighted regression models the values of w_i are constant, so that only one calibration has to be carried out to obtain a set of coefficient estimates. In this case w varies with i, a different calibration exists for every point in the study area. In this case, the parameter estimation formula could be written more generally as

$$a(i) = (x^t w(i)x)^{-1} x^t w(i) y \,. \tag{5}$$

Comparing this method and that of kernel regression and kernel density estimation, we can say the following: In kernel regression, y is modeled as a nonlinear function of x by weighted regression, with weights for the i th observation depending on the proximity of x and x_i, for each i with the estimator being

$$\tilde{a}(x) = (x^t w(x)x)^{-1} x^t w(x) y \,. \tag{6}$$

The main difference between the two methods is that in (6), kernel regression, the weighting system depends on the location in "attribute space" of the independent variables, whereas in geographically weighted regression (see (5)) it depends on location in geographical space. The output in (5) is typically a set of localized parameter estimates in x space so that highly nonlinear and nonmonotonic relationships between y and x can be modeled. The typical output in (6) is a set of parameter estimates that can be mapped in geographic space to represent nonstationarity or parameter "drift", [2].

From the above, w_i is a weighting scheme established on the proximity of i to the sampling locations around i, without an explicit relationship being stated. The choice of such a relationship will be considered in continuation. Firstly, consider the implicit weighting scheme of (2). Here

$$w_{ij} = 1, \tag{7}$$

for all i and j. Here j represents a specific point in space at which data are observed and i represents any point in space for which parameters are estimated. This means that, in the global model each observation has a weight one. An initial step toward weighting based on locality might be to exclude from the model calibration observations that are further than some distance d from the locality. This is equivalent to putting their weights to be zero, giving a weighting function by

$$\begin{cases} w_{ij} = 1, & d_{ij} < d \\ w_{ij} = 0, & otherwise \end{cases}. \tag{8}$$

The use of (8) allows efficient computation, since for every point for which coefficients are to be computed; only a subset (often quite small) of the sample points need to be included in the regression model. Hence, the spatial weighting function in (8) suffers the problem of discontinuity. As i vanes around the study area, the regression coefficients could vary drastically as one sample point moves into or out of the circular buffer around i and which defines the data to be included in the calibration for location i. Although instant changes in the parameters over space might genuinely occur, in this case changes in their estimates would be artifacts of the arrangement of sample points, rather than any underlying process in the phenomena under investigation. One way to address this problem is to make w_{ij} a continuous function of d_{ij}, where d_{ij} is the distance between i and j. In this case, it can be seen from (5) that the coefficient estimates would then vary continuously with i. A straightforward choice for the weight function w_{ij} might be

$$w_{ij} = \mathrm{e}^{-\beta d_{ij}^{2}}, \tag{9}$$

so that if i is a point in space at which data are observed, the weighting of that data will be unity and the weighting of other data will decrease according to a Gaussian curve as the distance between i and j increases. In the latter case the inclusion of data in the calibration procedure becomes "fractional." For example, in the calibration of a model for point i, if $w_{ij} = 0,5$, then data at point j contribute only half the weight in the calibration procedure as data at point i itself. For data far away from i the weighting will be asymptotically zero, effectively excluding these observations from the estimation of parameters for location i.

Adjustments of (8) and (9) may be made, having the computationally desirable property of excluding all data points greater than some distance from

i and also the desirable property of continuity. An example is the bisquare function given by

$$w_{ij} = \begin{cases} (1 - d_{ij}{}^2 / d^2)^2, & d_{ij} < d \\ 0, & otherwise \end{cases}. \tag{10}$$

This excludes points outside radius d, but tapers the weighting of points inside the radius, so that w_{ij} is a continuous and once differentiable function for all points less than d units from i.

Whatever the specific weighting function employed, the essential idea of Geographically weighted regression is that for each point i there is a "bump of influence" around i corresponding to the weighting function in a way that sampled observations close to i have more influence in the estimation of i's parameters than do sampled observations farther away.

The following problem occurs when use GWR: The estimated parameters are, in part, functions of the weighting function or kernel selected in the method. In (8)**,** for example, as d becomes larger, the closer will be the model solution to that of OLS and when d is equal to the maximum distance between points in the system, the two models will be equal. Equivalently, in (9) as β tends to zero, the weights tend to one for all pairs of points so that the estimated parameters become uniform and GWR becomes equivalent to OLS. Conversely, as the distance- decay becomes greater, the parameter estimates will increasingly depend on observations in close proximity to i and hence will have increased variance. The problem is therefore how to select an appropriate decay function in GWR. Consider the selection of β in (9), one possible solution is β to be chosen on a least squares criteria. If the error terms in (3) are assumed to be Gaussian, then this also fulfills a maximum likelihood criterion. Hence, the way to proceed would be to minimize the quantity

$$\sum_{i=1}^{n} (y_i - y_i^*(\beta))^2, \tag{11}$$

where $y_i^*(\beta)$ is the fitted value of y_i using a distance-decay of β. For the sake of finding the fitted value of y_i, it is necessary to estimate the a_{ik}'s at each of the sample points and then combine these with the z-values at these points. However, when minimizing the sum of squared errors suggested above, a problem is encountered. Let β was made very large so that the weighting of all points except for i itself become negligible. Hence the fitted values at the sampled points will tend to the actual values, so that the value of (11) tends to zero. This means that under such an optimizing criterion, the value of β tends to infinity, but clearly this degenerate case is not useful. First, the parameters of such a model, are not defined in this limiting case. Second, the estimates will fluctuate wildly throughout space in order to give locally good fitted values at each i.

The *cross-validation* (CV) approach suggested for local regression by Cleveland (1979) and for kernel density estimation by Bowman (1984), is a solution to this problem. Here, a score of the form

$$\sum_{i=1}^{n}(y_i - y_{\neq i}^{*}(\beta))^2$$

is used where $y_{\neq i}^{*}(\beta)$ is the fitted value of y_i with the observations for point i omitted from the calibration process. This approach has the desirable property of countering the wrap-around effect, since when becomes very large, the model is calibrated only on samples near to ***i*** and not at i itself. Plotting the CV score against the required parameter of whatever weighting function is selected will therefore provide guidance on selecting an appropriate value of that parameter. If it is desired to automate this process, then the CV score could be maximized using an optimization technique such as a Golden Section search, [2].

3. MODELLING WITH GEOGRAPHICALLY WEIGHTED REGRESSION

According to the theoretical settings, experience, available research and data made available from the *Register of Leases and Real Estate Prices*, a set of proposed explanatory variables has been identified that are considered to determine the market value of the real estate. Despite the good reasons for including any available real estate data as variables in the model, it was found that some of the explanatory variables were statistically significant and some were statistically insignificant. For this reason, statistical tests have been conducted to make a number of possible combinations of proposed input explanatory variables, requiring models that best explain the dependent variable and thus perform the model specification. The analysis of the proposed explanatory variables gave the results shown in the table below. Also, through the statistical analysis, multicollinearity is calculated between the explanatory variables, i.e., VIF value. In which the value taken as a limit is the value 7.5, i.e., if the VIF value is less than 7.5 there is no multicollinearity between the explanatory variables.

The following table shows the result for significance and multicollinearity based on the analysis of the explanatory variables.

Table 1: Result of the analysis of variables

Summary of variable significance				Multicollinearity
Variable	**Significant**	**Negative**	**Positive**	**VIF**
Area	100	0	100	1.69
Garage (area)	100	0	100	1.19
Distance to closet mall	100	100	0	4.17
Age	98.07	100	0	1.67
Elevator	87.67	0	100	1.64
Distance to closest university	85.09	99.14	0.86	2.97
Distance to school	79.93	0	100	1.33
Balcon area	74.53	0.02	99.98	1.19
Floor number	73.79	0	100	1.21
High quality interior	69.43	0	100	1.03
Distance to closest park	65.00	62.99	37.01	3.71
Rooms	60.68	18.84	81.16	1.47
Own heating system	60.21	8.77	91.23	1.93
Distance to closest hospital	54.98	16.58	83.42	1.90
Distance to closest kinder garden	44.19	42.45	57.55	1.38
Distance to city centre	43.51	48.20	51.80	2.56
Basement area	39.99	77.79	22.21	1.30
Communal heating system	33.04	24.30	75.70	2.32
Distance to closest bus station	22.33	72.18	27.82	1.33

The results obtained from the analysis of the explanatory variables show that there is a high significance of certain structural, but also spatial characteristics for the real estate that is subject to transaction. It can also be noted that we do not have a redundant explanatory variable, i.e., there is no multicollinearity between the explanatory variables. In the process of defining an appropriate model, it is necessary to experiment with different variables to explain the value of the real estate. It is important to be aware that the coefficients of the explanatory variables (and their statistical importance) may change radically depending on the combination of variables we include in the model.

For the purposes of the research, two models were created with GWR, while for assessing the quality of the created models, the statistical parameters R^2, adjusted $R^2$and Akanke's Information Criterion (AICc) were used. R^2 and adjusted R^2 are statistically derived from the regression equation to quantify model performance. The value of R^2 ranges from 0 to 1. If the model explains the dependent variable perfectly R^2 is 1.0. As an example, if you get a value of R^2 of 0.49, it can be interpreted with the words: "the model explains 49 percent of the variations in the dependent variable". Adjusted R^2 is always slightly lower than the value for R^2, as it reflects the complexity of the model (number of variables). Consequently, the adjusted R^2 is a more accurate measure of model performance. The Akaike information criterion (AIC) is an estimator of prediction error and thereby the relative quality of statistical models for a given set of data. AIC estimates the relative amount of information lost by a given model: the less information a model loses, the higher the quality of that model.

Model 1

Model 1 created with GWR is specified only with structural features of residential property. As explanatory variables for which statistical tests showed the greatest signification are as follows: Area, Garage (area), Balcony area and Age. Using these explanatory variables, the first model for which the following statistical indicators are obtained is formed in the table below, through which we can see the success of the model.

Table 2: Results of the analysis – GWR for Model 1

OID	VARNAME	VARIABLE	DEFINITION
0	Bandwidth	2137.860192	
1	ResidualSquares	54169319433	
2	EffectiveNumber	21.807198	
3	Sigma	8063.135068	
4	AICc	17821.981138	
5	R2	0.794736	
6	R2Adjusted	0.78961	
7	Dependent Field	0	PRICE_EU
8	Explanatory Field	1	AREA
9	Explanatory Field	2	AREA_BALCO
10	Explanatory Field	3	AREA_GARAG
11	Explanatory Field	4	AGE

Table 3: Correlation coefficient between the projected prices by Model 1 and the actual purchase prices for the control group points

Correlations

		price_eu	Predicted
price_eu	Pearson Correlation	1	,899**
	Sig. (2-tailed)		,000
	N	95	95
Predicted	Pearson Correlation	,899**	1
	Sig. (2-tailed)	,000	
	N	95	95

**. Correlation is significant at the 0.01 level (2-tailed).

The correlation analysis between the appraised market value of the residential property that has been sold, obtained with model 1 and the actual purchase price performed in transactions for the control group of transactions, calculated with Pearson the correlation coefficient in the SPSS software for this model is 0.899, i.e., 89.9 %.

Model 2

Model 2 created with GWR uses the same structural and explanatory variables as Model 1 and supplemented by three spatial explanatory variables that the analysis showed were statistically significant: Distance to the closest mall, Distance to the closest hospital and Distance to the closest university. By applying all these explanatory variables, the following statistical indicators shown in the following table are obtained:

Table 4: Results of the analysis –GWR for Model 2

OID	VARNAME	VARIABLE	DEFINITION
0	Bandwidth	2137.860192	
1	ResidualSquares	47401929637.900002	
2	EffectiveNumber	31.111344	
3	Sigma	7585.142586	
4	AICc	17724.392795	
5	R2	0.82038	
6	R2Adjusted	0.813815	
7	Dependent Field	0	PRICE_EU
8	Explanatory Field	1	AREA
9	Explanatory Field	2	AREA_BALCO
10	Explanatory Field	3	AREA_GARAG
11	Explanatory Field	4	AGE
12	Explanatory Field	5	DIST_MAL
13	Explanatory Field	6	DIST_HOS
14	Explanatory Field	7	DIST_UNI

The correlation analysis between the appraised market value of the residential property that has been sold, obtained with model 2 and the actual purchase price performed in the transactions for the control group points, calculated with Pearson correlation coefficient in the SPSS software for this model is 0.906, i.e., 90.6%.

Table 5: Correlation coefficient between the projected prices by Model 2 and the actual purchase prices for the control group points

Correlations

		price_eu	Predicted
price_eu	Pearson Correlation	1	,908**
	Sig. (2-tailed)		,000
	N	95	95
Predicted	Pearson Correlation	,908**	1
	Sig. (2-tailed)	,000	
	N	95	95

**. Correlation is significant at the 0.01 level (2-tailed).

When calibrating mass valuation models where spatial regression models are used, they have a variable value that varies depending on the location. In order to register this variation, spatial data in raster data format is used. Hence, a significant advantage in using the GWR model and applying GeoIS is the ability to create a series of raster layers of variable coefficients. This allows the identification of spatial variations within the research area, which can help in effective decision making. Such records can provide an excellent insight into the key parameters that affect the value of the property in a particular area. For example, the age of the property can have a significant negative impact on the value of the property in newly developed areas where most of the properties are completely new, and on the other hand it can have a positive impact in an old part of the city where older buildings have architectural features and historical significance. In order to emphasize the importance of these models, the results of the age factor of the building will be presented. As expected, the age of the building is inversely proportional to the value of the property, i.e., the older construction reduces the value of the property due to obsolescence, deterioration and depreciation. The analysis of the raster data model of the coefficient for the age of the building showed that the impact of this factor varies through the field of research and less impact (lower coefficients) this factor is observed in the central area of the city, while the impact of the age of the building increases as we move away from the central urban area, to the settlements of Karposh, Aerodrom, where new buildings are being built and the demand for new buildings is higher.

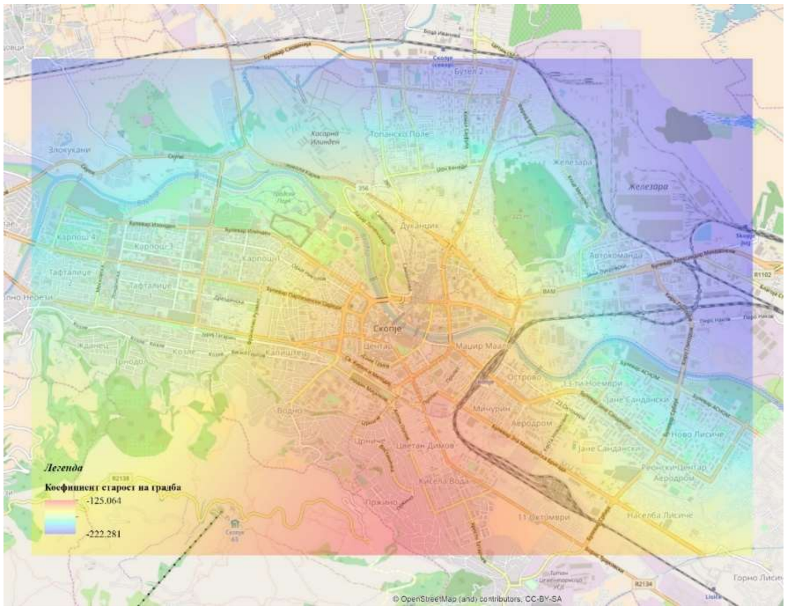

Figure 1: Value of the coefficient before the variable age

The analysis of the raster data model of the coefficient for the impact of the garage surface showed that this impact is greater in the municipalities of Centar and Karposh, while in the municipalities of Aerodrom and Chair, that impact is less, as expected, due to the existence of more and larger parking spaces.

Figure 2: Values of the coefficient in front of the variable area of the garage

4. CONCLUSIONS

Based on the results obtained from quality control of the established models for mass valuation we can conclude that both models meet the statistical checks and have a satisfactory accuracy of market value prediction. However, although they have satisfactory accuracy, it is necessary to emphasize the difference between the number and type of explanatory variables that these models incorporate and how they affect the end result.

Table 6: Comparison of mass valuation models performance

	Model 1 - GWR	*Model 2 - GWR*
Coefficient of determination – R^2	79.5%	82.0%
Akaike Information Criterion – AICc	17822	17724
Pearson correl.	89.9%	90.8%
Input data	Non-spatial	Spatial
Number of explanatory variables	4	7

Model 2 has higher R^2 coefficient, which means that the created model fits much better in the data. A higher percentage shows that the dependent variable (the value of the residential property) is better explained by the selected independent variables, while this percentage is lower in Model 1. Also, the AICc value of the first model is lower than the one of Model 2.

As for the accuracy of the prediction, which is calculated as the correlation coefficient between the projected prices of the control transactions that were omitted from the creation of the models and the actual prices of their purchase, Model 2 has a higher Pearson correlation factor than Model 1.

The results show that the use of Geographically weighted regression (GWR) in predicting market real estate values is a great basis for developing mass valuation models. In doing so, the incorporation of spatial explanatory variables can have a positive impact on real estate mass valuation models.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

ACKNOWLEDGMENTS

This research is partially supported by IMU-CDC grant, CDC Project Grants 2022.

References

[1] *Appraisal Foundation, Uniform standards of professional appraisal practice (USPAP),* Appraisal Foundation, Washington, D.C., 2016-2017.
[2] C. Brunsdon, A. S. Fotheringham, M. E. Charlton, *Geographically weighted regression: A method for exploring spatial nonstationarity*, Geographical Analysis 28, (1996), 281–298.
[3] J. Eckert, *Property Appraisal and Assessment Administration, International Association of Assessing Officers*, Chicago, Illinois, 1990.
[4] A. S. Fotheringham, T. M. Oshan*, Geographically weighted regression and multicollinearity: dispelling the myth*, Journal of Geographical Systems, 18 (4), (2016), 303-329.
[5] A. S. Fotheringham, R. Crespo, J. Yao, *Geographical and temporal weighted regression (GTWR)*, Geographical Analysis, 47 (4), (2015), 431–452.
[6] S. Fotheringham, C. Brunsdon, M. Charlton, *Geographically Weighted Regression: the analysis of spatially varying relationships*, John Wiley & Sons Ltd., West Sussex, England, 2002.
[7] R. Harris, G. Dong, W. Zhang, *Using contextualized geographically weighted regression to model the spatial heterogeneity of land prices in Beijing, China*, Transactions in GIS, 17 (6), (2013), 901–919.
[8] *IVSC, International Valuation Standards*, IVSC, London, United Kingdom, 2017.
[9] T. Kauko, M. d'Amato, *Mass Appraisal Methods. An International Perspective for Property Valuers*, Wiley-Blackwell, West Sussex, United Kingdom, 2008.
[10] Y. Leung, C. L. Mei, W. X. Zhang, *Statistical tests for spatial nonstationarity based on the geographically weighted regression model*, Environment and Planning A, 32 (1), (2000), 9–32.
[11] J. Liu, Y. Yang, S. Xu, *A geographically temporal weighted regression approach with travel distance for house price estimation*, Entropy, 18 (8), (2016), 303.
[12] W. R. Tobler, *A computer movie simulating urban growth in the Detroit region*, Economic Geography, 46, (1970), 234-240.

Faculty of Civil Engineering Skopje, Ss. Cyril and Methodius University in Skopje, Skopje, N. Macedonia
E-mail address: malijanska@gf.ukim.edu.mk

Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, Skopje, N. Macedonia
E-mail address: ksanja@feit.ukim.edu.mk

Faculty of Civil Engineering Skopje, Ss. Cyril and Methodius University in Skopje, Skopje, N. Macedonia
E-mail address: gorgi.gorgiev@gmail.com

Faculty of Civil Engineering Skopje, Ss. Cyril and Methodius University in Skopje, Skopje, N. Macedonia
E-mail address: pesevski@gf.ukim.edu.mk

Faculty of Civil Engineering Skopje, Ss. Cyril and Methodius University in Skopje, Skopje, N. Macedonia
E-mail address: velinovd@gf.ukim.edu.mk