DOI: 10.22616/j.balticsurveying.2021.15.008

RESIDENTIAL REAL ESTATE PRICE MODELLING THROUGH THE METHOD OF THE GEOGRAPHICALLY WEIGHTED REGRESSION: GOMEL CITY CASE STUDY

Zhukovskaya Natalia, Popko Olga
Belarusian State University

Abstract
One of the most challenging tasks in modelling of house pricing is to take into account the location factors. Geographically Weighted Regression (GWR) as a local regression model is an extremely effective instrument for spatial data analysis. The aim of the study is to model the relationships between a residential real estate price (per sq.m) and both building and location characteristics for Gomel using GWR. The data of the Belarus' National Cadastral Agency on real estate transactions (apartments) in Gomel in 2019 are used as initial. The global Moran I index has been used to estimate a spatial autocorrelation of the dependent variable (price per square meter of residential real estate). Several factors having the impact on the apartment sale prices have been determined. Independent variables having been used in analysis can be divided into building characteristics and spatial characteristics. The building characteristics section includes the number of rooms within the property, property area (square meters), building age, number of floors in the building, floor of the property. The spatial characteristics group contains proximity to city center, recreation areas, supermarkets, bus stops, healthcare and educational facilities. A regression model of housing price in Gomel has been developed. Mapping variable regression coefficients allows exploring spatial features of the impact of the different explanatory variables on the property price. Geographically weighted regression modelling has revealed the pricing peculiarities inherent for certain areas of the city.
Key words: Geographically Weighted Regression, spatial autocorrelation, spatial models, residential real estate market, price modelling.

Introduction

A number of qualitative and quantitative methods are used to assess the impact of property characteristics on the prices and value of real estate. Qualitative methods are more commonly used in weakly developed real estate markets. Quantitative methods generate more well-founded results, but their application presupposes the fulfillment of many formal and statistical requirements. Many studies have used a hedonic modeling approach (essentially a traditional multiple regression model formulation) to analyse the property market (Chau, Chin, 2003). Regression analysis is a statistical technique for determining one or several independent variables' impact on a dependent variable. The classical regression models used in real estate market analyses cannot take into account spatial autocorrelations of analysed data and spatial heterogenity of price formation process (Cellmer, 2012). Application of Geographically Weighted Regression (GWR) to real estate market analysis allows account spatial heterogeneity (non – stationarity) of analyzed phenomenon as is spatial distribution of prices (Cellmer et al., 2020). GWR is an extension of ordinary least squares regression that modells relationships as they vary across space by estimating regression coefficients locally at spatially referenced data points. Geographically weighted regression is being increasingly used in identifying price determinants (Dziauddin et.al., 2015; Tomal, 2020), pricing modeling (Cao et.al., 2019; Wang et.al., 2020) and real estate market segmentations (Manganelli et.al., 2014).

The object of investigation – residential real estate market of Gomel.

Gomel is the second largest city in Belarus after Minsk and is characterized by fairly high housing prices. Compared to Minsk (the capital of Belarus), the real estate market is less active, more predictable, there are no sharp price rises.

The aim of investigation – to model the relationships between a residential real estate price (per sq.m) and both building and location characteristics for Gomel during 2019 using GWR.

Methodology of research and materials

In the present study we have used data from the Belarus' National Cadastral Agency. The real estate transactions (apartments) in Gomel in 2019 have been analyzed. The total number was 1,046 purchase and sale transactions. The transaction database has been formed on the basis of information from the land plot price register of the state land cadaster. Recorded prices are the sale prices of completed transactions. Transaction data includes information about such characteristics of property as its address, construction year, wall material (bricks, large prefabricated panels, monolithic reinforced concrete, lightweight concrete blocks), building number of floors, property floor, property number of rooms, property area (square meters). The sample of purchase and sale transactions is limited to 2019 year. In 2019, there were no significant price changes in the residential real estate market. The modelling

of residential real estate pricing through the method of GWR was carried out in several steps: identifying and modelling the spatial factors influence on a price, statistical data analysis, assessment of spatial autocorrelation, conducting Geographically Weighted Regression.
Both internal and external factors affecting the residential real estate price were preliminarily identified and prepared. Internal factor group includes such characteristics as the property area, building age, number of floors in the building, floor of the property, wall material, number of property rooms. The external or spatial factor group contains proximity to the city center, recreation areas, supermarkets, bus stops, healthcare and educational facilities. The spatial factors models were constructed in the environment of ArcGIS 10.8 with using Euclidean distance tool. Units of measurement are meters. The data of the State Land Cadastre and OpenStreetMap project were used as a vector spatial basis.
Statistical data analysis includes the variable distribution assessment with the help of histograms and fitting criterions (the Kolmogorov-Smirnov test, the Shapiro-Wilk test), calculating descriptive statistics, nonparametric correlation analysis.
Spatial Autocorrelation Analysis was implemented with global and local Moran's I statistics (Anselin, 1995).
The Geographically Weighted Regression analysis was used to model the relationship between an apartment price and some building, property and location characteristics. GWR is a non-stationary technique that models spatially varying relationships. A GWR model can be written as (Fotheringham et al., 1998, 2002):

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (1)$$

where y_i is the value of the outcome variable at the coordinate location i where (u_i,v_i) denotes the coordinates of i, β_0 and β_j represents the local estimated intercept and effect of variable j for location i, respectively.
The Geographically Weighted Regression model was implemented by the ArcGIS Desktop Software (Version: 10.8; Type: Advanced).

Discussion and results
Analysis of purchase and sale transactions in Gomel secondary housing market has showed that apartments prices (measured in USD per square meters) were variable, ranging from 250 to 750 USD, with an average of 544 USD. A Histogram of housing prices with a normal distribution curve is shown in Figure 1. According to the histogram and Goodness-of-Fit criteria, the data are consistent with the normal distribution. Therefore, the raw data were used for geographical weighted regression.

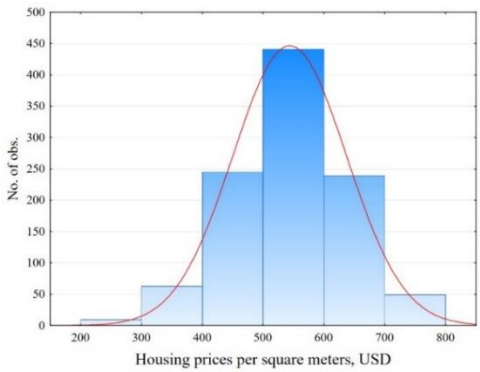


Fig. 1. Histogram of apartment prices (per square meters)

Considering the spatial distribution of apartment unit prices (measured in USD per square meters), two types of spatial association are tested: global spatial autocorrelation and local spatial association. Global spatial autocorrelation has been estimated by Moran's I statistic. It appears that apartment unit prices are positively spatially autocorrelated at p = 0.001 significance level which allows to use the geographically weighted regression for price modelling. The identification of local spatial association has been carried out by the means of local Moran's I. The results of the global Moran's I and LISA analysis are illustrated in Figure 2.
A large high–high spatial cluster is observed in the northwestern part of Gomel, mainly within Central administrative district of the city. A relatively small high-high spatial cluster is located within the Shvedskaya Gorka microdistrict in the southern part of the city while some low–low spatial clusters have been found in the middle and eastern areas of Gomel.

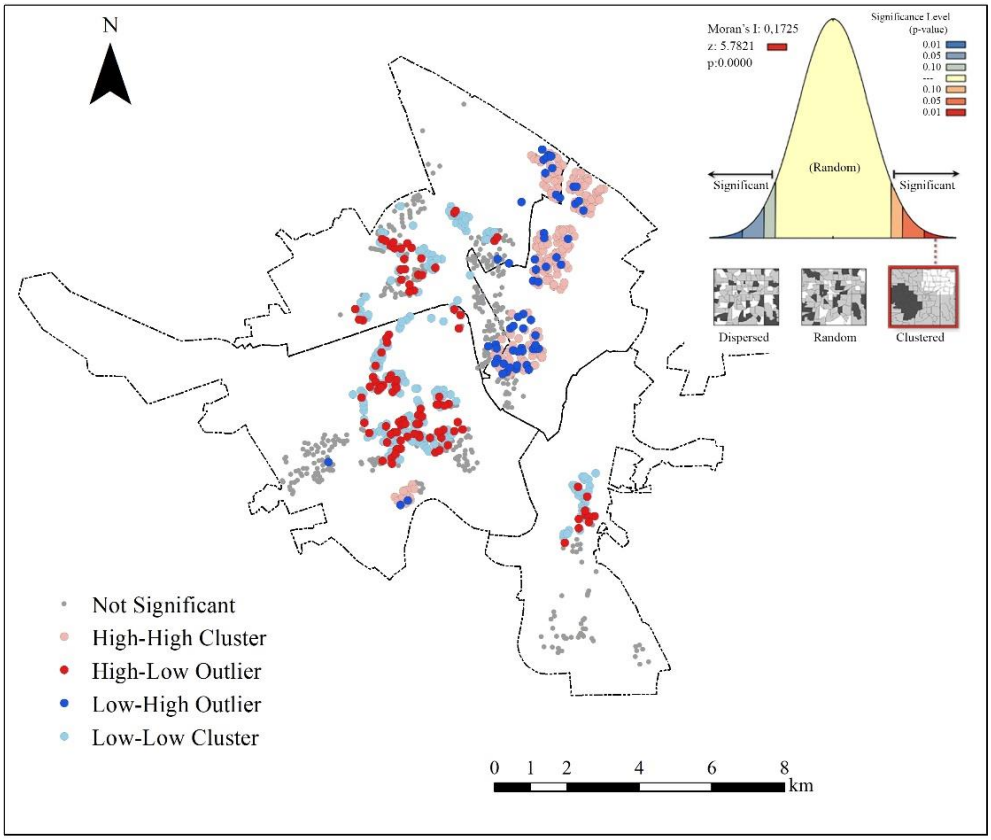


Fig. 2. Spatial clusters and spatial outliers map of apartment unit prices in Gomel

The relationships between apartment prices per m² (sqm) and building and location characteristics have been investigated with Spearman correlation coefficients. The results are summarized in Table 1.
The correlations between price per m² and eight indicators are significant: property area, building age, number of floors in the building, floor of the property, wall material and proximity to recreation areas, healthcare and educational facilities (Table 1). The characteristics having the greatest impact on the price are: building age and number of floors in the building.

Table 1

Spearman correlation coefficients between apartment prices per m²
and building and location characteristics

Variables	r
Property area (square meters)	0.064
Number of rooms	-0.039
Building age	-0.327
Number of floors in the building	0.285
Floor of the property	0.175
Wall material	0.120
Proximity to the city center	-0.030
Proximity to recreation areas	-0.141
Proximity to supermarkets	0.044
Proximity to healthcare facilities	0.182
Proximity to educational facilities	0.078
Proximity to bus stops	-0.005

Significant coefficients at the significance level 0.05 are determined in bold

Prior to establishing the final list, the independent variables were subjected to preliminary exploratory regression. There are evaluated all possible combinations of explanatory variables looking for the Ordinary Least Squares (OLS) models passing all necessary diagnostics. The summary of variable significance is presented in table 2.

Table 2

Independent variable significance (exploratory regression analysis)

Variables	Significance, %
Building age	100
Proximity to healthcare facilities	100
Proximity to recreation areas	100
Proximity to the city center	92
Proximity to supermarkets	86
Number of floors in the building	88
Wall material	82
Floor of the property	65
Number of rooms	41
Proximity to educational facilities	31
Property area (square meters)	30
Proximity to bus stops	7

To model the relationships between apartment price (per sq.m) and some building and location characteristics the Geographically Weighted Regression technique has been implemented. The explanatory variables in the model were property area (square meters), number of floors in the building, floor of the property, wall material, proximity to the city center, recreation areas, supermarkets, healthcare facilities. The inclusion of explanatory variables in the GWR model was carried out based on correlations between independent and dependent variables, the significance of the independent variables estimated by exploratory regression analysis and the absence of multicollinearity among independent variables. The Gaussian kernel is used for GWR model and the kernel type is fixed. The bandwidth is 950 meters. Table 3 presents the basic diagnostic statistics of the GWR model.

Table 3

Basic diagnostic statistics of the GWR model

Diagnostic content	Value
Number of observations	1046
Bandwidth	945
Residual squares	10099.18
Sigma	111.23
AICc	10.06
R^2	0.77
Adjusted R^2	0.74

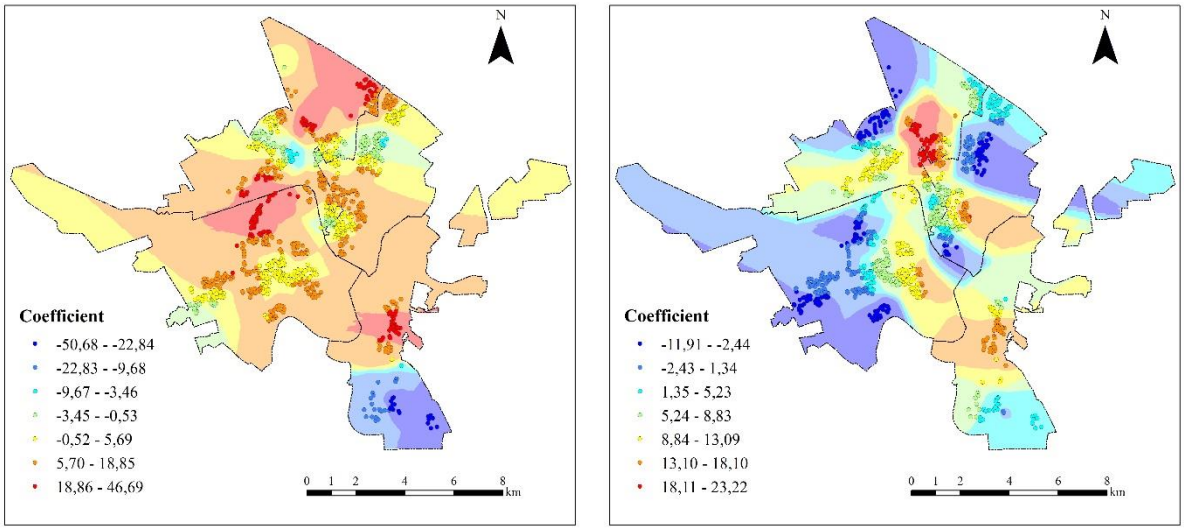
Geographically weighted regression lets the regression parameters to differ locally by disseminating location-wise parameter estimates for all independent variables (Fotheringham et al., 2002). Table 4 shows the results of the GWR model parameter estimation. A local determination coefficient R^2 varies from 0.05 to 0.97.

Table 4

Results of the GWR model parameter estimation

Variable	Min	Mean	Max
Intersept	-982.209	525.059	1187.275
Property area (square meters)	-2.649	-0.684	1.487
Building age	-7.883	-1.001	4.502
Number of floors in the building	-11.905	5.460	23.224
Floor of the property	-11.963	0.568	11.859
Wall material	-50.676	7.055	46.687
Proximity to the city center	-0.094	0.001	0.198
Proximity to recreation areas	-0.151	-0.014	0.329
Proximity to supermarkets	-0.491	0.009	0.184
Proximity to healthcare facilities	-0.134	0.010	0.242

The GWR allows visualizing and exploring the spatial variability of explanatory variables in the model. The model parameter estimates of explanatory variables may vary significantly over a geographical area. Spatial distribution of the coefficients estimates for variables most affecting the apartment`s price are presented at Figure 3. The greatest variability was observed for the variable “Wall material”. The direction and degree of the parameter influence varies throughout the city areas.



a **b**

Figure 3. The spatial distribution of the GWR coefficients:
(**a**) wall material; (**b**) number of floors in the building

The Moran's I-statistic has been calculated to measure the (global) spatial autocorrelation for GWR residuals. Small value of Moran's I (-0.0306) indicates a residuals spatial random distribution, which in turn suggests a good GWR fit (Figure 4).

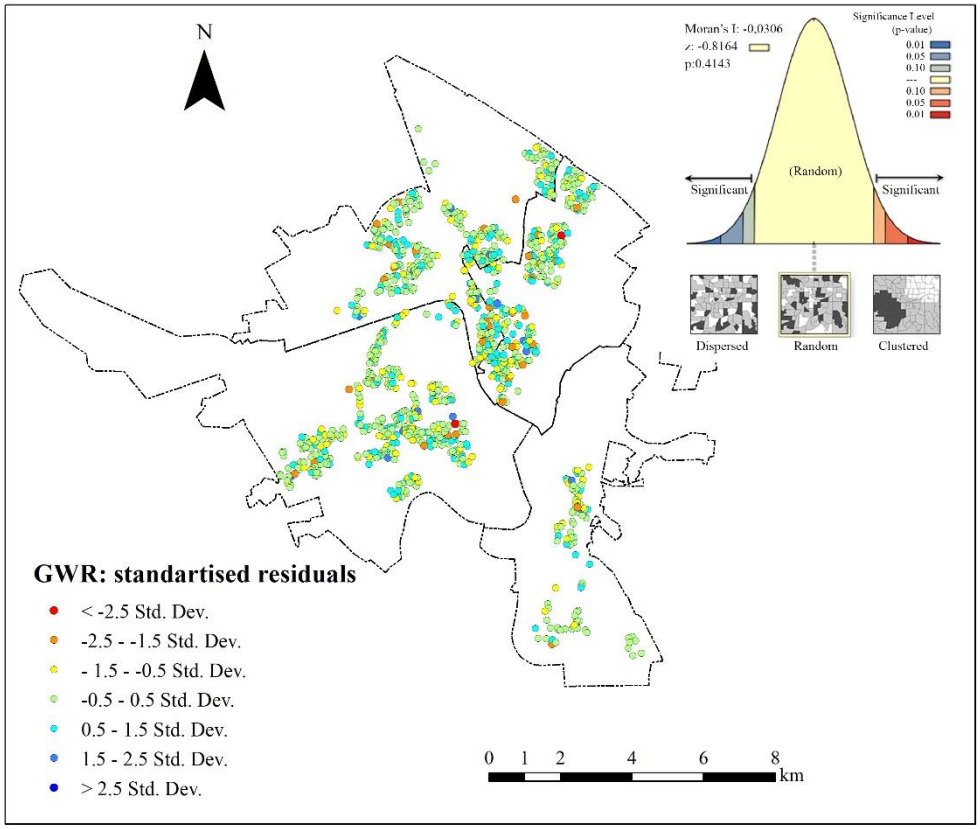


Fig. 4. Standard residual of Geographically Weighted Regression for apartment unit prices in Gomel

Conclusions and proposals

In this study, we investigated the relationships between residential real estate price (per sq.m) and both building and location characteristics for Gomel during 2019 using GWR. Geographically Weighted Regression is local regression technique which takes into account spatial heterogeneity. Several factors having the impact on the apartment sale prices have been determined. The final GWR model includes explanatory variables based on two categories: building and apartment characteristics (property area (square meters), building age, number of floors

in the building, floor of the property, wall material) and location characteristics (proximity to city center, recreation areas, supermarkets, healthcare facilities). A regression model of housing price in Gomel has been developed. The main output from GWR is a set of localised parameter estimates and associated diagnostics. The GWR adjusted R^2 has been 0.74 (R^2 is 0.77). AICc has given a value of 10.06. Mapping the GWR model coefficients allows exploring the spatial features of relationships between residential real estate price (per sq.m) and explanatory variables. The spatial autocorrelation of GWR residuals for our model have resulted in a Moran's I value of −0.03 ($p = 0.414$), indicating a residuals spatial random distribution, which in turn suggests a good GWR fit.

References

1. Anselin L. (1995). Local Indicators of Spatial Association – LISA. Geographical Analysis, Volume 27, pp. 93–115.
2. Cao K., Diao M., Wu B. (2019). A Big Data–Based Geographically Weighted Regression Model for Public Housing Prices: A Case Study in Singapore. Annals of the American Association of Geographers, Volume 109, Issue 1, pp. 173–186. DOI: 10.1080/24694452.2018.1470925.
3. Cellmer R., Cichulska A. Bełej M. (2020). Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression. ISPRS International Journal of Geo-Information, Volume 9, № 6. DOI:10.3390/ijgi9060380.
4. Cellmer R. (2012). The Use of the Geographically Weighted Regression for the Real Estate Market Analysis. Folia Oeconomica Stetinensia, Volume 11, Issue 1. DOI:10.2478/v10031-012-0009-6.
5. Chau K.W., Chin T. (2003). A Critical Review of Literature on the Hedonic Price Model. International Journal for Housing Science and Its Applications, Volume 27, pp. 145-165.
6. Dziauddin M.F., Ismail K., Zainudin O. (2015). Analysing the local geography of the relationship between residential property prices and its determinants. Bulletin of Geography. Socio-economic Series, Volume 28, pp. 21–35. DOI:10.1515/bog-2015-0013.
7. Fotheringham A.S., Brunsdon C., Charlton M. (2002). Geographically weighted regression: the analysis of spatially varying relationships. Chichester: Wiley. 269 p.
8. Fotheringham A.S., Charlton M.E., Brunsdon C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A, Volume 30, Issue 11, pp.1905–1927.
9. Manganelli B., Pontrandolfi P., Azzato A., Murgante B. (2014). Using geographically weighted regression for housing market segmentation. International Journal of Business Intelligence and Data Mining (IJBIDM), Volume 9, № 2, pp. 161–177. DOI:10.1504/IJBIDM.2014.065100.
10. Tomal M. (2020). Modelling Housing Rents Using Spatial Autoregressive Geographically Weighted Regression: A Case Study in Cracow, Poland. ISPRS International Journal of Geo-Information, Volume 9, № 6. DOI:10.3390/ijgi9060346.
11. Wang D., Li V.J., Yu H. (2020). Mass Appraisal Modeling of Real Estate in Urban Centers by Geographically and Temporally Weighted Regression: A Case Study of Beijing's Core Area. Land, Volume 9, № 5. DOI: 10.3390/land9050143.

Information about authors

Natalia Zhukovskaya, PhD, Associate Professor of the Soil Science and Geographical Information Systems Department of Belarusian State University. Address: Republic of Belarus, 220030, Minsk, Leningradskaya 16, Soil Science and Geographical Information Systems Department, Belarusian State University, phone +375296694248, e-mail: natazhuk@gmail.com.

Olga Popko, MA student of the Soil Science and Geographical Information Systems Department of Belarusian State University, Cadaster and GIS specialist at the National Cadastral Agency. Address: Republic of Belarus, 220030, Minsk, Leningradskaya 16, Soil Science and Geographical Information Systems Department, Belarusian State University, phone +375 297352113, e-mail: popko555@gmail.com.