UDC 51.77

ECONOMETRIC MODELING OF THE PRICE OF RESIDENTIAL REAL ESTATE BY THE METHOD OF GEOGRAPHICALLY WEIGHTED REGRESSION

Aitkozha Zhanibek Eldosovich
jan2602@mail.ru
2-nd year master student OP 7M06105 – "Mathematical and computer modeling"
L.N. Gumilyov ENU, Nur-Sultan, Kazakhstan
Scientific adviser - Abdenova G.A.

Introduction

The development of the economy and market relations in Kazakhstan has led to the formation of a real estate market in the country. At the same time, the factors that influenced the formation of the price of an apartment at the initial stage of development began to fade into the background, and therefore it becomes necessary to identify and measure new patterns in the market under consideration [1, 2]. Historically, this problem was solved in the framework of classical linear regression. Subsequently, dummy variables were included in the model, which made it possible to take into account the spatial differentiation (heterogeneity) of the contribution of regressors to the price formation depending on the geographical location of the objects. In the framework of this work, it is proposed to turn to the consideration of the possibility of using a new method for Kazakhstani practice - geographically weighted regression to study prices in the residential real estate market.

Geographically weighted regression takes it one step further than traditional methods by estimating the regression coefficients at each point in Cartesian space, thus allowing relationships that vary across space to be identified and measured. This approach expands the understanding of how the belonging of a population unit to specific geographic coordinates affects the relationship between regressors and real estate prices.

Formulation of the problem

In this regard, the purpose of this study is to develop an econometric model and analyze spatial differences in the price of one-room apartments presented on the secondary housing market in Semey. For the study, information was collected on one-room apartments in the secondary housing market. The choice of the object of study is explained by the increased demand for this category and a significant number of objects put up for sale per unit of time.

Initially, 350 objects were selected, after checking the information and rejecting anomalous observations (with an excessively low or overpriced price), 299 one-room apartments remained in the sample. The material collection period covered the range from November 25, 2013 to December 13, 2013. The information contained in the sales announcements allowed us to generate the following variables:

y_i – dependent variable denoting the price of a one-room apartment in the secondary housing market in Semey, million tenge;

x_i is an independent variable that characterizes the effect of the size of the total area of a one-room apartment on its price, sq. m.;

d_i is a dummy variable reflecting the "prestigiousness" of the location of an apartment in a multi-storey building, which is defined as follows:

Prestige = the floor on which the apartment is located / the number of floors in the house (1)

As a result of calculations for each object, values are obtained that vary in the range from zero to one. Obviously, apartments located on the middle floors should be more popular among buyers and, as a rule, have a higher price. Therefore, apartments located in the range from 0.3 to 0.7 will be coded 1, and apartments located "on the edges" will be coded 0:

$$d_i = \begin{cases} 1, & if\ prestige \in [0{,}3;\ 0{,}7] \\ 0, & in\ other\ cases \end{cases} \quad (2)$$

Information about the address of the house where the one-room apartment for sale is located, together with the capabilities of such Internet resources as Yandex maps and Google maps, made it possible to introduce two variables that characterize the position of the object in space:

$north_i$ – quantitative variable denoting the latitude at which the i-th object is located;

$east_i$ – quantitative variable denoting the longitude at which the i-th object is located.

Also, when modeling the influence of factors on the price of sq. m of housing, the following independent variable will be used:

$Dist_i$ – a variable that characterizes the degree of proximity of the house in which the apartment is being sold to the conditional geographical center, which is used as the "zero" kilometer of the city. In our case, these are the coordinates of the Detsky Mir store, which is located on the central Shakarim Avenue. We estimated the distance using the most common measure of the proximity of objects - the Euclidean distance.

Solution Methods

The study of the influence of socio-economic factors on the price of one-room apartments in Semey was carried out according to the following scheme:

1) a matrix of initial data is formed containing information on 299 objects and 6 variables (y_i , x_i, d_i , $north_i$, $easti_i$, $distance_i$);

2) the value of differentiation of objects by districts of Semey is estimated, for this a graphical method and cluster analysis (k-means method) are used;

3) geographically weighted regression (GWR) is evaluated, the hypothesis is tested: the GWR describes the data much better than the global model (the general model obtained on the basis of the least squares without taking into account positioning);

4) the interpretation of the obtained parameters for the GWR is carried out, while the graphical method is used.

Research results

According to the selected stages of the study, a multidimensional grouping of objects of observation was carried out using the k-means method, the coordinates of houses were taken as variables. The result was the division of the population into 5 clusters (Fig. 1). The first cluster included objects located in the city center, mainly in houses built in the 1960s. The second cluster - the “Vokzal” and “Silikatny districts” - is characterized by an increased crime situation. Apartments in the third cluster - the "Left Bank" area, are "new buildings", mostly cheap social housing in inexpensive panel houses. The fourth cluster is the "Akimat" district, which includes apartments in one of the "prestigious" residential areas with a developed infrastructure. The fifth cluster - microdistricts "Central Market", "Rainbow", is also characterized as "prestigious", includes apartments in panel houses of the 1980s. the buildings.

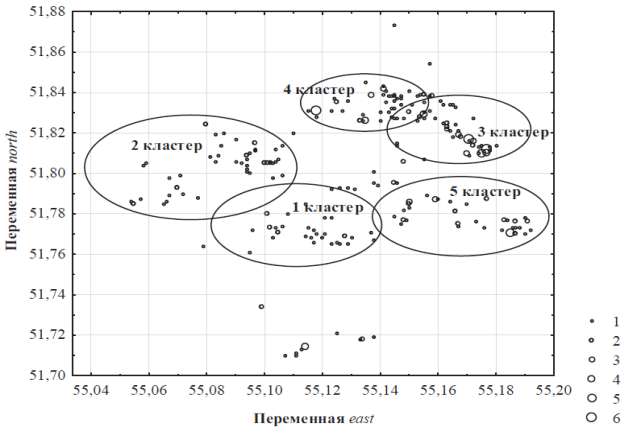

Fig.1. The results of splitting the set of space objects into clusters
(points on the graph - the frequency of occurrence of houses in the sample)

To confirm the difference in prices, we performed a one-way analysis of variance, resulting in an actual Fisher F-test value of 3.28 with a p-level of 0.01. It follows that the hypothesis of equality of means is rejected, and it can be concluded that the location of the house influences the formation of the price.

Another argument in favor of the presence of individual effects in the formation of the price of a one-room apartment is the assessment of the influence of the $dist_i$ regressor on the dependent variable y_i:

$$\hat{y} = 860{,}51 + 29{,}62x_i + 281{,}67dist_i; \quad R^2 = 0{,}58 \qquad (3)$$

(t) (15,32) (18,47) (0,56)

Fisher's criterion $F_{(2;\ 296)} = 201{,}12$ exceeds the tabular value ($F_{table\ (0{,}05;\ 2;\ 296)} = 3{,}026$), which indicates the significance of the equation as a whole. In parentheses are the calculated values of the t-criterion to test the hypothesis about the significance of the coefficients of the resulting equation. Significant, as follows from equation (3), is the free term and the regression coefficient in front of the independent variable at the 10% level, since the actual values of the t-test exceed $t_{table} = 2{,}59$.

The parameter with the variable $dist_i$ was not statistically significant, hence we can conclude that the conditional "zero" point has a slight influence on the price of an apartment (the price does not change when moving from the center to the outskirts), which is explained by the presence of local centers.

The inclusion of a dummy variable in the model, reflecting the prestige of the floor on which the apartment for sale is located, also did not bring the desired result, the parameter with the variable d_i was obtained not statistically significant.

An attempt was also made to introduce dummy variables into the multiple regression model that characterize the belonging of an object to one of the selected clusters, while the following model was used:

$$d_{1i} = \begin{cases} 1, & if\ the\ object\ belongs\ to\ 1\ cluster \\ 0, & otherwise \end{cases};$$

$$d_{2i} = \begin{cases} 1, & if\ the\ object\ belongs\ to\ 2\ cluster \\ 0, & otherwise \end{cases};$$

$$d_{3i} = \begin{cases} 1, & if\ the\ object\ belongs\ to\ 3\ cluster \\ 0, & otherwise \end{cases};$$

$$d_{4i} = \begin{cases} 1, & if\ the\ object\ belongs\ to\ 4\ cluster \\ 0, & otherwise \end{cases};$$

The regression model has the following form:

$$\hat{y} = 926{,}96 + 29{,}56x_i - 36{,}39d_{1i} - 97{,}68d_{2i} - 50{,}45d_{3i} - 43{,}09d_{4i} \qquad (4)$$

(t) (14,96) (17,72) (-0,96) (-2,58) (-1,47) (-1,13)

Fisher's criterion $F_{(5;\ 293)} = 82.71$ exceeds the tabular value ($F_{table\ (0.05;\ 5;\ 293)} = 2.244$), which indicates the significance of the equation as a whole. Significant, as follows from equation (4), are the free term and the regression coefficient before the regressor d_{2i} at the 10% level, since the actual values of the t-criterion exceed $t_{table} = 2.593$.

Since the parameters obtained with the regressors d_{1i} , d_{3i} , d_{4i} are not statistically significant, it can be argued that multiple regression is inconsistent when modeling the influence of factors on the price of one-room apartments.

Let's move on to the consideration of geographically weighted regression. Note that the HWR calculations were carried out in the STATA 11 package, which does not provide for the output of the model itself in an interval form, the researcher has access only to the global model, tests for the feasibility of using the HWR and parameter estimates at each point in space. In this regard, we turn to the consideration of the listed characteristics.

When using geographically weighted regression, it is necessary to solve the question: does the HWR model characterize the considered dependence much better than the general (global) regression model estimated on the basis of the least squares? As noted by Y. Leung [3] and A. Fotheringham [4], geographically weighted regression describes any dependence better than the global least squares model. However, in practice, simple models are generally preferred over more

complex ones if there is no significant improvement over the latter. To solve the problem of choosing between the considered models, the Monte Carlo test is used.

The Monte Carlo test compares the difference between the residual sums of squares from the LSM and HWR models with the residual sum of squares of the HWR model [5]. In our case, the bandwidth test showed the following results: statistic value – 0.030, p-significance level – 0.008.

Relying on the given values, it can be argued that, according to the HWR estimates, the residual sum of squares has significantly decreased relative to the LSM estimates. From this we conclude that geographically weighted regression is preferable.

Analyzing the obtained results, we see that in the northwestern part of the city a point with coordinates east = 55.092, north = 51.805 stands out: it is characterized by the highest theoretical price - more than 3 million tenge, which is explained by the construction of a complex of brick multi-storey buildings in this zone. Prices on the outskirts of the city do not have a clear distinction. So, there are also apartments costing less than 1700000 tenge. and about 2500000 tenge. An interesting feature is the location of inexpensive apartments in the center. In addition, one can trace the drift of the rising price in the south. In addition, the use of HWR allows you to trace the increase in the cost of an apartment due to an increase in area by 1 sq. m at each individual point.

Findings

The theoretical significance of the study is to illustrate the failure of global (general) regression models, which are often used by analysts to study geographic differences and tend to "mask" the true nature of the relationship between the result and explanatory variables. Because geographically weighted regression estimates the regression coefficients for each population unit over the entire study area, it is able to more powerfully highlight geographic differences in the relationships between the explanatory variables and the price of one-bedroom apartments.

Thus, geographically weighted regression should be an important addition to the researchers' toolbox. The HWR will allow assessing the extent to which the impact of various factors that determine the process of price formation in the real estate market depends on the geographical location of the units of observation, and determine what information (factor) can be used to make decisions in the housing market.

References

1. Tsypina Yu. S., Tsypin A. P. Statistical methods in the study of mortgage lending in Russia // New University. Ser. Economy and law. 2012. No. 6. P. 10–13

2. Kashkova Yu. N., Loginov D. V. Prospects for the use of geographically weighted regression to account for spatial dependencies in geological and geophysical data. TsKR Rosnedra. 2011. No. 6. S. 20–23.

3. Fotheringham A., Brunsdon C., Charlton M. Geographically Weighted Regression. Chichester : John Wiley & Sons Ltd, 2002. 269 p.

4. Leung Y, Mei C. L, Zhang W. X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model // Environment and Planning. 2000. Vol. 32, № 1. P. 9–32.

5. Brunsdon C., Fotheringham A., Charlton M. Some notes on parametric significance tests for Geographically Weighted Regression // Journal of Regional Science. 1999. Vol. 39, № 3. P. 497–524.

УДК 519.683.8

ПРИМЕНЕНИЕ ГЛУБОКОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В ДИАЛОГОВЫХ СИСТЕМАХ

Хлопук Василий Сергеевич
ibasileus@yandex.ru