ADAPTIVE METHODS OF TERRITORIAL HETEROGENEITY RECORDING DURING REAL ESTATE PRICE SIMULATION*

© 2009 O.S. Balash, A.V. Kharlamov **

Keywords: spatial correlation, adaptive model, geographically weighted regression, real estate price simulation.

In this article the authors consider the problem of econometric analysis of regional and spatial data. The spatial dependence may cause the standard econometric techniques to become inappropriate. The method of geographically weighted regression is discussed.

Territorial heterogeneity should be taken into account during socio-economic processes simulation.

The method of geographically weighted regression is an example of adaptive approach to territorial data analysis. Adaptive methods assume that the coefficients in the simulated model equation are revised continuously so that the last (or in case of territorial data - the closest) observations have the highest weight. As the result, our model has a constantly changing structure.

The equation of model of geographically weighted regression:

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

where the pair of variables (u_i, v_i) represent the position of a point i, $i=\overline{1,n}$; y_i is the value of the observed depended variable; $x_{i1},\ldots,x_{ip}$ are independent deterministic regressors, $k=\overline{1,p}$, p is the number of regressors; $\beta_k(u_i,v_i)$ are the unknown coefficients that must be estimated, $k=\overline{0,p}$; ε_i are random errors[1].

The least-squares method can be used to estimate the coefficients in point i. Not all observation data but only the data of neighboring points to point i are used so that local features can be detected with this model. It is assumed that regression models of neighboring points are similar but can vary in territory. The proximity degree is considered in weights w_{ij}. The coefficient evaluation vector for every point i is equal to:

$$\hat{B}(u_i,v_i) = (X^T W(u_i,v_i)X)^{-1} X^T W(u_i,v_i)Y \,,$$

where $W(u_i,v_i)$ is the diagonal matrix of weight coefficients, the matrix size is $(n\times n)$:

$$W(u_i,v_i) = \begin{bmatrix} w_{i1} & 0 & \ldots & 0 \\ 0 & w_{i2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & w_{in} \end{bmatrix}.$$

Matrix element w_{ij}, $i,j=\overline{1,n}$ determines the degree of neighbors j impact on point i. The matrix of weight coefficients is estimated for every point.

The coefficients are estimated for all measurements and the results are put down into the parameter evaluation matrix:

$$\hat{B} = \begin{bmatrix} \hat\beta_0(u_1,v_1) & \hat\beta_1(u_1,v_1) & \ldots & \hat\beta_p(u_1,v_1) \\ \hat\beta_0(u_2,v_2) & \hat\beta_1(u_2,v_2) & \ldots & \hat\beta_p(u_2,v_2) \\ \ldots & \ldots & \ldots & \ldots \\ \hat\beta_0(u_n,v_n) & \hat\beta_1(u_n,v_n) & \ldots & \hat\beta_p(u_n,v_n) \end{bmatrix},$$

where the row with number i represents the coefficient evaluation vector in point (u_i,v_i), $i=\overline{1,n}$:

$$\beta(i) = (X^T W(i)X)^{-1} X^T W(i)Y \,.$$

Weighting schemes can be either fixed or adaptive. Methods of borough-specific calibration, moving window method, methods of fixed or adaptive weighting functions can be used to evaluate weight coefficients[2].

If the borough can expose the features unique to territorial districts, then for all points, which belong to one district, the value of weight matrix element is assumed to be equal to 1, otherwise it is assumed to be equal to 0:

$$w_{ij} = 1, \text{ if } (i,j) \in A;$$

* This article was financially supported by Russian Humanitarian Foundation (project 08-02-27209 а/в).

** Olga S. Balash, PhD in Economics, associate Professor, Department of Higher Mathematics and Information Technologies, Russian State University of Trade and Economics, Saratov branch; Alexander V. Kharlamov, associate Professor, Department of Probability Theory and Stochastic Processes, Saratov State University.

$$w_{ij} = 0 \text{, if } (i, j) \notin A.$$

If the territorial districts have been formed historically and do not represent the natural stratifications of the objects, then discrete weights are evaluated with regard to the distance between objects. In addition the category of the closest neighbor is determined regarding to the preset value of marginal remoteness, i.e. some fixed distance *b*. The weight is assumed to be equal to 1, if the distance between the objects d_{ij} does not exceed the preset distance *b*, otherwise it is assumed to be equal to 0:

$$w_{ij} = 1 \text{, if } d_{ij} < b;$$

$$w_{ij} = 0 \text{, if } d_{ij} \geq b.$$

The value of d_{ij} is calculated as distance between 2-D points. This is the so called "fixed moving window method", where the value of distance *b* is fixed, the distance *b* is called window breadth or bandwidth.

The discrete approach to weight evaluation considers the territorial heterogeneity, but it is assumed that the impact of all neighbors located in the pass band is equal. In many cases the impact of neighbors decreases as the distance increase. That's why the closer neighbors are given more weight than those further away. The method, where weights are set up regarding to the continuous variation of distance between the tested objects, is more universal. Most commonly used the Gauss weighting schemes:

$$w_{ij} = \exp\left(-\frac{\alpha}{2}\left(\frac{d_{ij}}{b}\right)^2\right),$$

where d_{ij} is the distance between the point *i* и its neighbor *j*, and *b* is the bandwidth.

Bi-quadrate and tri-cube weighted functions are alternatively used:

$$w_{ij} = \begin{cases} (1-(d_{ij}/b)^2)^2, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases},$$

$$w_{ij} = \begin{cases} (1-(d_{ij}/b)^3)^4, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases}.$$

The impact of neighbors that are close to the point is practically equal to 1 and it decreases as they approach the pass band borders.

Let's use the method of geographically weighted regression to estimate a model of one-roomed flats prices in the city of Saratov.

The data was obtained from sales figures of secondary housing market of one-roomed flats. The sample size is 1813 objects.

The dependent variable y is the price of room (in thousand of rubles), the regressors are: x_1- living area, m^2; x_2- kitchen area, m^2; x_3 - the extra space, m^2; x_4- the logarithm of distance, ln(*m*); x_5- location of the room on the first floor ; x_6- location of the room on the top floor; x_7- small-storey building; x_8- five-storey building; x_9- brick house; x_{10}- good or excellent condition of the building; x_{11}- the presence of balcony or stanza.

The area around General Post Office was selected as the centre of the city of Saratov.

The global linear regression model created on basis of the source data can be represented as the following equation:

$$y = 1180{,}61 + \underset{(1{,}04)}{13{,}04}x_1 + \underset{(1{,}36)}{10{,}38}x_2 + \underset{(0{,}79)}{11{,}17}x_3 -$$
$$-\underset{(2{,}62)}{116{,}40}x_4 - \underset{(5{,}70)}{36{,}82}x_5 - \underset{(5{,}34)}{28{,}19}x_6 - \underset{(10{,}99)}{122{,}10}x_7 -$$
$$-\underset{(5{,}06)}{30{,}43}x_8 + \underset{(5{,}03)}{20{,}88}x_9 + \underset{(4{,}20)}{19{,}22}x_{10} + \underset{(5{,}30)}{16{,}87}x_{11}.$$

All coefficients of variables as well as the whole model have been found significant. The coefficient of determination R^2=0,7 shows that the model can explain only 70% of the present dependence.

In order to use the method of geographically weighted regression, the false coordinate of objects were taken from the digital database "All Cities of Russia" and this information was added to input data.

The "tri-cube" function was used to set up the weight matrix, and Akaike criterion was taken as "window" breadth optimization criterion:

$$AIC_C = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\frac{n+\nu_1}{n-2-\nu_1} \to \min,$$

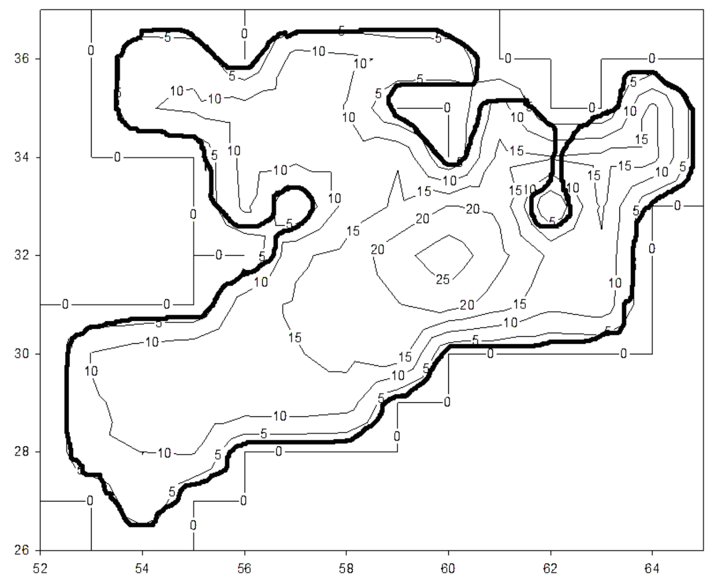
where $\hat{\sigma}$ is the estimated standard deviation,

$$\nu_1 = tr(S).$$

The following results were achieved by using the method of geographically weighted regression.

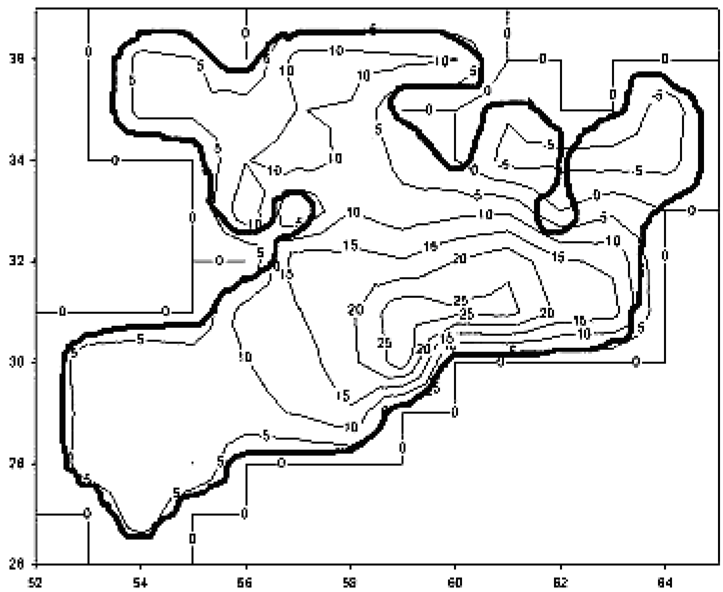
The optimal number of closest neighbors needed to achieve the minimum of the Akaike criterion is 295. The coefficient of determination R^2=0,8.

The analysis of estimated coefficient value obtained for each regressor is shown hereafter.



Living area

Fig. 1. **The estimated coefficient values to "living area" regressor**



Kitchen area

Fig. 2. **The estimated coefficient values to "kitchen area" regressor**

The variation of estimated coefficients is shown on fig. 1. The centre of the city corresponds to the following coordinates: X=61, Y=32.

The estimated coefficients are significant over the whole territory. The most peculiar point of the central part of the city is the square X=60, Y=32 where the most expensive apartments (near 30 thousand rubles per square meter) are located. This point is surrounded by the district where the living area cost exceeds 20 thousand rubles per square meter. The city outskirts, where one square meter of the living area costs about 10 thousand rubles, are also clearly defined. The decrease of price can be traced from the center of the city to Leninsky district.

The estimated coefficient values to "kitchen area" regressor are shown on fig. 2.

Figure 2 shows that there are zones where the kitchen area is the most valuable. The highest value is tracked in square X=59, Y=31 as well as in neighboring squares. It should be noted that the most expensive living area and the most expensive kitchen area are located in different squares but the highest value of both regressors is tracked in the same center. Another feature of the central districts is that the extra square meter of the kitchen area costs more the extra square meter of the living area, for example: in square X=57, Y=32 the price of extra square meter of kitchen area and living area is 18,2 и 14,1 thousand rubles respectively. Zones with relatively cheap kitchens are located in the city outskirts.

The analysis of coefficient values for other regressors also shows that the city can be divided into zones with almost the same values of the coefficients. There is a square in the centre of the city with the high cost of the extra space that exceeds the cost of living area and kitchen area. Hereby, the model of geographically weighted regression can reproduce the fact that the zones with the highest prices of kitchen area, living area and non-living area do not coincide.

The aforementioned calculations prove that the method of geographically weighted regression can expose the features unique to certain parts of the city which are not included in the global model and it allows taking into account the peculiarities of certain district development.

1. *Anselin L.* Spatial econometrics: methods and models, Kluwer Academic Publisher, 1988.

2. *Fotheringham A.S., Brunsdon C., Charlton M.* Geographically Weighted Regression, John Willey & Sons, 2002.