УДК 338.5

АДАПТИВНЫЕ МЕТОДЫ УЧЕТА ТЕРРИТОРИАЛЬНОЙ НЕОДНОРОДНОСТИ ПРИ МОДЕЛИРОВАНИИ ЦЕН НА НЕДВИЖИМОСТЬ*

© 2009 О.С. Балаш, А.В. Харламов**

Ключевые слова: пространственная корреляция, адаптивные методы, географически взвешенная регрессия, моделирование цен на недвижимость

Рассматривается проблема эконометрического анализа региональных и пространственных данных. Пространственная взаимозависимость может приводить к тому, что предпосылки стандартных эконометрических методов не выполняются. Описывается применение метода географически взвешенной регрессии для моделирования цен на жилую недвижимость на примере г. Саратова.

При моделировании социально-экономических процессов необходимо учитывать их территориальную неоднородность.

Метод географически взвешенной регрессии можно рассматривать как применение адаптивного подхода к анализу территориальных данных. Адаптивные методы предполагают, что коэффициенты модели пересматриваются непрерывно, так что наибольший вес имеют последние, а в территориальном случае - ближайшие наблюдения. В итоге получают модель с непрерывно меняющейся структурой.

Модель географически взвешенной регрессии имеет вид

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

где пара переменных (u_i, v_i) представляет координаты точки (местоположение) i, $i=\overline{1,n}$; y_i - значение наблюдаемой зависимой переменной; $x_{i1},\ldots,x_{ip}$ - независимые детерминированные регрессоры, $k=\overline{1,p}$, p - число регрессоров; $\beta_k(u_i,v_i)$ - неизвестные коэффициенты, подлежащие оценке, $k=\overline{0,p}$; ε_i - случайные ошибки [1].

Для вычисления оценок коэффициентов в местоположении i может использоваться метод наименьших квадратов. В целях выявления местных особенностей используются не все имеющиеся наблюдения, а только соседние с i. Предполагается, что регрессионные модели для соседних точек схожи, но могут варьироваться по территории. Степень близости учитывается с помощью весов w_{ij}. Вектор оценок коэффициентов для каждого местоположения i равен:

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

где $W(u_i,v_i)$ - диагональная матрица весовых коэффициентов размерности $(n \times n)$:

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

Элемент матрицы w_{ij}, $i,j=\overline{1,n}$ определяет степень влияния соседей j на зависимости в местоположении i. Матрица весовых коэффициентов вычисляется для каждого местоположения.

Так как расчеты коэффициентов проводятся для всех измерений, то в результате получают матрицу оценок параметров:

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

где i-я строка представляет собой вектор оценок коэффициентов в точке (u_i,v_i), $i=\overline{1,n}$:

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

Для определения весовых коэффициентов используют методы административно-тер-

* Статья подготовлена при финансовой поддержке Российского гуманитарного фонда (проект 08-02-27209а/в).

** Балаш Ольга Сергеевна, кандидат экономических наук, доцент Саратовского института Российского государственного торгово-экономического университета; Харламов Александр Владимирович, ст. преподаватель Саратовского государственного университета.

риториального деления, движущегося окна, фиксированных и адаптивных ядер [2].

Если административное деление раскрывает специфические закономерности, присущим административным единицам, то для точек, принадлежащих одному району, элемент весовой матрицы полагают равным единице и нулю в противном случае:

$$w_{ij} = 1 \text{, если } (i, j) \quad ;$$

$$w_{ij} = 0 \text{, если } (i, j) \quad .$$

Если административные районы сформированы исторически и не отражают естественное расслоение объектов, то дискретные веса определяются с учетом расстояния между объектами. При этом задают предельно допустимую удаленность, то есть некоторое фиксированное расстояние *b*, относительно которого определяют категорию ближайшего соседа. Вес принимают равным единице, если расстояние d_{ij} между объектами не превосходит заданного расстояния *b*, и нулю в противном случае:

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

$$w_{ij} = 0 \text{, если } d_{ij} \ge b.$$

Значение d_{ij} вычисляется как расстояние между точками на плоскости. Это так называемый метод движущегося фиксированного окна, *b* фиксировано и называется шириной окна или полосы пропускания.

Использование дискретного подхода при определении весов учитывает территориальную неоднородность, но влияние соседей, попавших в полосу пропускания, считается одинаковым. Во многих случаях влияние соседей уменьшается с увеличением расстояния. Поэтому более близким соседям придают больший вес, чем дальним. Подход, в котором веса строятся с учетом непрерывного изменения расстояния между исследуемыми объектами, называют ядерным. Наиболее часто применяют ядра Гаусса:

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

где d_{ij} - расстояние между местоположением *i* и соседом *j*, а *b* - ширина полосы пропускания.

Альтернативными вариантами являются ядра би-квадрат и три-куб:

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

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

Влияние соседей в непосредственной близости к местоположению практически равно единице и уменьшается при приближении к границе полосы пропускания.

Применим метод географически взвешенной регрессии для построения модели ценообразования на рынке недвижимости однокомнатных квартир г. Саратова.

Информационной базой послужили данные о продажах однокомнатных квартир на вторичном рынке жилья (http://www.ks.sarbc.ru) за январь 2006 года. Численность выборки составила 1813 объектов.

Зависимая переменная *y* - цена квартиры (тыс. руб.), регрессоры: x_1 - жилая площадь, м^2; x_2 - площадь кухни, м^2; x_3 - дополнительная площадь, м^2; x_4 - логарифм расстояния, ln(*м*); x_5 - расположение на первом этаже; x_6 - расположение на последнем этаже; x_7 - дом малой этажности; x_8 - пятиэтажка; x_9 - кирпичный дом; x_{10} - в хорошем или отличном состоянии; x_{11} - имеются балкон или лоджия.

В качестве центра г. Саратова выбран район Главпочтамта.

Глобальная линейная регрессионная модель, построенная по исходным, данным имеет вид

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

Все коэффициенты при переменных оказались значимыми, как и вся модель в целом. Коэффициент детерминации $R^2 = 0{,}7$ показывает, что модель объясняет только 70% имеющейся зависимости.

Для применения географически взвешенной регрессии в исходные данные были добавлены условные координаты объектов, полученные с помощью электронной базы данных "Все города России".

При построении весовой матрицы использовалась функция "три-куб", в качестве критерия оптимизации ширины "окна" - критерий Акаике (Akaike):

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

где $\hat{\sigma}$ - оценка стандартного отклонения, $\nu_1 = tr(S)$.

Географически взвешенный метод дал следующие результаты.

Оптимальное число ближайших соседей, дающее минимум критерия Акаике, равно 295. Коэффициент детерминации $R^2 = 0{,}8$.

Проанализируем значения полученных оценок коэффициентов при каждом регрессоре. Вариация оценок коэффициентов при регрессоре “жилая площадь” по территории города изображена на рис. 1. Координаты центра города соответствуют значениям $X = 61$, $Y = 32$.

Оценки коэффициентов значимы на всей территории. В центральной части города выделяется квадрат x = 60, y = 32 с самыми дорогими квартирами, практически по 30 тыс. рублей за квадратный метр. Рядом с ним можно выделить район, стоимость метра жилой площади в котором превышает 20 тыс. рублей. Четко выделяются окраины города, где цена квадратного метра жилой площади

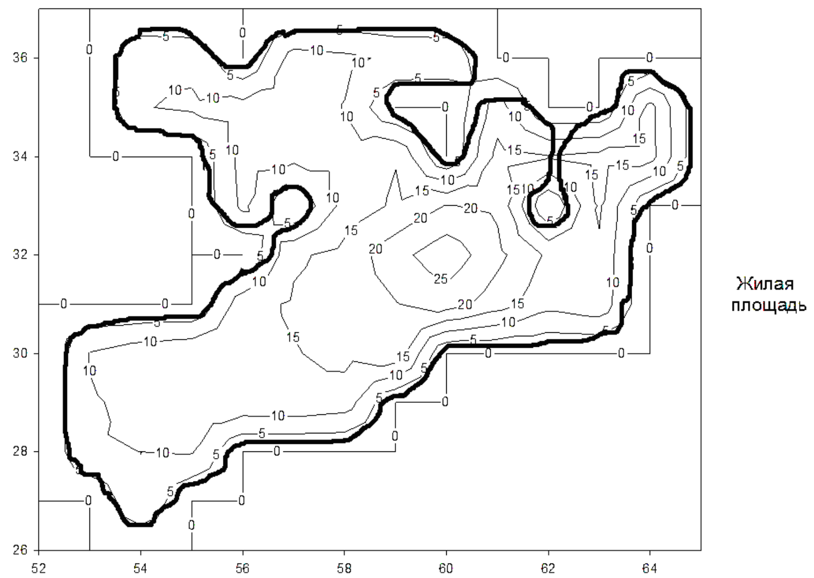

Рис. 1. **Значения оценок коэффициента при регрессоре “жилая площадь”**

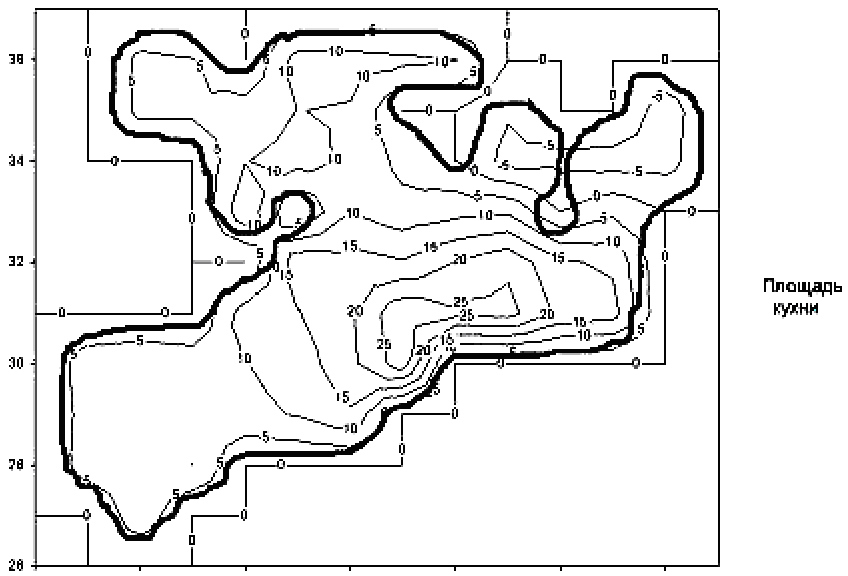

Рис. 2. **Значения оценок коэффициента при регрессоре “площадь кухни”**

порядка 10 тыс. рублей. Можно проследить дрейф убывающей цены от центра в направлении Ленинского района.

Оценки коэффициентов при регрессоре "площадь кухни" представлены на рис. 2.

На рис. 2 можно выделить зоны, где очень высоко ценится размер кухни. Самая высокая стоимость в квадрате $X = 59$, $Y = 31$, а также в соседних квадратах. Надо отметить, что самая дорогая жилая площадь и дорогая кухня попадают в разные квадраты, хотя высокая стоимость по обоим регрессорам оказывается в общем центре. Еще одной особенностью центральных районов является то, что дополнительный метр кухни стоит дороже дополнительного метра жилой площади, так для координат $X = 57$, $Y = 32$ стоимость равна 18,2 и 14,1 соответственно. На окраинах города расположены зоны с относительно дешевыми кухнями.

Для остальных регрессоров также отчетливо прослеживается разделение территории города на зоны с приблизительно схожими значениями коэффициентов. В центре города выделяется район с высокой стоимостью дополнительной площади, превышающей цену квадратного метра кухни и жилой комнаты. Таким образом, географически взвешенная регрессионная модель отражает то, что границы районов с наиболее высокими ценами квадратных метров кухни, жилой и нежилой площади различны.

Проведенные расчеты показывают, что географический подход позволяет выявить особенности, присущие отдельным районам города, которые нивелируются в глобальной модели, и гибко учитывать специфику застройки отдельных районов.

---

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

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