УДК 518:91

А.В. Харламов

ГЕОГРАФИЧЕСКИЕ МОДЕЛИ В СЛУЧАЕ ПРОСТРАНСТВЕННОЙ НЕСТАЦИОНАРНОСТИ

В задачах пространственного моделирования для больших территорий, условия применения классической регрессионной модели [1]:

$$Y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \sigma^2 I,$$

как правило, не выполняются в силу естественной пространственной неоднородности или пространственной нестационарности. Пространственная нестационарность обусловлена тем, что объекты, которые находятся поблизости друг от друга, как правило, имеют много общего и в силу этого строятся регрессионные модели с учетом пространственной автокорреляции, отличающиеся от классической модели с постоянной дисперсией, которая в случае нормальности распределения ошибок представима в виде

$$Y \sim N(X\beta, \sigma^2 I).$$

Модель пространственной автокоредяции в общем случае имеет следующий вид

$$Y \sim N(\mu, A),$$

где $\mu = X\beta$ и $A = V(\varepsilon)$ — дисперсионная матрица произвольного вида.

При анализе таких моделей используют два метода: условной авторегрессии и одновременной авторегрессии [2].

В модели условной авторегрессии предполагается, что переменная y_i зависит не только от регрессоров, но также от своих соседних значений, и ее распределение можно представить как условное нормальное:

$$y_i | \{y_j, j \neq i\} \sim N(\mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j), \tau^2).$$

Условное распределение зависит не только от набора регрессоров x_{ik}, $i = \overline{1;n}$, $k = \overline{1;p}$, но также от ошибок прогнозирования, которые берутся с соответствующими весовыми коэффициентами c_{ij}. Значения весов могут быть рассчитаны одним из возможных способов. Матричное представление модели имеет вид

$$Y \sim N(\mu, (I - C)^{-1}\tau^2).$$

Одновременная авторегрессионная модель определяется следующим образом:

$$y_i \sim N(\mu_i + \sum_{j=1}^{n} c_{ij}(y_j - \mu_j), \tau^2).$$

Здесь рассматриваются уже безусловные распределения и вся система представима как система одновременных уравнений. Матричное представление данной модели имеет следующий вид:

$$Y \sim N(\mu, (I - C)^{-1}(I - C^T)^{-1}\tau^2).$$

Рассмотренные модели пространственной автокорреляции были предложены как вариант моделирования при наличии автокорреляции в ошибках.

Проведем сравнительный анализ различных методов моделирования для данных на вторичном рынке жилья однокомнатных квартир г. Саратова, полученных с сайта еженедельника газеты «Квадратный метр» (http://www.ks.sarbc.ru) за январь 2006 года. В моделях использовались следующие переменные:

y – цена квартиры, тыс. руб.;

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.

Глобальная линейная регрессионная модель, построенная по этим данным, имела следующий вид:

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

в скобках указаны стандартные ошибки. Все коэффициенты при переменных оказались значимыми, как и вся модель в целом. Коэффициент детерминации, равный $R^2 = 0,7008$ показывает, что модель объясняет 70% имеющейся зависимости.

Анализ ошибок глобальной модели (предполагается, что наличие автокорреляционной зависимости непременно проявляется в остатках классической регрессии) показал наличие автокорреляционной зависимости и значение коэффициента пространственной автокорреляции $\rho = 0,79$.

Построение авторегрессионных моделей дало следующие результаты.

Условная авторегрессионная модель —

$$y = 1102,99 + \underset{(1,00)}{13,19}\,x_1 + \underset{(1,27)}{9,93}\,x_2 + \underset{(0,75)}{9,89}\,x_3 - \underset{(3,40)}{106,53}\,x_4 - \underset{(5,28)}{34,98}\,x_5 - \underset{(4,97)}{28,03}\,x_6 - \underset{(10,68)}{124,56}\,x_7 - \underset{(4,76)}{28,73}\,x_8 + \underset{(4,69)}{14,02}\,x_9 + \underset{(3,85)}{18,33}\,x_{10} + \underset{(4,82)}{12,53}\,x_{11} + \underset{(0,07)}{0,97}\,y^*$$

и одновременная авторегрессии модель —

$$y = 1117,09 + \underset{(1,00)}{13,30}\,x_1 + \underset{(1,28)}{10,09}\,x_2 + \underset{(0,76)}{10,08}\,x_3 - \underset{(3,50)}{108,54}\,x_4 - \underset{(5,37)}{36,14}\,x_5 - \underset{(4,99)}{28,19}\,x_6 - \underset{(10,64)}{125,51}\,x_7 - \underset{(4,79)}{28,46}\,x_8 + \underset{(4,85)}{14,74}\,x_9 + \underset{(3,89)}{18,44}\,x_{10} + \underset{(4,94)}{12,87}\,x_{11} + \underset{(0,05)}{0,64}\,y^*.$$

Сравнивая все три модели, убеждаемся, что оценки коэффициентов в глобальной регрессионной модели и в моделях с пространственной автокорреляцией практически не различаются. Отметим, что коэффициент авторегрессии в условной модели принимает большее значение, чем в одновременной, что обусловливается спецификой методики их построения. Можно сделать вывод, что применение авторегрессионных моделей не дает какого-либо преимущества по сравнению с классической регрессией. Построенные модели только выявили пространственную нестационарность в данных, никак ее не объясняя.

Теперь можно отметить, что использование коэффициента авторегрессии в глобальной модели не дает того эффекта, который можно получить при географическом подходе, когда распределение получаемых остатков носит практически случайный характер. Анализ географически взвешенной модели по тем же данным [3] показывает, что географически поход дает более наглядное представление о наличии или отсутствии пространственной нестационарности в изучаемой области. Более того, этот поход позволяет находить объяснение пространственной зависимости и выявлять специфические особенности, присущие отдельным районам города.

Анализ остатков географической регрессии показывает значительное уменьшение пространственной зависимости, коэффициент автокорреляции уже принимает значение, равное $\rho = 0,36$.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Greene C.W.* Econometric Analise. 3^{rd} edition. New Jersey, 1997

2. *Fotheringham A.S., Brunsdon C., Charrlton M.*. Geographically weighted regression the analysis of spatially varyng relationships. University of Newcastle, 2002

3. *Харламов А.В.* Географический метод в построении регрессионных моделей // Математика. Механика: Сб. науч. тр. Саратов: Изд-во Сарат. ун-та, 2007. Вып.9 С. 100-103.