

ПОСТАНОВКА ВОПРОСА МОДЕЛИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ С УЧЕТОМ ТЕРРИТОРИАЛЬНОЙ НЕОДНОРОДНОСТИ

Салиева А.К.

Оренбургский государственный университет, г. Оренбург

Аннотация: Пространственно варьирующие данные характеризуются существенной территориальной привязкой и служат источником неоднородности анализируемых статистических процессов или явлений. Это объясняет необходимость использования специальных моделей, поэтому в рамках настоящей работы предлагается обратиться к рассмотрению возможности применения нового для российской практики метода – географически взвешенной регрессии.

Ключевые слова: пространственная неоднородность, географически взвешенная регрессия, глобальные модели, локальные модели

При моделировании процессов, имеющую большую территориальную неоднородность, классическая модель регрессии не всегда является наилучшей и может дать неверный результат. Классические модели предполагают, что исходные данные являются регрессионно-однородными по всей исследуемой совокупности, а полученные в результате оценивания функции описывают изменения, характерные в среднем для всей совокупности данных. При этом оценки коэффициентов модели вычисляются по статистическим данным, не зависящим от конкретного местоположения объекта и, возможно, несущим его специфические особенности. Но на практике зачастую исходные данные являются пространственно зависимыми, то есть характеризующие изменения взаимосвязей признаков в пространстве.

Пространственная структура измеряемого отношения обеспечивает выявление признаков, для которых существует возможное исключение из модели. Скрытое влияние таких факторов может оказаться достаточно существенным, поэтому учет данных признаков значительно улучшит точность глобальной модели. Это объясняет необходимость использования специальных моделей при статистическом исследовании зависимостей, изменяющихся в пространстве [1].

Постановка новых задач требует от эконометрики новых методов выявления и измерения скрытых (латентных) зависимостей. Методы пространственного анализа в последнее время получили широкое распространение. Например, инструменты локального анализа были разработаны для обычных сводных статистических данных, а также для анализа пространственной зависимости как количественных, так и качественных показателей. Как правило, для удобного представления и интерпретации, к анализу локальной пространственной статистики часто привязывают карты и другую графику [2].

Локальный анализ можно рассматривать как процедуру, конечная цель которой заключается в создании глобальной модели, не проявляющей никакой существенной пространственной неоднородности. По существу, локальное моделирование используется для выявления неточностей глобальной модели; только при условии отсутствия существенного пространственного изменения изучаемых процессов, можно принять результаты глобальной модели. Кроме того, если устранить проблему неправильной спецификации глобальной модели путем добавления одной или нескольких переменных (например, при отсутствии данных) не представляется возможным, локальное моделирование позволяет включить эти эффекты с помощью пространственно различных оценок параметров [3].

При анализе пространственно варьирующих данных достаточно хорошо зарекомендовал себя метод географически взвешенной регрессии (ГВР), представляющий собой локальную форма линейной регрессии, которая используется для моделирования отношений, варьирующихся в пространстве [4, 5]. Метод ГВР – это некоторое обобщение модели с фиктивными переменными, дающее модель (1) с непрерывно меняющейся структурой [6].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) \cdot x_{ik} + \varepsilon_i, \quad (1)$$

где (u_i, v_i) – местоположение i -го объекта, т. е. координаты i -й точки;

y_i – значение результативного признака;

x_{ik} – значение k -й объясняющей переменной для i -го объекта;

$\beta_k(u_i, v_i)$ – неизвестные коэффициенты;

ε_i – регрессионные остатки;

$i = 1, 2, \dots, n$; $k = 1, 2, \dots, p$.

Коэффициенты $\beta_k(u_i, v_i)$ будут оценены обычным способом, следующие новые подмножества «соседних» точек уже будут оценены по аналогии. Эти оценки будут иметь некоторую степень смещения, так как коэффициенты будут проявлять некоторое отклонение по всем локальным взаимосвязанным подмножествам. Чем больше размер локальных взаимосвязей подмножества, тем ниже стандартные ошибки оценок коэффициентов; но необходимо учесть, что расширение этого подмножества увеличивает вероятность ввода в наклон коэффициента. Для уменьшения данного эффекта можно применить корректировку: предполагая, что чем дальше точки подмножества от i -го объекта, тем больше вероятность, что они имеют различные коэффициенты.

По существу, модель позволяет измерить отношения, присущие модели вокруг каждого местоположения i -го объекта. Поэтому метод взвешенных наименьших квадратов дает основу для понимания принципа работы географически взвешенной регрессии. В ГВР в соответствии с близостью наблюдений к местоположению i -го объекта можно сказать, что взвешивание наблюдения не является постоянным, однако зависит от i [3].

Оценки коэффициентов модели $\beta_k(u_i, v_i)$, вычисляются отдельно в каждой пространственной точке i . Для выявления локальных особенностей учитывают влияние измерений в соседних точках, исходя из их расстояния к местоположению i -го объекта. Степень близости задается весом w_{ij} .

Вектор оценок коэффициентов для точки i имеет вид:

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y, \quad (2)$$

где $W(u_i, v_i)$ – диагональная матрица весовых коэффициентов размерности $(n \times n)$, элементы которой определяют степень влияния соседей j на зависимость в местоположении i .

Оценку уравнения (2) получают в результате применения метода взвешенных наименьших квадратов. Однако географически взвешенная регрессия исключает постоянство матрицы весов, используя вариацию в зависимости от местоположения i -го объекта. Следовательно, весовая матрица должна быть вычислена для каждой точки i .

Отметим, что локальные оценки параметров могут быть получены для любой точки в пространстве [3]. Таким образом, в результате получают матрицу оценок параметров:

$$B = \begin{pmatrix} \beta_0(u_1, v_1) \beta_1(u_1, v_1) \dots \beta_p(u_1, v_1) \\ \beta_0(u_2, v_2) \beta_1(u_2, v_2) \dots \beta_p(u_2, v_2) \\ \dots \\ \beta_0(u_n, v_n) \beta_1(u_n, v_n) \dots \beta_p(u_n, v_{n1}) \end{pmatrix}. \quad (3)$$

На сегодняшний день существуют различные методы расчета весовых коэффициентов, описанные в трудах зарубежных и отечественных ученых.

Если территория имеет административное деление, раскрывающее специфические закономерности, характерные отдельным или всем административным единицам, то это учитывают в весовых коэффициентах. Для точек, относящихся к району A с местоположением i , элемент матрицы принимаем равным единице, в противном случае – равным нулю:

$$\begin{aligned} w_{ij} &= 1, \text{ если } (ij) \in A, \\ w_{ij} &= 0, \text{ если } (ij) \notin A. \end{aligned} \quad (4)$$

Метод движущегося окна применяют при исторически образованных административных районах, не характеризующих естественное расслоение объектов. Дискретные веса рассчитывают, учитывая расстояния между данными объектами. Отметим, что при методе движущегося окна необходимо определить предельно допустимую удаленность – фиксированное расстояние b , относительно которого определяют категорию ближайшего соседа (5).

$$\begin{aligned} w_{ij} &= 1, \text{ если } d_{ij} < b, \\ w_{ij} &= 0, \text{ если } d_{ij} \geq b. \end{aligned} \quad (5)$$

Расстояние между объектами определяют как расстояние между точками на плоскости. Величина b является фиксированной и называется шириной окна (или полосы пропускания); подробнее данный аспект рассмотрим ниже.

Подход, в котором веса строятся с учетом непрерывного изменения расстояния между исследуемыми объектами, называют ядерным, а веса, которые характеризуются убывающими функциями расстояния – ядрами. Рассмотрим наиболее распространенный и часто используемый критерий – ядро Гаусса:

$$w_{ij} = \exp\left(-\frac{\alpha}{2}\left(\frac{d_{ij}}{b}\right)^2\right), \quad (6)$$

где b – фиксированная ширина полосы пропускания;
 α – масштабный коэффициент.

Ядра би-квадрат являются альтернативой ядрам Гаусса:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & \text{если } d_{ij} < b \\ 0 & \text{если } d_{ij} \geq b \end{cases}. \quad (7)$$

Би-квадрат представляет собой непрерывное изменение веса в пределах полосы пропускания, причем за ее границей значение будет равно нулю. Вычислением с непрерывно меняющимися весами также служат ядра три-куб:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^3\right)^3, & \text{если } d_{ij} < b \\ 0, & \text{если } d_{ij} \geq b \end{cases}. \quad (8)$$

Однако здесь убывание носит «более крутой» характер. Большой вес возникает у наиболее близкого окружения заданной точки и быстро убывает, приближаясь к границе полосы пропускания. Ядра с неизменной шириной полосы пропускания дают хороший результат, при измерениях на равномерной решетке. Но, как правило, на практике наблюдения расположены по территории неравномерно. В данном случае использование зафиксированной ширины полосы пропускания может привести к недостаточному объему данных на слабо заполненных территориях, что повлечет за собой неустойчивость оценок коэффициентов. Кроме того, это приведет к огрублению зависимости на территориях с высокой плотностью наблюдений. Данные недостатки помогает избежать использование адаптивных ядер.

Выделяют различные способы поиска оптимальной ширины полосы пропускания:

1 Простейшим вариантом является одиночный размер полосы пропускания. Исследователь в праве самостоятельно определить конкретный размер полосы пропускания.

2 Золотое сечение. В большинстве случаев поиск золотого сечения дает возможность эффективно определять оптимальный размер полосы пропускания. Для автоматического поиска оптимального размера полосы пропускания есть и второй вариант – поиск интервала.

3 Поиск интервала является простым перебором с помощью обычного размера пропускной способности в пределах заранее определенного диапазона. По сравнению с методом золотого сечения, поиск интервала является более интуитивным и надежным методом.

Отметим, что не исключено неадекватное завышение дисперсии прогноза по сравнению с наблюдаемой переменной. Это обеспечивает появление локальной проблемы коллинеарности, вызывающей нестабильный прогноз. Проблему можно решить, используя более большую пропускную способность.

В случае золотого сечения и интервальных параметров поиска, оптимальная ширина полосы может быть найдена в качестве нижнего предела данной категории. В таком случае, рекомендуется использовать меньший нижний предел поиска полосы пропускания и повторить процедуру выбора ширины окна. В методе золотого сечения и интервальных поисков, оптимальный размер полосы пропускания определяется путем сравнения показателей отобранных модели с различным пропускным размером.

Кроме того, моделирование методом географически взвешенной регрессии требует применение критериев, по которым сравниваются модели, полученные с использованием различной ширины полосы пропускания:

1 AICс и AIC является наиболее подходящим с точки зрения статистических предсказаний для локального моделирования регрессии Гаусса. Отметим, что степень свободы, как правило, невысокого порядка. Классический AIC имеет тенденцию к выбору меньшей полосы пропускания, что обеспечивает сглаживание географически различных коэффициентов.

2 BIC / MDL является опцией, выбирающей большие размеры полосы пропускания. Индикатор используется для утверждения степени сложности процесса, не с целью статистического прогнозирования ненаблюдаемых результатов, а обеспечивая возможность проведения анализа.

3 CV – другим классическим, надежным показателем является кросс проверка (CV), которая применима только к гауссовой модели.

Однако выполнение эконометрических исследований ГВР в российской практике на сегодняшний день достаточно проблематично и требует ускорения внедрения на отечественный рынок таких программных продуктов как: MatCad, SigmaPlot, ArcGIS, STATA, GWR4, R [7]. К сожалению, в России в настоящее время мало внимания уделяется определению «справедливой» оценки

социально-экономических процессов; не развита система достоверных информационных баз данных.

Метод географически взвешенной регрессии должен стать важным дополнением в арсенале инструментов исследователей. ГВР позволит им оценить, насколько воздействие различных факторов, определяющих формирование социально-экономических явлений и процессов, зависит от географического расположения единиц наблюдения, и установить, какая информация (фактор) может быть использована для принятия управленческих решений.

Список литературы

- 1 Харламов, А. В. *Статистический анализ пространственно варьирующих данных методом географически взвешенной регрессии на примере рынка жилья г. Саратова: дис. ... канд. экон. наук / А. В. Харламов. – Саратов, 2012. – 161 с. – Библиогр.: с. 123-137.*
- 2 Mennis, J. *Mapping the Results of Geographically Weighted Regression / J. Mennis // The Cartographic Journal. – 2006. – №2. – Pp. 171–179.*
- 3 Fotheringham, A. *Geographically Weighted Regression / A. Fotheringham, C. Brunsdon, M. Charlton. John Willey & Sons, 2002. – 269 pp.*
- 4 Харламов, А. В. *Исследование динамики цен на жилую недвижимость методом географически взвешенной регрессии / А. В. Харламов // Известия Саратов. ун-та. Серия: Экономика. Управление. Право. – 2011. – № 2. – С. 25-29.*
- 5 Носов В.В. *Эконометрическое моделирование цены однокомнатной квартиры методом географически взвешенной регрессии / В.В. Носов, А.П. Цыпин // Известия Саратовского университета. Новая серия. Серия Экономика. Управление. Право. - 2015, Т. 15, выпуск 4. – С. 381-387.*
- 6 Стебунова О.И. *К вопросу о применении метода географически взвешенной регрессии / О.И. Стебунова // В сборнике: Университетский комплекс как региональный центр образования, науки и культуры Материалы Всероссийской научно-методической конференции (с международным участием). 2013. - С. 1554-1557.*
- 7 Салиева, А. К., Цыпин, А. П. *Эконометрическое моделирование цен на вторичном рынке жилья крупных городов ПФО / А. К. Салиева, А. П. Цыпин // Международная студенческая олимпиада по статистике: сб. науч. тр. – Москва: МЭСИ, 2015. – С. 100-104.*